Optimizable and implementable aggregate response modeling for marketing decision support

Sönke Albers

Kühne Logistics University, Brooktorkai 20, 20457 Hamburg, Germany

ARTICLE INFO

Article history:
First received in 1, November 2011 and was under review for 1½ months
Available online 23 March 2012
Area Editors: Gary L. Lilien and Marnik G. Dekimpe

ABSTRACT

The methodological discussion on the calibration of aggregate marketing response models has shifted away from how to obtain usable input for optimization toward how to avoid biases in statistical estimation. The purpose of this article is to remind researchers that such calibration is performed either to support managers in their marketing-mix decisions or to create general knowledge that leads to a better understanding of marketing relationships and thus indirectly supports decisions. Both goals require response models that are optimizable. The models must also be implementable if actual decision support is the objective. Herein, I identify several aspects for which these requirements are not always fulfilled: First, the appropriateness of the chosen functional form of the marketing response models is rarely discussed, although different forms imply quite different optimal solutions. Second, endogeneity is taken into account by structural equations, even though we lack sufficient information on how managers reach their decisions. Third, estimation methods for response models are often evaluated based on goodness-of-fit, while an assessment of their usefulness for subsequent optimization is neglected. Therefore, I provide recommendations for improving the current practice by better specifying the response function and undertaking more simulation-based evaluations of the best estimation method for use in subsequent optimization. With respect to implementation, usability can be facilitated using spreadsheets and heuristics. Moreover, gaining generalizable and replicable knowledge requires better documentation of results, which can be achieved through providing elasticities and as many details as are necessary to replicate a study, thereby enabling faster learning.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

At the beginning of my scientific career, I was fascinated by the field of operations research because its methods allowed for improving managerial decision-making. After working on scheduling problems (Albers, 1980) with only a small potential for improvement, I moved into the field of marketing, which offered more challenging and interesting problems with much greater opportunities for impact. After designing algorithms for better product positioning (Albers, 1979; Albers & Brockhoff, 1977), I realized through collaboration with managers that the development of optimization models will only lead to implementable results if one also appropriately calibrates the sales response functions on which all of the models are based. I found sales management to be an under-researched area with great potential for improvement, but I also learned that researchers face very skeptical sales managers who question whether the proposed models will really help. I was involved in a wide range of problems regarding the deployment of effort and organization of responsibilities that lead to support of sales managers’ territory planning (Skiera & Albers, 1998, 2008). From this experience, I came to realize that the focus of the majority of researchers in marketing has shifted away from supporting managers toward estimating interesting response phenomena with ever more sophisticated algorithms but often few managerial implications. Thus, the estimation of sales response functions has received much more attention than has the subsequent use of the resulting parameter estimates for decision-making.

Many marketing studies focus on estimating the effectiveness of marketing instruments, such as advertising or personal selling. Such knowledge is needed to assess performance, evaluate policies, and optimize the marketing-mix. If the findings are published appropriately, practitioners and academics can learn from them and prepare for assessments when no data collection is possible. Effectiveness analyses can be based on aggregate sales response functions that model sales or market share at the aggregate market level or, in case of the buying or choice behaviors of customers, at the disaggregate level. I will concentrate my discussion on aggregate sales response functions. In response to the aforementioned shift in research, the purpose of this article is to remind researchers that the calibration of response functions should directly support managers in their marketing-mix decisions and/or produce proper insights into marketing relationships. Both purposes require response models that are optimizable. If actual decision support is the objective, then the models should also be
implementable. These requirements are not always fulfilled because researchers often construct aggregate response models that do not provide opportunities for optimization and are difficult to implement for decision support. In addition, researchers sometimes do not offer results that can be used to produce generalizable insights or that allow for easy replication. Therefore, I will derive several recommendations for improving current practice toward distinguished marketing, as postulated by Leeflang (2011).

My first observation is that response models in marketing are not always constructed for later optimization. Sometimes, these models consist of linear or even exponential relationships, or they use quadratic relationships even when the turning point of the resulting parabola is only supported by a few observations. Therefore, I discuss how the use of moving averages and visual inspection can help researchers to better determine the appropriate functional form. Moreover, how to address possible biases arising from endogeneity is currently a popular topic. However, the recipe to use structural models that model not only the demand side but also the supply side is questionable when we do not know how managers reach their decisions. Thus, I discuss what kind of research on the decision-making behavior of marketing managers is sought. Furthermore, the predominant evaluation criteria, which are based on statistical considerations such as bias avoidance and goodness-of-fit, may neglect the usefulness of the results for subsequent optimization. Therefore, I discuss how simulation-based research can guide researchers to choose the best estimation methods.

My second observation is that optimization models are rarely used in practice (Lilien, 2011), unless they are implemented in tools, such as spreadsheets, that the decision-maker can readily master. The manager must also understand the logic of the solution structure, which is realized when using simple heuristics. With the help of two examples that actually supported real-world decisions, I discuss possible ways to persuade managers to use models.

My third observation is that studies often provide results that cannot be used beyond the specific purpose of that particular study. Gaining generalizable knowledge requires a better documentation of results. Rather than reporting the significance and value of regression coefficients, the authors of publications on marketing variables should be asked to provide elasticities that represent informative and generalizable results (Farley et al., 1995; Tellis, 1988). Thus, I state requirements on providing sufficient details on the derivation of estimates and other outcomes so that other researchers can easily reproduce their results, thereby enabling faster learning.

In the following, Section 2 of this paper examines the interrelationship between the estimation of aggregate response functions and the subsequent optimization of marketing spending. Subsection 2.1 concentrates on recommendations for the suitable choice of the functional form; Subsection 2.2 addresses how to avoid pitfalls in taking endogeneity into account; and, finally, Subsection 2.3 discusses the problem that achieving a good fit does not guarantee the best estimates for subsequent optimization. Section 3 provides two examples that show that using numerical optimization to solve optimization problems may be unacceptable to managers who want to understand the logic of a solution and consequently favor easy-to-understand heuristics. Section 4 criticizes the poor documentation found in many articles, showing that their results can neither be used for meta-analyses, nor easily reproduced or thoroughly assessed. Section 5 provides conclusions and recommendations.

2. Optimizable aggregate response modeling

Marketing research is an applied discipline and should therefore provide methods and knowledge to directly support the marketing-mix decisions of marketing managers (Lilien, 2011). Alternatively, it should create general knowledge that leads to a better understanding of marketing relationships and thus indirectly supports managers’ decisions. If we concentrate on direct support, we must discuss not only the statistical aspects of estimating aggregate marketing response models but also the implications for the subsequent use of these estimates for optimization. In the following, I critically discuss neglected areas with regard to specifying response models and their estimations, and I make recommendations to improve the current state of the art.

2.1. Functional form of the response function

Franses (2005a) discusses several important diagnostics for econometric models before using them in policy simulations. These diagnostics include multicollinearity, autocorrelation, heteroscedasticity, distributional assumptions of the error term, functional form, sample selection, static versus evolving data, heterogeneity, and, in particular, endogeneity. The motivation is that violations of the linear model should not lead to biased estimates or significance tests. Although the importance of the correct functional form is stressed in textbooks such as those by Lilien and Kotler (1983, p. 67 et seq.), Leeflang et al. (2000, p. 50) and Hanssens et al. (2001, p. 113), marketing literature pays surprisingly little attention to it. For example, Franses (2005a) could not identify a single article in the Journal of Marketing Research from 1998 to 2003 in which diagnostics for evaluating the appropriate functional form were discussed. To obtain more insights, I discuss the difficulty of choosing the appropriate functional form on the basis of an example from the pharmaceutical industry. The data are quite typical and displayed in Fig. 1.

Fig. 1 displays 5007 observations of sales depending on visits (calls) of sales persons with a set of physicians (also called detailing) across geographical areas called nanobricks. In Germany, sales data are not available for individual physicians, but only for sets of at least 6 physicians in 5007 extra constructed geographically defined nanobricks, to preclude any individual inference. In contrast, call data are available for individual physicians because these data have to be internally reported (see Skiera & Albers, 2008 for a more detailed description of this type of data). To make the observations comparable, I divided the sales and call data by the number of physicians per nanobrick to arrive at the average sales per physician and the average frequency of calls per physician. The company wanted to determine the influence of the frequency of calls (visits per physician that took place over a period of three years) on the average sales per physician in the aforementioned 5007 geographically defined nanobricks. The rationale is that different sales people apply different frequency policies, and the company wanted to see according to which functional relationship this would lead to different levels of sales per physician. Fig. 1 displays a dense cloud of points from which one can identify only a vague relationship between higher frequency and increased sales. However, it is difficult to infer the exact relationship.
Thus, the specification of the functional relationship warrants closer investigation.

The first consideration is that a functional form must be chosen that has reasonable properties for the effect of marketing instrument spending on sales. This consideration implies that functions should always exhibit diminishing marginal returns, at least after a certain point. Otherwise, the optimum would be zero or infinity, which is implausible. While this rule was frequently violated in the 1970s, response models are generally nonlinear now. However, a few papers that use linear sales response functions can still be found in leading journals (e.g., Mizik & Jacobson, 2004). It is sometimes argued that a linear function is a good approximation of the neighborhood of current behavior and at least allows for a conclusion about whether one should increase or decrease spending. However, as they do not provide information on how much one should alter spending, linear functions are inferior to nonlinear functions, which can be conveniently estimated now with standard statistical software. Furthermore, proponents of linear functions argue that the purpose of response function estimation is not only to achieve optimization but also to gain generalizable insights into marketing response, for which approximation is sufficient. What appears to be plausible at first glance is actually misleading. If approximation is sufficient, why are some researchers criticizing the failure to take endogeneity into account, for example? In both cases, we know that we will obtain biased estimates, so it is unclear why some biases should be accepted while others are not.

Even as it is not advisable to use linear functions, it is strange to observe that convex functions are occasionally specified. This phenomenon can be observed for count data where distributions such as the negative binomial are used and the expected effect on mean sales is specified as an exponential function (Venkataraman & Stremersch, 2007). This specification implies “the more the better” at any level, which is counterintuitive for marketing instrument spending.

When researchers acknowledge that sales response functions should exhibit diminishing marginal returns, they still have many options. Textbooks such as the one by Lilien and Kotler (1983, p. 67 et seq.) and extended in Lilien et al. (1992, p. 650 et seq.) and later by Leeflang et al. (2000, pp. 66–83) discuss several alternative functional forms, which are shown in Table 1 (note that the names of the functions vary across textbooks).

Researchers often specify the functional form for the response function to be multiplicative (also called log–log) (Hanssens et al., 2001, p. 102). This method is popular because it implies a constant elasticity that facilitates interpretation (see Table 1) and can be linearized by taking logarithms. However, studies rarely discuss the fact that linearization implies a different distribution for the error term, namely, log-normal instead of normal. This lack of discussion is unfortunate because estimating with a normal (which is only possible with nonlinear least-squares estimation, or NLS) or a log-normal error term (with an ordinary least-squares (OLS) estimation based on logs) leads to different estimates and thus to different optimal solutions. In Table 2, the two methods are compared by providing the optimal values for the frequency (x) based on a typical margin of 75% (Sridhar et al., 2012) and a typical cost per call of €30.

### Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Function</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic</td>
<td>[ y = a + b \cdot x + c \cdot x^2 ]</td>
<td>[ b \cdot x + 2c \cdot x^2 ]</td>
</tr>
<tr>
<td>Constant elasticity</td>
<td>[ y = a \cdot e^x ]</td>
<td>b</td>
</tr>
<tr>
<td>(or log–log or multiplicative)</td>
<td>[ y = a + b \cdot \ln(x) ]</td>
<td>b</td>
</tr>
<tr>
<td>Semi-logarithmic</td>
<td>[ y = a \cdot a^{(1-e^{-bx})} ]</td>
<td>[ (a-y) \cdot bx ]</td>
</tr>
<tr>
<td>Modified exponential</td>
<td>[ y = \exp(a + b/(1 + x)) ]</td>
<td>[ a \cdot b \cdot (1 + x)^2 ]</td>
</tr>
<tr>
<td>Log-reciprocal</td>
<td>[ y = \exp(a + b/(1 + x)) ]</td>
<td>[ a \cdot b \cdot (1 + x)^2 ]</td>
</tr>
</tbody>
</table>

The difference in results for the optimal x-values (frequency of calls) for the different functional forms estimated either linearly or non-linearly may be extreme, as is shown in Table 2. When we observe such divergent optimal values, we must question what is correct. The reason for the divergence is that for estimations in logs, larger values play a less important role in the sum of squared errors. Very often, estimation in logs requires that a constant be added to all values; otherwise, zero values would preclude taking logs. In my example, I added a 1 to sales and a 0.01 to frequency, but it remains unclear what the influence of this practice is. Not knowing which functional form is correct, it would be helpful to test the prediction quality. In most cases, this test is performed with the help of holdouts. However, this is not a reliable testing method for functional forms. The holdout cases are often structurally equivalent, which means that fitting and testing their validity would yield similar results. A true test is only possible for cases that are rather different from the sample, sometimes with disappointing results. Perhaps holdout validation would be worthwhile only for assessing the danger of overfitting, but overfitting is an issue only for intervals that are not supported by data—not for any interval with a sufficient number of data points.

Another popular way to apply linear models for estimation and account for diminishing marginal returns is to work with a quadratic response function that can be estimated linearly in the linear and quadratic terms. The problem with this function is that it will always provide a point at which the dependent variable (sales) reaches a maximum and then decreases. Some researchers find this attractive because they believe that they can test for decreasing effectiveness (e.g., supersaturation) (Hanssens et al., 2001, p. 112). For example, Manchanda and Chintagunta (2004) find that because of the negative sign of the quadratic term, detailing effectiveness decreases after a certain point, and they argue that at some point of over-detailing, more visits do not lead to more sales. Homburg et al. (2011) argue that there must be an optimal level of sales person–customer orientation in sales encounters and find support for their argument in the significant negative coefficient for the quadratic term. However, such strong conclusions warrant further investigation.

The distribution of the observed data points is essential for the interpretation of decreasing effectiveness, as is suggested by quadratic response functions. If the majority of the points are normally distributed around the mean, then the distribution of the points in the middle mass strongly determines the shape of the function because their error terms are more frequent than the error terms for the tails of the distribution. Thus, it can happen that the point of maximum sales lies in the top decile (i.e., in the 10% of outside values). Of course, the shape of the curve is then not really determined because a regression only through these more extreme points would provide quite different results. A conclusion that sales are indeed decreasing can only be derived if the quadratic function better predicts this outer interval than a nonlinear function, which is concave and has no supersaturation by definition. In our case, the quadratic relationship displayed in Fig. 1 has positive signs for both terms, the linear and quadratic, which imply a convex relationship where the optimal value of x is infinite and thus implausible. Therefore, we omit this functional form from further consideration.

What is the best function? If we rely on statistical goodness-of-fit criteria, we will choose the one with the highest $R^2$. However, that
does not automatically mean that we can use the parameter values for optimization. This situation becomes apparent when we look at Fig. 2, which exhibits the shape of all of the functional forms (estimated differently) provided in Table 1.

We see that there are three types of functional forms in Fig. 2. The non-linearly estimated shapes of the constant elasticity and the modified exponential form are almost linear. In contrast, a moving average shows that the function flattens considerably after x > 40. The log-reciprocal comes closest to it. The constant elasticity function as estimated in logs with OLS is much more concave than the semilog function because residuals of originally smaller values get relatively higher weights. The fit of the functions decreases as the more concave it becomes. As this is in contrast to the more intuitively plausible higher weights. The

estimated in logs with OLS is much more concave than the semilog

log-reciprocal comes closest to it. The constant elasticity function as

fl

age shows that the function

response function, we may identify two sources of error: the wrong

However, if we take into account different possible shapes of the re-

pro

fi

form, we

—

The results are better described using

inverted V-shaped

functions is inadequate, and

and too few degrees of freedom. Kolsarici and Vakratsas (2011)

apply spline regressions to discover the shape of the advertising re-

tion by n linear functions connected to each other so that any func-

example, with the help of the MARS software (Salford Systems, 2008). The advantage of using a spline regression is that it provides the ability to derive flexible functional forms. However, this advantage comes at the cost of more parameters being required to describe the function. The more splines are chosen, the fewer degrees of freedom remain available. This leads to a difficult trade-off between fit and too few degrees of freedom. Kolsarici and Vakratsas (2011) apply spline regressions to discover the shape of the advertising response function. They find that the standard nomenclature based on “concave,” “convex,” and “S-shaped” functions is inadequate, and the results are better described using “hockey-stick,” “V-shaped,” and “inverted V-shaped” functions and their combinations. In our case, the estimation leads to a system of equations suggesting a convex shape that is not trustworthy (see Fig. 2). Other non-parametric methods include Kernel regression and local polynomial smoothing. Van Heerde et al. (2001) find Kernel regression superior to polynomial smoothing and spline regression, but Kolsarici and Vakratsas (2011) find the reverse. Of course, non-parametric methods need large sample sizes, which are not always available. The smaller the sample size, the more suitable are the parametric functions.

In my experience, visual inspection of the data, with the help of moving averages, is an important step. This visualization can be achieved even with the help of spreadsheet software such as Excel. The result is plotted in Fig. 1. We see from the moving average that the true relationship flattens at the end, which the estimated functions do not take into account. The reason is that the mass of observations lies in the lower range of the frequency of calls (median 12.75). Of course, moving averages depend on the data window applied. In my experience, the appropriate window depends on the number of observations. The goal must be to specify a wide-enough data window so that a more or less smooth function results. In this case, I used a window of 250 data points to compute the moving average.

In the end, visual inspection based on moving averages should guide in the selection of the best functional form. There is no simple recipe for how to arrive at the appropriate function. Following a combination of model diagnostics and the marketing scientist’s intuition is often the best way to choose the appropriate function (Blattnberg & Hoch, 1990). In our case, I selected a few points from the moving average that show that the function evolves more or less linearly up to a frequency of 40 and then is almost flat (see Fig. 3). In my experience with data sets for sales force response functions, such a shape— with diminishing elasticity because of saturation—occurs rather often. Therefore, I prefer to work with a functional form that allows for many different shapes but is particularly attractive because it can directly take into account diminishing elasticities. Surprisingly, this functional form is not commonly discussed in textbooks and is thus

![Fig. 2. Estimated shapes for different functional forms.](image-url)
Endogeneity bias has long been discussed in economics as a potential threat to obtaining consistent parameter estimates in a response function. Failure to account for endogeneity can potentially bias the parameter estimates of the marketing-mix variables. If this happens, both major uses of these models—i.e., for diagnostic and optimization purposes—produce misleading results that can seriously affect the outcomes of marketing decisions (Villas-Boas & Winer, 1999).

With an increasing number of reports stressing the danger of the endogeneity bias, reviewers have placed this bias on their agenda (Shugan, 2004). This becomes evident from Franses (2005a), although he states that he did not intend to specify a set of rules that would make it easier for editors and reviewers to reject articles (Franses, 2005b, p. 27). The endogeneity bias technically refers to a variable as being endogenous if it is correlated with the error term. In principle, endogeneity can arise in static models as a result of measurement error, autocorrelated errors, omitted variables, and sample selection errors, whereas in dynamic models, simultaneity plays an important role in causing endogeneity.

In their seminal article, Chintagunta et al. (2006) argue that data frequently encompass only some marketing instruments but not all of them. Very often, the data are correlated so that the omitted variables lead to an overestimation of the effects of the included marketing instruments. This bias is confirmed in meta-analyses (Albers et al., 2010; Henningsson et al., 2011; Sethuraman et al., 2011). The question then becomes whether one can do something to address this bias. Of course, better data (Shugan, 2004) or experiments (Chintagunta et al., 2006) would provide the necessary insights, but managers are often reluctant to conduct field experiments because they want to avoid negative customer feedback if some customer groups are treated less favorably as part of an experiment.

In principle, endogeneity can be corrected by instrumenting the included marketing variables (see Subsection 2.2.1) with the help of variables that correlate with the included ones but are exogenous to the omitted ones. However, the success of this procedure depends on finding truly exogenous instrumental variables. If this is not possible, variations of a certain marketing variable based on the incomplete response function can only be evaluated under the assumption that the other correlated variables (for which data are not available) are also similarly altered. In this case, the effectiveness of the respective marketing instrument is inconsistently estimated because it also captures the effects of the other marketing instruments. However, the result can still be used for policy insights as long as we assume that the other marketing instruments would be altered accordingly based on the hypothesized correlation (similarly Villas-Boas & Winer, 1999, footnote 2).

Endogeneity is also present when the independent variables are characterized by non-random strategic behavior (Manchanda et al., 2004). These authors find only a very loose relationship between calls by sales people and sales. However, by establishing a separate input equation for this relationship to take into account the fact that the sales people may have their own private information (which is unobservable) on the responsiveness of their calls, the authors were able to derive “stronger” estimates that showed better holdout predictive validity. However, the holdout samples used in marketing studies are usually not systematically different from the estimation samples (Van Heerde et al., 2005, p. 20). Therefore, the usefulness of such estimates for a major change in policy cannot be assessed. We do not know whether the incorporation of private information in an additional input equation is really better than what the data show. Given the subjectivity of this approach, one is reminded of the so-called decision calculus approach (Little, 1970 and Lodish, 1971), which advocates the use of subjective estimates. Apparently, this approach lost acceptance because today’s more detailed data favor the more objective econometric derivation of estimates. However, with the introduction of additional input equations based on the subjective specification of the dependence of input, it seems as if we are in part returning to decision calculus.

Finally, note that endogeneity is not something that is either present or not. Rather, endogeneity is always present to a certain degree. Therefore, we only need to correct for endogeneity if it is present at too high a degree. It is clear that pricing decisions depend on the behavior of competitors, which is the classic argument for the treatment of endogeneity. However, it is less likely that data on marketing spending variables are highly endogenous when the variables are allocated to many units, as is the case when sales people must allocate calls across hundreds of small sales coverage units or when the company must align territories based on elasticity estimates (see Skiera & Albers, 1998). This is particularly true if the data are reported for short time intervals such as months because the optimal call

---

**Fig. 3.** Shape of response function as estimated from moving averages.
frequency in sales forces is often specified on a yearly basis. In this case, the sales people usually consider which calls are convenient in certain months and only try to achieve optimal performance levels over the course of the whole year. Thus, we may observe a substantial randomness of calls across time but not across customers. As a consequence, critical levels of endogeneity for marketing spending data may be given only at more aggregate levels (over time) but not at disaggregate levels.

At first glance, it seems imperative that we correct for the endogeneity bias to obtain the correct optimization results. However, on closer inspection, it is unclear whether correcting for endogeneity bias will always help with subsequent optimization. Therefore, I critically discuss proposals to address endogeneity in Subsections 2.2.1 (instrumental variable approach) and 2.2.2 (structural models). In Subsection 2.2.3, I discuss the prerequisites for these proposals, namely, more detailed research on how the supply side is determined by managers.

### 2.2.1. Instrumental variable approach

One way to correct for the endogeneity bias is to replace an endogenous independent variable with its estimates based on exogenous variables. If we have the following relation:

$$y = a_0 + a_1 x_1 + a_2 z_1 + \epsilon$$

and \(x_1\) is endogenous, we need exogenous variables (i.e., instrumental variables) that are correlated with the endogenous variable but not with the error term. Denote this additional exogenous variable as \(z_2\). Thus, \(x_1\) has to be instrumented on the basis of all exogenous variables, the variables that are exogenous but correlated with \(y (z_1)\), and the instrumental variable \((z_2)\), which is correlated with \(x_1\) but not with the error term. This is performed with the estimation shown below:

$$x_1 = b_0 + b_1 z_1 + b_2 z_2 + \nu$$

and \(x_1\) in Eq. (2) is then replaced with its predicted values estimated from Eq. (3). The estimation can be performed with Two-Stage Least Squares (2SLS). The estimates are identical for linear functions that take \(y\) into Eq. (2) as an additional variable according to the control function approach (Petrin & Train, 2010). The control function approach has the advantage of enabling a direct test of the degree of endogeneity, which can be performed by testing the significance of the coefficient for \(\nu\). In this case, we assume that we have linear functions. However, if we have nonlinear functions such as logits, as it is often occurring in marketing, it is less clear how to specify the instrumental variables (Andrews & Ebbes, 2011).

#### 2.2.1.1. Exogenous instruments

Although the instrumental variable approach is theoretically appealing, its application is far from trivial and leads to some nontransparency with respect to the consequences of using instrumental variables. Several studies (e.g., Danaher et al., 2008) do not report the results of the first stage of the estimation—namely, the regression of the endogenous variables on the instrumental variables—and sometimes do not even report the instruments themselves. Consequently, the reader cannot assess the success of the replacement with the instrumented variable. For example, the higher the incremental \(R^2\) is for including the instrumental variables in this first stage, the less defensible is the claim that these variables themselves are uncorrelated with the disturbances (Ebbes et al., 2005, p. 366). Therefore, it can happen that the instrumental variables originally assumed to be exogenous are not actually exogenous. Consequently, if the incremental \(R^2\) is low, it is unclear whether the instrumental variables are sufficiently strong. With unsuitable instruments at the extremes of the \(R^2\) scale, it might be worthwhile to investigate whether the ideal instrument should exhibit an intermediate \(R^2\) at 0.5. In the case of weak instruments, the coefficients are not estimated consistently (Bound et al., 1995), and the consequence of this inconsistency might even be worse than the consequence of not correcting for endogeneity (Ebbes et al., 2009). Thus, it is desirable that the incremental \(R^2\) be reported. Further, Staiger and Stock (1997) propose a rule of thumb that the \(F\)-value for the additional inclusion of instruments in the first stage should be greater than 10. However, this \(F\)-test does not evaluate whether the instruments are correlated too much with the error and do not remove the endogeneity. Thus, the \(F\)-test is only a necessary but not a sufficient condition.

In the case of weak instruments, the instrumental variable (IV) approach “may simply be replacing one possible source of specification error with another” (Bronnenberg et al., 2005, p. 24). Chintagunta et al. (2006, pp. 609–610) further note: “All too often, variables are picked to be instruments without much formal justification. The principal advantage of this approach is that it allows the researcher to remain agnostic about the nature of firm behavior that generates the correlation between the observed and unobserved factors.” Fong et al. (2011) show that not all instruments are equally suitable. In an investigation of which instruments come closest to the experimental results for price sensitivity, they find that commodity prices perform poorly, but wholesale prices are well suited.

#### 2.2.1.2. Lagged instruments

Sometimes researchers use lagged variables as instruments (Villas-Boas & Winer, 1999). This approach is appropriate as long as the data have no dynamic effects that may hold for prices, as applied by Villas-Boas and Winer (1999). However, we generally find a certain degree of carry-over for variables that are based on marketing spending, such as advertising or selling effort. In these cases, lagged variables as instruments may still be correlated with the error. Other instruments besides lagged variables are rarely adopted for marketing-mix variables. For example, if we take wholesale prices as exogenous instrumental variables for price, the question arises as to why these prices are exogenous. Every wholesaler will try to set his prices in such a way that, given the margin of the retailer, he will achieve a profit maximum. Thus, which instrument is truly exogenous? For example, Luan and Sudhir (2010) use more than 35 instrumental variables to instrument an advertising budget; however, none of the variables were truly exogenous because all were derived from within the movie business. In his concluding editorial as editor of Journal of Marketing, Kohli states that “perhaps only God may not be endogenous, but it would be difficult to convince some reviewers of that” (2011, p.4).

#### 2.2.1.3. Latent instruments

If there are no observed instrumental variables available, Ebbes et al. (2005) and Ebbes et al. (2009) suggest using latent discrete instrumental variables, where group classifications of the observations for the latent instrument and regression parameters are estimated jointly while controlling for correlated errors. This approach has been applied recently by Zhang et al. (2009), Grewal et al. (2010), and Sonnier et al. (2011). Although this method has been shown to be somewhat robust against departures of assumptions, it breaks down in particular if the distribution of the unobserved instruments and the endogenous variable is normal (Ebbes et al., 2009, p. 454).

Another question is how to demonstrate that the use of instrumental variables provides better results. The goodness-of-fit criterion does not represent a meaningful metric because it must be weaker if we replace an endogenous variable with its instrumented variable. Ebbes et al. (2011) investigate whether it is possible to assess the suitability of instrumental variables with holdout validation. They find this to be impossible because holdout validity is always better for regressors of biased OLS than it is for corrected OLS with IV.

Luan and Sudhir (2010) distinguish between intercept and slope endogeneity. They argue that the typical instrumental variable approach, as it has been described in this paper, only solves the problem...
of intercept endogeneity bias. However, it can still lead to biased estimates if the values of one endogenous variable are chosen non-randomly according to some type of strategic behavior, as discussed by Manchanda et al. (2004). Luan and Sudhir (2010) call this effect slope endogeneity in which the marketing-mix variables correlate with the slope coefficients. To correct for slope endogeneity, Luan and Sudhir (2010) suggest using a control-function approach that introduces an interaction term of the error from the first stage of instrumentation with the endogenous variable into the second stage. They show that this term is significant and yields consistent estimates for the endogenous variable.

Returning to the original purpose of obtaining consistent estimates that are useful for optimization, it is important to approximate the magnitude of bias that can be corrected with instrumental variables. This approximation should be realized by reporting not only the results of the final instrumental variable approach but also the inconsistent OLS estimates. This detailed information is important because the use of different instrumental variables always leads to different elasticities, thereby implying different optimal values. The reports should be complemented by sufficient test results that show the existence of endogeneity before and after instrumentation. Only then will the user gain an impression of how dangerous it is to use the coefficients in subsequent optimization. Therefore, I recommend the following:

R2: Articles should provide better reasoning for the choice of instrumental variables, a better report of the results of the first stage estimation, tests for endogeneity and exogeneity in the case of several competing instruments, and a comparison of the estimates obtained with simple OLS and those obtained with 2SLS or the control function.

Bascle (2008) provides a good overview (with a checklist) of procedures in Stata that can be used to test the various assumptions put into the instrumental variable approach. From this, we can gain a better understanding of all of the effects caused by the use of instrumental variables. Currently, it seems that if an author uses instrumental variables, the reviewers are satisfied, but the use of these variables can have severe consequences that have not been sufficiently investigated.

2.2.2. Structural models

Another method for addressing endogeneity is to specify a structural model wherein the endogenous variables are explained by other variables, and a system of equations is simultaneously estimated. In the estimation of the effectiveness of marketing instruments, several authors suggest specifying the levels of the marketing instruments depending on some kind of optimization behavior by the managers. The assumption here is that the managers have private information that allows them to optimize their decisions. In this case, the levels of the independent but endogenous variable—e.g., detailing—are modeled in a way that follows the optimality condition of allocation, which, in turn, depends on the effectiveness parameters to be estimated. Thus, these parameters are estimated for the original response function while simultaneously attempting to reproduce the behavior of the manager. This procedure appears to be appropriate on a theoretical basis, but its appropriateness depends on our knowledge of the behavior of marketing managers. If managers already optimize their spending, further recommendations for optimization cannot be derived because the parameters would already reflect optimality. However, “the parameter estimates (and even the ability to identify some parameters) rest on these underlying assumptions. If these assumptions are incorrect, the resultant parameter estimates are biased” (Lehmann et al., 2011, p. 162). Indeed, the next section will show that managers often do not behave optimally. Their competitive behavior is typically not characterized by striving for a Nash equilibrium. Rather, it is more competitive than Nash equilibria would suggest (e.g., Leeflang & Wittink, 1996; Marks & Albers, 2001). Thus, if we incorrectly specify the underlying behavior, the derived estimates may be biased even more. Strangely enough, academic work on managerial decision-making in marketing is scarce (Wierenga, 2011). Consequently,

R3: More research is needed to justify the various assumptions on how managers actually behave in their decision-making when using structural models.

2.2.3. Research on management behavior

To more effectively eliminate endogeneity bias, we need to fully understand the determinants and logic on which the so-called supply side depends. This information would provide insight into either the appropriate instrumental variables that can be used to instrument the endogenous variable, or the correct specification of supply-side equations. For example, if we are concerned with the effect of marketing budgets on sales, we need to know exactly how the budgeting process is performed by managers, which involves two dimensions. First, if we suspect that these budgets depend on exogenous factors, we need to conduct empirical research on which factors are the truly relevant drivers. Preferably, the results should be fairly general so that a multitude of studies can be based on their results. Second, if we suspect that the budgeting follows a strategic behavior (or a certain logic) that suggests working with structural equations, it is necessary to specify exactly which rationale the managers use to determine their budgets. The rationale behind strategic behavior can be optimal or heuristic behavior. Only after this rationale is defined would we be able to specify the supply-side equation correctly. Otherwise, the estimated parameters would be seriously mis-specified. Alternatively, it is suggested to build the model by assuming multiple types of strategic behavior and subsequently picking the one with the best fit (e.g., Otter et al., 2011). However, as long as we do not have hard evidence of the type of strategic behavior, we may run into the danger of overfitting the model.

An early project to gain more insight into budgeting behavior was ADVISOR (Lilien & Little, 1976). A database of budgets for the promotion and sales force of 125 industrial products was analyzed by Lilien (1979) with the help of a log–log regression model. Lilien shows substantial effects of the size of the market, previous sales, higher sales force budgets for more complex products, and decreasing budgets as the product life cycle advances. Later, Balasubramanian and Kumar (1990) and Alilwadi et al. (1994) use a broad sample of products from all sectors to investigate whether the ratio of advertising budget to sales varies according to growth and share. Whereas the first group of authors finds an influence, the second group does not, and each group claims that the analysis of the other is flawed (Alilwadi et al., 1997, and Balasubramanian & Kumar, 1997). Note, however, that these studies were not conducted for the purpose of finding the best instrumental variables. In fact, share and growth are also not exogenous, and they would call for the best specification of a supply-side equation. In fact, we observe that the majority of research occurs on the demand side, whereas the supply side is very much neglected. In my opinion, this observation can be attributed to the fact that this type of research does not test theories, which is something that reviewers frequently request (Lehmann et al., 2011, p. 157). This shortcoming leads to the following recommendation:

R4: Marketing science needs more fact-based studies that provide descriptive material on how managers behave. These data would create a basis for developing theories about managers’ behavior.

To provide a sense of the many under-researched questions regarding managers’ behavior in determining budgets for various marketing instruments, I refer to a recent study by Fandrich and Albers (2010), who investigate the marketing budgets for all products in the US pharmaceutical sector. They use the same very broad database as Fischer and Albers (2010). Although Fischer and Albers (2010)
omit products with either no marketing spending over the whole period or less than 1% market share, even for the 2831 remaining products that can be considered to be important market players, observations show that more than 65% have zero budgets (65% for detailing, 85% for journal advertising, and 97% for DTCA). Zero budgets can be due to pulsed strategies, but for many products, there is no budget at all, which means that we may have missed an important aspect of extant budgeting behavior. Apart from pulsing (e.g., Doganoglu & Klapper, 2006), there must be reasons for spending nothing at all, and these reasons have rarely been incorporated into studies that model the supply side of budget levels.

Another question is whether one can assume the same behavior for all managers on the supply side. The investigation by Fandrich and Albers (2010) with the help of latent class regression shows that there are segments that behave very differently, for example, with respect to the influence of the product life cycle on the budget level. Finally, Fandrich and Albers (2010) tested the optimality of the budgeting decisions of managers, which is often assumed in supply-side equations. They found that many budgets are suboptimal. Except for Mantrala et al. (2007), who found that the majority of budgets in the newspaper business were near optimal, the findings by Fandrich and Albers (2010) are in line with many other studies, e.g., Hanssens et al. (2001, p. 363) and Naik et al. (2005). This finding supports the argument that in structural modeling, it is not reasonable to assume optimal behavior on the supply side. Otter et al. (2011) developed a test based on Bayesian estimation for whether the inputs show indications of strategic behavior. The test might be useful in determining whether to accept supply-side equations. Thus, I recommend:

R5: When specifying instrumental variables or supply-side equations in structural models, the respective models must allow varying behaviors among individuals that can produce results contrary to those produced when optimal behavior is assumed.

2.3. Usefulness of estimates for optimization

The ruling paradigm in estimating the parameters of sales response functions is to optimize a goodness-of-fit measure. For simple regressions, this may be the minimization of the sum of squares of the estimation error, which is equivalent to maximizing the R². In more advanced cases, the likelihood of reproducing values is maximized. As noted in Section 2.1, it is debatable whether the estimated parameter values should be based more on the mass of points around the mean or more on the extreme values. Weighted regression is one possible solution to this problem, but it implies that a modified version of the goodness-of-fit objective is retained.

A more comprehensive way to evaluate the estimation methods for aggregate sales response functions is to rely on the usefulness of the resulting coefficients for subsequent uses, such as optimization or deriving recommendations, as proposed by Propp and Albers (2009). Thus, we must consider the whole process, from estimating the parameter values to using them in decision models. For example, allocation decisions depend on reasonable estimates of the parameter values of all allocation units. If just one parameter value is heavily over- or underestimated, then the other units will be incorrectly allocated less (or more) of the budget. In this case, a goodness-of-fit measure that is typically based on the average fit across units is not helpful. Instead, it is better to base the choice of the estimation method on results from published simulation studies that have already evaluated the superiority of estimation methods based on the results of subsequent optimization in a wealth of systematically varied data situations.

Using a computer simulation study, Propp and Albers (2009) evaluated the ability of certain estimation techniques—e.g., ordinary linear regression (OLS), fixed-effects regression, Hierarchical Bayes (HB), Maximum Simulated Likelihood (MSL) and Empirical Bayes methods—to provide useful estimates for a subsequent allocation task. Simulation experiments are effective methods for producing generalizable results (Leeflang, 2011). From an assumed true sales response function, Propp and Albers generated 80 different data situations that varied with respect to the number of allocation units, the number of observations per unit, the heterogeneity across units and marketing instruments, the data quality (i.e., the randomness of the data), and the carry-over intensity. To assess the influence of randomness, three replications were generated based on common random numbers. The results show that in cases with very poor data, heuristic approaches to allocation—which were not based on estimates—outperformed all econometric methods in terms of simulated expected profit. The authors observed that both fixed effects and MSL methods are very robust with respect to weak information in the data.

Although the parameter recovery of the fixed effects model, as measured by the root mean squared error (RMSE), may not be as good as approaches that explicitly account for heterogeneity across marketing instruments, the fixed effects model still delivers quite satisfactory allocation solutions. Thus, (average) goodness-of-fit may be misleading. In cases of strong heterogeneity, MSL delivers very good results. However, this method has the drawback of being more complicated to implement because it is not commonly offered by standard statistical software packages. Additionally, the model is very sensitive to the number of random draws for the simulation. The same problem occurs in HB procedures, which are very sensitive to the prior variance assigned to them. MSL performs better, but HB offers slight advantages when the data are heterogeneous and have a small error. For the Empirical Bayes and OLS methods, the authors find very high sensitivity with regard to the magnitude of the error and the number of individual observations. Interestingly, heuristics work well. In particular, if the cost/benefit ratio is taken into account, the proportional-to-sales heuristic is very good.

The findings by Propp and Albers (2009) provide evidence for the conjecture that robustness is more important than estimating individual heterogeneity when little information is available in the data. As a consequence, researchers should be aware that choosing the method that leads to the best fit may not necessarily be the most appropriate choice with regard to subsequent allocation decisions; thus,

R6: When concerned with the estimation of coefficients to be used in a subsequent optimization task, the appropriate estimation method should be chosen based on results from published simulation studies that evaluate what can be expected for specific data situations with respect to the method’s usefulness for optimization.

3. Implementable aggregate response modeling

Most studies that provide estimates on the effectiveness of marketing instruments display the values of these parameters and then perform some type of optimization. If the problem is complex, as is the case for many budgeting and allocation tasks, numerical optimization is performed to provide the manager with an optimal solution that s/he should implement. My experience has shown me that managers rarely implement such solutions. Instead, they prefer to make use of tools that they can master. Managers face a high degree of uncertainty in achieving success. To reduce their perceived risk, they want to fully understand the logic of the solutions or recommendations that they receive (Little, 1970). Numerical optimization does not allow for that understanding. Thus, if we really want to support the decisions of managers, we should invest more time in providing simple heuristics that are proven to be close to the optimum.

In keeping with Little’s (1970) request for models that are easy to control, the derivation of solutions must be implemented in a tool that managers can use. Spreadsheets are such an instrument; thus, decision support should be made available in spreadsheet tools. Fader et al. (2005, p. 282) follow Albers (2000) in noting, “the use
of marketing models in actual practice is becoming less of an exception and more of a rule because of spreadsheet software. The textbook "Marketing Engineering" by Lilien and Rangaswamy (2004) is also based on the idea that the basic concepts can be better understood if implemented in spreadsheets. Such software also allows for easy what-if analyses.

To better illustrate the usefulness of heuristics and implementations in spreadsheets, I will discuss two examples of projects that achieved acceptance by managers. In the first project, to support the pricing decision for checking accounts at a bank (Albers et al., 2010), we used choice-based conjoint analysis to estimate the parameters of the preference functions of the bank's customers. Of course, this analysis deals with disaggregate response functions, but the conclusions can be easily transferred to the case of aggregate response functions. Based on the preference functions of more than 300 customers, we derived a recommended menu of three two-part-tariffs to better skim the customers’ willingness-to-pay. However, the vice president in charge did not accept this solution because if its implementation caused problems, he would lose his job. Therefore, he asked not only for the optimal solution but also for similarly profitable alternative solutions, so that he could identify solutions with a profit potential nearly as high as the optimal one but with lower risk. Thus, our choice simulator was constructed (using Excel) in such a way that the manager could evaluate solutions similar to what was proposed for multi-criteria decision-making. In particular, many solutions could be generated and ranked according to criteria defined by the manager, including profit, loss of customers, and the risk of non-differentiated tariffs. On this basis, the manager finally chose a solution that he was comfortable with and which on implementation fortunately led to substantial profit improvements.

In the second project, supporting allocation decisions for marketing budgets across countries, instruments, and products, Fischer et al. (2011) provide a heuristic that is conveniently implemented in Excel and easily understood by managers. Their procedure was implemented by Bayer and provided a profit improvement potential of more than 500 million Euros. The basic idea is to first derive the optimality condition for a problem. If this does not provide a closed-form solution, then find ways to replace variables in their optimum with current values, so that by iteratively applying it, one approaches the optimum. This approach makes use of the Banach fixed-point theorem proving that an iterative sequence, where values are sequentially replaced by values closer to the fixed point, will converge to the fixed point, which in our case is the true optimum (Granas & Dugundji, 2003). For the case of the general allocation problem of allocating money to marketing units and instruments, Fischer et al. (2011) derive the following optimal solution:

$$\text{Optimal Budget}_{k_{int}} = \sum_{\text{Countries}} \sum_{\text{Products}} \sum_{\text{Activities}} \text{Optimal Allocation Weight}_{k_{int}} \times \text{Total Budget}_{n} \times \sum_{n=N} \text{Optim. Allocation Weight}_{k_{int}} \times \text{Total Budget}_{n}$$

$$\text{Optimal allocation weight}_{k_{int}} = \frac{\text{Profit contribution}_{n} \times \text{Optimal unit sales}_{n} \times (\text{Optimal mktg. elasticity}_{k_{int}} + \text{Optimal growth elasticity}_{k_{int}})}{1 + \text{Discount rate} - \text{Marketing carryover}_{k_{int}}}$$

Eqs. (4) and (5) do not provide an implementable solution because the values of the unit sales and the elasticities depend on the optimal budgets. Therefore, other researchers typically engage in numerical optimization in similar cases. However, this optimality condition may be approximated if one replaces the optimal values with the actually chosen or observed values from the previous period. If subsequently applied, the values for sales and elasticities must be updated. As a result, the solution converges to the fixed point of the optimal solution.

Given the simple structure of the solution—i.e., that the budgets must be allocated proportionally to the current profit contribution margin, the current sales, the elasticity of the marketing instruments, and a growth multiplier replacing the optimal growth elasticity—the solution was easily understood and accepted. The main effect was to transform the political processes of negotiating for budgets into a fact-driven process of discussing the effectiveness of marketing instruments and growth potentials of products in different countries.

To improve the usability of the estimated aggregate response models for optimization, I, therefore, call for the following:

R7: To facilitate the use of results from the estimation of aggregate response models, the subsequent step of optimization should be based on easily understandable heuristics (if a simple, optimal, closed-form solution is unavailable) that are close-to-optimum and implemented in familiar tools such as spreadsheets.

4. Better documentation for generalizable results

4.1. Elasticities

In the last few decades, many authors have reported the results of their estimation of response functions to be able to show certain effects. Given this wealth of literature, researchers are interested in drawing generalizable conclusions about the effectiveness of various marketing instruments. Such generalizations have been provided in the form of meta-analyses of elasticities (e.g., Albers et al., 2010; Henningsen et al., 2011; Sethuraman et al., 2011). Unfortunately, many publications report only the values of the coefficients of their response function, and their significance but not the corresponding elasticities. This is acceptable only if the means of the dependent and independent variables are provided and the elasticity can be calculated for these values. However, in the case of interaction terms, the elasticities cannot be calculated unless the means of the interactions are also provided. Otherwise, the coefficient values do not provide valuable information because any generalizable conclusion depends on the context of the study. To arrive at a body of generalizable insights, it is better to report comparable results such as elasticities. Therefore, I ask editors to require these elasticities as a condition for accepting a paper for publication. This would broaden our understanding of the sales response of certain marketing instruments.

If elasticities are to serve as generalizations of effectiveness, they must be truly comparable. However, this is rarely the case. As ratios of relative changes, elasticities often depend on the values of the independent variables, which in our case are the marketing instruments. We know that elasticities typically decrease as the values for marketing instruments (spending) increase. Thus, if we do not take this into account, we are “comparing apples and oranges.” As elasticities are the most important elements of the formulas for optimal budget values, such as those derived from the Dorfman–Steiner theorem (Dorfman & Steiner, 1954), we should not only report elasticities at current values (if they are not constant) but also at the optimal value of the independent variable. Even more informative would be to obtain elasticity values at the quartiles. Consequently, if editors requested these values, we would be able to collect a wealth of truly comparable information.
Asking for comparable elasticities also allows for plausibility checks. If researchers only report the coefficients, reviewers are not able to evaluate the plausibility of the values of these estimates. We have already acquired a body of knowledge about elasticities in meta-analyses that can serve as prior information for the comparison of new results. This knowledge could help to avoid the publication of articles that appear to be technically correct but actually provide strange results that cannot be justified. To facilitate a growing body of knowledge, I therefore recommend the following:

R8: Editors should require that elasticities be supplied as a condition for accepting the paper for publication. The elasticities should be evaluated for the optimum of the function (rather than at the current values) to allow for comparisons across studies.

4.2. Reproducible results

Articles in our field often lack reproducible results because the data and estimation codes are not made available. This contrasts with many natural sciences such as physics, where a result is only trusted if someone else is able to reproduce it. Although this has been discussed already in Marketing in the 70s (Permut, Michel, & Joseph 1976), not much progress has been made.

First, making data available to the public can considerably speed up the progress in developing appropriate estimation models and techniques because it would then be possible to demonstrate a superior method based on exactly the same data. In the 1990s, we observed substantial progress in the estimation of sales response from scanner panel data after several data sets were made available by Nielsen. In the same way, tournaments such as the one on predicting churn from telecommunication data—where everybody could download the same data set and predict churn, and the organizers had data on the true churn—have elevated our understanding of the performance of certain forecasting techniques (Neslin et al., 2006).

Second, if self-developed software is used, it should be shared so that correctness checks can be performed and numerical instabilities discovered. Re-analyses have often shown that numerical results are not reproducible (Anderson et al., 2008). Software tools also often need certain settings to run. For example, in Hierarchical Bayes, priors and the numbers of draws must be defined but are not always reported, so it is unclear whether a more inexperienced user would be able to achieve the same results. Other software requires settings for convergence precision that are also very rarely reported but can influence the results considerably.

Finally, making data and software publicly available reduces the likelihood that researchers will commit the misbehavior of reporting manipulated results because everything can be checked. Hopefully, we will not see in marketing a case like that of the psychology professor who has recently been suspended for fabricating data for his publications (Levitt et al., 2011).

Given these supporting arguments, some economics journals have introduced the rule of making data and estimation codes available on request. However, this practice has shown that researchers are often unable to obtain this material when they ask for it. McCullough et al. (2006, 2008) used the (mandatory) archives of the Journal of Money, Credit and Banking and the Federal Reserve Bank of St. Louis Review to attempt the replication of 62 (113) studies but succeeded only 14 (29) times. Therefore, enforcement is a key issue.

Of course, it is easier in economics than in marketing to enforce a policy that makes everything available in a mandatory archive. In marketing, we often have proprietary or commercially offered data sets. Therefore, we should distinguish between various cases. First, estimation codes (if not commercially available) or simulation data can and should be provided without difficulty. Second, if anyone conducts a meta-analysis, it is not problematic to publish the data because it has already been published. The latter has been realized in the journal BuR—Business Research for data used in the meta-analysis by Henningsen et al. (2011). Third, if somebody has collected data in a survey, the question arises of whether there are privacy issues in cases that data can be traced back to someone—as is apparently possible for social network data (Wondracek et al., 2011)—and when combinations of control variables such as size, location, and industry allow readers to infer the identity of firms. Similarly, when the data come from individual companies, the managers may be concerned that their competitors could gain confidential information or that researchers engaged by the competitors would inform them of something valuable. Sometimes, this concern can be addressed by disguising the data appropriately, but it more often precludes such data sets from becoming public. Another strategy is to work with an embargo on the data for a certain number of years (e.g., three, five or ten years). This scheme would at least enable researchers to reproduce the results at some future time. In cases with non-publishable data, it should be possible to have a third party check the data— with the help of the original researcher at the premises of the original investigator—in such a way that the checker cannot access the whole data set but can still view its structure. Of course, making research results more reproducible is a very difficult process, with many barriers to overcome, and some of my proposals may appear unrealistic at the moment. However, if editors begin to steer their editorial policy in the direction that articles are less likely to be accepted if the results are not reproducible, our field will benefit in the long-run by speeding up scientific progress (Anderson et al., 2008). None of my proposals should be construed as mistrust, but rather as efforts to progress science by checking and improving everything. In addition, if authors using code, programs, or data from other authors properly cite their sources, there will be a positive incentive to make that information publicly accessible. Therefore, I recommend the following:

R9: The editors of journals should engage in discussions on how to better enable the reproducibility of results. Private computer codes should always be made available in an online archive. Articles should not be published unless material that can be made public without restrictions is actually made available beforehand. Editors should encourage the publication of articles that report differing results of attempts to reproduce the results of previous studies.

I am convinced that following these recommendations will lead to the creation of a new culture of science with a much faster methodological progress.

5. Conclusions

Marketing research has provided many interesting results that have not been applied often because the studies have not taken into account the aspects that managers consider to be important. If the estimation of aggregate marketing response models is performed to support managers in their marketing-mix decisions, such research must discuss not only the statistical issues of fitting and avoiding bias but also the need to yield optimizable models.

In my view, a thorough analysis of the functional form of the response model is often neglected, yet it is very important. With the help of a typical example from sales management, I show that it is not trivial to arrive at the appropriate functions and that a misspecification has substantial implications for subsequent optimization. Researchers should be aware that the typical evaluation criteria of statistical fit or significance of polynomial terms can be misleading. For applications, researchers as well as practitioners should rely more on visual inspection with the help of moving averages. Combining subjective knowledge and insights from plots is recommended to accurately specify the functional form. From my experience, a rather unknown functional form with a diminishing elasticity in the exponent often captures the relationship well.
In stark contrast to their neglect of the correct specification of the functional form, researchers are more often concerned with avoiding all sorts of biases. Reviewers apparently take the endogeneity bias as the most important one with regard to rejecting papers, although it is not clear what consequences the methods for correcting this bias might have on subsequent optimization. Researchers should be very careful in supporting their models with predictive holdout validity measures because one of the methods, instrumental variables, generally leads to worse holdout validity than does the inclusion of the biased endogenous variable. Structural equations as an alternative should also be used with care because we have limited knowledge of how managers reach decisions. Thus, I call for more studies that investigate the actual behavior of managers in setting their marketing-mix. This implies that reviewers should accept more studies that are descriptive and do not necessarily test a theory.

Researchers should take into account that response models may successfully reproduce sales by utilizing the response function but may nevertheless be problematic if the models lead to false optima. This shortcoming can occur even with sophisticated methods such as Hierarchical Bayes and Maximum Simulated Likelihood when they are applied to smaller data sets because a single misestimated parameter can ruin the whole allocation task. Thus, researchers should use simulations to investigate which estimation method is most suitable in certain data conditions.

Researchers should also be rewarded for successful implementations, which will shift the focus away from sophistication and toward facilitating the use of models by integrating the optimization into tools mastered by decision-makers, such as spreadsheets. In addition, researchers should provide a solution structure that can be understood by managers, as is the case for heuristics. Researchers should, therefore, develop more powerful heuristics that can be implemented in spreadsheet software.

Gaining generalizable knowledge requires the proper documentation of results. Therefore, rather than reporting only the significance and the values of the regression coefficients, the authors of publications on marketing variables should be asked to provide elasticities at the optimal values as informative and generalizable results. They should also provide sufficient details on the derivation of estimates and other results so that the results may easily be replicated by others, thereby enabling faster learning. Ideally, editors should require authors to provide data and estimation codes in a publicly accessible archive. Of course, disclosure agreements and the exclusive use of commercial data or software must be respected. However, following these guidelines will enable a new culture of science to emerge and realize faster methodological progress.

Acknowledgments
I wish to thank Marnik Dekimpe, Peter Ebbes, Peter Leefflang, Don Lehmann, Gary Lilien, Demetrios Vakratsas, Harald Van Heerde, Berend Wierenga, and my group of former academic students, in particular, Christian Barrot, Silke Boenigk, Michel Clement, Marc Fischer, Karen Gedenk, Manfred Kraft, Dominik Papiès, Bernd Skiera, and Franziska Völkner, for their valuable comments on previous versions of this manuscript, and my current doctoral students, in particular, Thomas Fanndrich, Michael Kiechert, and Sina Henningsen, for their support.

References


The role of consumer self-control in the consumption of virtue products

Danit Ein-Gar a,⁎, Jacob Goldenberg b, Lilach Sagiv b

a Recanati Graduate School of Business, Tel-Aviv University, Ramat Aviv 69978, Israel
b School of Business Administration, The Hebrew University of Jerusalem, 91905 Israel

A R T I C L E   I N F O
Article history:
First received in 5 May 2010, and was under review for 5½ months
Available online 7 March 2012
Area editor: Luk Warlop

Keywords:
Self-control
Virtue products
Health behavior
Sunscreen lotion

A B S T R A C T
Virtue products (such as sunscreen lotion and dental floss) promise future benefits and, at the same time, carry immediate and ongoing usage costs. Although consumers acknowledge the benefits of virtue products, they find it difficult to consume them on a daily basis. This research focuses on a key problem in the consumption of virtue products—ongoing use—and identifies ways to help consumers maintain ongoing consumption.

We propose and show that products' attributes (in terms of future versus present benefits) and consumers' dispositional self-control interact to shape the consumption of virtue products. In two field experiments that use different product categories—dental floss and sunscreen lotion—we show that low self-control participants consume a virtue product whose product description highlights a present benefit more than they consume a virtue product whose description highlights a future benefit. Among high self-control participants the reverse effect was observed. In a third study we show the same pattern of results when willingness to pay is measured.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Virtue products (e.g., sunscreen lotion, dental floss, condoms, gym workouts and car seatbelts) are products whose usage is associated with future benefits and immediate costs (Read, Loewenstein, & Kalyanaraman, 1999; Wertenbroch, 1998). These costs include not only purchase costs, which are common to all products, but also psychological, physical and emotional costs that are experienced during consumption and that lead many people to consume virtue products less often than they should (e.g., Arthey & Clarke, 1995; DellaVigna & Malmendier, 2006; Wichstrøm, 1994). The present research investigates how a virtue product's attributes (in terms of present benefits versus future benefits) interact with the consumer’s dispositional self-control to influence product consumption.

Our research focuses on a key problem in the consumption of virtue products: ongoing use. Consumers find it difficult to consume virtue products on a daily basis. For example, DellaVigna and Malmendier (2006) show that health-club members paid a monthly fee that reflected an expectation to visit the club more than seven times a month, but they actually visited, on average, less than four and a half times per month. Similarly, research on sunbathing and the use of sunscreen indicates that consumers fail to use sunscreen adequately, even though they are aware of the potential damage caused by exposure to the sun (Arthey & Clarke, 1995; Wichstrøm, 1994).

Frequent consumption of virtue products is difficult for several reasons. The benefits gained from using these products are experienced not immediately but in the distant future. Moreover, the consumption of such products can be physically or emotionally costly, time-consuming, painstaking, and in many cases unpleasant (e.g., a dental checkup).

To encourage consumers to purchase and use virtue products, traditional marketing approaches have usually highlighted the essence of these products and the benefits they provide. Thus, for example, campaigns for dental floss illustrate the importance of protecting one's teeth and gums from plaque build-up, and these campaigns also emphasize product attributes, such as the strength of the floss, that help achieve the goal of a beautiful smile. We argue that the main essence of a virtue product is typically associated with a benefit that is experienced in the future. Past research suggests that future-focused messages may be effective in encouraging consumers to purchase virtue products because these are products that people feel they “should” use, and people are more prone to spend money on “should” products when making decisions that apply to the distant future (for a review, see Milkman, Rogers, & Bazerman, 2010). However, we conjecture that the persuasiveness of such approaches is limited in establishing ongoing consumption because these approaches fail to change the basic premise that the cost, no matter how small, is to be experienced immediately and throughout consumption (i.e., in the present), whereas the benefit, no matter how important, is to be experienced sometime in the distant future. More specifically, we suggest that emphasizing the future benefit appeals only to some consumers, depending on the consumers’ levels of self-control.

In two field studies measuring the actual, ongoing consumption of sunscreen lotion (Study 1) and dental floss (Study 2), we hypothesize and show that among consumers with low self-control, a virtue...
product description that highlights a present benefit yields higher consumption than does a virtue product description highlighting a future benefit. We find the opposite pattern for consumers with high self-control: a product description that highlights a future benefit yields higher consumption of the virtue product than does a description highlighting a present benefit. Study 3 expands these findings and shows that consumer willingness to pay for a virtue product is affected in a similar manner.

We begin by laying the theoretical basis for our proposed model. We define the construct of dispositional self-control and discuss its relationships to time focus (present versus future). We then draw from the literature on congruency effects to establish our benefit-congruency hypotheses. Finally, we report findings of three studies demonstrating how dispositional self-control and product attributes influence consumption of virtue products and discuss the implications.

2. Theoretical background

Past studies refer to self-control as the ability to delay gratification (Funder, Block, & Block, 1983; Metcalfe & Mischel, 1999; Mischel, Shoda, & Rodriguez, 1989), avoid being impulsive (Ainslie, 1975), avoid procrastination (Ariely & Wertenbroch, 2002; Lay, 1986; O’Donoghue & Rabin, 1999; Steel, 2007) and override short-term goals that stand in the way of long-term goals (Fishbach, Friedman, & Kruglanski, 2003; Fishbach & Shah, 2006; Muraven & Baumeister, 2000). Integrating these studies into a general theoretical framework, we suggest that self-control can be viewed as a process reflecting an inner struggle and an intentional effort that individuals invest to override the desire to perform actions or inactions that promise immediate gratification in the present, yet at the same time promote future negative outcomes. Ein-Gar and colleagues (Ein-Gar, Goldenberg, & Sagiv, 2008; Ein-Gar & Sagiv, 2011; Ein-Gar & Steinbart, 2011) have suggested that such actions can be categorized into two general types: “doing wrong” (impulsive, self-satisfying actions that harm one’s future well-being, such as eating a cake while on a diet, buying products one cannot afford), and “not doing right” (actions of procrastinating about what needs to be done and thus, once again, risking one’s future well-being, such as failing to exercise, use sunscreen or write an important paper).

Like many other psychological constructs (e.g., emotions (Kahneman, Diener, & Schwarz, 2003) and anxiety (Spilberger, 1996)), self-control can be viewed as both a trait (i.e., a stable, individual attribute) and a state (i.e., affected by the immediate context). Most of the research has studied self-control as a state, inferring it on the basis of how individuals perform tasks demanding self-control (e.g., Ariely & Wertenbroch, 2002; Mischel et al., 1989; Muraven & Baumeister, 2000; Vohs & Faber, 2003). In contrast, a growing body of literature has investigated self-control as a stable, personal attribute (notable examples include McCabe, Cunnington, & Brooks-Gunn, 2004; Mischel et al., 1989; O’Gorman & Baxter, 2002; Tangney, Baumeister, & Boone, 2004; Turner & Piquero, 2002). These studies provide evidence for the stability of self-control as an individual difference. For example, Mischel, Shoda, and Rodriguez (1989) showed that a four-year-old’s ability to resist temptations predicted achieving high grades in school later between the ages of six and twelve. More recently, Duckworth and Seligman (2005) found that eighth-grade students’ self-discipline predicted (above and beyond intelligence) grades, school attendance, and high school selection at older ages. These studies suggest that dispositional self-control is a stable personality aspect with a powerful role in shaping people’s behavior.

In the current research, we adapt the perspective that self-control is an inherent personality trait. We suggest that dispositional self-control predicts the actual, ongoing consumption of virtue products, and that this relationship depends on the congruency between the product’s benefit and the consumer’s self-control. Specifically, we propose that consumers with high self-control are more likely to consume a virtue product when the product offers future-focused benefits than when present-focused benefits are offered, whereas consumers with low self-control are more responsive when present-focused rather than future-focused benefits are offered.

Past studies have pointed to the relationships between self-control and time orientation, showing that individuals who focus on the distant future have higher levels of conscientiousness, higher impulse control, lower sensation seeking and a stronger focus on the future consequences of their actions, compared with those who focus on the near future (Strathman, Gleicher, Boninger, & Edwards, 1994; Zimbardo & Boyd, 1999). Individuals who focus on the distant future show less impulsive and more controlled behavior in consumption-related activities such as smoking, drinking and drug use (Keough, Zimbardo, & Boyd, 1999), impulsive and compulsive buying, and credit card use (Nenkov, Inman, & Huland, 2008). Thus, the literature has established a basis for the notion that having a future-oriented time perspective is related to a self-controlled behavior.

This idea is further supported by a pilot study we conducted that directly investigated the relationships between dispositional self-control and focus measures. One hundred and twenty-one participants (Mage = 26.52% females) completed a self-control measure (the Dispositional Self-Control scale (DSC), Ein-Gar et al., 2008; Ein-Gar & Steinbart, 2011), and three self-reported measures of time orientation: the consideration of future consequences (Strathman et al., 1994), the elaboration of potential outcomes (Nenkov et al., 2008) and time perspective (Zimbardo & Boyd, 1999). As expected, respondents with higher levels of self-control were more likely to consider the future consequences of their actions (r = .54), to consider the positive and negative potential outcomes of their actions (r = .42), and to hold a future time perspective (r = .61) but not a hedonic present perspective (r = -.51; for the full list of correlations, see Appendix B).

Following the above, consumers with high self-control tend to have a long-term time orientation. They are motivated to enhance their long-term well-being, and their goals are future-oriented. Thus, they are likely to value and to focus their attention on those benefits that are experienced in the future, and that are expected to enhance future well-being. Consequently, consumers with high self-control are expected to be more persuaded by messages that highlight the future rather than the present benefits of consuming virtue products. Individuals with low self-control, in contrast, have a short-term time orientation; they focus on their current well-being rather than their future well-being. Their goals are present-oriented and strongly associated with immediate gratification and the inability to resist temptations. Accordingly, consumers with low self-control are likely to value those benefits that are experienced immediately and that immediately affect their well-being, and they are expected to be more persuaded by messages highlighting present benefits as opposed to future benefits. Consumers with low self-control are less occupied with the consideration of future outcomes, whether good or bad; therefore, they are less likely to be attentive to benefits related to future experiences. This implies that traditional marketing approaches that highlight virtue products’ essences (i.e., their future benefits) are less likely to be effective for promoting consumption among such consumers.

What approach, then, might encourage the consumption of virtue products among individuals with low self-control?

If the focus of consumers with low self-control is indeed on the present, the key to influencing such consumers lies in introducing present benefits associated with consuming virtue products. Specifically, we suggest that introducing an additional product attribute (or emphasizing an existing attribute) whose benefit can be experienced immediately and continuously would be effective for encouraging consumers with low self-control to consume a virtue product.
Furthermore, for this attribute to be meaningful in terms of changing consumers' expected (and actual) experience of the product, not only must this attribute provide an immediate benefit, but this benefit should also be ongoing and dependent on consumption.

This is necessary to ensure that, unlike a one-time handout (e.g., a gift added to the product), the benefit is perceived by the customer as part of the product itself.

We further reason that the added present-benefit attribute can be peripheral and does not necessarily have to be essential: it should merely offset the immediate cost associated with consuming the virtue product. Thus, for example, in a dental floss campaign, we propose that instead of focusing on a message that one's teeth will be healthier with frequent flossing, the persuasion message might emphasize that the dental floss has a refreshing mint flavor. Although mint flavor is a peripheral attribute and is less dominant and important in comparison with the product's essence (i.e., a means of achieving healthy teeth), it provides an immediate and ongoing benefit (i.e., the enjoyable taste and smell of mint) that depends on ongoing consumption and offsets, in a sense, the immediate and ongoing cost associated with flossing.

Adding a present benefit to product consumption often provides the consumer with an enjoyable consumption experience (Chitturi, Raghunathan, & Mahajan, 2008), which could serve as an alternative explanation for increased consumption among consumers with low self-control. Thus, for example, dental floss with mint flavor may be more enjoyable to use than regular dental floss. In designing our studies, we aimed to show that an added present benefit affects consumption above and beyond its impact on the enjoyment of consumption.

Research has given a great deal of attention to different benefit-congruency effects such as the one proposed. Chandon, Wansink, and Laurent (2000), for example, explored the congruency between product type (utilitarian or hedonic) and benefit type (monetary or non-monetary). Work within the persuasion literature has focused on message-receiver congruency (e.g., Fabrigar & Petty, 1999). Wheeler, Petty, and Bizer (2005) tested a benefit-congruency effect from a dispositional perspective, focusing on consumers’ levels of extraversion (Experiment 1) and need for cognition (Experiment 2). A growing body of research is focusing on benefit-congruency effects related to consumer regulatory goals. These studies explore congruency effects when a message’s appeal matched the consumer’s regulatory state (e.g., Cesario, Grant, & Higgins, 2004; Chang & Chou, 2008; Chernev, 2004) or regulatory disposition (e.g., Cesario & Higgins, 2008; Latimer et al., 2007; Latimer et al., 2008; Zhao & Pechmann, 2007). For example, Latimer et al. (2007) demonstrate that information about a virtue act—a physical activity—framed in terms of prevention elicited positive feelings toward this act among people with a prevention-goal orientation, whereas the same activity framed in terms of promotion elicited positive feelings among people with a promotion-goal orientation. Similar results were found for messages encouraging other virtue acts such as the intake of fruit and vegetables (Latimer et al., 2008). In general, studies in the regulatory focus domain show that marketing messages that use a benefit-congruency approach enhance individuals’ positive cognitive and emotional reactions; when there was a fit between consumer regulatory goals and the message appeal, overall persuasion increased (Chang & Chou, 2008), and the value consumers inferred from their choices or actions increased (Avnet & Higgins, 2006) as did their positive feelings and confidence about the choices made (Cesario & Higgins, 2008). Chernev’s (2004) work on goal-attribute compatibility further reveals the underlying process for these congruency effects by demonstrating that specific product attributes that are compatible with consumers’ goals are given more weight in the consumer’s evaluation process.

In this same vein, we suggest that a virtue product’s message appeal should go hand in hand with consumers’ self-control in creating benefit-congruency experiences. The goals of consumers with low self-control are embodied in their current existences. Although the essence of a virtue product is the future benefit that it offers, the addition of a present benefit can enhance the product’s fit with the present-time orientation of consumers with low self-control. Consumers with high self-control, in contrast, focus on their long-term goals and well-being, and therefore a virtue product that offers future benefits is congruent with their orientation. Present benefits, however, are incongruent with the orientation of consumers with high self-control. Adding present benefits may, therefore, result in lower overall congruency as compared with offering only future benefits. Therefore, although it may seem counterintuitive, we suggest that consumers with high self-control may be less likely to consume virtue products when a present benefit is offered together with a future benefit than when only a future benefit is offered.

In sum, we hypothesize that the time focus of the product’s attribute (whether present or future) will interact with consumers’ dispositional self-control to shape the consumption of virtue products.

Previous studies measure the consumption behavior of virtue products by observing whether consumers cashed in coupons for free samples of a product (e.g., sunscreen) or reported the intention to use a product (e.g., Detweiler, Bedell, Salovey, Pronin, & Rothman, 1999; Rothman, Salovey, Antone, Keough, & Martin, 1993). However, these measures have two shortcomings. First, as the consumption of virtue products requires ongoing effort, such consumption is not well-reflected in these measures (e.g., Detweiler et al., 1999, p. 194, with regard to sunscreen). Second, these measures provide little indication for consumer behaviors when they actually need to pay for the product. In the current paper, we try to address these two issues. In the first two studies, we test ongoing consumption in natural settings, whereas in Study 3 we test willingness to pay. Specifically, in Study 1, participants were given a dental floss product that either did or did not include an additional present benefit. In Study 2, all participants were given the same sunscreen lotion, and the experimental conditions differed in the type of benefit highlighted (present vs. future). In both studies, we tested our hypotheses for the ongoing consumption of the product. Study 3 was designed to test whether similar effects would be observed for participants’ willingness to pay for the product and to rule out alternative explanations for the findings of Studies 1 and 2.

3. Study 1

In Study 1, we measured the ongoing consumption of dental floss over several weeks, in a natural, everyday consumption environment. We hypothesized that consumers with low self-control will consume a virtue product that provides an immediate benefit more than a virtue product with no present benefit. Consumers with high self-control, in contrast, will consume a virtue product that offers only a future benefit more than a product that offers both future and present benefits. We further expected the interaction effect to go above and beyond the effect of consumption enjoyment.

3.1. Method

3.1.1. Participants

Undergraduate students (n = 111; M_age = 26, 53% females) participated in the study in exchange for a package of Plackers dental floss (containing 30 units)1 and were offered course credit and a raffle ticket (the winning ticket holder had a choice of one of the following three

1 Placontrol manufactures single-use flossing aids, which are sold worldwide. Their Tuffloss dental floss, which was used in this study, is one of Placontrol’s licensed technologies and is imported to the local market by Dentalon, Inc.
prizes: a TV, a coffee maker, and a DVD player; all prizes were priced at $180).

3.1.2. Procedure and design

Participants were informed that they were part of a two-session market research study on Plackers dental floss, conducted in collaboration with a respectable international manufacturer. Participants were randomly assigned to one of two experimental conditions (present- or future-benefit focus).

All participants first completed a self-reporting self-control measure. They then read a description of a virtue product and responded to pre-consumption attitude questions based on the product description (as part of the “market research” cover story). Next, participants received the product for personal use. The product was a package of Plackers dental floss, manufactured with patented “Tuffloss,” that (according to the manufacturer) increases floss strength and durability. Two weeks later, participants came back to the laboratory. They reported their overall enjoyment of the product, and we collected their packages of Plackers to assess their usage of the product. The Plackers packages were returned to the participants, who received their course credit and raffle tickets. Participants were then debriefed and thanked.

3.1.3. Instruments

3.1.3.1. The self-control measure. To measure self-control, we used the Dispositional Self-Control scale (DSC; Ein-Gar et al., 2008; Ein-Gar & Steinhart, 2011). The DSC conceptualizes self-control as the combination of overcoming “doing wrong” and overcoming “not doing right” impulses. This scale is the first to measure these two aspects of self-control. It includes 17 context-free items measuring self-control as a general and stable trait-like attribute. Participants report their agreement with each statement on a scale of 1 (“does not describe me at all”) to 5 (“describes me very much”). For the full list of items, see Appendix A.

The scale has been validated in several ways. First, confirmatory factor analysis (CFA) has verified the conceptualization of self-control, showing that “doing wrong” and “not doing right” are two latent factors of self-control (N = 1902, χ² (df = 107) = 1083.06, p < .001, RMSEA = .07, NFI = .90, CFI = .93; Ein-Gar & Sagiv, 2011). Second, construct validity was confirmed, relating the DSC to conceptually relevant constructs. Third, concurrent predictive validity was established, showing that DSC measurements correlate negatively with measurements of deviant behavior (for the full list of correlations, see Appendix B). In addition, self-control, as measured by the DSC, has predicted performance in self-control-demanding tasks. For example, in a recent study, Ein-Gar and Steinhart (2011) used this scale to predict persistence in monotonous, boring tasks and to predict impulsive buying and the consumption of hedonic food. (For complete details of the scale construction and validation, see Ein-Gar et al., 2008; Ein-Gar & Sagiv, 2011; Ein-Gar & Steinhart, 2011.)

3.1.3.2. Present/future benefit focus. Although all participants received Plackers dental floss manufactured with Tuffloss, we used different product descriptions to highlight benefits differently. Participants were each assigned at random to one of the following two conditions.

The future-focus condition highlighted the Tuffloss attribute, emphasizing its future benefit:

“The product is Plackers Dental Floss. This product employs patented Tuffloss®, the preferred floss for prevention of future dental problems: Tuffloss is seven times stronger than nylon and does not wear or tear. Tuffloss does not shred during use. Its efficacy in ensuring future tooth and gum health considerably exceeds that of other flosses.”

The present-focus condition did not mention the Tuffloss patent and highlighted “mint flavor” as an attribute with a present benefit:

“The product is Plackers Dental Floss. This dental floss cleans your teeth and, in addition, features a breath-refreshing mint flavor. Thus, with one action, you can clean your teeth, keep them healthy, and freshen your breath with mint.”

These two manipulations were pre-tested twice. In the first pre-test (n = 65), we conducted a pairwise t-test comparing the two attributes. According to participants’ ratings on a scale of 1 (present benefit) to 5 (future benefit), the mint flavor attribute was perceived by participants as more of a present benefit (mean rating (M) = 1.20, STD = .59), and the Tuffloss attribute was perceived as more as a future benefit (M = 2.65, STD = 1.2; t(154) = −10.05, p < .001). The second pre-test (n = 63) confirmed that the descriptions used in the manipulations (present benefit and future benefit) were preferred over a generic product description (containing no added special attribute), (χ² present benefit (1) = 4.63; p < .05; χ² future benefit (1) = 7.4; p < .01). The descriptions of the two conditions did not significantly differ from each other in their attractiveness (t(151) = −.55, NS).

Each participant in the present- and future-focus conditions received a 30-unit package of Plackers. All packages looked identical; however, only participants in the present-focus condition received mint-flavored Plackers.

Along with the product, participants received a self-report log, in which they were asked to document each time they used the product. The self-report log was intended to reinforce the manipulation. To that end, the product description (corresponding to the participant’s assigned condition) appeared at the top of each page in the log.

3.1.3.3. Product usage. In the second session, which took place 2 weeks after the first, the experimenter counted the number of units remaining in the packages. Use of the Plackers dental floss was evaluated according to the average daily number of units used by each participant. To ensure that participants did not anticipate in advance that the number of dental floss units they used would be measured, they were asked in the first session to bring the package with them to the second session under the pretext that the experimenter wished to test whether the graphics on the package wear off after frequent use, as part of the market research.

3.1.3.4. Enjoyment. To measure enjoyment, we asked participants to respond to three items in which they indicated how much they liked the product, enjoyed using it, and thought it was good. An index of enjoyment was calculated for each participant by averaging the three items (α = .91).

3.2. Results

We tested our hypothesis with a hierarchical regression to predict the average daily use of dental floss. In the first step, participants’ self-control, time-focus condition (present vs. future) and enjoyment were entered as predictors of product usage, and the model was significant (F(3,107) = 7.86; p < .01). Enjoyment positively predicted the average daily use of dental floss (β = .42, p < .01). No other significant effect emerged. Overall, the model explained 42% of the variance. In the second step, the interaction between the time-focus condition and self-control was added. This second model was significant as well (F(4,106) = 9.69, p < .01). The interaction was a significant predictor of product usage (β = 1.48, p < .01), explaining an additional 8.7% of the variance (p < .01).
To further interpret our findings, we conducted a univariate ANOVA and planned contrasts. Participants were designated as having high or low self-control according to median split. We conducted a 2×2 treatment condition (self-control: high vs. low) × 2 time-focus condition (present vs. future) ANOVA to predict participants’ consumption. Enjoyment was added as a covariate. The findings revealed a significant main effect of enjoyment (F1,106 = 25.0, p < .01). No other main effect emerged (time-focus condition: F1,106 = .28, NS; self-control: F1,106 = .30, NS). As hypothesized, the interaction between self-control and time focus was significant (F1,106 = 10.52, p < .01; see Fig. 1). Thus, the findings supported our hypothesis, showing that the consumer’s self-control interacts with the time focus of product benefits (present vs. future) in predicting product usage, and that this effect goes above and beyond the effect of consumption enjoyment.

Planned contrasts revealed, as expected, that participants with high self-control used more units in the future-focus condition (i.e., the description emphasizing Tuffloss; M = .67 units, STD = .50) than in the present-focus condition (i.e., the description emphasizing mint flavor; M = .46 units, STD = .32; t(1,107) = 2.12, p < .05). In addition, as hypothesized, participants with low self-control used more units in the present-focus condition (M = .68 units, STD = .41) than in the future-focus condition (M = .44 units, STD = .23; t(1,107) = 2.10, p < .04). Thus, the findings fully supported our hypothesis.

### 3.2.1. Conclusions

The findings of Study 1 indicate that a virtue product’s time focus interacts with a consumer’s level of self-control in predicting the consumer’s consumption. More specifically, our findings are consistent with our reasoning that congruency between self-control and type of benefit increases consumption. Among participants with low self-control, adding a present benefit (a mint flavor) resulted in more consumption as compared with highlighting a future benefit. Conversely, participants with high self-control used the product more if a future benefit was highlighted than if a present benefit was added.

Past studies suggest that utilitarian product benefits and hedonic benefits play different roles in consumer satisfaction and delight (Chitturi et al., 2008). Taste, as a specific type of hedonic benefit, was found to impact consumers’ actual experience of products (e.g., Raghunathan, Naylor, & Hoyer, 2006) and to influence consumers’ judgment of product quality (e.g., Warlop, Ratneshwar, & van Osselaer, 2005). In addition, the literature on emotions suggests that consumers’ actual and anticipated emotions play a significant role in shaping their preferences (e.g., Phillips & Baumgartner, 2002; Pollai, Hoelzl, & Possas, 2010; Shiv & Huber, 2000; Wang, Novemsky, & Dhar, 2009). We ruled out the possibility that enjoyment was the only driver of increased consumption, showing that the interaction between self-control and the time focus of the product’s benefit affected consumption above and beyond the effect of enjoyment. To further rule out enjoyment as an alternative explanation, in Study 2 we tested our hypothesis with an attribute that provides a present benefit that lacks a hedonic aspect. Again, we controlled for enjoyment.

Study 2 was also designed to generalize our findings in several ways. First, in Study 1, participants in the two experimental conditions used somewhat different products (the product in the present-focus condition was mint-flavored and the product in the future-focus condition was not). In contrast, in Study 2, all participants used the same product, and different benefits were highlighted in each experimental condition. Second, in Study 1, participants’ familiarity with the product and its essential benefits was expected to be low: a pre-test we conducted prior to running this study showed that only about 2% of the sample population was familiar with the product category of Plackers dental floss. Moreover, we expected low expertise with Plackers specifically as it is imported to the local market without any marketing communication efforts and is sold both in very small amounts and in only a handful of stores. Thus, in Study 2, we sought to investigate the effect on a product category that participants were familiar with.

### 4. Study 2

Marketing campaigns for virtue products often emphasize product essence, which generally entails an important future benefit. Unlike Study 1, the product’s essence was described in both the present focus and the future focus conditions and in the same way in this study.

In the future-focus condition, participants were exposed to a description highlighting the product’s essence—its primary future benefit. In the present-focus condition, participants were exposed to the same description used in the future-focus condition alongside a description of an additional, present benefit. This was done for the purpose of showing how a traditional marketing approach, which focuses on the product’s generic, primary essence, can induce the same effect as a message focusing on a specific future benefit. Because the essence of a virtue product is an outcome in the future, traditional marketing campaigns that focus on product essence are actually future-focused messages. Therefore, the same pattern of consumption that emerged in Study 1 for present versus future benefit messages should be observed for present-benefit versus essence-focused messages.

We expected consumers with high self-control to consume the product more when the product’s essence (which is a future benefit) was highlighted than when a present benefit was highlighted because present benefit is less congruent with their time orientation. Consumers with low self-control, however, are less likely to be influenced by an appeal emphasizing a product’s essence because it is future-oriented. Therefore, we expected consumers with low self-control to consume the product more when its appeal emphasized an attribute with a present benefit rather than the product’s essence. In addition, this study used a less hedonic present benefit as compared with Study 1.

#### 4.1. Method

**4.1.1. Participants**

Female undergraduate students (n = 71; Mage = 24) participated in this experiment in exchange for a tube of facial sunscreen lotion (retail value: $30 per unit). Course credit and a raffle ticket. (The prize was a $100 gift certificate at a retail fashion chain.)

---

2 There were no differences in enjoyment between the experimental conditions (t(1,109) = −.59; NS).

---

3 Anna Lotan Laboratories manufactures skin care products such as such as the facial sunscreen lotion used in Study 2 for use and sale in beauty salons. The company also makes private-label cosmetics such that are sold in pharmacies and department stores in Europe.
4.1.2. Procedure

As in Study 1, this study consisted of two sessions, and participants were told that they were participating in a market research study. In the first session, participants began by completing the DSC scale. They then read a description of the product and were asked to evaluate it (as part of the “market research” cover story). To ensure that the product category used in this study (facial sunscreen lotion) is characterized by high product experience we conducted a pre-test. We tested the evaluation of facial sunscreen lotion among participants from the same population as that of the study. On a 5-point scale ranging from 1 (“not at all”) to 5 (“very much”), respondents rated how important sunscreen lotion was and the frequency with which they used it. These two measures of category involvement and general usage provide some indication of product experience. t-Tests against the scale’s midpoint (3) showed that participants had high product experience (timportance (47) = 4.44; tusage (47) = 4.15; both p < .001). Results suggest that sunscreen lotion is a product category with high product experience.

Participants were each assigned at random to one of the two following experimental conditions. The future-focus condition highlighted the product’s essential nature, which in itself includes a future benefit:

“The product you are receiving is facial sunscreen lotion. This lotion will protect your face from sunburn and future skin damage. This lotion is suitable for everyday use, for all skin types, and is approved by the Health Ministry.”

The present-focus condition highlighted a present benefit in the form of moisturizing ingredients. Participants in this condition received the same description that was used in the product essence condition, yet with an addition highlighting the moisturizing ingredients:

“The product you are receiving is facial sunscreen lotion. This lotion will protect your face from sunburn and future skin damage. This lotion is suitable for everyday use, for all skin types, and is approved by the Health Ministry. To this special sunscreen lotion we added moisturizing ingredients. Thus, in one action, you can nurture your skin and moisturize it, and protect it from sun damage.”

It is important to note that participants in both conditions actually received the exact same product; that is, the sunscreen lotion did contain moisturizing ingredients, but this information was highlighted only in the present-focus condition.

This manipulation was validated in the pre-test described in Study 1. We conducted a pairwise t-test to compare the moisturizing ingredients attribute to an attribute with a future benefit (anti-wrinkle ingredients). Moisturizing ingredients received a significantly lower score on a scale from 1 (present benefit) to 5 (future benefit) (mean rating (M) = 2.06, STD = .97) than did anti-wrinkle ingredients (M = 4.17, STD = .82; t1,64) = −13.31, p < .001). Thus, the moisturizing feeling experienced when putting on the lotion is perceived as a present benefit.

As in Study 1, participants read the product description, answered an attitude questionnaire as part of the cover story, and received the product for their personal use along with a self-report log.

In the second session, which took place three weeks later, the experimenter weighed participants’ tubes of lotion to evaluate product usage. For each participant, average daily use of the lotion was calculated. As in Study 1, participants answered three survey items measuring “enjoyment” (α = .79). Finally, the tubes of lotion were returned to the participants, who then received their course credit and raffle ticket and were debriefed.

4.2. Results

We tested our hypothesis with a hierarchical regression to predict the average daily use of sunscreen. In the first step, participants’ self-control, time-focus condition (present vs. future) and enjoyment were entered as predictors. This model was insignificant (F1,67 = 97; NS). When the interaction of self-control and time focus was added as a predictor in the second step, the model became significant (F3,60 = 4.89; p < .01), indicating that the interaction contributed significantly to the consumption prediction (β = −.58, p < .01) and explaining an additional 19% of the variance (p < .01).

To further test our hypotheses, we conducted an ANOVA with planned contrasts. To that end, participants were designated as having high or low self-control based on median split. We conducted a 2 (self-control: high vs. low) × 2 (time focus: future benefit (essence) vs. present benefit) ANOVA to predict the consumption of the sunscreen lotion; enjoyment was added as a covariate.4 Consistent with the regression results, the interaction between self-control and time focus was significant (F1,66 = 22.51, p < .01). No main effect was found for either self-control (F1,66 = .80, NS), time focus (F1,66 = .70, NS) or enjoyment (F1,66 = .74, NS).

Participants with high self-control used the lotion more when only the product’s future benefit (essence) was highlighted (average consumption (M) = .23 ml per day, STD = .17) than when an additional present benefit (moisturizing ingredients) was highlighted (M = .14 ml per day, STD = .08; t1,67 = 2.61, p = .02). Participants with low self-control used the lotion more when a present benefit was highlighted (M = .24 ml per day, STD = .11) than when only the future benefit was highlighted (M = .09 ml per day, STD = .06; t1,67 = 4.02, p < .001). Fig. 2 presents the average daily consumption of the facial sunscreen lotion as a function of self-control and time focus.

As in Study 1, we used participants’ reported enjoyment of the product as a covariate. As expected, the findings suggest that consumption was not affected by an enjoyable aspect, in this case (F1,66 = .74, NS). This is consistent with our reasoning that sunscreen is a product with no special enjoyment aspect. Hence, the impact of the present benefit on the consumption of the lotion among participants with low self-control cannot be attributed to enjoyment derived from the consumption experience and is attributable to the timing of the benefit.

4.2.1. Conclusions

Study 2 indicates further support for the benefit-congruency hypothesis by showing that emphasis on a virtue product’s essence constitutes a future-focused appeal. Thus, participants with high self-control were more responsive to such an appeal than to an appeal focusing on a present benefit, whereas participants with low self-control showed the opposite response. Unlike Study 1, all participants in Study 2 received the same product, differing only in the product’s description, which highlighted an existing present benefit in the present-focus condition but not in the future-focus condition. In addition, whereas we focused on a product that was unfamiliar to the participants in Study 1, Study 2 focused on a product of which participants had experience and knowledge and showed the same pattern of results.

Our findings indicate that consumers with high self-control consumed the product less when more benefits were offered (present and future) compared to when only a future benefit was highlighted. These findings may seem surprising. However, they are consistent with our reasoning that consumption is higher when the product message is congruent with individual goals (in this case, long-term focus for participants with high self-control). The added present benefit is incongruent with the future-orientation of participants with high self-control, resulting in a less attractive product, which leads to less consumption. Participants with high self-control were more likely to consume the product whose description emphasized only a future benefit because this product offered higher overall benefit-congruency.

4 There were no differences in enjoyment between the experimental conditions (t1,60 = .24; NS).
Combined together, Studies 1 and 2 provide consistent support for our hypotheses. Study 3 was designed to further evaluate the robustness of our findings by ruling out additional alternative explanations. First, in Studies 1 and 2, more information and more benefits were provided in the present-focus condition than in the future-focus condition. This is especially true for Study 2, in which participants were knowledgeable about the product, and the future benefit was presented clearly in both conditions. To rule this out as an alternative explanation, in Study 3 we provide more information in the future-focus condition.

Second, in Studies 1 and 2, the present benefits were described in a promotion-oriented manner. Additionally, in both studies, the present condition included a peripheral benefit. Thus, it could be argued that as compared with participants with high self-control, participants with low self-control are either more influenced by peripheral benefits or more sensitive to promotion information, especially in a prevention context as in the case of virtue products. To rule out these two alternative explanations, Study 3 was designed to relate the present-focus manipulation to the primary essence of the product, and we used a prevention-oriented description.

5. Study 3

In this study, we go one step further in testing the robustness of our theoretical model by showing that benefit-congruency affects not only actual consumption but also consumers’ willingness to pay. Additionally, in this study, we manipulated time focus by emphasizing different types of benefits, aiming to rule out the alternative explanations discussed above. In Studies 1 and 2 we conducted field experiments, thus providing ecological validity, whereas in Study 3 we conducted the experiment in a more controlled environment.

5.1. Method

5.1.1. Participants

Participants (n = 315; Mage = 36, 72% females) volunteered to complete an online survey and in return were included in a raffle for an Amazon.com gift certificate of $25.

5.1.2. Procedure

This study consisted of only one session. In that session, participants first completed the self-control measure, completed a filler task, and then read a description of facial sunscreen lotion. Finally, participants were asked to indicate how much they were willing to pay for the product. As in Study 2, the beginning of the product description, which explained the product’s essence and future benefit, was the same in both conditions:

“The product you are receiving is facial sunscreen lotion. This lotion will protect your face from sunburn and future skin damage. This lotion is suitable for everyday use, for all skin types, and is approved by the Health Ministry.”

However, the ending of the description varied across the two following experimental conditions.

The future-focus condition highlighted an additional attribute with a future benefit:

“This special sunscreen includes anti-wrinkle ingredients. These ingredients help prevent future skin damage such as wrinkles and pigmentation spots.”

The present-focus condition highlighted an existing attribute that was framed as a present benefit:

“This special sunscreen contains SPF 60! This extremely high SPF will protect your skin from UVB rays, which are short waves, thus preventing short-term sun damage such as sunburn.”

These manipulations were validated in a pre-test similar to that described in Study 1. We carried out a t-test against the scale’s middle score (of 3) to test the manipulation. Participants’ average rating of the SPF 60 attribute was significantly lower than the scale’s middle score ($M = 2.03$, $STD = 1.27$; $t_{(1,91)} = -7.3$, $p < .001$). The anti-wrinkle attribute received a significantly higher score than the scale’s middle score ($M = 3.85$, $STD = 1.11$; $t_{(1,80)} = 7.24$, $p < .001$). Thus, SPF 60 is perceived as a present benefit, whereas the anti-wrinkle attribute is seen as a future benefit.

Finally, participants were asked in an open-ended question to indicate the amount of money they would be willing to pay for the product.

5.2. Results

We tested our hypothesis with a hierarchical regression to predict participants’ willingness to pay. In the first model, participants’ self-control and the time-focus condition were entered as predictors. In the second model, the interaction of self-control and time focus was added as a predictor. Neither model was significant (Model 1: $F_{1,91} = 1.45$; both NS). However, results show that in the second model, time focus and the interaction were significant ($\beta_{timefocus} = -.62$; $\beta_{interaction} = .64$; both $p < .05$).

To further test our hypothesis, we carried out an ANOVA and planned contrasts. To that end, participants were designated as having high or low self-control based on median split. We conducted a 2(self-control: high vs. low) $\times$ 2(time focus: future vs. present benefit) ANOVA to predict willingness to pay. As hypothesized, the interaction between self-control and time focus was significant ($F_{1,132} = 7.17$, $p < .01$). No main effect was found for either self-control or time focus.

Fig. 3 presents the average price participants indicated as a function of self-control and time focus. Participants with low self-control were willing to pay more after reading a present-focused description ($M = \$20.19$, $STD = 7.48$) than after reading a future-focused description ($M = \$17.83$, $STD = 7.67$; $t_{(1,312)} = 2.07$, $p < .04$). Participants with high self-control were willing to pay more after reading a future-focused description ($M = \$20.76$, $STD = 8.31$) than after reading a present-focused description ($M = \$18.5$, $STD = 6.43$). This difference, however, was only marginally significant ($t_{(1,312)} = 1.74$, $p = .08$).

5.2.1. Conclusions

The results of Study 3 provide additional support for the benefit-congruency effect. A present-benefit focus affected participants with low self-control more than a future-benefit focus did, and the opposite effect occurred among participants with high self-control.
Whereas Studies 1 and 2 evaluated daily usage, which is a habitual behavior, Study 3 shows that the effect took place when participants were asked to indicate a price they were willing to pay for the product, a process that might call for calculation and deliberate thinking. In addition, Study 3 rules out the promotion focus, the significance of the benefit (peripheral or central), and the number of benefits as alternative explanations. The results of this study reproduced the pattern found in Studies 1 and 2, when the future-focus condition offered more benefits and was framed in a prevention-oriented manner.

6. General discussion

This research focused on virtue products, which are characterized by sub-optimal consumption. We suggested that the interaction between consumer self-control and product attributes can influence the consumption of such products. Among consumers with low self-control, a product description that added an attribute (Study 1) or highlighted an existing attribute (Studies 2 and 3) offering an immediate benefit resulted in more consumption compared with a product description that highlighted a future benefit only. Conversely, among participants with high self-control, highlighting an attribute that provides a future benefit (Studies 1 and 3) or merely mentioning the product’s essence (which entails a future benefit, or Study 2) resulted in more consumption compared with highlighting a present benefit. These findings are in line with the benefit-congruency theory and suggest that when product benefits are congruent with consumers’ time focus—whether future for consumers with high self-control or present for consumers with low self-control—overall responsiveness increases in terms of both actual consumption (Studies 1 and 2) and willingness to pay (Study 3).

Past studies on self-control have typically measured induced effects of situational cues on behaviors that call for self-control (e.g., Ariely & Wertenbroch, 2002; Fishbach & Trope, 2005; Kivetz & Simonson, 2002; Muraven & Baumeister, 2000). Our research extends the scope of examination by investigating the role of the interaction between a personality trait (i.e., dispositional self-control) and a situational factor (i.e., time-focused product description) in shaping actual behavior.

We show that highlighting a present- or a future-focused benefit can encourage consumption and that the effect depends on consumers’ dispositional self-control. This effect was observed for products with which consumers had either little (Study 1) experience or knowledge or a great deal (Studies 2 and 3) of experience and knowledge. In addition, the effect occurred regardless of whether more benefits were offered in the present-focus condition (Studies 1 and 2) or in the future-focus condition (Study 3), whether the present-focus condition offered a promotion-oriented, enjoyable benefit (Study 1) or a prevention-oriented, non-enjoyable benefit (Study 3), and whether the present benefit was peripheral to the product’s main purpose (Studies 1 and 2) or primary and part of the product’s main essence (Study 3).

Because we did not measure the consumption of a control group, we do not have a baseline estimate of consumption. Therefore, future studies might test whether a present-focused message is indeed less appealing than a future-focused message to consumers with high self-control, and whether it actually reduces consumers’ liking and usage of the product.

In the current research, dispositional self-control affected consumer responsiveness to different product features. It would be interesting to explore whether induced levels of self-control yield the same results. Thus, for example, future research could test whether consumers respond differently to product features when depleted (i.e., having few self-control resources) versus when not depleted (i.e., having sufficient resources for applying self-control). We argue that the ongoing consumption of virtue products is a complex behavior by which individuals express different aspects of self-control (i.e., “doing wrong” aspects and “not doing right” aspects), and as a result, dispositional self-control can provide insights into this consumption behavior.

In the current research, we measured participants’ usage after one time period. It would be interesting in future research to carry out a longitudinal study to learn about differences in usage patterns between high and low self-control segments. In addition, the environment in our studies was somewhat different from consumers’ usual consumption environment. For example, participants received the products for free. It would be interesting to test whether our findings still hold for purchasing decisions. We took one step in this direction in Study 3, in which we tested and reproduced the effect on consumer willingness to pay. Future research could provide more insight regarding the point of purchase phase. For example, past research has discussed the controversial issue of how distributing small gifts or handouts at the moment of decision-making affect the consumer’s behavior (e.g., Raghubir, 2004). We speculate that for virtue products, a one-time gift or handout given at the moment of purchase may influence the purchase decision but will not facilitate ongoing usage, especially in the segment with low self-control. One-time handouts and attributes that provide ongoing present benefits are not identical, but may be compatible. It is possible that changes in the purchase decision are influenced more by a one-time handout, whereas a change in ongoing consumption is influenced more by a present versus a future benefit.

Although this research focuses on virtue products, the proposed conceptualization of benefit-disposition congruency may apply to vice products as well. We theorized (and found) that highlighting a future benefit increases consumption of virtue products among consumers with high self-control, whereas highlighting a present benefit increases consumption among consumers with low self-control. A similar effect might apply to vice products, such that highlighting a future benefit of a vice product may increase consumption among consumers with high self-control. Research (Kivetz & Keinan, 2006) has shown that providing consumers with a long-term perspective can result in increased orientation toward vice products. The authors interpret these findings as showing that when thinking about the future, people fear they might regret not indulging enough and hence choose hedonic options in the present. This behavior might be even more prominent among consumers who have high self-control and are more likely to anticipate experiencing such regret in the future. Our findings suggest an additional process that might lead to the same outcome, wherein consumption of vice products increases not because people anticipate regret for not indulging but because the product highlights features that are congruent with consumers’ time.
orientations. An example for such a product is facial make-up (a vice product) containing anti-aging ingredients (a future benefit). In this sense, adding future benefits to vice products makes these products more “virtuous.” Future research could further explore this idea.

Our findings have practical implications for policy-makers and profit-maximizing firms in that adding a seemingly small and secondary present benefit to “virtue” products can enhance consumption among consumers with low self-control. Such present benefits might seem negligible in comparison to the essence of the virtue product; however, as long as the present benefit offsets the cost, counters it, or even serves as a mental excuse for enduring the present cost, it has the potential to promote consumption among consumers with low self-control. Note that adding such benefits does not necessarily entail substantial increases in manufacturing costs. For example, according to the manufacturer, the cost of adding moisturizing ingredients to the facial sunscreen lotion used in Study 2 was negligible (3%), yet emphasizing these ingredients to consumers significantly increased the actual consumption in a specific segment (31.25%). These numbers demonstrate how small changes sometimes make a big difference. However, it is also important to control for the counter effect that such an emphasis may create for other segments. Our findings suggest that a present-benefit focus might not be as appealing to consumers with high self-control. It would be worthwhile to explore whether, among consumers with low self-control, a future-focused message is simply less favorable than a present-focused message or whether a future-focused message actually creates a negative response of reduced consumption, and whether the opposite effect occurs among consumers with high self-control. This research takes one step toward a better understanding of the problematic consumption of virtue products. Virtue products are an important part of our everyday lives. Although consumers acknowledge these products’ important benefits, they tend to display suboptimal consumption patterns. Based on the benefit-congruency rationale, we suggest that matching the product’s highlighted attributes with the consumer’s self-control (i.e., the sensitivity of consumers with low self-control to present benefits and the sensitivity of consumers with high self-control to future benefits) may increase consumers’ ability to establish the ongoing consumption of such products.

Acknowledgments

This research was supported by grants from the K-mart International Center of Marketing and Retailing, the Davidson Center and by the Reccanati Fund at the School of Business Administration at the Hebrew University of Jerusalem. The authors are grateful for the valuable comments received from the editor, area editor and two reviewers. We wish to thank Dan Ariely, Gita V. Johar, Liat Levontin, Dudi Mazursky, Oded Netzer, Shaul Oreg, Yael Steinhart, and Gal Zauberman for their suggestions and support during different stages of the research.

Appendix A. The Dispositional Self-Control scale (DSC)

Adapted from: Ein-Gar, Goldenberg and Sagiv (2008)

I usually succeed in overcoming temptations.
Usually, when something tempts me, I manage to withstand it.
Even when something exciting happens to me, I do not get carried away by my feelings or act without thinking.
Even when stressed, most of the decisions I make are considered and calculated.
I rarely act impulsively.
I am able to work effectively toward long-term goals, while resisting temptations along the way.
People can trust me to stay on schedule even if I am busy and under a lot of pressure.

It is important for me to finish all of my tasks on time, even if I do not feel like doing them.
I never delay work that needs to be done, even if I am busy.
I tend to finish assignments right away, even if they are unpleasant.

* I do many things on the spur of the moment.
* People say I often make up my mind without thinking things through.
* I often act without thinking through all of the alternatives.
* I often make spontaneous and rather hasty decisions.
* I tend to postpone completing unpleasant tasks.
* When I need to run errands, I usually put them off until the last minute.
* I sometimes postpone tasks that I have to do until it is almost too late.

Items marked * are reverse coded.

Appendix B. A table of correlations between the DSC and other, related constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Scale source</th>
<th>Alpha</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time orientation constructs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consideration of future consequences</td>
<td>Strathman et al. (1994)</td>
<td>.54**</td>
<td>3</td>
</tr>
<tr>
<td>Elaboration of potential outcomes</td>
<td>Nenkov et al. (2008)</td>
<td>.33**</td>
<td>3</td>
</tr>
<tr>
<td>General</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time perspective</td>
<td>Zimbardo and Boyd (1999)</td>
<td>.61**</td>
<td>3</td>
</tr>
<tr>
<td>Future</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present hedonic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present fatalistic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personality constructs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-control</td>
<td>Tangney et al. (2004)</td>
<td>.68**</td>
<td>2</td>
</tr>
<tr>
<td>Five factor model (Big 5)</td>
<td>Saucier (1994)</td>
<td>.60</td>
<td>2</td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procrastination</td>
<td>Lay (1986)</td>
<td>.71**</td>
<td>2</td>
</tr>
<tr>
<td>Impulsiveness (UPPS)</td>
<td>Whiteside and Lynam (2001)</td>
<td>.70**</td>
<td>2</td>
</tr>
<tr>
<td>(short version)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premeditation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urgency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensation seeking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perseverance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral constructs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buying impulsiveness</td>
<td>Rook and Fisher (1995)</td>
<td>(-.41**)</td>
<td>1</td>
</tr>
<tr>
<td>Frugality</td>
<td>Lastovicka et al. (1999)</td>
<td>(.35**)</td>
<td>1</td>
</tr>
<tr>
<td>Driving behavior (errors, violations and lapses)</td>
<td>Westerman and Haigney (2000)</td>
<td>(-.21**)</td>
<td>2</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>Saunders et al. (1993)</td>
<td>(-.19**)</td>
<td>2</td>
</tr>
<tr>
<td>(AUDIT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive behavior</td>
<td>Driscoll, Campbell, and Muncer (2005) (short EXPAGE)</td>
<td>(-.29**)</td>
<td>2</td>
</tr>
</tbody>
</table>

Notes:

Time orientation constructs: These three measures indicate the extent to which the respondent engages in thoughts that are future or present in nature. Future thoughts are positively correlated with
self-control, whereas present thoughts are negatively correlated with self-control. **Personality constructs:** These four scales measure general personality-aspects, a self-control measure (different from the DSC), the five-factor model scale, general impulsiveness and general procrastination scales. DSC is positively correlated with another self-control measure; it is also positively correlated with aspects of the five-factor model relating to being responsible, doing the right thing and persisting in a task (i.e., Conscientiousness, the Perseverance facet of UPPS) yet negatively correlated with impulsivity and hasty decision-making (i.e., the Urgency aspect of UPPS) and procrastination. **Behavioral constructs:** These five scales measure behaviors that are strongly related to self-control. Self-control is negatively correlated to harmful behaviors (i.e., impulsive buying, risky driving, alcohol consumption and aggression) and positively correlated to beneficial behaviors (i.e., frugality). **Data sources:**


3. **Data reported in the pilot study.**

**References**


Modeling coexisting business scenarios with time-series panel data: A dynamics-based segmentation approach

Catarina Sismeiro a,⁎, Natalie Mizik b, Randolph E. Bucklin c

a Imperial College Business School, Imperial College, London, United Kingdom
b Kenan-Flagler School of Business, University of North Carolina, Chapel Hill, United States
c UCLA Anderson School of Management, United States

A R T I C L E   I N F O

Article history:
First received in 26, November 2009 and was under review for 4½ months
Available online 17 March 2012

Area Editor: Harald J. Van Heerde

Keywords:
Sales force
Segmentation
Marketing-mix effectiveness
Econometric methods
Time-series modeling

A B S T R A C T

At a given point in time, individual consumers may be in different stages of the product adoption or consumption cycle. As a result, different types of behavioral patterns may coexist within a single product market. Existing segmentation approaches typically do not address long-term dynamics in customer response and do not adequately capture this phenomenon. We develop an approach for modeling the coexistence of multiple dynamic behavioral patterns (business scenarios) within a single product market. We apply this approach to physician panel data on drug prescriptions and direct-to-physician promotions. We find markedly different responses across physician segments. For firms that track customer-level marketing activity and sales over time, market segmentation based on dynamic scenarios can provide a new tool for efficient targeting. The proposed approach is straightforward to implement and is scalable to very large samples and continuous testing.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The concept of business scenarios in the context of time-series modeling was first discussed by Dekimpe and Hanssens (1999). In their study, the authors presented four possible alternative scenarios for a given market (Fig. 1): “business as usual” (in which both performance and marketing variables are stationary), “escalation” (in which only the marketing variables are evolving), “hysteresis” (in which only the performance measure is evolving), and “evolving business practice” (with evolving performance and marketing variables). Depending which scenario is detected in the market under analysis, alternative formulations for vector autoregressive models would be appropriate, which, when estimated, lead to different strategic conclusions.

In this paper, we propose that, due to differences in customers and firm behavior, multiple business scenarios may coexist within a single product market. Just as products, industries, and markets may be at different stages of their life cycles, individual consumers may be at different stages of their consumption life cycles and may be subject to targeted marketing actions that differ significantly (in terms of intensity and scope) from those aimed at other consumers. Several prominent theories provide support for such a phenomenon. For example, new product diffusion and adoption theories rely on the existence of distinct consumer segments that learn about and adopt the product at different points in time (Bass, 1969; Rogers, 2003). Firms have access to a broad set of tools to customize marketing actions at the individual level in terms of intensity, message, and even the medium used.

Despite the likely coexistence of multiple business scenarios in a market, the modeling of distinct dynamic consumer responses and dynamic firm behavior has not been incorporated into existing segmentation methodologies. With this paper, we develop and illustrate an approach that uses time-series panel data to investigate whether different dynamic business scenarios might exist concurrently across a firm’s customer base and what the data imply strategically. If these business scenarios do coexist, some customers (or segments) might present the firm with “business as usual,” while others present the perils of “escalation” or the opportunities of “hysteresis” or “coevolution” scenarios. To study this phenomenon, we relax the common, though typically implicit, assumption that the time-series properties of the data are uniform across panelists.

Our approach allows us to effectively identify distinct dynamic patterns at an individual level and to investigate the differences in response dynamics across customers. It consists of two steps. In the first step, we test for the order of integration in the data using unit-root tests at the disaggregate level. We conduct two sets of unit-root tests, one for the outcome variable and one for the marketing covariates. Then, using the test results, we assign individuals to one of four groups. In the second step, we specify an appropriate panel vector autoregressive model (PVAR) for each group, and we estimate...
separate PVAR models to investigate potential differences in response dynamics across groups.

To test our approach, we use individual physician-level time-series panel data on prescribing and direct-to-physician promotion (DTP) from a pharmaceutical prescription drug market. Differences across doctors in age, experience, practice size and type, risk aversion and adoption timing could lead to differences in the level of time-series integration for individual prescribing (i.e., prescription data series might be evolving or stationary depending on these factors). Different levels of marketing activity targeted by pharmaceutical companies across physicians can also give rise to evolution versus stationarity of DTP activity at the individual level. Thus, different dynamic business scenarios might be present across physicians.

Empirically, we find that multiple dynamic scenarios coexist within a single drug market, each with markedly different response magnitudes and patterns. We assess the performance of the proposed method relative to alternative segmentation approaches and present evidence that common segmentation variables used by the industry fail to detect meaningful response differences across groups. Tests using cross-sectional and longitudinal holdout samples support the superiority of the proposed dynamic business scenario-based segmentation, which produces not only better in-sample fit but also better out-of-sample fit. Hence, we believe that the proposed segmentation approach provides a useful new tool for enhancing the productivity of marketing resources through better targeting at the individual level.

2. Motivation

In markets where sales and/or marketing activity may be evolving, econometricians have emphasized the importance of handling potential nonstationarity in time-series data. These “persistence modeling” methods (1) employ unit-root tests to ascertain the stationary versus evolving nature of the data and (2) estimate appropriate vector autoregressive (VAR) models given the integration level of the data to assess the market dynamics and long-run effects. The application of VAR-based persistence models to marketing data has yielded key insights about dynamic response in a series of studies conducted on aggregate-level data (e.g., Dekimpe & Hanssens, 1995, 1999; Bronnenberg, Mahajan, & Vanhonacker, 2000; Pauwels, Hanssens, & Siddarth, 2002; Horváth & Franses, 2003). Dekimpe and Hanssens (1999), for example, have demonstrated how time-series methods applied to aggregate historical data can identify different dynamic business scenarios: business as usual, escalation, hysteresis, and evolving business practice (or coevolution). Implications for return on marketing spending, marketing strategy, and profitability can differ dramatically across these scenarios. The possibility that different dynamic business scenarios may coexist within a single market, however, has not been examined.

Another branch of marketing science has extensively studied disaggregate-level data in consumer panels, based largely on UPC scanner data for packaged goods. This stream has emphasized the study of short-run response to marketing activity across individual
decision makers and the potential for segmentation in individual-level response (e.g., Kamakura & Russell, 1989; Bucklin, Gupta, & Siddarth, 1998; Kamakura & Wedel, 2000). Despite the emphasis on short-run response in individual-level studies, a number of authors working with individual scanner panel data have also attempted to examine dynamic effects or long-run changes in response parameters (e.g., Papati & Krishnamurthi, 1996; Mela, Gupta, & Lehmann, 1997). These studies have found empirical evidence for long-run changes in the nature of consumer response, but they have not investigated the time-series properties of the underlying series.

Pauwels et al. (2002) applied VAR-based persistence modeling to UPC scanner data and found that evolution (i.e., the presence of unit roots) was uncommon in aggregate-level scanner data, a phenomenon they attributed to the maturity of most products studied. The link between product maturity and a business scenario may also explain the results of Bronnenberg et al. (2000). In their study of new product sales and distribution, the authors analyzed aggregate-level data for the launch phase of the ready-to-drink tea category and found empirical patterns consistent with coevolution for sales and distribution.

We contend that just as different products might be at different stages of the life cycle, so might individual consumers or markets. Indeed, most theories of new product diffusion and new product adoption rely on the presence of consumer segments that learn about and adopt the product at different times (Bass, 1969; Rogers, 2003). Dekimpe, Parker, and Sarvary (2000), for example, showed that technological products may be at different points of the life cycle across countries depending both on the time of the initial adoption and the level of current adoption. Thus, product sales can have different dynamics across countries. Parsons (1975), Lieberman (1987), and Osinga, Leeflang, and Wieringa (2010) reported that marketing-mix responsiveness can vary across the stages of the life cycle. However, when using aggregate-level time-series data for dynamic scenario and response analysis, the investigator is required to treat all the firm’s customers within a single scenario classification, which is a significant limitation. If analysis is restricted to aggregate-level data, there is no opportunity to observe heterogeneity in the dynamics of firm and customer behavior.

Pauwels et al. (2005) explicitly called for more research into the problems associated with cross-sectional heterogeneity in VAR-based models. Lim, Currim, and Andrews (2005) presented a VAR-based analysis of scanner panel data that took a step in this direction. They first assigned panelists to segments determined a priori based on panelists’ brand loyalty (loyals versus switchers) and category usage rates (heavy versus light users). The authors then aggregated data to the segment level. The estimated segment-specific VARs revealed differences in response and adjustment period across segments. Perhaps due to the mature product categories involved, unit roots showed the aggregate time series to be almost universally stationary (similar to the results of Pauwels et al., 2002). This constrained the analysis of response dynamics to differences within the “business as usual” scenario.

In addition to preventing the identification of coexisting business scenarios, estimating aggregate-level VAR models presents another limitation: the difficulty of handling the typically large number of endogenous variables and/or lag effects. A lack of degrees of freedom has led researchers to impose a variety of restrictions on the structure of VAR models to limit the number of parameters. Panel data can be used to increase the degrees of freedom under appropriate pooling assumptions (e.g., fixed effects, Horváth & Wieringa, 2008). Thus, PVAR models offer the advantage of imposing fewer restrictions on lag structure and explanatory variables.

3. Modeling approach

Granger and Newbold (1974) first highlighted the “spurious regression” problem, in which unrelated unit-root series appear related with a very high probability if conventional estimation methods are employed. This finding stimulated interest in appropriately modeling time-series dynamics, and it led to a multitude of research studies testing dynamic patterns and addressing the advantages and disadvantages of various modeling approaches to control for the dynamic properties of the series (e.g., Plosser & Schwert, 1977; Stock & Watson, 1988; Nelson & Kang, 1984). More recently, researchers in econometrics have begun scrutinizing the methods involved in the analysis of panel data and examining how these data can be used to improve the study of time-series dynamics (e.g., Baltagi & Kao, 2000). New methodological advancements in panel data analysis provide a foundation for researchers in marketing to integrate the two streams of research: (1) the study of segmentation with panel data and (2) the study of long-run dynamic effects with (potentially) evolving time-series data.

Our approach builds on these advancements. Methodologically, our approach is based on unit-root tests and PVAR modeling techniques. The proposed method is applicable to data in which different dynamic business scenarios are thought to coexist. It is not limited to individual-level data but can be applied where cross-sectional units are, for example, markets or geographical regions.

3.1. The proposed procedure

Our approach proceeds as follows. First, we conduct unit-root tests for the outcome and marketing variables for each panelist at the individual level. Based on the unit-root test results, we classify each panelist into one of the four dynamic business scenarios: evolving business practice (coevolution), hysteresis, escalation, or business as usual. We then specify the PVAR models for each group such that variables enter either in levels (if they were deemed stationary) or in differences (if evolving), depending on the unit-root test results (e.g., Campbell & Perron, 1991). Fig. 1 depicts the four groups and the corresponding stylized representations of the impulse response functions (IRFs) of marketing activity and sales response.

In Fig. 1, Group 1 is the “evolving business practice” scenario. Here, all variables are evolving, and all enter the PVAR model in first differences. Group 4 is the “business as usual” scenario: all variables are stationary and enter the PVAR model in levels. Group 2 is the “hysteresis” scenario (evolving in sales, stationary in marketing activity), and Group 3 is the “escalation” scenario (stationary in sales, evolving in marketing activity). Both Group 2 and Group 3 are modeled with mixed PVARs. The outcome variables (e.g., sales) enter the model in levels for Group 3 and in differences for Group 2. Marketing activity variables enter in levels for Group 2 and in differences for Group 3. To address potential endogeneity in marketing activity and response, we treat all variables for all groups as endogenous.

Next, we ascertain the presence of cross-sectional heterogeneity in the data. We conduct Hausman (1978) specification tests for the presence of individual-specific fixed effects. When necessary, the PVAR model is specified to incorporate fixed effects. All PVAR models include time-specific indicator variables to control for unobserved time-specific effects and to ensure that the models are robust to structural changes. The time-specific intercepts also serve as equilibrium-correction terms (e.g., Theil, 1961, Clements & Hendry, 1999).

We test each PVAR model for the appropriate number of lags using the Schwarz criterion (SC). After estimation of the appropriate PVAR models for each scenario, we obtain IRFs and gauge the effect of marketing effort on the performance variable. To provide a benchmark for our results, we estimate a pooled PVAR model across all panelists, in levels and in differences. (Later, we discuss other benchmark alternatives, including a pooled Bayesian random-effects formulation.)

---

1 A number of different unit-root tests has been advanced in the time series literature (Maddala, 1992). Our segmentation approach relies on the outcomes of a unit-root test but does not depend on the specific unit-root test employed.
Finally, we assess the performance of the dynamic scenario-based segmentation in cross-sectional and longitudinal holdout samples.

4. Research on pharmaceutical marketing dynamics

Recent studies on pharmaceutical marketing have used different data as well as different methods to investigate the effects of marketing activity on physicians’ prescribing decisions (see Kremer, Bijmolt, Leeﬂang, & Wieringa, 2008; Leeﬂang & Wieringa, 2010; Manchanda et al., 2005 for comprehensive reviews). Although some of these studies incorporate dynamics, to some extent, in their modeling, we seldom see the simultaneous investigation of heterogeneity and the time-series properties of the data. For example, using aggregate data, Dekimpe and Hanssens (1999) employed VAR models to investigate the long-run effects of changes in marketing activity and to draw implications for profitability. Narayan, Manchanda, and Chintagunta (2005) also used aggregate-level data on new prescriptions to investigate temporal differences in the role of detailing and other marketing expenditures. These studies did not deal with potential heterogeneity in the data.

Studies conducted at the disaggregate level typically investigate the role of heterogeneity in response across physicians. For example, Manchanda and Chintagunta (2004) used a Poisson model to study the effect of marketing activity on the number of prescriptions written by physicians on a quarterly basis. Manchanda, Rossi, and Chintagunta (2004) extended the analysis to incorporate potential endogeneity in prescribing and marketing activity but did not investigate the time-series properties of the data. Mizik and Jacobson (2004) employed fixed-effects instrumental variable estimation to address both heterogeneity and endogeneity in physician response to direct-to-physician marketing activities. Finally, Narayanan and Manchanda (2005), Janakiraman, Dutta, Sismeiro, and Stern (2008), and Janakiraman, Sismeiro, and Dutta (2009) used physician panel data to estimate an individual-level model of prescription choice within a therapeutic class. Although these studies did not investigate the evolving nature of the corresponding prescription and marketing mix series, these authors find significant differences across physicians in response to pharmaceutical detailing, in choice state-dependence, and in learning rates.

In sum, existing disaggregate-level pharmaceutical models have accounted for some dynamic effects (e.g., carryover effects, Bayesian learning, and lagged dependent variables), but they have not provided a full analysis of the time-series properties of the data. In addition, to the best of our knowledge, previous research has not investigated the time-series properties of the data at the individual physician level and has not yet attempted to model the coexistence of multiple dynamic business scenarios.

5. Empirical application

5.1. Data

Our data come from an anonymous panel of 5000 U.S. physicians tracked monthly over a period of approximately 2 years (October 2001–August 2003) for a single drug. Due to conﬁdentiality requirements, the identity of the drug and the company are masked. For each physician in the panel, the data set includes general demographic information (age, gender, and year of graduation), the number of new prescriptions written, and the number of sales calls (details) and samples received each month.

To focus on decision makers who were at least minimally involved with the product, we have excluded physicians with fewer than two prescriptions and with less than one sales call per year. To avoid the undue inﬂuence of outliers, we also removed physicians with extremely high levels of prescribing activity (top one-half of 1%). According to the collaborating ﬁrm, the activity assigned to these physicians might reﬂect large group practices or hospitals rather than representing individual doctors. Our ﬁnal study sample consisted of 3942 physicians (79% of the original sample). We randomly assigned two-thirds (2628) of physicians to the estimation sample and the remaining one-third (1314) to a holdout sample.

Table 1 presents a summary of descriptive statistics for the ﬁnal sample of 3942 physicians, including both estimation and holdout samples. The mean number of new prescriptions written per month was 3.33, the mean number of details was 3.09, and the mean number of samples received was 15.71. The range statistics show that our data sample includes a wide cross-section of physicians, with panelists varying signiﬁcantly in experience, level of prescribing, and attention from pharmaceutical sales representatives. The variable of ‘years since graduation’ is based on the number of years between the physician’s graduation year and the ﬁrst year of our sample period. Physicians were also assigned a national decile value by the company based on their past prescribing volume. A decile assignment of 7, for example, indicates that the doctor is in the 70th percentile for new prescription volume. The average decile value was 6.17. According to the collaborating ﬁrm, the characteristics of our sample are representative of the general population of physicians prescribing this drug.

Fig. 2 graphs the aggregate time series of key variables over the study period. An examination of the time-series patterns in Fig. 2 suggests possible evolution in the aggregate series for prescriptions, detailing, and sampling. Augmented Dickey–Fuller (ADF) tests performed on the aggregate data shown in Fig. 2 are all close to the 95% critical value, with prescriptions marginally evolving and stationary marketing activity.

5.2. Unit-root tests

As the ﬁrst step of our approach, we conducted individual-level ADF unit-root tests on all 3942 physicians in the sample. We tested prescription and marketing activity variables (details and samples).3 We relied on the ADF test to classify our panelists into the evolving or stationary conditions for several reasons. First, previous research has shown that the data-generating process for drug prescriptions has a high autoregressive order (e.g., Mizik & Jacobson, 2004 report six signiﬁcant lags for the ﬁrst-differenced prescriptions series for all three drugs in their study), and our data exhibit signiﬁcant lag structure. The ADF test has been shown to perform well in Monte Carlo studies under these conditions.

Table 1

<table>
<thead>
<tr>
<th>Summary of descriptive statistics.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td>New prescriptions</td>
</tr>
<tr>
<td>Details</td>
</tr>
<tr>
<td>Samples</td>
</tr>
<tr>
<td>Physician national deciles</td>
</tr>
<tr>
<td>Years since graduation</td>
</tr>
</tbody>
</table>

Legend: Descriptive statistics for prescriptions, details, and samples reflect monthly values per physician and are computed across the 3942 physicians in the study sample. Physicians were assigned a national decile by the company based on their prescribing volume. A decile assignment of 7, for example, indicates that the doctor is in the 70th percentile for new prescription volume. The variable “years since graduation” uses 2001, the first year of our sample, as the reference year.

2 It is important to note that the prescription data at the individual level are not count data. Because prescription size (i.e., the number of pills) varies signiﬁcantly across prescriptions, prescriptions are tracked and recorded as the number of “standardized” prescription units issued in a given month. In addition, caution is needed in the interpretation and use of averages for evolving variables because population means are not deﬁned for l (1) processes (evolving series).

3 As a sensitivity test, we replicated our analysis using the KPPSS test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). We report the results in the Sensitivity analyses section.
conditions because it increases in power as the number of lags increases (Harris, 1992; Haug, 1996). Second, the null hypothesis in ADF tests is the presence of a unit root, and the consequences of modeling a stationary process as having a unit root are less severe than the opposite. The negative consequences of ignoring unit roots involve false inferences and incorrect conclusions, whereas treating a stationary process as a unit root (and potentially overdifferencing the data) only leads to a decrease in the efficiency of the estimates (Plosser & Schwert, 1977). Thus, the negative consequences of underdifferencing significantly outweigh the consequences of overdifferencing.

Finally, we chose the ADF because it has been shown in simulation studies to have very close, albeit slightly more conservative, test statistics for count data (our detailing and sampling data are integer counts) and to perform well in short time-series data (i.e., small sample sizes similar to our data). Hellström (2001), for example, conducted a large-scale simulation study and reported small sample distributions for the ADF test statistic for count data. He specifically examined short time series (T = 25) and derived approximation equations to enable calculation of critical values for any values of time-series length (T) and drift. Given our data characteristics, the Hellström (2001) corrections suggest very small deviations from the classic ADF statistic.

The ADF tests we performed for each physician included an intercept and a trend. The number of lags used for each physician was selected based on the SC. Given the test results, we classified each physician into one of four groups. Panelists were classified as evolving in the number of new prescriptions if we failed to reject the null hypothesis of a unit root at the .05 significance level. Otherwise, they were classified as stationary. Evolving prescription behavior was found in the physician-level data for 22% of the doctors, whereas 78% was stationary. For the marketing activity variables, we classified panelists as evolving when, for at least one of the marketing variables (detailing or sampling), we failed to reject the null hypothesis of a unit root (in such cases, all marketing variables are modeled in differences). Marketing activity was classified as evolving for 52% of the doctors and as stationary for 48%. In addition, we found that each of the four business scenario groups was populated by a large, albeit unequal, number of physicians. Group 1 (coevolution) had 12%, Group 2 (hysteresis) had 10%, Group 3 (escalation) had 40%, and Group 4 (business as usual) had 38% of the physicians.

Table 2 presents the sample means for the characteristics of the resulting four business scenario groups. We see no significant differences in the number of monthly new prescriptions or experience (as measured by years since graduation). Table 2 also presents the national decile rankings for physicians assigned to each of the four groups. Pharmaceutical firms use national decile rankings of prescription volume as a key component in targeting marketing efforts to physicians (Manchanda et al., 2004). Table 3 shows the percentage of each group's membership corresponding to the seven different decile ratings available (numbered 4 through 10). For example, in Group 1, about 21.2% of the doctors fall into decile 4, whereas 2.7% fall into decile 10. We see few differences in decile rankings across the four groups.

The results from Tables 2 and 3 suggest that readily available descriptive information does not predict the classifications produced by the unit-root tests. If the segmentation based on the unit-root classifications carries through to meaningful differences in response dynamics, it could open new opportunities for segmentation and targeting. For example, at first glance, Group 2 (hysteresis) would seem to be the most interesting and profitable for marketers to identify and target. Indeed, it may only be a temporary increase in marketing activity to produce a permanent increase in new prescription activity. We note, however, that the unit-root classifications (evolving outcomes, stationary marketing) simply make it possible for hysteresis to occur but do not ensure that it occurs (evolution in performance variables might not be the result of specific marketing actions but might be derived from other phenomena, such as learning from usage).

In contrast, Group 3 panelists may be the least attractive because the scenario involves permanent increases in marketing effort with only temporary increases in prescribing. Incidentally, Group 3 is the largest group in our data sample. Its size might partly reflect the recent intensely competitive environment in the pharmaceutical industry. This phenomenon has been referred to as the “PSR arms race,” in which firms field large numbers of new pharmaceutical sales representatives (PSRs).

5 In previous studies using VARX specifications in the presence of evolving series, researchers excluded the cross-sections with non-stationary series from their analyses. For example, Horváth, Leeflang, Wieringa, and Wittink (2005) excluded two stores showing evolving series for some of the variables and used the 24 remaining stores with stationary series in their fixed-effects VARX model.

5 This approach is parsimonious in that it holds the number of resulting segments to four. If desired, this approach can be extended to all possible combinations of evolving versus nonevolving behavior for all variables. In our sample, there are approximately 2047 physicians with evolving marketing mix (those in Group 1 and Group 3). For 1174 of these physicians (i.e., 57%), only detailing is evolving, for 427 physicians (21%) both marketing variables are evolving, and for the remaining 446 physician (22%), the only evolving marketing variable is samples. A very similar pattern was observed within Group 1 and Group 3.
dropped the group subscript). In Group 1 (coevolution), all variables are evolving, and therefore all series are specified in differences. In Group 4 (business as usual), all series are stationary, and thus all equations are specified in levels. Groups 2 and 3 are mixed PVAR models with evolving series entering in differences and stationary entering in levels.

Group 1 (coevolution)⁶:

\[
\begin{align*}
\frac{\Delta Q_{it}}{\Delta S_{it}} &= \begin{bmatrix} \alpha_{Q0} \\ \alpha_{S0} \\ \alpha_{QS} \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \beta_{11}^j \\ \beta_{21}^j \\ \beta_{31}^j \end{bmatrix} \times \frac{\Delta Q_{i,t-j}}{\Delta S_{i,t-j}} \\
&+ \left[ \begin{array}{c} \sum_{t=1}^{T-1} \delta_{Qt} \\ \sum_{t=1}^{T-1} \delta_{St} \end{array} \right] \times \Delta \text{Time}(t) + \begin{bmatrix} \varepsilon_{Qit} \\ \varepsilon_{Sit} \end{bmatrix},
\end{align*}
\]

Group 2 (hysteresis):

\[
\begin{align*}
\frac{\Delta Q_{it}}{\Delta S_{it}} &= \begin{bmatrix} \alpha_{Q0} \\ \alpha_{S0} \\ \alpha_{QS} \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \beta_{11}^j \\ \beta_{21}^j \\ \beta_{31}^j \end{bmatrix} \times \frac{\Delta Q_{i,t-j}}{\Delta S_{i,t-j}} \\
&+ \left[ \begin{array}{c} \sum_{t=1}^{T-1} \delta_{Qt} \times \Delta \text{Time}(t) \\ \sum_{t=1}^{T-1} \delta_{Qt} \times \text{Time}(t) \\ \sum_{t=1}^{T-1} \delta_{St} \times \text{Time}(t) \end{array} \right] + \begin{bmatrix} \varepsilon_{Qit} \\ \varepsilon_{Sit} \end{bmatrix},
\end{align*}
\]

Group 3 (escalation):

\[
\begin{align*}
\frac{Q_{i,t}}{\Delta S_{i,t}} &= \begin{bmatrix} \alpha_{Q0} \\ \alpha_{S0} \\ \alpha_{QS} \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \beta_{11}^j \\ \beta_{21}^j \\ \beta_{31}^j \end{bmatrix} \times \frac{Q_{i,t-j}}{\Delta S_{i,t-j}} \\
&+ \left[ \begin{array}{c} \sum_{t=1}^{T-1} \delta_{Qt} \times \text{Time}(t) \\ \sum_{t=1}^{T-1} \delta_{Qt} \times \Delta \text{Time}(t) \\ \sum_{t=1}^{T-1} \delta_{St} \times \Delta \text{Time}(t) \end{array} \right] + \begin{bmatrix} \varepsilon_{Qit} \\ \varepsilon_{Sit} \end{bmatrix},
\end{align*}
\]

In the above equations, \( Q_{i,t} \) denotes the number of new prescriptions by physician \( i \) in month \( t \), and \( D_{i,t} \) and \( S_{i,t} \) denote, respectively, the number of details and samples received by physician \( i \) in month \( t \). \( \Delta \) denotes the difference operator. The \( p \) matrices of the parameters \( \beta \) depict the effects of past prescriptions, detailing, and sampling. The number of lags, \( p \), in each group is selected to minimize the SC in the PVAR estimation (in each group, SC might lead to the use of a different number of lags). We use Hausman specification tests to assess the presence of fixed effects in our level equations. We denote individual-specific intercepts as \( \alpha \) and uniform intercepts as \( \alpha_0 \). The subscripts \( Q, D, \) and \( S \) identify the intercepts for prescriptions, detailing, and samples, respectively. \( \varepsilon_{Qit}, \varepsilon_{Dit}, \varepsilon_{Sit} \) are the error terms.

To model the time-specific effects, we adopt a time indicator specification commonly employed in the time-series models of regime changes. Our indicator variable \( \text{Time}(t) \) is equal to zero before time period \( t \) and unity from time \( t \) on. In the difference equations, this time indicator variable is differenced (\( \Delta \text{Time}(t) \)) and takes the familiar form of a standard time-period dummy variable equal to 1 if the time period is \( t \) and zero otherwise. This specification of time indicators accommodates a very general structure of exogenous shocks with various dynamic effects and allows for robust estimation without prior knowledge of where structural breaks might be occurring (e.g., Clements & Hendry, 1999).

---

**Table 2**

Summary of descriptive statistics by group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Description</th>
<th>New Prescriptions</th>
<th>Marketing</th>
<th>Number of New Prescriptions</th>
<th>Details</th>
<th>Samples</th>
<th>National Deciles</th>
<th>Years since graduation</th>
<th>Number of Physicians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Evolving business practice</td>
<td>Evolving</td>
<td>Evolving</td>
<td>3.62</td>
<td>3.02</td>
<td>15.30</td>
<td>6.18</td>
<td>19.15</td>
<td>476</td>
</tr>
<tr>
<td>Group 2</td>
<td>Hysteresis</td>
<td>Evolving</td>
<td>Stationary</td>
<td>3.66</td>
<td>2.92</td>
<td>15.69</td>
<td>6.24</td>
<td>19.75</td>
<td>375</td>
</tr>
<tr>
<td>Group 3</td>
<td>Escalation</td>
<td>Stationary</td>
<td>Evolving</td>
<td>3.27</td>
<td>3.08</td>
<td>15.33</td>
<td>6.16</td>
<td>20.26</td>
<td>1571</td>
</tr>
<tr>
<td>Group 4</td>
<td>Business as usual</td>
<td>Stationary</td>
<td>Stationary</td>
<td>3.22</td>
<td>3.16</td>
<td>16.23</td>
<td>6.17</td>
<td>20.36</td>
<td>1520</td>
</tr>
</tbody>
</table>

---

**Table 3**

Percentage of physicians by group and decile.

<table>
<thead>
<tr>
<th>Decile</th>
<th>Group 1 (%)</th>
<th>Group 2 (%)</th>
<th>Group 3 (%)</th>
<th>Group 4 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>21.2</td>
<td>24.0</td>
<td>23.7</td>
<td>24.1</td>
</tr>
<tr>
<td>5</td>
<td>18.3</td>
<td>16.0</td>
<td>17.2</td>
<td>18.0</td>
</tr>
<tr>
<td>6</td>
<td>18.7</td>
<td>17.1</td>
<td>19.0</td>
<td>18.1</td>
</tr>
<tr>
<td>7</td>
<td>18.1</td>
<td>15.5</td>
<td>15.7</td>
<td>13.8</td>
</tr>
<tr>
<td>8</td>
<td>13.2</td>
<td>13.1</td>
<td>12.0</td>
<td>12.4</td>
</tr>
<tr>
<td>9</td>
<td>7.8</td>
<td>11.2</td>
<td>8.8</td>
<td>9.7</td>
</tr>
<tr>
<td>10</td>
<td>2.7</td>
<td>3.2</td>
<td>3.6</td>
<td>3.8</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

---

⁶ We tested for and did not find cointegration in Group 1 in our data.
and, consistent with prior research (Mizik & Jacobson, 2004), detected the presence of significant effects. The summary of the estimated models provides a comparison of the specification and parameter estimates across different groups.

Table 4 summarizes the estimated models.

<table>
<thead>
<tr>
<th>Variable specification</th>
<th>Period effects</th>
<th>Cross-sectional fixed effects</th>
<th>Number of lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prescriptions</td>
<td>Marketing</td>
<td>Prescriptions</td>
<td>Marketing</td>
</tr>
<tr>
<td>Segmented Group 1</td>
<td>Difference</td>
<td>Difference</td>
<td>ΔTime(t)</td>
</tr>
<tr>
<td>Evolving business practice Group 2</td>
<td>Difference</td>
<td>Level</td>
<td>ΔTime(t)</td>
</tr>
<tr>
<td>Hysteresis Group 3</td>
<td>Level</td>
<td>Difference</td>
<td>Time(t)</td>
</tr>
<tr>
<td>Escalation Group 4</td>
<td>Level</td>
<td>Difference</td>
<td>Time(t)</td>
</tr>
<tr>
<td>Business as usual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled Differences</td>
<td>Difference</td>
<td>Level</td>
<td>ΔTime(t)</td>
</tr>
<tr>
<td>Levels</td>
<td></td>
<td></td>
<td>ΔTime(t)</td>
</tr>
</tbody>
</table>

Legend: Time(t) represents an indicator variable that is equal to zero before time period t and unity from time t on; ΔTime(t) represents the first difference of Time(t) and is the same as monthly dummy variables. The number of lags for each final model was selected based on the Schwarz criterion. The variables entering the models in differences do not include cross-sectional fixed effects because these are canceled out through first-differencing.

5.4. PVAR model estimation and testing

Testing, estimation, and model selection for the PVARs proceeded as follows. For each group, we first determined the appropriate number of lag terms to include (p) by minimizing the SC. Table 4 provides a summary of the specifications for each of the PVAR models estimated. The last column of the table shows that the selected PVAR models for groups 1, 2, 3, and 4 had lag lengths of 7, 4, 8, and 5, respectively. The final pooled PVAR model in levels had 7 lags, and the pooled PVAR model in differences had 11 lags.

Next, we conducted group-specific Hausman specification tests and, consistent with prior research (Mizik & Jacobson, 2004), detected the presence of significant fixed effects in the levels data. As such, our PVAR models differ across the four groups in that we have an individual-level intercepts specification for all equations in Group 4, for marketing activity equations in Group 2, and for the prescribing equation in Group 3. Due to the differencing of evolving series, any physician-specific fixed effects are eliminated, and the remainder of the equations are specified with uniform intercepts (i.e., without fixed effects).

Table 5 reports SC values for each group model as well as the pooled models. The group-level PVAR models are estimated separately for each group, but we also report the aggregate SC values to facilitate comparison with the pooled PVAR models. Conversely, the pooled PVAR models are estimated across all physicians in the estimation sample. Again, to facilitate comparison, we report the SC values produced by the pooled models specifications within each physician group. The results in Table 5 show that the group-level models are preferred to the pooled models in all cases. The group-level PVAR models have the minimum SC values both overall and for each group taken separately.

We also compared the performance of the group-level models with the pooled models using cross-sectional and longitudinal holdout samples. Table 6 reports results for the cross-sectional holdout. Using the estimates of PVAR parameters from the estimation sample, we predict the number of new prescriptions for physicians in the holdout sample. The table provides two measures of predictive validity, root mean squared error (RMSE) and mean absolute deviation (MAD). The group-level PVAR models produce lower RMSE and MAD values than pooled PVAR models, with one exception. For Group 3, the MAD for the group-level model is equal to the corresponding MAD of the pooled model in differences. The RMSE, however, still favors the group-level model for the physicians in Group 3.

Table 7 presents results for the longitudinal holdout tests. We created a longitudinal holdout sample containing the last 2 months of the data for each physician. We re-estimated our models using the remaining data points (i.e., excluding the holdout time periods for each individual) and used the estimates for prediction. Again, we see clear evidence of superior performance for the dynamic scenario-based segmentation because it generates better quality forecasts.

In sum, the model comparison results consistently support the group-level PVAR models over the pooled PVAR models. For the estimation sample, the SC is minimized by the group-level modeling approach, for all four groups separately and for the estimation sample taken as a whole. For the holdout samples, the forecast errors, as measured by RMSE and MAD, are lower for the group-level models than for the pooled models. Collectively, these results suggest superior fit and forecasting in time-series panel data for a model that incorporates the potential for coexisting business scenarios.

5.5. Nature of dynamic response

To examine whether our group-level approach produces substantive differences in dynamic response patterns, we computed impulse response functions (IRFs) for the effect of changes in detailing and marketing activity on prescribing behavior. The IRFs within each physician segment were then averaged across segments within each physician group to facilitate model comparison across the four business practice scenarios in Table 4. We also compared the performance of the group-level models with the pooled models using cross-sectional and longitudinal holdout samples. Table 6 reports results for the cross-sectional holdout. Using the estimates of PVAR parameters from the estimation sample, we predict the number of new prescriptions for physicians in the holdout sample. The table provides two measures of predictive validity, root mean squared error (RMSE) and mean absolute deviation (MAD). The group-level PVAR models produce lower RMSE and MAD values than pooled PVAR models, with one exception. For Group 3, the MAD for the group-level model is equal to the corresponding MAD of the pooled model in differences. The RMSE, however, still favors the group-level model for the physicians in Group 3.

Table 7 presents results for the longitudinal holdout tests. We created a longitudinal holdout sample containing the last 2 months of the data for each physician. We re-estimated our models using the remaining data points (i.e., excluding the holdout time periods for each individual) and used the estimates for prediction. Again, we see clear evidence of superior performance for the dynamic scenario-based segmentation because it generates better quality forecasts.

In sum, the model comparison results consistently support the group-level PVAR models over the pooled PVAR models. For the estimation sample, the SC is minimized by the group-level modeling approach, for all four groups separately and for the estimation sample taken as a whole. For the holdout samples, the forecast errors, as measured by RMSE and MAD, are lower for the group-level models than for the pooled models. Collectively, these results suggest superior fit and forecasting in time-series panel data for a model that incorporates the potential for coexisting business scenarios.

5.5. Nature of dynamic response

To examine whether our group-level approach produces substantive differences in dynamic response patterns, we computed impulse response functions (IRFs) for the effect of changes in detailing and marketing activity on prescribing behavior. The IRFs within each physician segment were then averaged across segments within each physician group to facilitate model comparison across the four business practice scenarios in Table 4. We also compared the performance of the group-level models with the pooled models using cross-sectional and longitudinal holdout samples. Table 6 reports results for the cross-sectional holdout. Using the estimates of PVAR parameters from the estimation sample, we predict the number of new prescriptions for physicians in the holdout sample. The table provides two measures of predictive validity, root mean squared error (RMSE) and mean absolute deviation (MAD). The group-level PVAR models produce lower RMSE and MAD values than pooled PVAR models, with one exception. For Group 3, the MAD for the group-level model is equal to the corresponding MAD of the pooled model in differences. The RMSE, however, still favors the group-level model for the physicians in Group 3.

Table 7 presents results for the longitudinal holdout tests. We created a longitudinal holdout sample containing the last 2 months of the data for each physician. We re-estimated our models using the remaining data points (i.e., excluding the holdout time periods for each individual) and used the estimates for prediction. Again, we see clear evidence of superior performance for the dynamic scenario-based segmentation because it generates better quality forecasts.

In sum, the model comparison results consistently support the group-level PVAR models over the pooled PVAR models. For the estimation sample, the SC is minimized by the group-level modeling approach, for all four groups separately and for the estimation sample taken as a whole. For the holdout samples, the forecast errors, as measured by RMSE and MAD, are lower for the group-level models than for the pooled models. Collectively, these results suggest superior fit and forecasting in time-series panel data for a model that incorporates the potential for coexisting business scenarios.

5.5. Nature of dynamic response

To examine whether our group-level approach produces substantive differences in dynamic response patterns, we computed impulse response functions (IRFs) for the effect of changes in detailing and marketing activity on prescribing behavior. The IRFs within each physician segment were then averaged across segments within each physician group to facilitate model comparison across the four business practice scenarios in Table 4. We also compared the performance of the group-level models with the pooled models using cross-sectional and longitudinal holdout samples. Table 6 reports results for the cross-sectional holdout. Using the estimates of PVAR parameters from the estimation sample, we predict the number of new prescriptions for physicians in the holdout sample. The table provides two measures of predictive validity, root mean squared error (RMSE) and mean absolute deviation (MAD). The group-level PVAR models produce lower RMSE and MAD values than pooled PVAR models, with one exception. For Group 3, the MAD for the group-level model is equal to the corresponding MAD of the pooled model in differences. The RMSE, however, still favors the group-level model for the physicians in Group 3.

Table 7 presents results for the longitudinal holdout tests. We created a longitudinal holdout sample containing the last 2 months of the data for each physician. We re-estimated our models using the remaining data points (i.e., excluding the holdout time periods for each individual) and used the estimates for prediction. Again, we see clear evidence of superior performance for the dynamic scenario-based segmentation because it generates better quality forecasts.

In sum, the model comparison results consistently support the group-level PVAR models over the pooled PVAR models. For the estimation sample, the SC is minimized by the group-level modeling approach, for all four groups separately and for the estimation sample taken as a whole. For the holdout samples, the forecast errors, as measured by RMSE and MAD, are lower for the group-level models than for the pooled models. Collectively, these results suggest superior fit and forecasting in time-series panel data for a model that incorporates the potential for coexisting business scenarios.
sampling on prescriptions. We follow previous research (Dekimpe & Hanssens, 1999; Nijs, Dekimpe, Steenkamp, & Hanssens, 2001) to derive generalized impulses to compute IRFs. These generalized impulses do not impose any particular ordering on the effect of the endogenous variables. To compute the standard errors of the IRF estimates, we use a bootstrap procedure repeated 250 times (see Srinivasan, Pauwels, Hanssens, & Dekimpe, 2004 for a discussion).

We report the impulse response functions for new prescriptions with respect to a one-standard-deviation change in detailing in Fig. 3 and for new prescriptions with respect to a change of one standard deviation in sampling in Fig. 4. The solid lines give the IRFs, and the dashed lines give their 95% confidence intervals.

The patterns of dynamic response represented by the IRFs are broadly similar for detailing and sampling and are consistent with the nature of the business scenario classifications pertaining to each group (Dekimpe & Hanssens, 1999). For Group 1, the coevolution scenario, the responses over time are positive, significant, and permanent for both detailing and sampling. Group 4, the “business as usual” scenario, shows significant short-run response in both cases but returns to zero after about six periods (we note that only the immediate effect is significant in this group). In Group 3, escalation, a non-significant response to both detailing and sampling is well contained within the error bands except in the case of the immediate effect of sampling. For Group 2, hysteresis, we also find that response is not significantly different from zero. For detailing, the IRF values are positive for most periods, but the error bands always include zero. This suggests that the sales calls made to physicians in this group, on average, are not producing sufficient short-run response to capitalize on the potential for a permanent effect from evolution in the prescriptions series. In other words, we do not find a marketing-based hysteresis in our data (i.e., the evolving nature of the sales series in this group does not seem to be caused by detailing and sample marketing effort). Indeed, evolution in sales can be due to factors other than marketing stimuli (e.g., physicians learning through usage). Our finding of the absence of marketing-induced hysteresis is consistent with prior research in that hysteresis is a very rarely observed phenomenon.²⁹

Our finding that detailing and sampling have a limited impact on prescriptions (albeit significant in some of the smaller groups) is also consistent with prior research documenting modest effectiveness of detailing and sampling. For example, Kremer et al. (2008) surveyed the literature and noted that most recent studies report very modest effectiveness of detailing and sampling (some report no effect, and some studies report negative effects). Some earlier studies that did not control for individual-specific heterogeneity in levels of prescribing (e.g., Wittink, 2002) reported much higher effectiveness for detailing, although such results might have been driven by omitted variable bias (these studies did not account for individual-level heterogeneity via a fixed-effects formulation).

5.6. Comparison with pooled models

The IRFs for the pooled PVAR models show quite different patterns of response (bottom two panels of each figure). For details and samples, the pooled model in levels shows a significant and positive short-run effect that reverts to zero after about eight periods. Most importantly, the pooled model in differences estimates the effects of detailing and sampling to be not only positive and significant but also permanent. This is in stark contrast to the dynamic response implied by the group-level models (we find a significantly smaller group of physicians for whom detailing and sampling have permanent effects). The reason for these differences is that we do not impose the same dynamic patterns on all physicians. The group-level models allow impulse response functions to capture different dynamics at the segment level, revealing heterogeneity in the response dynamics within the same market.

5.7. Elasticities

To better understand the magnitude of the effect sizes for each group and for the sample as a whole, we compute detailing and sampling elasticities at different time points (1, 6, and 12 months). These elasticities are presented in Table 8. We begin by noting that the magnitude of the elasticities differs markedly across the four groups. Consistent with the IRF patterns, both detailing and sampling elasticities are largest for Group 1 and smallest for Group 3, while Groups 2 and 4 fall in between. The detailing elasticities for Group 3 are negatively signed but are very close to zero. These group-level results for the elasticities provide further support for the segmentation and targeting potential of business scenarios. The IRFs for the pooled PVAR models show quite different patterns of response (bottom two panels of each figure). For details and samples, the pooled model in levels shows a significant and positive short-run effect that reverts to zero after about eight periods. Most importantly, the pooled model in differences estimates the effects of detailing and sampling to be not only positive and significant but also permanent. This is in stark contrast to the dynamic response implied by the group-level models (we find a significantly smaller group of physicians for whom detailing and sampling have permanent effects). The reason for these differences is that we do not impose the same dynamic patterns on all physicians. The group-level models allow impulse response functions to capture different dynamics at the segment level, revealing heterogeneity in the response dynamics within the same market.

Table 6

Model comparisons for cross-sectional holdout sample: out-of-sample forecast errors for new prescriptions.

<table>
<thead>
<tr>
<th></th>
<th>Group-level models</th>
<th>Pooled model: Levels</th>
<th>Pooled model: Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAD</td>
<td>RMSE</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evolving practice</td>
<td>3.398</td>
<td>2.344</td>
<td>3.418</td>
</tr>
<tr>
<td>Group 2</td>
<td>3.769</td>
<td>2.576</td>
<td>3.792</td>
</tr>
<tr>
<td>Hysteresis</td>
<td>3.202</td>
<td>2.238</td>
<td>3.217</td>
</tr>
<tr>
<td>Group 3</td>
<td>3.083</td>
<td>2.227</td>
<td>3.099</td>
</tr>
<tr>
<td>Escalation</td>
<td>3.240</td>
<td>2.279</td>
<td>3.257</td>
</tr>
</tbody>
</table>

Legend: RMSE is the root mean square forecast error, and MAD is the mean absolute forecast error.

Table 7


<table>
<thead>
<tr>
<th></th>
<th>Group-level models</th>
<th>Pooled model: Levels</th>
<th>Pooled model: Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAD</td>
<td>RMSE</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evolving practice</td>
<td>3.836</td>
<td>2.777</td>
<td>4.201</td>
</tr>
<tr>
<td>Group 2</td>
<td>3.905</td>
<td>2.710</td>
<td>4.143</td>
</tr>
<tr>
<td>Hysteresis</td>
<td>3.162</td>
<td>2.176</td>
<td>3.183</td>
</tr>
<tr>
<td>Group 3</td>
<td>3.175</td>
<td>2.199</td>
<td>3.194</td>
</tr>
</tbody>
</table>

Legend: RMSE is the root mean square forecast error, and MAD is the mean absolute forecast error.

²⁹We thank Dominique Hanssens for very helpful discussions on the nature of the hysteresis scenario and the frequency of its occurrence. We note in the directions for future research that further within-group heterogeneity in response may exist within dynamic scenarios, but this issue is beyond the scope of the present study.
suggested possible evolution in the prescribing series, the choice of a pooled model in differences might be justified. However, the elasticity results we present show that a pooled model could suggest very different implications for market response and resource allocation.

6. Sensitivity analyses

We undertook several sensitivity analyses to ensure the robustness of our results. In this section, we briefly summarize these analyses and our findings. We note that all additional analyses are fully supportive of our proposed approach and suggest the superiority of dynamic business scenario-based segmentation in identifying groupings with meaningful response differences over common alternatives used in the industry. Full results of these additional analyses are available from the authors upon request.

6.1. Alternative classification heuristics

We examined alternative segmentation heuristics and compared their performance to that of the proposed dynamic segmentation.

We tested two different a priori groupings, one based on deciles (typically used by the industry) and one based on a combination of deciles, gender and years of experience (i.e., accounting for all physician-specific information available). For the first segmentation benchmark, we performed a median split of physicians based on their deciles (ranging from 4 to 10 and reflecting previous physician prescribing levels; the median decile was 6). We segmented physicians into a group of high deciles (deciles greater than 6) and one of low deciles (deciles less than or equal to 6). For the second segmentation benchmark, we used the k-means clustering algorithm and the Gower distance to cluster physicians using the three physician descriptors.10 The three-cluster solution was deemed the best based on the silhouette criterion (further details available upon request). This solution segmented doctors into one group with all the female physicians in the high deciles group, one group with physicians in the low deciles group and one group with physicians in the middle deciles group. Full results of these additional analyses are available from the authors upon request.

10 The Gower metric easily combines discrete and continuous variables in one single measurement, which is important in our specific application. In our case, deciles range from four to ten and reflect previous prescription levels, gender is a dummy variable that takes the value 1 for males and 0 for females, and years of experience ranges from 1 to 55 years for our physician sample.
physicians (who are less experienced and are typically in lower deciles) and two groups of male physicians, one with higher levels of past prescribing and the other with lower levels.

For both segmentation alternatives, we find no systematic relation between these alternative segments and the proposed dynamic-based groupings. The four dynamic groups are uniformly distributed across the segments, further strengthening the results in Table 3. In addition, we used panel unit-root tests to test for unit roots in each segment and concluded that no unit roots were present. We then estimated PVAR models in levels for each of the segments, and although unit root tests indicated no unit roots, we estimated a first-difference specification for completeness. As before, we relied on SC to determine the number of lags, allowed for time-specific effects, and included individual-level fixed effects in all models estimated in levels (detailed results available upon request). Finally, we estimated the elasticities for the new segments and computed the in- and out-of-sample performance of these two alternative segmentation benchmarks.

Our cross-sectional and longitudinal holdout tests clearly favor the business scenario-based segmentation (see Table 9 for detailed results). Similar results were found for the in-sample performance comparison. More importantly, we found no meaningful differences in response across segments, contrary to the results of our dynamic-segmentation approach. Thus, in our context, these alternative approaches (and the variables underlying the segmentation) were unable to reveal meaningful diagnostic information to managers, although significant differences were indeed present in the market. These results further highlight the substantive benefits of examining and addressing dynamic response using the proposed segmentation based on dynamic business scenarios.

6.2. Alternative estimation: Bayesian approach

In large samples, Bayesian methods are computationally intensive, and, as such, can be less practical. However, it might be argued that Bayesian methods are feasible and beneficial in moderate-size samples because they can better depict individual-level heterogeneity in response, even in a time-series context. To assess these arguments and to compare the performance of our two-step segmentation procedure against Bayesian estimation, we undertook additional tests.
We re-estimated the PVAR models using random-effects Bayesian models in which all model parameters were allowed to be physician-specific. Because of the spurious regression problem, we re-estimated the models in differences with all variables—prescriptions, detailing and samples—entering the models in differences. We used conjugate and uninformative priors and estimated the models using standard Markov chain Monte Carlo (MCMC) methods and convergence diagnostics.

Our in- and out-of-sample fit results suggest that a non-segmented random-effects approach (i.e., a single model specified across all panelists allowing for intercept and response heterogeneity) is not appropriate if different dynamic scenarios coexist in the data. Our proposed model out-performed the Bayesian alternative, especially in longitudinal holdout (detailed results available from the authors upon request). One possible explanation for this result is that traditional Bayesian estimation shrinks individual estimates toward a prior and a common mean (or means) without any a priori screening for dynamic patterns (the first step in unit-root testing). This means that the Bayesian procedure may mask or blend dynamic patterns. Thus, for many panelists, the Bayesian estimates yield a dynamic profile that is similar to the largest group, which, in our case is Group 3, the escalation scenario. Given the existing differences in the market, the shrinkage leads to significantly worse out-of-sample predictive performance.

### 6.3. Alternative computation of elasticities

The elasticities presented in Table 8 (for both the segmented and pooled models) were computed with respect to the initial exogenous shocks of detailing and sampling following the standard approach. That is, the elasticity numerator takes into account the impact on prescriptions from subsequent changes of the marketing variables (i.e., it is based on the generalized IRF), but the elasticity denominator does not incorporate the accumulated marketing changes. These cumulative changes, however, can contribute significantly to the total cost of the original marketing action, particularly when marketing is evolving.

To incorporate the full marketing cost considerations in the elasticity computation, we recomputed the 6-month and 12-month detailing elasticities, including the effect of a detailing shock on future detailing.

Overall, these new elasticities tend to be lower than the ones reported in Table 8, but they follow similar patterns, retaining the significant differences across groups. The recomputed elasticities also confirm that pooled models produce different implications for market response. For example, the recomputed detailing elasticities for the pooled model in differences are .082 and .110 for the 6-month and the 12-month windows, respectively. The corresponding recomputed values obtained for the segmented approach are .046 and .055, which are significantly lower.

### 6.4. Alternative formulation for time-specific effects

We chose not to use time trends in our models because trends impose the restriction of constant change from one period to the next. Instead, we opted for a more flexible formulation of time-specific effects. The levels of prescribing and marketing effort differ from period to period and are contemporaneously correlated simply due to the different number of work days in a given month. Our sensitivity tests show that failure to explicitly model these time-specific differences results in spurious relationships and biased estimates. We

---

**Table 8**

Elasticities for group and pooled models.

<table>
<thead>
<tr>
<th>Group</th>
<th>Detailing elasticity</th>
<th>Sampling elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 month</td>
<td>6 months</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evolving business practice</td>
<td>.014</td>
<td>.325</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hysteresis</td>
<td>.032</td>
<td>.095</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escalation</td>
<td>-.004</td>
<td>-.008</td>
</tr>
<tr>
<td>Group 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business as usual</td>
<td>.029</td>
<td>.111</td>
</tr>
<tr>
<td>All groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.014</td>
<td>.088</td>
</tr>
<tr>
<td>Pooled in levels</td>
<td>.020</td>
<td>.145</td>
</tr>
<tr>
<td>Pooled in differences</td>
<td>.021</td>
<td>.221</td>
</tr>
</tbody>
</table>

Legend: The “All groups” elasticities are computed as a weighted average of the group elasticities. We first compute the generalized IRFs of prescriptions due to a shock of one standard deviation of detailing or sampling (these will depend on the estimated PVAR parameters). We then compute the percentage change variation in prescriptions with respect to the mean prescriptions and the percentage change in detailing or sampling with respect to the detailing and sampling means. We take their ratios to obtain the elasticity. We estimate the standard deviation of detailing or sampling (these will depend on the estimated PVAR parameters). We then compute the percentage change variation in prescriptions with respect to the detailing and sampling means. We take their ratios to obtain the elasticity. We re-estimated the PVAR models using random-effects Bayesian models in which all model parameters were allowed to be physician-specific. Because of the spurious regression problem, we re-estimated the models in differences with all variables—prescriptions, detailing and samples—entering the models in differences. We used conjugate and uninformative priors and estimated the models using standard Markov chain Monte Carlo (MCMC) methods and convergence diagnostics.

Our in- and out-of-sample fit results suggest that a non-segmented random-effects approach (i.e., a single model specified across all panelists allowing for intercept and response heterogeneity) is not appropriate if different dynamic scenarios coexist in the data. Our proposed model out-performed the Bayesian alternative, especially in longitudinal holdout (detailed results available from the authors upon request). One possible explanation for this result is that traditional Bayesian estimation shrinks individual estimates toward a prior and a common mean (or means) without any a priori screening for dynamic patterns (the first step in unit-root testing). This means that the Bayesian procedure may mask or blend dynamic patterns. Thus, for many panelists, the Bayesian estimates yield a dynamic profile that is similar to the largest group, which, in our case is Group 3, the escalation scenario. Given the existing differences in the market, the shrinkage leads to significantly worse out-of-sample predictive performance.

### 6.3. Alternative computation of elasticities

The elasticities presented in Table 8 (for both the segmented and pooled models) were computed with respect to the initial exogenous shocks of detailing and sampling following the standard approach. That is, the elasticity numerator takes into account the impact on prescriptions from subsequent changes of the marketing variables (i.e., it is based on the generalized IRF), but the elasticity denominator does not incorporate the accumulated marketing changes. These cumulative changes, however, can contribute significantly to the total cost of the original marketing action, particularly when marketing is evolving.

To incorporate the full marketing cost considerations in the elasticity computation, we recomputed the 6-month and 12-month detailing elasticities, including the effect of a detailing shock on future detailing.

Overall, these new elasticities tend to be lower than the ones reported in Table 8, but they follow similar patterns, retaining the significant differences across groups. The recomputed elasticities also confirm that pooled models produce different implications for market response. For example, the recomputed detailing elasticities for the pooled model in differences are .082 and .110 for the 6-month and the 12-month windows, respectively. The corresponding recomputed values obtained for the segmented approach are .046 and .055, which are significantly lower.

### 6.4. Alternative formulation for time-specific effects

We chose not to use time trends in our models because trends impose the restriction of constant change from one period to the next. Instead, we opted for a more flexible formulation of time-specific effects. The levels of prescribing and marketing effort differ from period to period and are contemporaneously correlated simply due to the different number of work days in a given month. Our sensitivity tests show that failure to explicitly model these time-specific differences results in spurious relationships and biased estimates. We

---

We would like to thank Professor Els Gijsbrechts for her comments on the alternative elasticity computations. We do not report a similar analysis for sampling because it is typically treated as having zero marginal cost in the industry.
verified the superior performance of our specification by re-estimating all group-level models, replacing time-specific effects with linear and quadratic trends and examining the cross-sectional and longitudinal out-of-sample predictive performance. The results showed that our specification of time-specific effects produced the lowest RMSE and MAD in all cases (e.g., for the cross-sectional holdout, including a linear trend produced a RMSE of 3.470 and a MAD of 2.967; for the longitudinal holdout, the RMSE was 3.374, and the MAD was 2.396; the inclusion of the quadratic trend produced even worse results).

6.5. Alternative unit-root tests

Finally, to assess the robustness of our findings to a particular unit-root test, we re-ran our analyses based on segments formed using the KPSS test. We obtained similar results. Specifically, we found that (1) the proportion of physicians assigned to each business scenario group changed little when using KPSS instead of ADF, (2) the KPSS-based segments also outperformed the pooled models in terms of fit, and (3) the segment elasticities between the ADF and KPSS-based groupings were very similar.

7. Discussion

In cases where marketing outcomes and marketing effort are potentially evolving, researchers have used persistence modeling techniques to identify and study the nature of the business scenario that characterizes time-series data (Dekimpe & Hanssens, 1995, 1999). These scenarios (business as usual, escalation, hysteresis, and evolving business practice) can have radically different implications for management and for the efficient allocation of scarce marketing resources.

We propose and illustrate an approach to modeling potentially coexisting business scenarios within the same product market. In doing so, we combine persistence modeling techniques with the possibility that customers can be segmented using the individual dynamic properties of prescribing and marketing. First, we conduct unit-root tests on the outcome and marketing activity variables at the panelist level and classify these series, for each panelist, as either evolving or stationary. Using this classification, we group panelists into the four business scenario groups and specify and estimate appropriate PVAR models for each group. We use IRFs to study the dynamic properties of the data and compute elasticities.

The proposed approach assesses the response to marketing effort and simultaneously addresses the dynamic properties of marketing effort, performance, and dependencies (feedback) among all series. It allows for a comprehensive assessment of the returns on marketing. For example, in the escalation scenario (evolving marketing effort and stationary outcomes), any (short-run) benefits of marketing will be eventually overshadowed by the escalation of marketing and ever-increasing spending.

We illustrate the approach using physician panel data provided by a pharmaceutical company. The segmentation obtained from the unit-root tests shows that each of the four business scenarios is populated by a sizable proportion of doctors. We also find that the PVAR models estimated at the group level provide a better in-sample and holdout fit to the data than the alternative models. These findings document the ability of a multiple scenario modeling approach to better represent the data than conventional benchmark models.

Most importantly, in addition to producing superior fit, the IRFs and dynamic response elasticities derived from the multiple scenario approach highlight significant differences across the groups. These differences suggest potential benefits of using the proposed approach in practice. Using the approach, firms can draw important new implications for segmentation, targeting, and marketing resource allocation.

7.1. The role of additional factors

The approach we propose can easily be extended to include additional factors that might influence physician prescribing, including other promotional tools or competitive marketing efforts. Unfortunately, we did not have information on other marketing efforts or competitive detailing and sampling activity in our data set. The competitive information is unavailable to the pharmaceutical firm providing the data. Several recent academic studies have also lacked competitive information and have presented a variety of compelling arguments for proceeding without it (e.g., Manchanda & Chintagunta, 2004; Manchanda et al., 2004; Mizik & Jacobson, 2004). Prior research has shown that the level of a physician’s prescribing is the major determinant of the frequency of detailing visits. Because we control for physician-specific effects, the major common source of bivariate correlation between own and competitive marketing effort is removed. Given the body of empirical evidence on this issue, the results we report are not likely to be significantly affected by the lack of competitive marketing data.

7.2. Managerial use of the dynamic business scenarios segmentation

Our sensitivity analyses show that business scenario-based groups do not align with traditional targeting used in the industry (physician demographics and decile rankings) and that common segmentation heuristics do not create meaningful groupings with differentiated response or differentiated relevant behavioral dynamics. Many businesses track customer relationship data and transaction data, but they often do not have access to reliable demographic descriptors of the customers. Our findings suggest that meaningful and actionable segments can be constructed using the dynamic business scenario segmentation approach based on activity-only data and can be used for targeting and marketing activity allocation.

The approach is easy to implement and is highly scalable to large databases and continuous testing. Because tests can be run and models re-estimated quickly using standard methods, managers can easily obtain updated groupings and the associated IRFs. The results can be updated every time new data become available or when major events occur in the category, such as new drug entry or loss of patent protection. Gonzalez, Sismeiro, and Dutta (2008) demonstrate that events such as the loss of patent protection and new drug entry can lead to significant changes in the marketing policy of firms.

In the case of the pharmaceutical industry, new information on prescribing and marketing effort is updated every month. Testing and estimation could be performed monthly, and physicians’ group assignments could be monitored over time. Indeed, customer membership in a group is unlikely to be stable over a long time period. We would expect customers to migrate from positive evolving consumption, to stationary, and then to negative evolving as the product moves through the adoption, maturity, and decline stages of its life cycle, respectively. That is, in the initial stages of the product life cycle, we would expect consumers to migrate from Groups 1 and 2 into Groups 3 and 4 and then back (assuming that product use is not abruptly discontinued).

Dynamic business scenario segmentation may offer greater benefits if it is implemented as a continuous process rather than as a one-time exercise. Continuously reclassifying customers and tracking changes in classification on a routine basis can enable managers to draw specific implications for when and to whom to accelerate, decelerate, or even stop marketing activity. For example, in our empirical illustration, we find a large group of physicians populating the escalation scenario (Group 3) and generating zero (insignificantly negative) response to marketing effort. Decelerating effort to physicians in this
group can free up resources that could be deployed for physicians in the “evolving business practice” scenario (Group 1) or in the “business as usual” scenario (Group 4), groups that show positive response. Clearly, escalation of marketing spending to non-responsive physicians is not appropriate unless a new and more effective message can be developed. Thus, when the movement of a physician from Group 1 into Group 3 is noticed, it may be advisable to decelerate marketing effort.

We note that such considerations could be subject to the Lucas critique. Because we are using PVAR models, if the acceleration and deceleration of marketing activities are within the typical ranges for a particular group (i.e., within the relevant data range used for the VAR estimation) and do not alter the long-term properties of the series, the problems associated with the Lucas critique are not significant. Indeed, as discussed by van Heerde, Dekimpe, and Putis (2005), the Lucas critique is relevant in the VAR framework when radical policy shifts are studied.

We further note that our segment-based estimation and cross-sectional holdout tests may provide guidance (although far from a complete analysis) regarding possible outcomes of marketing activity regime shifts. Because we find that the patterns and estimated relationships continue into the cross-sectional holdout period, we can use our estimates to predict physicians’ responses to marketing regime shifts.

An important caveat is that our business scenario segmentation is partially driven by consumer response and partially driven by firm marketing efforts. Hence, changes in marketing effort allocation that shift panelists into a different scenario could change the long-term properties of the response series and therefore should be analyzed with care. For example, escalation of marketing efforts to physicians in Group 2 may have no effect on the dynamic properties of the prescription series and thus may move Group 2 physicians to Group 1, resulting in significantly improved response. However, if the regime shift in marketing activity spurs a regime shift in response series for some physicians, it would move them to Group 3 and would result in decreased responsiveness. Similarly, escalating marketing efforts to Group 4 may move these physicians to Group 3 and result in a waste of marketing resources. The hope, of course, would be that a sustained, persistent marketing effort could spur evolution in decreased responsiveness. Clearly, escalation of marketing spending to non-responsive physicians is not appropriate unless a new and more effective message accompanies the change in prescribing and move physicians to Group 1.

Because we did not formally test the process and the consequences of shifting marketing effort regimes (from stationary to evolving and vice versa), further research and careful experimentation (preferably through field experiments) are needed to better understand these processes. For example, future studies can examine physicians’ switching patterns across scenarios in response to changing marketing activity regimes. In sum, we view research into regime switching as an important topic for future study.12

8. Conclusion

We propose a new approach for segmenting customers based on the coexistence of multiple business scenarios in a single product market. This approach promises advantages in identifying more responsive targets and aiding resource allocation. In our empirical application, using data from a pharmaceutical drug market, the proposed multiple business scenario approach produced superior in-and out-of-sample fit and highlighted large differences across physician business scenario groupings. These differences (and segments) did not align with traditional targeting variables used in the industry (e.g., physician demographics and decile rankings) and could not be captured by other common segmentation tools. Our results demonstrate that the proposed approach can be used by firms to better understand the underlying dynamic market structure and better target their marketing efforts.

There are several limitations to our study that highlight opportunities for future research. First, we rely on unit-root tests to determine the stationary or evolving nature of each panelist’s outcome and marketing activity variables. To the extent that the power of these tests is weak, there is a classification risk. Development of better diagnostic tests would improve the reliability of our dynamic segmentation approach. Current research on unit-root tests using wavelets shows promise in increasing power, especially when analyzing near unit-root alternatives (e.g., Fan & Gençay, 2010). Although wavelets have been previously used in marketing to study time-series properties in the frequency domain (Lemmens, Croux, & Dekimpe, 2007), their use in unit-root testing is less developed and is a promising area for future research.

Further improvements to this segmentation approach might be developed by incorporating heterogeneity in individual-level response dynamics within each business scenario. Indeed, an optimal allocation of resources requires a careful study of how spending should be allocated across segments and also within a given segment (e.g., based on individual-level response). While our sensitivity tests suggest that initial unit-root testing is always essential for the proper modeling of dynamic data, Bayesian estimation of individual-level response within business scenarios might offer additional insights and refinements to targeting strategy. The choice between classical and Bayesian estimation in the second phase should be determined by context, sample size, and the need for scalability.

Another interesting direction for future research is to address the applicability of the Dorfman–Steiner theorem for marketing resource allocation in non-stationary environments (our Groups 2, 3 and 4). One way of approaching this may be to consider an extension of the models of Doraszelski and Markovich (2007) and of Dubé, Hitsch, and Manchanda (2005) to incorporate the possibility of dynamic business scenarios. These authors model the demand and supply side of the market and simulate the outcome of a dynamic game using the Markov perfect equilibrium (MPE) concept, allowing for dynamic competitive behavior (in price and advertising), market entry and exit, and time-dependent solutions. Because these papers did not consider dynamic business scenarios, we feel that this presents a very interesting opportunity for future research.

Finally, future research into modeling customer migration from one business scenario to another over time (e.g., from evolving practice to business as usual) and factors explaining the differences in the propensity to switch can generate academically and managerially valuable insights.

References


12 We would like to stress the conceptual distinction between our focus on regime shifts (stationary versus evolving nature of the series) and the research on modeling transition processes among hidden states (e.g., inactive, infrequent, frequent), as in Montoya, Netzer, and Jedidi (2010).


Measuring willingness to pay as a range, revisited: When should we care?

Florian Dost a,⁎, Robert Wilken b,1

a ESCP Europe Business School Berlin, Heubnerweg 8-10, D-14059 Berlin, Germany
b ESCP Europe Business School Berlin, Heubnerweg 8-10, D-14059 Berlin, Germany

A R T I C L E   I N F O

Article history:
First received in 18, November 2010 and was under review for 6½ months
Available online 8 March 2012

Area Editor: Russell S. Winer

Keywords:
Willingness to pay as a range
ICERANGE
BDM mechanism
Demand functions
Monte Carlo simulation

A B S T R A C T

Recent research has conceptualized consumers’ willingness to pay (WTP) as a range rather than as a single point. However, there are important gaps in this research stream: The existing method to measure WTP as a range, ICERANGE, features restrictive assumptions and is rather complex, such that it hampers real-world applications. Furthermore, it is unclear what has been measured in the past with point-based methods, compared with WTP ranges; thus, researchers cannot evaluate “traditional” WTP measurements. Most importantly, why should anyone even care about WTP ranges? In making pricing decisions, aggregate-level information is common, and the add-on information contained in individual WTP ranges would seemingly become obsolete when averaging it across consumers. This article addresses all three issues: We show empirically that traditional point-based methods reveal the midpoint of WTP ranges. Our proposed range-based method, which is simpler and less restricted than ICERANGE, achieves comparable performance. We use a Monte Carlo simulation to show that, except for in rather artificial conditions, point-based methods fail to reproduce the revenue-maximizing prices identified by range-based methods. Together, these results deliver a compelling argument for the use of range-based methods to elicit WTP in real-world applications.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Pricing has been a continuous focus of marketing research. Demand functions provide key information for optimal pricing decisions and contain aggregate information about consumers’ individual willingness to pay (WTP). To estimate unbiased demand functions, researchers require a valid method to measure WTP (e.g., Cameron & James, 1987; Gijsbrechts, 1993; Jedidi, Jagpal, & Manchanda, 2003; Miller, Ho¨stetter, Krohmer, & Zhang, 2011; Voelckner, 2006). Various applied methods include those that employ transaction data (experimental and non-experimental), survey data (e.g., conjoint analysis), and instruments that incorporate bids (e.g., auctions, lotteries). The BDM (Becker, DeGroot, & Marschak, 1964) lottery combines the advantages of methods that rely on either transactions or on survey data, as discussed in detail by Wertenburg and Sikera (2002).

Using this BDM lottery, Wang, Venkatesh, and Chatterjee (2007) proposed measuring WTP as a range rather than as a single point. They argued that the common definitions of a reservation price (which often appears as a synonym for WTP)—such as “the price at or below which a consumer will demand one unit of the good” (Varian, 1992, p. 152), “the price at which a consumer is indifferent between buying and not buying” (Moorthy, Ratchford, & Talukdar, 1997, p. 265), or “the minimum price at which a consumer would no longer purchase” (Hauser & Urban, 1986, p. 449)—are equivalent only if consumers make rational choices and are certain about their preferences and about product performance. However, in the far more likely cases of bounded rationality, preference uncertainty, and performance uncertainty, WTP should be conceptualized and measured as a range of prices.

Thus, Wang et al. (2007) proposed ICERANGE, a method that incorporates three prices: the floor price, the indifference price, and the ceiling reservation price. Each price corresponds to one of the previous WTP definitions and is linked to choice probabilities of 0.5, 1, and 0, respectively. The difference between the ceiling and floor reservation price is the WTP range, which exists according to the authors’ empirical demonstration. Their method performs better than point-based methods in terms of predictive performance.

However, existing knowledge about WTP as a range and about its uses remains limited. We identify three major areas of inquiry. The first relates to the ICERANGE procedure itself and its underlying theoretical assumptions: ICERANGE may be complex for many respondents because it implicitly assumes that respondents can state their reservation prices for any given purchase probability within their individual WTP ranges. Strictly speaking, this assumption contradicts the general finding of consumers’ preference uncertainty. The procedure also assumes that purchase probability decreases linearly between the floor and the ceiling price. However, it seems more reasonable to predict that when prices are closer to the boundaries...
of the WTP range, a consumer will grow even more certain, whereas the highest uncertainty occurs in the middle of the interval, implying some S-shaped, instead of linear, function. Finally, ICERANGE features two lotteries, and respondents must estimate the potential outcomes of the second before even completing the first, which may be overtaxing. In summary, the characteristics of ICERANGE are restricted in both theoretical and practical senses.

As a second concern, no research has determined how range-based WTP estimates relate to point-based estimates at the individual consumer level. This relationship is important if we want to evaluate what has been measured previously with point-based methods—that is, the floor price, the indifference price, or the ceiling price or some other reservation price. Assuming bounded rationality, when studies measure WTP as a point, consumers should tend to average their WTP, independent of the underlying WTP definition. Such an averaging procedure implies a specific WTP definition that does not necessarily equal the assumed one. For example, imagine a point-based study that assumes that WTP relates to a choice probability of 1. Imagine further that the comparison between point- and range-based methods reveals that the former estimates indifference prices (i.e., WTP values with a choice probability of .5 instead of 1). Ignoring these differences in choice probabilities may lead to biased demand at a given price and, thus, less-than-optimal pricing decisions.

Finally, and most importantly, though conceptualizing WTP as a range instead of a single point is often theoretically reasonable, the empirical benefit of range-based methods is unclear. More simply, why should we even care about WTP as a range from an empirical perspective? The only criterion provided by Wang et al. (2007) relates to the predictive performance of individual WTP. We go a step further and ask when it is necessary to use range-based WTP estimates to derive revenue-maximizing prices on the individual consumer and aggregate levels.

With this study, we aim to address all three aspects. First, we propose a procedure for measuring WTP as a range, called “BDM-Range,” which is simpler and more flexible regarding the assumptions about WTP distributions than ICERANGE is, although it performs just as well. Second, we use this method (and compare it with ICERANGE as the state of the art) to demonstrate empirically that the point-based BDM method (Becker et al., 1964; see also Wertenbroch & Skiera, 2002) elicits the “expected” WTP, which in this case equals the midpoint of the WTP range. Third, based on this observation, we discuss implications for pricing decisions at the individual consumer level. With a Monte Carlo simulation, we determine when it is advisable to use range-based, instead of point-based, WTP estimates for an aggregate pricing decision. The results offer rules of thumb for deciding when WTP range estimates are preferable to point-based estimates, according to parameters that we discuss in detail (e.g., sample size).

In the remainder of this article, we outline the ICERANGE procedure (Wang et al., 2007) and then develop our simplified BDM-Range variant and highlight why and how it addresses existing shortcomings. We demonstrate the performance of BDM-Range in three quantitative empirical studies and extend these results based on a qualitative study, after which we offer a Monte Carlo simulation that provides a systematic comparison of the point-based and range-based methods with regard to pricing decisions on the aggregate level of demand functions. We conclude with a discussion and avenues for further research.

2. Range-based methods to measure willingness to pay and theoretical background

2.1. ICERANGE: method and restrictions

As developed by Wang et al. (2007), ICERANGE combines the lottery-based BDM method with the idea of a range, in which choice probability as a function of price decreases linearly. Similar to BDM, ICERANGE directly asks for WTP; a lottery ensures that the respondent has an incentive to indicate his or her true WTP because he or she must bear the consequences, in the form of a purchase.

ICERANGE elicits three different reservation prices: floor, indifference, and ceiling. If a lottery price, drawn randomly, is lower than the floor reservation price (i.e., the price below which a respondent would definitely buy), then the respondent must purchase at the lottery price. If the lottery price is higher than the ceiling reservation price (i.e., the price beyond which the respondent would definitely not buy), then the respondent cannot purchase. If the lottery price lies between two adjacent reservation prices, then a second lottery determines whether the respondent must buy. In this second lottery, the probability that determines the obligation to buy (which we call the “winning probability”) equals the choice probability of the price from the first lottery. Thereby, choice probability is modeled to decrease linearly between the floor (100% choice probability) and indifference (50%) prices and again linearly between the indifference and ceiling prices (0%).

We offer an example for illustration (for a graphical representation, see Wang et al., 2007, p. 206): Imagine that a respondent states 1, 2, and 4 EUR as his or her floor, indifference, and ceiling prices, respectively. The drawn price is 3 EUR. ICERANGE would require the respondent to participate in a second lottery, with a \((4−3)/(4−2) \times 50\% = 25\%\) winning probability. The respondent would also have the option to state three identical reservation prices, in which case ICERANGE is equivalent to BDM. Without repeating the detailed discussion, we note that Wang et al. (2007, p. 204) can show that “incentive compatibility is maintained over the entire range.”

As perhaps its most important contribution, ICERANGE measures WTP as a range. The BDM mechanism, which provides the basis of ICERANGE, is reliable and valid (e.g., Ding, 2007; Ding, Grewal, & Liechty, 2005; Dong, Ding, & Huber, 2010; Lusk & Schroeder, 2004; Miller et al., 2011; Wertenbroch & Skiera, 2002) and not subject to overbidding biases, as are Vickrey auctions (Kagel, 1995; Wertenbroch & Skiera, 2002). Accordingly, ICERANGE “avoids strategic bias and incentive incompatibility bias” and “predicts choice significantly better than multiple extant methods” (Wang et al., 2007, p. 210). However, we propose to modify it by solving potential problems related to its procedure and associated assumptions.

The first potential problem relates to the incorporation of two lotteries, such that the respondent simultaneously must estimate the potential outcomes of both. The first lottery, borrowed from the BDM mechanism, determines whether the respondent must purchase the product at a price below the reservation price or may not purchase at all; it also indicates whether a second lottery is necessary (i.e., if the drawn price falls within the WTP range). Because the mechanism of the potential second lottery depends on the stated reservation prices from the first step, the respondent must consider two steps at once, instead of responding sequentially, which could demand more cognitive effort and possibly even overtax the respondent. Additionally, respondents might want to minimize this double risk (which may appear unfair) or might simply not understand the process of economic punishment. These effects could decrease method acceptance in surveys.

The second potential problem relates to the second lottery, which takes place only if the first lottery draws a price within the

\[2\text{ More formally, if the first lottery determines a price between the floor price (FP) and indifference price (IP), then the obligation to buy at the lottery price (LP) occurs with probability } \left(\frac{IP−LP}{IP−FP}\right) \times 50\% = 50\% \text{ (Wang et al., 2007, p. 204). If the first lottery produces a price between the IP and ceiling price (CP), then the probability is } \left(\frac{CP−LP}{CP−IP}\right) \times 50\%. \text{ Both expressions can be derived through a simple application of the intercept theorem.}\]
WTP range. The probability of being obliged to buy at that price (i.e., winning probability) relates directly to the choice probability associated with the randomly drawn price. However, this link could affect the underlying incentive alignment. Strictly speaking, the underlying assumption is appropriate only if (1) the respondent can state an exact reservation price for any probability within his or her WTP range, and (2) the experimenter can prescribe the actual distribution of choice probabilities. However, these conditions may contradict the basic assumption of the concept of WTP as a range —namely, the existence of uncertainty. Uncertainty means that consumers have only limited cognitive access to their own preferences (and hence WTP), and their preferences can be unknown or unstable (e.g., Bettman, Luce, & Payne, 1998; Gregory, Liechtenstein, & Slovic, 1993; March, 1978). Cognitive capabilities are also not sufficient for more than bounded rationality (Simon, 1955), so when asked for their WTP, consumers must determine their latent preferences through thought and experience (e.g., Plott, 1996). We assume, in turn, that stated WTP follows some distribution around the "true," yet latent, WTP (see also Park, MacLachlan, & Love, 2011).

We extend ICERANGE as follows. First, to simplify the procedure, we unlink the two lotteries. Second, we reformulate the second lottery, such that it does not imply a specific WTP distribution within the WTP range. This scenario should be more in line with consumer uncertainty and, thus, with the key argument for conceptualizing WTP as a range, instead of as a single point.

2.2. BDM-RANGE: simplified and theoretically improved variant of ICERANGE

Similar to ICERANGE, our procedure is based on the BDM lottery, extended to WTP as a range. Therefore, we refer to it as BDM-Range. This method asks for two reservation prices, floor and ceiling. Again, respondents may state two identical reservation prices, in which case BDM-Range is equivalent to BDM.

The first lottery is the same as the first lottery in ICERANGE, but the second lottery differs. If the randomly drawn price from the first lottery falls within the range (and thus the uncertainty interval), then the respondent has a second chance to think about his or her preference. If the respondent feels that the drawn price is not attractive enough, then he or she can reject the offer to buy at that price; if he or she finds the price worthwhile, then he or she retains the possibility to buy. However, as in the BDM lottery, it is necessary to include the possibility of not being permitted to buy to avoid strategic answers. A coin toss thus determines whether the respondent may buy at the drawn price (which equals a winning probability of 50%).

This mechanism is easier to grasp than the method in ICERANGE and does not assume a particular WTP distribution in the uncertainty interval. Rather, the uncertainty a respondent may have about a particular price in the WTP range is reflected in his or her individual decision to reject or retain the possibility to buy.

We return to our example for illustration (i.e., a respondent states 1, 2, and 4 EUR as his or her floor, indifference, and ceiling prices, respectively; the drawn price is 3 EUR). Whereas ICERANGE would require the respondent to participate in a second lottery, with a 25% probability of a purchase obligation, BDM-Range offers an additional question regarding whether the respondent wants to buy at a price of 3 EUR. Only if the respondent has a choice probability of 25% at a price of 3 EUR will ICERANGE be fair. In contrast, BDM-Range does not assume a specific choice probability and simply asks the respondent to choose, according to his or her actual choice probability at that price.

Fig. 1 summarizes both methods and the differences in their procedures.

In terms of incentive alignment, BDM-Range provides respondents with an incentive not to overstate their floor reservation prices because otherwise they might be obliged to buy at higher prices. Understatements of the floor reservation price also induce punishment because a respondent might want to buy, choose to do so, but get turned down by the coin flip. If they understate their ceiling reservation prices, respondents might not get the chance to buy products they want at those prices; however, overstating their ceiling reservation prices is not punished. Consequently, the method is not fully incentive compatible because there is one price that may be overstated without economic punishment.3

This property may seem undesirable, but we argue it is not problematic for two reasons. First, a respondent who overstates his or her ceiling reservation price does not gain anything economically, so overstating this price is not a strictly dominant strategy. Second, despite the lack of punishment for overstating the ceiling reservation price, the uncertainty the respondent faces implies that a higher ceiling price increases the probability of a second lottery, which increases the risk of making a wrong decision. We also can demonstrate empirically that the mechanism does not evoke overstated ceiling prices.

Thus BDM-Range is theoretically less restrictive and simpler than ICERANGE; we attempt to prove empirically that these benefits do not come at the cost of predictive validity, compared with ICERANGE.

2.3. Relationship between range-based and point-based methods to elicit WTP

A traditional definition of WTP claims that it is the maximum price a buyer is willing to pay for a given quantity of a good (e.g., Wertenbroch & Skiera, 2002). In rational choice theory, the choice probability at a price that is equal to or less than WTP is 1, whereas the choice probability is 0 for higher prices (e.g., Hanemann, 1984). Managerial pricing decisions rely on this link between price and choice probability. However, range-based methods to measure WTP assume that uncertainty prompts the WTP to follow some distribution, so optimal pricing decisions might depend on the properties of the distribution. It is therefore necessary to explore the relationship among single-point based methods, range-based methods, and individual choice probabilities.

We argue that, when explicitly asked for floor and ceiling reservation prices, respondents try to state individual confidence intervals (Dubourg, Jones-Lee, & Looness, 1997) or WTP ranges that reflect their decision-making ambiguity (Ariely, Loewenstein, & Prelec, 2003). Therefore, they reveal variance in their individual WTP distributions. We further argue that when asked about a single WTP, consumers try to state a kind of "best guess" of WTP; thus, this single-point WTP can represent an average over their experiences and retrieved preferences. Theoretically, this "average" corresponds to the first (non-central) moment of their individual WTP distribution (expected WTP).

This averaging procedure is in line with findings from behavioral or experiential pricing literature, including price judgment concepts, such as the reference price or reference price range (e.g.,

---

3 To illustrate, we continue with our example: The respondent may choose to state a ceiling price of 6 EUR instead of 4 EUR (and thus does not tell the truth, for whatever reason). Now assume the lottery produces a price (LP) of 5 EUR. In ICERANGE, there is a [(CP − LP)/(CP − IP)] = 50% = 12.5% chance of being obliged to buy at 5 EUR, a price that exceeds the actual ceiling price. Thus, overstating true CPs can be dangerous. In BDM-Range, the respondent instead is asked whether he or she wants a 50/50 chance to buy at 5 EUR. If absolutely sure about his or her true CP (which is 4 EUR), then the respondent simply rejects this offer and definitely does not buy the product. Thus, there is always a chance to exit the game by saying "no" to the second lottery.
3. Empirical studies

3.1. General overview of quantitative empirical studies

We conducted three quantitative empirical studies to test our proposed method, BDM-Range, and to relate our results to those elicited from ICERANGE, as well as to investigate the relationship between these two range-based methods and the point-based standard BDM.

Study 1 used a glass of caffè latte as the stimulus, with no purchases; we called it a hypothetical incentive-compatible setting. Study 2 replicated this study in a real choice setting, at the point of purchase. Study 3 replicated Study 2 with a low-priced service (a ticket for a 90-minute guided tour of the historic underground in a European capital).

The reason to include a hypothetical setting (Study 1) is that real BDM with expensive, durable products or with innovations is difficult to implement and can even prevent the researcher from conducting a survey that necessitates real purchases. We believe it should not affect the relationship across the methods we compare because the point-based BDM is part of both range-based methods. However, a hypothetical bias is likely in such a setting (Cummings & Taylor, 1999; Ding et al., 2005; Voelckner, 2006), as we acknowledge in our subsequent empirical analyses.

We chose caffè latte as the first product for several reasons. First, it is a convenience product, a desirable property in a point-of-purchase setting. Second, we can assume some degree of product familiarity, which should have a positive impact on demand for most respondents. Product familiarity also helps reduce choice uncertainty (Bettman et al., 1998), which creates rather strict testing conditions. If we could demonstrate the relevance of WTP ranges for well-known products (Studies 1 and 2), then this concept would be even more relevant for products associated with higher levels of uncertainty. To test this prediction, we replicated the real-choice study with an unfamiliar service, for which we could assume a higher level of uncertainty and, thus, expect larger WTP ranges (Study 3).
3.2. Procedure and measures

Across three studies, we used ICERANGE and BDM-Range as range-based methods, and we included the foundational standard BDM procedure as the well-known and comprehensively validated point-based approach (Wertenbroch & Skiera, 2002). In Studies 2 and 3, the ICERANGE procedure involved two prices (floor and ceiling), which simplified the task and created an even more rigorous comparison with BDM-Range in terms of its applicability.5

Each study used a between-subjects design, and participants were assigned randomly to the three methods. In addition to the WTP elicitation, in each study, we included an option to revise answers (Wertenbroch & Skiera, 2002), a control choice task, and items to measure perceived method difficulty, using the four-item effort index by Menon, Raghurir, and Schwarz (1995). For an additional measurement of how participants evaluated the different elicitation methods in a real choice setting, we used a “perceived method fairness” item related to the actual transactions (Studies 2 and 3), as we detail in Appendix A.

All lottery mechanisms used the same price distribution (unknown to respondents), ranging from .59 EUR to 4.59 EUR (Studies 1 and 2) or from 7.00 EUR to 15.00 EUR (Study 3), which we extrapolated from the interval of existing prices for the study items. The expected WTP for the BDM-Range condition (and in Studies 2 and 3, for ICERANGE) was estimated as the midpoint of the floor and ceiling reservation prices, which assumed the WTP distribution was symmetric in the uncertainty interval. All studies ended with a ceiling reservation price, which assumed the WTP distribution

3.2.1. Procedure and measures

Across three studies, we used ICERANGE and BDM-Range as range-based methods, and we included the foundational standard BDM procedure as the well-known and comprehensively validated point-based approach (Wertenbroch & Skiera, 2002). In Studies 2 and 3, the ICERANGE procedure involved two prices (floor and ceiling), which simplified the task and created an even more rigorous comparison with BDM-Range in terms of its applicability.5

Each study used a between-subjects design, and participants were assigned randomly to the three methods. In addition to the WTP elicitation, in each study, we included an option to revise answers (Wertenbroch & Skiera, 2002), a control choice task, and items to measure perceived method difficulty, using the four-item effort index by Menon, Raghurir, and Schwarz (1995). For an additional measurement of how participants evaluated the different elicitation methods in a real choice setting, we used a “perceived method fairness” item related to the actual transactions (Studies 2 and 3), as we detail in Appendix A.

All lottery mechanisms used the same price distribution (unknown to respondents), ranging from .59 EUR to 4.59 EUR (Studies 1 and 2) or from 7.00 EUR to 15.00 EUR (Study 3), which we extrapolated from the interval of existing prices for the study items. The expected WTP for the BDM-Range condition (and in Studies 2 and 3, for ICERANGE) was estimated as the midpoint of the floor and ceiling reservation prices, which assumed the WTP distribution was symmetric in the uncertainty interval. All studies ended with a ceiling reservation price, which assumed the WTP distribution was symmetric in the uncertainty interval. All studies ended with the voluntary providing of sociodemographic data.

3.3. Samples

Study 1 used a consumer sample (n = 168) and was administered with an online survey that described approaching a coffee bar. No purchases took place. The lottery mechanism thus had no monetary consequences for the participants. However, we distributed three online food store coupons as incentives to encourage participation. Eight respondents were excluded due to missing data or obvious lack of interest (i.e., judged by completion time and answering pattern), leaving n = 160 data sets for our analyses.

Study 2 took place at a coffee bar in a coworking facility over a period of 12 successive working days. The actual prices displayed for a glass of cappuccino and similar products were removed at the time of purchase. Customers who ordered a cappuccino were asked to participate; only 4 customers insisted on buying the product without participating. An experimenter explained that the price for a cappuccino was not fixed and that the survey included a lottery that would determine the price. Participants were also made aware that they would be bound to the lottery outcome, including an obligation to pay or the possibility of not being allowed to buy the product, but that they would definitely not pay more than they actually wanted. We excluded respondents if they had confronted the same WTP elicitation method more than once, though 21 respondents participated twice and undertook two different methods. A MANOVA on measured reservation prices showed no significant influence of encountering another method previously, compared with whether it was the first time a respondent participated in the experiment (p > .39 for all methods). Together, this procedure created a sample size of 119 responses. All data were collected by an experimenter, using a personal computer at the coffee bar.

Study 3 was administered online again; tickets for the underground tour regularly sell online. Consumers in a European capital were asked to participate via e-mail, and the invitation indicated that in the survey to follow, they would have a chance to purchase a ticket for a 90-minute guided tour, but only at a price they personally accepted and expressed themselves as willing to pay. After an illustrated description of the tour, participants voluntarily provided their names and e-mail addresses, so we could contact them in case of a purchase. In total, we collected n = 124 completed responses, of which 44 included an obligation to purchase. The tickets were distributed by e-mail, along with the request to pay the individual price generated by the survey. Five respondents refused to do so, leaving n = 119 questionnaires for our analyses.

In each study, MANOVAs (Pillai’s Trace: p > .50) for age (p > .84), sex (p > .17), and income (p > .33) showed that the three experimental groups did not differ in terms of their sociodemographics; thus, we could assume that our comparisons of WTP elicitation methods did not suffer from unbalanced samples.

3.4. Results of the quantitative studies

This section is organized according to our main research questions, to highlight the similarities and differences across Studies 1–3. Furthermore, we summarize all the statistics we discuss in Table 1.

3.4.1. Predictive performance and internal validity

To assess the predictive performance of each WTP elicitation method, we calculated shift-in-choice likelihood (SCL) values. Individual SCL was the difference between an observed choice (1 if the control task was a choice and 0 otherwise) and the predicted choice likelihood that depended on the previously administered WTP elicitation method (see Inman et al., 1990). Therefore, this SCL could be positive (underestimation of choice probability) or negative (possibly deviating from the “1—hit rate” calculation; Wang et al., 2007). In the aggregate level, SCL was the average of individual SCL across respondents, so under- and overestimation might balance, which was obviously undesirable. We thus considered the absolute values of individual SCL. Larger values indicated worse predictive performance, in that they implied a biased estimation of choice probabilities.

In Studies 1–3, BDM-Range achieved lower absolute SCL values than ICERANGE, although these differences were not significant (pairwise t-tests, p > .11). Standard BDM outperformed ICERANGE in Studies 2 and 3 and BDM-Range only in Study 3 (p < .05).

To assess internal validity, we adopted Wertenbroch and Skiera’s (2002) method. We derived the empirical demand function (= observed demand) by first calculating, for each respondent, his or her WTP as the midpoint of individual WTP range (= ceiling price minus floor price, divided by 2). Observed demand as a function of price p, then, was the proportion of respondents who bought at a price of p ≤ WTP (see also Miller et al., 2011). Logit analyses of the purchase probabilities, Pr(buy|p) = exp(a + b × p) / [1 + exp(a + b × p)], revealed predicted demand. We then could compare the correlations of the observed and predicted demand using Fisher Z-transformed correlations. All parameter estimates, Pearson’s r, and the respective Z-transformations appear in Appendix B.

In each study, we found high correlations of observed and predicted demand for any method (r > .98), but the range-based methods revealed better fit and thus higher internal validity than BDM in Studies 1 and 2 (p < .001). Study 3 yielded comparable levels of internal validity for all elicitation methods. The ICERANGE and BDM-Range approaches never differed significantly (p > .22).

3.4.2. Estimation of WTP ranges

In each study, both range-based methods revealed positive WTP ranges, in that the floor and ceiling prices differed significantly

---

5 Wang et al. (2007, p. 210) showed that ICERANGE can be simplified to elicit two instead of three reservation prices.
(p < .001 for any comparison), but neither differed significantly across methods, as was also the case for expected WTP (p > .20). Therefore, across settings (hypothetical, real) and products (caffè latte, ticket for a guided tour), we provide evidence of the existence and stability of a WTP range, independent of the elicitation procedure.

This result is less evident than it seems. Although both methods used similar questions, the different lottery mechanisms could have produced different results. The finding that ICERANGE and BDM-Range produced the same ceiling prices demonstrates that, though BDM-Range does not punish an overstatement of ceiling prices economically, it is not a problem in an empirical sense.

To compare WTP ranges across settings (Studies 1 and 2) and products (Studies 2 and 3), we also had to consider potentially different levels of expected WTP. This difference was quite obvious across products. For the research settings, we observed significantly higher expected WTP in a hypothetical setting than in a real choice setting, independent of the elicitation method (p < .01), in support of a hypothetical bias. Therefore, we compared relative WTP ranges, calculated as the WTP range divided by expected WTP.

Relative WTP ranges were larger in the hypothetical than in the real incentive-compatible setting (ICERANGE: .67 vs. .37, p = .03; BDM-Range: .49 vs. .31, p < .001), which implied that the WTP range (not just WTP itself) was affected by a hypothetical bias. In the real choice setting, relative WTP ranges were much larger for the service than for the convenience product (ICERANGE: .67 vs. .37, p < .001; BDM-Range: .68 vs. .31, p < .001), as we expected because of the higher preference and product uncertainty associated with the rather unfamiliar service, compared with the familiar glass of caffè latte.

3.4.3. Perception of methods

Regarding perceived method difficulty, we compared the means over all scale items. In Studies 1 and 2, BDM-Range was not perceived as easier than ICERANGE but just as easy as BDM (p > .16 in any comparison); ICERANGE seemed slightly more difficult than BDM (p < .09). In Study 3, BDM-Range was perceived as the easiest method (p < .01), although BDM-Range is obviously more elaborate than standard BDM.

Table 1
Empirical results from different WTP elicitation methods.

<table>
<thead>
<tr>
<th>Study 1: Hypothetical incentive-aligned study (caffè latte)</th>
<th>ICERANGE (n = 46)</th>
<th>BDM-Range (n = 56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (EUR)</td>
<td>3.13 (1.13)</td>
<td>2.40 (1.02)</td>
</tr>
<tr>
<td>(SE) (EUR)</td>
<td>2.37 (1.03)</td>
<td>3.14 (1.24)</td>
</tr>
<tr>
<td>Floor</td>
<td>3.27 (1.15)</td>
<td>3.87 (1.24)</td>
</tr>
<tr>
<td>Indifference</td>
<td>4.18 (1.15)</td>
<td></td>
</tr>
<tr>
<td>Ceiling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study 2: Real incentive-aligned (point-of-sale) study (caffè latte)</th>
<th>ICERANGE (n = 40)</th>
<th>BDM-Range (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (EUR)</td>
<td>2.62 (.51)</td>
<td>2.22 (.57)</td>
</tr>
<tr>
<td>(SE) (EUR)</td>
<td>2.07 (.49)</td>
<td>2.62 (.61)</td>
</tr>
<tr>
<td>Floor</td>
<td>2.54 (.73)</td>
<td>3.03 (.78)</td>
</tr>
<tr>
<td>Expected</td>
<td>3.01 (.73)</td>
<td></td>
</tr>
<tr>
<td>Ceiling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study 3: Real incentive-aligned (point-of-sale) study (service)</th>
<th>ICERANGE (n = 44)</th>
<th>BDM-Range (n = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (EUR)</td>
<td>9.35 (4.24)</td>
<td>6.34 (3.39)</td>
</tr>
<tr>
<td>(SE) (EUR)</td>
<td>6.06 (3.23)</td>
<td>9.61 (4.42)</td>
</tr>
<tr>
<td>Floor</td>
<td>9.02 (3.60)</td>
<td>12.89 (6.05)</td>
</tr>
<tr>
<td>Indifference</td>
<td>11.97 (4.70)</td>
<td></td>
</tr>
<tr>
<td>Ceiling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: SCL = shift-in-choice likelihood (see Inman, McAlister, & Hoyer, 1990; Wang et al., 2007). SE = standard error. Lower scores for difficulty and fairness support a specific method. Difficulty: 1 = “is not difficult”, …, 5 = “is very difficult”. Fairness: 1 = “is fair”, …, 5 = “is unfair”.

(p < .001 for any comparison), but neither differed significantly across methods, as was also the case for expected WTP (p > .20). Therefore, across settings (hypothetical, real) and products (caffè latte, ticket for a guided tour), we provide evidence of the existence and stability of a WTP range, independent of the elicitation procedure.

This result is less evident than it seems. Although both methods used similar questions, the different lottery mechanisms could have produced different results. The finding that ICERANGE and BDM-Range produced the same ceiling prices demonstrates that, though BDM-Range does not punish an overstatement of ceiling prices economically, it is not a problem in an empirical sense.

To compare WTP ranges across settings (Studies 1 and 2) and products (Studies 2 and 3), we also had to consider potentially different levels of expected WTP. This difference was quite obvious across products. For the research settings, we observed significantly higher expected WTP in a hypothetical setting than in a real choice setting, independent of the elicitation method (p < .01), in support of a hypothetical bias. Therefore, we compared relative WTP ranges, calculated as the WTP range divided by expected WTP.

Relative WTP ranges were larger in the hypothetical than in the real incentive-compatible setting (ICERANGE: .67 vs. .37, p = .03; BDM-Range: .49 vs. .31, p < .001), which implied that the WTP range (not just WTP itself) was affected by a hypothetical bias. In the real choice setting, relative WTP ranges were much larger for the service than for the convenience product (ICERANGE: .67 vs. .37, p = .001; BDM-Range: .68 vs. .31, p < .001), as we expected because of the higher preference and product uncertainty associated with the rather unfamiliar service, compared with the familiar glass of caffè latte.

3.4.3. Perception of methods

Regarding perceived method difficulty, we compared the means over all scale items. In Studies 1 and 2, BDM-Range was not perceived as easier than ICERANGE but just as easy as BDM (p > .16 in any comparison); ICERANGE seemed slightly more difficult than BDM (p < .09). In Study 3, BDM-Range was perceived as the easiest method (p < .01), although BDM-Range is obviously more elaborate than standard BDM.
With our fairness measure, we found no differences across methods in Study 3 (p > .42), but we uncovered higher scores for BDM-Range than for BDM or for ICERANGE in Study 2 (p < .05). The higher fairness scores for BDM-Range may have emerged because it allowed respondents to choose whether they wanted to participate in a second lottery, whereas ICERANGE involved a deterministic procedure in which participants could not directly intervene. Presumably, respondents preferred to decide by themselves, rather than depend on a procedure that they felt they could not influence. Although higher fairness was only partly supported (in one of two studies), our overall results imply the applicability of BDM-Range in an actual transaction setting.

3.4.4. Relationship between point-based and range-based WTP estimates

In all three studies, pairwise t-tests showed that mean WTP elicited through BDM and the means of the expected WTP values elicited through either ICERANGE or BDM-Range were all similar (p > .47). The floor and ceiling reservation prices of any range-based method differed significantly from the point-based BDM estimates (p < .02).

Taken together, these results confirmed our proposition that point-based WTP elicitation methods (standard BDM) will estimate the expected WTP of individual WTP distributions (either ICERANGE or BDM-Range). This result held across research settings (hypothetical and real incentive-compatible; Studies 1 and 2), independent of the relative WTP size arising from different levels of uncertainty (familiar convenience product, low-priced but unfamiliar service; Studies 2 and 3). To substantiate this empirical observation, independent of the methods employed, we conducted a qualitative follow-up study.

3.5. Qualitative study of different WTP elicitation formats (range- versus point-based)

To collect the data for the qualitative study, we used an online survey, which n = 41 students completed. The coffee bar and the glass of café latte from Studies 1 and 2 provided the setting. We included closed- and open-ended questions about WTP. The closed-ended question asked, “Please state the maximum price you would pay for the glass of café latte’’ (point-based format), whereas in the range-based sequence, respondents indicated their floor and ceiling prices (“Please state the maximum price up to which you would definitely buy the café latte” and “Please state the price beyond which you would definitely no longer buy the café latte”). In response to an open-ended question, respondents indicated the process they used to arrive at their answers (“Please describe how you made your decision to state exactly this price exactly”). Each participant completed both question sequences, although we systematically varied the order, which had no effect on the frequencies of different answer categories (MANOVA, p > .18 for all six categories).

Our focus was the open-ended question. We developed a categorization system, according to our inductive extraction of 78 textbox answers, using Mayring’s (2000) framework in line with inductive processes in psychological text processing (van Dijk, 1980). After we defined our research scope based on a theoretical background (step 1), the material was coded and paraphrased by three independent coders. Categories were formulated and labeled (step 2). After 30% of the material had been coded, we revised the category scheme (step 3), and in a feedback loop, we applied the new categories to the original material (step 4). We then reduced our scheme to the main answer categories (Miles & Huberman, 1994; Strauss & Corbin, 1990) and coded the material again (step 5). Intercoder differences were resolved by reference to the majority opinion; the percentage of agreement was 85.6% (Bettman & Park, 1980). We eliminated any categories for which both frequencies (point- or range-based elicitation) were less than 10%.

This process resulted in three main answer categories: (1) “The maximum reservation price stated here is an average of some kind” and an obvious rip-off.” (Upper threshold-which price is a total robbery (n = 40))

<table>
<thead>
<tr>
<th>Answer categories with respect to the open question (with stated text examples)</th>
<th>Point-based format (n = 39)</th>
<th>Range-based format (n = 40)</th>
<th>MANOVA on category frequencies</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The maximum reservation price stated here is an average of some kind”</td>
<td>53.8%</td>
<td>15.0%</td>
<td>15.51</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>“I mentally took all prices I usually pay. I intuitively started with 2 EUR, and then I thought, I have also paid 2.5, 3, 3.5 or 4 EUR. But 4 EUR felt too expensive so I went back to 3.5 EUR.”</td>
<td>30.8%</td>
<td>75.0%</td>
<td>18.82</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>“I made a good average of prices already known to me.”</td>
<td>3 EUR is too much for a coffee 2.50 relatively cheap. So I took a rough average as my willingness to pay.”</td>
<td>3.5%</td>
<td>25.0%</td>
<td>6.39</td>
<td>.01</td>
</tr>
<tr>
<td>“The maximum reservation price stated is associated with negative feelings”</td>
<td>Every price above 5 EUR is rip-off! Even 5 EUR is much, but I could just take it</td>
<td>“23.0 EUR is still a fair price. Anything more than 3 EUR is too expensive for me.”</td>
<td>“Upper threshold-which price is a total robbery” and an obvious rip-off.”</td>
<td>Upper price-threshold: beyond this price I would feel short-changed.”</td>
<td>“Upper price: no coffee is worth more than 4 EUR.”</td>
</tr>
</tbody>
</table>

Notes: The point-based format tasks: “Please state the maximum price you would pay for the glass of café latte’’. The range-based format asks two questions: “Please state the maximum price up to which you would definitely buy the café latte’’ and “Please state the price beyond which you would definitely no longer buy the café latte’’. The frequencies of the answer categories do not have to add up to 100%. Frequency differences indicate differing processing behavior caused by the different elicitation formats. Significant differences (p < .05) are bold.

This finding reinforced our empirical observations from Studies 1-3 that single point-based WTP elicitation yielded an indifference price or expected WTP, even when we explicitly asked for the maximum price a consumer was willing to pay for a specific product (as in any standard BDM). The finding that more respondents associated negative feelings with the reservation price in the range-
...based, rather than in the point-based, format substantiated this observation.

3.6. Summary of the empirical results

Taken together, our studies shed light on the theoretical and empirical relationships across point-based and range-based WTP elicitation formats. We confirmed across all scenarios (hypothetical and real research settings; different relative WTP sizes) of the three quantitative studies and in the additional qualitative study that point-based WTP elicitation yielded the expected WTP of the range-based WTP distribution. For symmetric WTP distributions, the case we consistently observed here, expected WTP equaled the indifference price (which is at the same time the mean of floor and ceiling price). Wang et al. (2007, p. 202) claimed that point-based measures would depend on the type of elicitation question (i.e., asking respondents for an indifference threshold vs. asking for a maximum threshold). In contrast, we do not attribute this effect to the phrasing of the elicitation question but to the elicitation format itself (i.e., point-based vs. range-based). Our conclusion helps us to understand how extant point-based WTP elicitation relates to WTP ranges and provides a guideline for the results that any inclusion helps us to understand how extant point-based WTP elicitation should theoretically retrieve. We next discuss how our main result of the relationship between point- and range-based WTP elicitations might affect optimal pricing decisions.

4. Pricing decisions regarding WTP range

4.1. Pricing decisions on the individual consumer level

In subsequent analyses on optimal pricing decisions, we used the criterion of revenue (instead of profit) maximization. That is, we assumed variable costs of 0, which seemed justifiable for the goods used in Studies 1–3 (a glass of caffè latte, a ticket for a guided tour). Although this assumption is not generally applicable, it was consistent with our focus on demand estimation, based on WTP data: Demand is independent of costs.

The expected revenue for consumer i is \( r_i(p) = q_i(p) \times p \), where \( q_i(p) \) is the individual buying probability at \( p \), and \( p \) is the price of the sold unit. Individual buying probability (IBP) is the survival function \( q_i(p) = IBP(p \leq WTP_i) \), or the probability that a consumer’s WTP is equal to or greater than price \( p \) (Miller et al., 2011).

To compare pricing decisions with either point- or range-based WTP elicitation, imagine a (single) consumer with a WTP of 100 and a range of 80, which we can assume to be symmetrically distributed around 100 (see Fig. 2, upper panel). The point-based case assumes this consumer will buy with certainty at prices below 100 and definitely not buy at prices beyond 100. In contrast, the range-based case assumes the consumer will buy with certainty at prices below 60 (= 100 – 80/2), definitely not buy at prices beyond 140 (= 100 + 80/2), and buy at \( p = 100 \), with a probability of 50% (Fig. 2, middle panel). The revenue-maximizing price \( p_{\text{point}}^{\ast} \) in the point-based case is obviously 100, leading to a revenue \( r_{\text{point}}^{\ast} \) of 100. In the range-based case, this example yields a revenue-maximizing price \( p_{\text{range}}^{\ast} \) of 77.9 with a revenue \( r_{\text{range}}^{\ast} \) of 68.8 (Fig. 2, lower panel).

Our example indicates that conceptualizing an individual consumer’s WTP as a distribution of choice probabilities can lead to different pricing decisions than a point-based conceptualization. More importantly, it is possible to show that, for any symmetric unimodal WTP distribution, ignoring the individual WTP range leads to overpricing (a result for which Fig. 2 serves as an illustration). We provide a formal proof for the general result in Appendix C.1.

Fig. 2. Conceptual differences in individual consumer pricing.
Table 3
Illustrative example for pricing decisions on the aggregate (demand) level

<table>
<thead>
<tr>
<th>Segment 1 (homogeneous in WTP across consumers)</th>
<th>Segment 2 (heterogeneous in WTP across consumers)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point-based format</strong></td>
<td><strong>Range-based format</strong></td>
</tr>
<tr>
<td>Consumer 1</td>
<td>100</td>
</tr>
<tr>
<td>Consumer 2</td>
<td>100</td>
</tr>
<tr>
<td>Consumer 3</td>
<td>100</td>
</tr>
<tr>
<td><strong>Individual-level WTP</strong></td>
<td><strong>Individual-level buying probability at a price of 100</strong></td>
</tr>
<tr>
<td>Consumer 1</td>
<td>1</td>
</tr>
<tr>
<td>Consumer 2</td>
<td>1</td>
</tr>
<tr>
<td>Consumer 3</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Expected demand at a price of 100</strong></td>
<td></td>
</tr>
<tr>
<td>Consumer 1</td>
<td>1.5</td>
</tr>
<tr>
<td>Consumer 2</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Demand bias (basis: range-based demand)</strong></td>
<td></td>
</tr>
<tr>
<td>Segment 1 (homogeneous in WTP across consumers)</td>
<td>(3–1.5) / 1.5 = 1 (100%)</td>
</tr>
<tr>
<td>Segment 2 (heterogeneous in WTP across consumers)</td>
<td>(2–1.5) / 1.5 = 0.33 (33%)</td>
</tr>
</tbody>
</table>

4.2. Pricing decisions on the aggregate (demand) level

Optimal pricing decisions usually refer to an aggregate consumer level (e.g., Cameron & James, 1987; Jedidi et al., 2003; Leeflang & Wittink, 2000; Miller et al., 2011; Voelckner, 2006). It is necessary to use a valid and reliable WTP elicitation method on the individual level to derive valid demand curves, but it is not clear whether restricting individual WTP measures to the expected individual WTP values (as in point-based methods) makes a difference at the demand function level.

Using the central limit theorem, we could argue that a sufficiently large sample would diminish the differences between aggregating individual WTP distributions and aggregating the means of the individual WTP distributions. On the aggregate level of demand functions, point-based WTP elicitation methods then should yield the same results with respect to optimal prices as range-based methods.

However, if WTP variance across consumers is low, but individual-level ranges are considerable, then the individual-level differences will add up and alter the result on the aggregate level, such that we may expect biased demand on the aggregate level.

To illustrate this intuition, we extend the example from Section 4.1: Imagine two segments of three consumers each. The first segment has no WTP variance across consumers; it includes three identical consumers with considerable WTP ranges (with WTP = 100; WTP range = 80). The second segment, instead, is heterogeneous in WTP across consumers; it includes three consumers with different WTPs (60, 100, and 140, respectively) but still with identical WTP ranges of 80. As Table 3 shows, the expected demand at a price of 100 is 1.5 in both segments when range-based WTP elicitation is used. In contrast, expected demand is 3 in the homogeneous segment (2 in the heterogeneous segment) when a point-based method is used. This result shows a demand bias in both segments, with a much larger bias in the homogeneous segment (100% vs. 33%).

Therefore, the question remains: When does it make a difference to use point-based or range-based WTP estimates as the basis for optimal pricing decisions? Specifically, when do (1) WTP variance across consumers, (2) WTP variance on the individual level (related to WTP range), and (3) sample size matter? We used a Monte Carlo simulation study to answer these questions.

4.3. Simulation study design

To determine whether point- and range-based methods yield different optimal prices, according to the aggregate information of demand functions, we generated different samples of individual expected WTP as a starting point for both methods because our empirical studies suggest that point-based methods elicit the expected WTP of range-based methods. We then aggregated individual expected WTP into demand functions using WTP estimates from either a point-based or a range-based method. We used these demand functions to predict revenue-maximizing prices, which we could compare across both WTP elicitation methods, as well as to determine the bias in revenue prediction with point-based methods. The design factors included the sample size, n; the population-level standard deviation of expected individual WTP, σ_pop, which reflects variance in expected WTP across consumers; and the individual-level standard deviation of individual WTP, σ_pop,i, which reflects the WTP range.

In the overall study structure, as we depict in Fig. 3, we divided these steps into several substeps. We conducted the simulation in the R computing environment.

In Step 1a, we generated a WTP sample. Without loss of generality, we set the mean expected WTP to μ = 100 and assumed expected WTP in the population to be normally distributed with N(100; σ^2_pop). We used four levels of σ_pop—10, 20, 30, and 40—to account for different levels of heterogeneity in WTP across consumers (Studies 2 and 3 yielded values that corresponded to σ_pop ≈ 20 and σ_pop ≈ 40, respectively). A sample consisted of n individual expected WTPs; the six variations of n were 25, 50, 100, 200, 400, and 800.

In Step 1b, we generated WTP ranges (floor and ceiling prices). We applied 10 levels of individual-level WTP variation to each subject. We assumed that individual-level WTP was normally distributed around latent WTP, for each individual consumer i.7 We truncated at WTP = μ ± 2×σ_pop,i, where σ_pop,i is the individual-level standard deviation of consumer i. The floor price was set to F_P = WTP – 2×σ_pop,i, and the ceiling price was C_P = WTP + 2×σ_pop,i.8 We thus ensured that roughly 95% of all nontruncated, normally distributed, individual

---

7 Although we do not claim asymmetric distributions are impossible, the empirical results from Studies 1–3 suggest a symmetric distribution. For consistency, we assume such a distribution in the simulation.

8 Before truncation, the floor (ceiling) price relates to a choice probability of approximately 97.5% (2.5%). After truncation, it is 100% (0%), a minor correction. Wang et al. (2007) classified a choice probability of 10% for CP as small.
WTP values were within the individual WTP range. Without loss of generality, we used a constant level of \( \sigma_{\text{ind},i} \) for all individual consumers in a specific scenario. The 10 variations of \( \sigma_{\text{ind},i} \) were {0.0001, 2, 4, 6, 8, 10, 12, 14, 16, 18} (the Study 2 value corresponded to \( \sigma_{\text{ind},i} \approx 8 \); the Study 3 value corresponded to \( \sigma_{\text{ind},i} \approx 16 \))\(^8\), such that we had \( 4 \times 6 \times 10 = 240 \) scenarios.

In Step 2, we calculated aggregated demand for each of the 240 scenarios by adding all individual buying probability functions and transforming them into relative numbers, such that demand decreased from 1 to 0. Individual buying probability functions depended, in turn, on the method applied to estimate WTP. Condition 1 represented a point-based elicitation method. We assumed that it correctly retrieved the expected individual WTP, an assumption justified by Studies 1–3. The individual buying probability of the point-based methods could be represented as:

\[
IBP_i(p) = \begin{cases} 
1, & \text{if } p \leq \text{WTP}_i \\
0, & \text{if } p > \text{WTP}_i 
\end{cases}
\]

Condition 2 represented a range-based elicitation method. We assumed that it correctly retrieved the WTP range and that the individual WTP followed a truncated normal distribution. Here, IBP took the following form:

\[
IBP_i(p) = \begin{cases} 
1, & \text{if } p \leq \text{FP}_i \\
1 - F_p(p, \text{FP}_i, \text{CP}_i, \text{WTP}_i, \sigma_{\text{ind},i}), & \text{if } \text{FP}_i < p \leq \text{CP}_i \\
0, & \text{if } p > \text{CP}_i 
\end{cases}
\]

with the parameters as specified previously. Furthermore, \( F \) denotes the truncated normal cumulative distribution that depends on \( \Phi \), the cumulative distribution function of \( N(0,1) \):

\[
F \left( p, \text{FP}_i, \text{CP}_i, \text{WTP}_i, \sigma_{\text{ind},i} \right) = \Phi \left( \frac{p - \text{WTP}_i}{\sigma_{\text{ind},i}} \right) - \Phi \left( \frac{\text{FP}_i - \text{WTP}_i}{\sigma_{\text{ind},i}} \right) \quad \text{FP}_i \leq p \leq \text{CP}_i 
\]

The numerator of \( F \) refers to the cumulative probability up to any price \( p \), recognizing that the distribution starts at the floor price, \( \text{FP}_i \). The denominator is the standardization factor that ensures \( F \) is a cumulative distribution function.

Step 3 entailed fitting the demand functions. To parameterize demand in each scenario for each elicitation method, we used a logit model of the form \( d(p) = \exp(\alpha + \beta \times p) / [1 + \exp(\alpha + \beta \times p)] \) (Brown, Champ, Bishop, & McCollum, 1996; Miller et al., 2011; Wertenbroch & Skiera, 2002).

In Step 4, we calculated optimal prices and revenue. As argued previously, we assumed revenue maximization as the company’s goal. Using estimated demand from Step 3, as denoted by \( d(p) \), we determined the price \( p^* \) that maximized revenue, \( r(p) = d(p) \times p \). We also calculated the optimal revenue, \( r(p^*) \), and the revenue predicted for the point-based optimal price when using the range-based demand function (for details on solving the optimization problem using the Lambert W function, see Miller et al., 2011, web Appendix D).

With Step 5 we calculated price and revenue differences. That is, for each of the 240 scenarios, we calculated differences in optimal prices across both WTP elicitation methods. As a measure of revenue

\[
\text{Step 1a: Generating a WTP sample} \\
\text{A random sample of WTP with sample size } n \text{ is generated (normally distributed with } N(100; \sigma_{\text{pop}}^2). \text{ Population-level standard deviation: } \sigma_{\text{pop}} = \{10, 20, 30, 40\} \text{ Sample size: } n = \{25, 50, 100, 200, 400, 800\} \text{ \# of scenarios: } 4 \times x \times 6 \times 10 = 240.
\]

\[
\text{Step 1b: Generating WTP ranges} \\
\text{Floor and Ceiling Prices are generated for the sample at WTP}_i \pm 2 \times \sigma_{\text{ind},i}. \text{ Individual-level standard deviation: } \sigma_{\text{ind}} = \{0.001, 2, 4, 6, 8, 10, 12, 14, 16, 18\} \text{ \# of scenarios: } 10.
\]

\[
\text{Step 2: Calculating demand} \\
\text{Aggregated relative demand data is calculated for both types of WTP elicitation methods as the average over individual buying probabilities. \# of scenarios: } \text{240}. \text{ \# of repl. at } 1000 \text{ repl.}
\]

\[
\text{Step 3: Fitting demand functions} \\
\text{Logit functions are fitted to the relative demand data. \# of scenarios: } 1000 \text{ repl.}
\]

\[
\text{Step 4: Calculating optimal prices and revenue} \\
\text{Revenue maximizing prices and respective revenues are calculated based on the logit functions, as well as range-based revenue at the point-based price. \# of scenarios: } 240 \text{ \# of repl. at } 1000 \text{ repl.}
\]

\[
\text{Step 5: Calculating price and revenue differences} \\
\text{The difference in optimal prices and "revenue bias" measure the point-based method’s deviation from the theoretical superior range-based method. \# of scenarios: } 240 \text{ \# of repl. at } 1000 \text{ repl.}
\]

\[
\text{Step 6: Replications} \\
\text{All previous steps are replicated 1000 times. \# of scenarios: } 240 \text{ \# of repl. at } 1000 \text{ repl.}
\]

\[
\text{Step 7: Aggregating replication results} \\
\text{Average over price differences are calculated for both methods. \# of scenarios: } 240 \text{ \# of repl. at } 1000 \text{ repl.}
\]
Fig. 4. Simulation results (relative differences between point- and range-based WTP elicitation methods).

Notes: n denotes sample size; SD denotes standard deviation; mean population WTP is 100.
bias for the point-based method, we calculated the difference between the revenues for the point-based optimal price, according to (1) the (biased) demand function derived through the point-based method and (2) the demand function derived through the range-based method.

Step 6 involved replications of the previous steps, 1000 times. Finally, in Step 7 we aggregated the replication results. For each scenario, we calculated the average differences of optimal prices and the average revenue bias over all replications.10

**4.4. Simulation study results**

We illustrate the simulation results in Fig. 4. In the first and second rows, we display the price and revenue differences, respectively, from simulation Step 5. We use relative numbers, with the range-based values as the basis for comparison. In so doing we differentiate among various sample sizes and population-level heterogeneity in expected WTP. The complete data underlying these figures appear in Appendix D.

Across all tested scenarios, the maximum relative difference in optimal prices was 7.7%, with point-based \( p^*_{\text{point}} = 85.38 \) being larger than range-based \( p^*_{\text{range}} = 79.27 \). The maximum revenue bias amounted to 20.7% (with a revenue of 79.49 when point-based demand was used and a revenue of 65.87 when actual range-based demand was used). Both maximum differences occurred in scenario \( \sigma_{\text{pop}} = 10, n = 25, \) and \( \sigma_{\text{ind}} = 18. \) However, we note that in any simulated scenario, we found differences in the optimal prices across the WTP elicitation methods, as well as a revenue bias for the point-based approach (we provide a formal proof of these results in Appendix C.2).

Therefore, range-based methods generally should be preferred over point-based methods to estimate demand and derive optimal prices because conceptually, point-based methods yield biased demand on both individual and aggregate levels.

Formally, the differences across methods in Appendix C.2 depend on the Lambert W function, for which no expression in terms of elementary functions exists. Nevertheless, it is interesting to know how the values of the design factors used in our simulation “create” fairly systematic differences across the methods in Fig. 4. We focus on the revenue bias, which relates to the company’s decision criterion.

On a descriptive level, we first detected a sample size effect, in that greater (smaller) sample sizes led to a smaller (greater) revenue bias. Roughly speaking, doubling a sample size halved the reduction in revenue bias, which suggests an inverse proportional relationship: revenue bias \( \sim 1/n \). Furthermore, we observed an uncertainty effect, such that higher (lower) individual WTP ranges produced an increasing (decreasing) revenue bias. Specifically, Fig. 4 (second row) suggests revenue bias is roughly equal to WTP ranges squared. Heterogeneity in expected WTP across consumers also interacted with these two effects. In particular, the sample size effect increased with greater heterogeneity, whereas the uncertainty effect diminished when heterogeneity increased. Taken together, these observations suggest that revenue bias can be modeled as a function of the design factors, as follows:

\[
\text{revenue bias} = b_0 + \frac{b_1}{n} + b_2 \left( \sigma_{\text{ind}} \right)^2 + \frac{b_3}{n} \sigma_{\text{pop}} + \frac{b_4}{\sigma_{\text{pop}}} \left( \sigma_{\text{ind}} \right)^2 + \text{error}
\]

An ordinary least squares regression on the mean revenue-bias values for all 240 scenarios yields \( b_0 = .001, b_1 = .203, b_2 = -.0001, b_3 = .018, \) and \( b_4 = .007, \) with an R-square of .994. The model approximates the formally exact revenue bias very well. Finally, though \( b_2 \) is negative, for a given level of heterogeneity in WTP across consumers \( \sigma_{\text{pop}}, \) the overall effect of WTP ranges \( \sigma_{\text{ind}} \) remains positive.

From a practical perspective, we also might wonder if, in some conditions, revenue bias is negligible, such that point-based methods are sufficient. Considering our observations in combination and assuming a particular level of tolerated revenue biases, we derive some suggestions about when point-based methods are acceptable and when they must be avoided.

---

10 The code used in the R computing environment and the simulated data are available from the authors upon request.
Specifically, we assume a tolerance level of 1%, and we investigate extreme values for our three parameters, $c_{\text{pop}}$, $n$, and $\sigma_{\text{ind}}$. The results in Fig. 5 reveal when a point-based method is sufficient (i.e., maximum revenue bias of 1%).

That is, point-based methods are sufficient only if the WTP ranges are virtually nonexistent (i.e., close but not equal to 0; in our study, $\sigma_{\text{ind}} = 0.001$). Even this condition is not enough if heterogeneity in expected WTP across consumers is high (in which case a large sample size is needed).

We expect that consumers are at least somewhat uncertain about their preferences and product performances, so this result delivers a compelling argument for the use of range-based methods to elicit WTP. The assumption of certain levels for these kinds of uncertainty is especially likely for innovations, for which a high degree of performance uncertainty arises because of the novelty and lack of consumer experience with the product; fast-moving consumer goods, for which low involvement is likely and which may create ambiguity in the buying decision; or complex products, which are characterized by many attributes that increase the likelihood of attribute conflicts.

5. Conclusion

5.1. Contributions and implications

Similar to Wang et al. (2007), we argue that willingness to pay is a range and not a single point, according to both theoretical and empirical perspectives. In addition, we extend knowledge about WTP as a range in several directions. We shed light on past research that has described WTP as a single point. Through several studies, we provide empirical evidence that “traditional” point-based methods measure expected WTP and neglect individual uncertainty, which exists even for daily-use products, such as glasses of caffè latte. We thus contribute to a better understanding of prior studies that have applied point-based methods, and we embed range-based methods more clearly in WTP elicitation literature.

We obtain this result by not only applying ICERANGE as the state-of-the-art method to elicit WTP as a range but also by proposing a simplified method (BDM-Range). Similar to ICERANGE, this procedure is based on the well-established BDM mechanism; however, it is simpler by construction, sometimes perceived as fairer, and it is less restricted in terms of theoretical assumptions. It also reaches comparable levels of predictive performance and internal validity. Our method, therefore, is superior for practical applications that aim to measure consumers’ willingness to pay.

In line with the observation that point-based methods measure expected WTP, we show through a simulation that, in certain cases, measuring WTP as a range matters particularly for optimal pricing decisions. The existence of a WTP range means that biased decisions on the individual level emerge when researchers use a point-based method. These biased decisions then translate to the aggregate level. The differences between approaches in terms of revenue bias increase with greater WTP ranges and smaller samples; however, these effects interact with sample heterogeneity in expected WTP. We provide a model that relates revenue bias to these influential factors. On a more general note, pricing decisions based on point-based methods are acceptably close to pricing decisions derived from range-based methods only if the WTP ranges are small.

Our results thus provide several implications. We demonstrate that the concept of WTP as a range should be adopted by marketing managers in their pricing decisions. In contrast with traditional methods that rely on unrealistic assumptions (e.g., preference and choice certainty, rational choice), range-based methods avoid the potential for biased pricing on the individual and aggregate levels. Biased pricing and biased estimations of demand result in poor estimates of production capacities and an incorrect basis for profit calculations and investment decisions. Because this effect tends to grow with the variance caused by consumer uncertainty, our recommendation of range-based methods is even more pressing for new or unfamiliar products, as well as public goods. The need to estimate WTP as a range in these cases is further confirmed because, from a decision maker’s perspective, they are characterized by large investments and, thus, high risks.

The simulation results in Fig. 4 also can be used in market research to reveal when range-based methods are absolutely mandatory. The simulation results thus might help researchers evaluate previous WTP elicitation studies that have relied on point-based methods. Because its predictive performance and applicability (simplicity, perceived fairness) is comparable to standard BDM techniques, our BDM-Range approach facilitates the diffusion of the theoretically superior WTP-as-a-range concept in real-world applications.

We recognize that WTP estimates are not exclusive to pricing decisions. For example, WTP may be used as a proxy for consumers’ brand attitudes. In such cases, traditional point-based WTP estimates could contain sufficient information, yet even in these settings, WTP ranges reveal important insights because they capture an additional dimension. For example, they might reveal not only brand attitudes but also customer uncertainty about expressing those attitudes.

5.2. Further research

Our study is based on some assumptions that might be relaxed in additional research. For example, we assume no particular (but a symmetric) WTP distribution on the individual level, though we must make a specific, appropriate assumption for the simulation study. Although the normal distribution seems reasonable,11 and truncation of this distribution is admissible (because of the two-sided restrictions in real-world WTP values), we do not investigate actual WTP distributions empirically.

Another interesting research avenue would be to analyze the influence of costs on the differences between point- and range-based methods. Although we have argued in favor of revenue maximization, a more systematic variation of cost levels could extend the results of our simulation study.

We restricted our studies to lottery-based approaches. Wang et al. (2007) provided a justification for doing so, but it would be interesting to relate BDM-Range to other standard procedures, such as conjoint analysis, which might be open to an extension to a range-based variant (Schlereth, Eckert, & Skiera, 2011).

Finally, another suggestion for further research relates to a topic beyond the scope of our work. We suggest analyzing marketing activities that companies use to reduce consumer uncertainty about product performance (e.g., branding, product knowledge-enhancing communication), and we suggest investigating their influences on WTP range. Preference uncertainty is difficult to influence, but as an extension of our study, further research might explore factors, such as risk aversion, that relate to decision making under uncertainty and may help clarify why WTP ranges exist. This knowledge would be useful particularly for efforts such as socio-psychological segmentation, for which researchers could identify segments that are less predictable in terms of their choices or WTP.

11 Formally, the WTP distribution on the individual level is a model for WTP across a large number of statements expressed by a specific consumer. We assume that the effect of uncertainty on these WTP statements remains the same across all statements. According to the central limit theorem, this assumption can be applied with a normal distribution as the probabilistic model.
Appendix A. Constructs, items, and scales

<table>
<thead>
<tr>
<th>Scales used</th>
<th>Source</th>
<th>Items</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct</td>
<td></td>
<td>Items</td>
<td>Scale</td>
</tr>
<tr>
<td>Method</td>
<td>Menon et al. (1995)</td>
<td>How would you rate the level of difficulty of responding to this method?</td>
<td>5-point Likert scale, 1: Not at all difficult, 5: Very difficult</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How would you rate the amount of effort it took you to respond to this method?</td>
<td>5-point Likert scale, 1: No effort, 5: A lot of effort</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How would you rate the amount of time it took you to respond to this method?</td>
<td>5-point Likert scale, 1: No time, 5: A lot of time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How would you rate the amount of thought you had to put into responding to this method?</td>
<td>5-point Likert scale, 1: No thought, 5: A lot of thought</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How fair would you rate this price elicitation method?</td>
<td></td>
</tr>
<tr>
<td>Method fairness</td>
<td>Own measure</td>
<td>How fair would you rate this price elicitation method?</td>
<td></td>
</tr>
</tbody>
</table>

Appendix B. Internal Validity for BDM, ICERANGE, and BDM-Range in Studies 1–3

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample size n</th>
<th>(a^2)</th>
<th>(b^2)</th>
<th>Pearson's (r^2)</th>
<th>Fisher's Z(\text{c})</th>
<th>Method comparison</th>
<th>(p) values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1: Hypothetical incentive-aligned study (caffè latte)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDM</td>
<td>58</td>
<td>5.31</td>
<td>-1.62</td>
<td>.9931</td>
<td>2.84</td>
<td>BDM vs. ICERANGE</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ICERANGE</td>
<td>46</td>
<td>5.56</td>
<td>-1.73</td>
<td>.9984</td>
<td>3.55</td>
<td>ICERANGE vs. BDM-Range</td>
<td>.88</td>
</tr>
<tr>
<td>BDM-Range</td>
<td>56</td>
<td>5.15</td>
<td>-1.62</td>
<td>.9985</td>
<td>3.58</td>
<td>BDM-Range vs. BDM</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Study 2: Real incentive-aligned (point-of-sale) study (caffè latte)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDM</td>
<td>39</td>
<td>8.98</td>
<td>-3.41</td>
<td>.9822</td>
<td>2.36</td>
<td>BDM vs. ICERANGE</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ICERANGE</td>
<td>40</td>
<td>7.43</td>
<td>-2.93</td>
<td>.9977</td>
<td>3.39</td>
<td>ICERANGE vs. BDM-Range</td>
<td>.80</td>
</tr>
<tr>
<td>BDM-Range</td>
<td>40</td>
<td>7.78</td>
<td>-3.05</td>
<td>.9980</td>
<td>3.45</td>
<td>BDM-Range vs. BDM</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Study 3: Real incentive-aligned (point-of-sale) study (service)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDM</td>
<td>31</td>
<td>3.72</td>
<td>-3.39</td>
<td>.9937</td>
<td>2.88</td>
<td>BDM vs. ICERANGE</td>
<td>.06</td>
</tr>
<tr>
<td>ICERANGE</td>
<td>44</td>
<td>3.31</td>
<td>-3.8</td>
<td>.9975</td>
<td>3.34</td>
<td>ICERANGE vs. BDM-Range</td>
<td>.22</td>
</tr>
<tr>
<td>BDM-Range</td>
<td>44</td>
<td>3.10</td>
<td>-3.4</td>
<td>.9957</td>
<td>3.07</td>
<td>BDM-Range vs. BDM</td>
<td>.45</td>
</tr>
</tbody>
</table>

\[a\] Parameters of the predicted demand function \(Pr(buy|p) = \exp(a + b \times p) / [1 + \exp(a + b \times p)]\), based on WTP estimates of BDM, ICERANGE, or BDM-Range, respectively

\[b\] Pearson correlation between predicted demand and observed demand for any WTP elicitation method

\[c\] Fisher-Z transformed correlation between predicted demand and observed demand for any WTP elicitation method

Appendix C. Differences across WTP elicitation methods

C.1. Formal proofs on the individual level

C.1.1. Proof for differences in revenue-maximizing prices

To illustrate the idea behind the following proof, we refer to the numerical example used in Section 4.1. We imagine a (single) consumer with a WTP of 100, a floor price (FP) of 60, and a ceiling price (CP) of 140. The revenue-maximizing price derived through a point-based WTP elicitation is 100, as shown in Section 4.1. When we assume this distribution to be symmetric and unimodal. As this assumption is very general, we use upward estimation for the individual buying probability, which is a component of the revenue function. Given our assumptions, the highest possible individual buying probability for prices beyond the WTP of 100 relates to the (continuous) uniform distribution of WTP within the WTP range. This relationship results in a linearly decreasing individual buying probability, with \(q_i(p) = -\frac{(p - CP)}{CP - FP}) = (140 - p)/80\). The revenue at price \(p\) with this individual buying probability is \(r_i(p) = q_i(p) \times p = (140 - p) \times p / 80\). Maximizing this function yields an optimal price \(p^{opt}_{range}\) of 70 (= CP/2), which is obviously not greater than 100. Thus, the revenue-maximizing price derived through range-based WTP cannot be higher than 100; point-based WTP elicitation leads to overpricing, compared to range-based WTP elicitation.
The general proof beyond the numerical example reads as follows: The point-based revenue function on the individual level is
\[ r_{\text{point}}(p) = \begin{cases} p, & 0 \leq p \leq \text{WTP} \\ 0, & p > \text{WTP} \end{cases} \]

(C.1)

The range-based revenue function on the individual level is
\[ r_{\text{range}}(p) = \begin{cases} p, & 0 \leq p \leq \text{FP} \\ q_{\text{range}}(p) \cdot p, & \text{FP} < p < \text{CP} \\ 0, & p > \text{CP} \end{cases} \]

(C.2)

Hence, the revenue-maximizing price is obviously \( p^*_{\text{point}} = \text{WTP} \).

The range-based revenue function on the individual level is
\[ r_{\text{range}}(p) = q_{\text{range}}(p) \cdot p \]

(C.3)

We can show that the range-based revenue-maximizing price is lower than WTP, the optimal point-based price. We assume \( \text{FP} \neq \text{CP} \). The question is whether the range-based optimum can be between WTP and CP (revenue beyond CP is equal to 0). For WTP \( \leq p \leq \text{CP} \), we know
\[ r_{\text{range}}(p) = q_{\text{range}}(p) \cdot p < p \cdot \text{CP} \]

\[ \implies p = q_{\text{range}}(p) \cdot p \]

where \( q \) is a linear function above \( q \) that connects the points (WTP; 0.5) and (CP; 0). The right side of the inequality yields an optimal price of \( p^* = \text{CP}/2 \), which is less than or equal to WTP because \( \text{FP} \geq 0 \). Because \( r_{\text{range}}(p) < q_{\text{range}}(p) \cdot p \), the maximum of \( r_{\text{range}} \) must be lower than the maximum of \( g(p) \cdot p \) (which is \( p^* = \text{CP}/2 \)). Consequently, the range-based optimal price is lower than the point-based optimal price.

C.1.2. Proof for a revenue bias

The revenue bias of the point-based method refers to the difference between revenues for the point-based optimal prices, using either the point- or range-based demand (\( q_{\text{point}} = q_{\text{point}}(p) \cdot p \) or \( q_{\text{range}} = q_{\text{range}}(p) \cdot p \)). When WTP ranges exist, then the optimal point-based price (which equals WTP) is within this range. Therefore, individual buying probability at WTP is strictly below the “certainty level” of 1. More formally, \( q_{\text{range}}(p_{\text{point}} = \text{WTP}) < 1 < q_{\text{point}}(\text{WTP}) \), which implies a revenue bias (as \( q_{\text{point}} = q_{\text{point}}(p_{\text{point}}) \cdot p_{\text{point}} < q_{\text{range}} = q_{\text{range}}(p_{\text{point}}) \cdot p_{\text{point}} \)). Hence, point-based methods can lead to overestimation of optimal revenue.

C.2. Formal proofs on the aggregate level

C.2.1. Proof for differences in revenue-maximizing prices

To illustrate the idea behind the following proof, we refer to the numerical example used in Section 4.2. One segment was homogeneous in WTP across consumers and included three consumers with different WTPs (60, 100, and 140, respectively) but with identical WTP ranges of 80. Now, focus on two consumers, the one with the lowest WTP (60) and the one with the highest WTP (140). Using range-based WTP elicitation implies that the demand function starts to drop at a price of 20 (= 60–80/2); this is the floor price of the consumer with the lowest WTP and reaches a value of 0 at a price of 180 (= 140 + 80/2); this is the ceiling price of the consumer with the highest WTP. In contrast, measuring WTP with a point-based method implies that demand starts to drop later (at a price of 60; this is the lowest WTP value) and reaches a value of 0 earlier (at a price of 140; this is the highest WTP value). Consequently, point-based WTP elicitation yields a “steeper” demand function. This is the essential observation for the formal proof. In the case of many consumers, demand is represented by continuous functions, in which the information on “steepness” is represented by the slope parameter \( \beta \).

The formal proof reads as follows:

Because point-based WTP elicitation methods measure the expected WTP \( \mu \) of range-based methods, we can rewrite the point- and range-based demand functions, as follows:
\[ d_{\mu}(p) = \frac{1}{1 + \exp(-\beta(p - \mu))} \]

and
\[ d_{\beta}(p) = \frac{1}{1 + \exp(-\beta(p - \mu))} \]

where \( \beta \) and \( \beta' < 0 \). The difference between point-based and range-based demand, therefore, is expressed in the parameter \( \mu \); specifically, we can set \( \beta' = \gamma \times \beta \) with some constant \( \gamma > 0 \), but \( \gamma \neq 1 \) (\( \gamma = 1 \) would imply that WTP ranges do not exist, an assumption that contradicts the empirical findings). Using the results presented by Miller et al. (2011; see also web appendix E), the point-based revenue-maximizing price \( p^* \) can be expressed as
\[ p^* = -1 - \exp(-LW(\exp(-1 - \mu \beta) - 1 - \mu \beta)) \]

where \( LW \) denotes the Lambert W function. We now show that \( p^*(\beta') \) differs from \( p^*(\gamma \times \beta) \), the latter being the range-based optimal price. Assume that \( p^*(\beta') = p^*(\gamma \times \beta) \). Simple algebra yields
\[ 1 - \gamma - \gamma \cdot \exp(-LW(\exp(-1 - \mu \beta) - 1 - \mu \beta)) = 0 \]

With \( F(\beta') := -LW(\exp(-1 - \mu \beta) - 1 - \mu \beta) \), this equation simplifies to
\[ 1 - \gamma = \gamma \cdot \exp(F(\beta')) \cdot \exp(F(\gamma \cdot \beta')) \]

If this equation holds, then the derivative of the right side must be 0 because the left side is constant. Differentiating the right side with respect to \( \beta \) and setting the result to 0 yields
\[ \frac{\gamma - \gamma \cdot \exp(F(\beta')) \cdot \exp(F(\gamma \cdot \beta'))}{F(\beta')} = 0 \]

Integrating both sides of the final equation and using integration by substitution yields \( \gamma = 1 \), which contradicts our assumption. Consequently, \( p^*(\beta') \neq p^*(\gamma \times \beta) \).

C.2.2. Proof for a revenue bias

As in Appendix C.1, revenue bias equals \( p^* \times d_{\mu}(p^*) - p^* \times d_{\mu}(p^*) \). Using Eqs. (C.4) and (C.5) and recognizing that \( \gamma > 0 \) but \( \gamma \neq 1 \), we know that revenue bias can be 0 only for \( p^* = \mu \). This scenario is generally not the case, however, as Eq. (C.6) shows. Therefore, except for the specific parameters \( \mu \) and \( \beta \) that fulfill the equation
\[ \mu = -1 - \exp(-LW(\exp(-1 - \mu \beta) - 1 - \mu \beta)) \]

a revenue bias occurs.
## Appendix D. Simulation results ($p^*$, $r^*$, $\Delta p\%$, and $\Delta r\%$ in the different simulation scenarios)

<table>
<thead>
<tr>
<th>Design factors</th>
<th>Range-based elicitation</th>
<th>Point-based elicitation</th>
<th>Bias measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population-level SD</td>
<td>Sample size (n)</td>
<td>Individual-level SE</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>25</td>
<td>.0001</td>
</tr>
<tr>
<td>2</td>
<td>84.66</td>
<td>78.66</td>
<td>84.44</td>
</tr>
<tr>
<td>6</td>
<td>83.72</td>
<td>76.69</td>
<td>82.96</td>
</tr>
<tr>
<td>10</td>
<td>82.17</td>
<td>74.11</td>
<td>81.39</td>
</tr>
<tr>
<td>14</td>
<td>80.66</td>
<td>70.85</td>
<td>79.86</td>
</tr>
<tr>
<td>16</td>
<td>79.27</td>
<td>67.41</td>
<td>.0001</td>
</tr>
<tr>
<td>18</td>
<td>.0001</td>
<td>84.59</td>
<td>78.60</td>
</tr>
<tr>
<td>20</td>
<td>.0001</td>
<td>84.54</td>
<td>78.50</td>
</tr>
<tr>
<td>100</td>
<td>.0001</td>
<td>84.66</td>
<td>78.75</td>
</tr>
<tr>
<td>200</td>
<td>.0001</td>
<td>84.79</td>
<td>78.91</td>
</tr>
<tr>
<td>400</td>
<td>.0001</td>
<td>84.73</td>
<td>78.83</td>
</tr>
<tr>
<td>800</td>
<td>.0001</td>
<td>84.77</td>
<td>78.83</td>
</tr>
<tr>
<td>20</td>
<td>.0001</td>
<td>84.77</td>
<td>78.89</td>
</tr>
<tr>
<td>25</td>
<td>.0001</td>
<td>84.77</td>
<td>78.89</td>
</tr>
<tr>
<td>50</td>
<td>.0001</td>
<td>84.77</td>
<td>78.89</td>
</tr>
</tbody>
</table>

(continued on next page)
### Appendix D. (continued)

<table>
<thead>
<tr>
<th>Design factors</th>
<th>Range-based elicitation</th>
<th>Point-based elicitation</th>
<th>Bias measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population-level SD (σ&lt;sub&gt;pop&lt;/sub&gt;)</td>
<td>Sample size (n)</td>
<td>Individual-level SE (σ&lt;sub&gt;ind&lt;/sub&gt;)</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>2</td>
<td>79.63</td>
<td>67.84</td>
<td>80.26</td>
</tr>
<tr>
<td>4</td>
<td>79.37</td>
<td>67.39</td>
<td>80.07</td>
</tr>
<tr>
<td>6</td>
<td>79.19</td>
<td>66.96</td>
<td>80.01</td>
</tr>
<tr>
<td>8</td>
<td>78.92</td>
<td>66.57</td>
<td>80.25</td>
</tr>
<tr>
<td>10</td>
<td>78.81</td>
<td>65.38</td>
<td>80.12</td>
</tr>
<tr>
<td>12</td>
<td>78.75</td>
<td>64.74</td>
<td>80.22</td>
</tr>
<tr>
<td>14</td>
<td>78.63</td>
<td>63.97</td>
<td>80.27</td>
</tr>
<tr>
<td>16</td>
<td>78.57</td>
<td>62.98</td>
<td>80.09</td>
</tr>
<tr>
<td>18</td>
<td>78.50</td>
<td>61.93</td>
<td>79.74</td>
</tr>
<tr>
<td>20</td>
<td>78.46</td>
<td>60.87</td>
<td>79.86</td>
</tr>
<tr>
<td>22</td>
<td>78.41</td>
<td>59.79</td>
<td>79.62</td>
</tr>
<tr>
<td>24</td>
<td>78.38</td>
<td>58.65</td>
<td>79.50</td>
</tr>
<tr>
<td>26</td>
<td>78.36</td>
<td>57.44</td>
<td>79.55</td>
</tr>
<tr>
<td>28</td>
<td>78.34</td>
<td>56.24</td>
<td>79.50</td>
</tr>
<tr>
<td>30</td>
<td>78.32</td>
<td>55.04</td>
<td>79.62</td>
</tr>
<tr>
<td>32</td>
<td>78.31</td>
<td>53.84</td>
<td>79.68</td>
</tr>
<tr>
<td>34</td>
<td>78.30</td>
<td>52.64</td>
<td>79.75</td>
</tr>
<tr>
<td>36</td>
<td>78.29</td>
<td>51.44</td>
<td>79.79</td>
</tr>
<tr>
<td>38</td>
<td>78.28</td>
<td>50.24</td>
<td>79.84</td>
</tr>
<tr>
<td>40</td>
<td>78.27</td>
<td>49.04</td>
<td>79.88</td>
</tr>
<tr>
<td>42</td>
<td>78.26</td>
<td>47.84</td>
<td>79.92</td>
</tr>
<tr>
<td>44</td>
<td>78.25</td>
<td>46.64</td>
<td>79.96</td>
</tr>
<tr>
<td>46</td>
<td>78.24</td>
<td>45.44</td>
<td>80.00</td>
</tr>
<tr>
<td>48</td>
<td>78.23</td>
<td>44.24</td>
<td>80.04</td>
</tr>
<tr>
<td>50</td>
<td>78.22</td>
<td>43.04</td>
<td>80.08</td>
</tr>
</tbody>
</table>

---

**Appendix D. (continued)**

164

### Design factors

<table>
<thead>
<tr>
<th>Population-level SD</th>
<th>Sample size (n)</th>
<th>Individual-level SE</th>
<th>Individual-optimal elicitation</th>
<th>Point-based elicitation</th>
<th>Bias measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Optimal price (p\textsubscript{opt})</td>
<td>Maximum revenue (r\textsubscript{max})</td>
<td>Optimal price (p\textsubscript{opt})</td>
</tr>
<tr>
<td>8</td>
<td>78.43</td>
<td>60.24</td>
<td>78.82</td>
<td>61.14</td>
<td>60.24</td>
</tr>
<tr>
<td>10</td>
<td>78.41</td>
<td>59.99</td>
<td>78.77</td>
<td>61.15</td>
<td>59.99</td>
</tr>
<tr>
<td>12</td>
<td>78.32</td>
<td>59.46</td>
<td>78.65</td>
<td>60.91</td>
<td>59.46</td>
</tr>
<tr>
<td>14</td>
<td>78.42</td>
<td>59.14</td>
<td>78.69</td>
<td>60.94</td>
<td>59.14</td>
</tr>
<tr>
<td>16</td>
<td>78.52</td>
<td>58.89</td>
<td>78.73</td>
<td>61.09</td>
<td>58.89</td>
</tr>
<tr>
<td>8</td>
<td>78.63</td>
<td>58.54</td>
<td>78.75</td>
<td>61.17</td>
<td>58.54</td>
</tr>
</tbody>
</table>

#### Bias measures

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>14</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

#### Range-based elicitation

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>14</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

#### Point-based elicitation

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>14</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

#### Relative price difference

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>25</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>14</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

#### Relative revenue bias

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>14</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

#### Relative price difference

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>14</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

#### Relative revenue bias

<table>
<thead>
<tr>
<th></th>
<th>Relative price difference Δp (in %)</th>
<th>Relative revenue bias Δr (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

(continued on next page)
### References


Mayring, P. (2000). Qualitative content analysis. *Qualitative Social Research Forum*, 1(2) http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm


Measurement of consumer preferences for bucket pricing plans with different service attributes

Christian Schlereth*, Bernd Skiera 1

Goethe University Frankfurt, Faculty of Business and Economics, Department of Marketing, Grueneburgplatz 1, 60323 Frankfurt am Main, Germany

A R T I C L E   I N F O

Article history:
First received in 5, August 2010 and was under review for 5½ months
Available online 8 March 2012
Area Editor: David Soberman

Keywords:
Pricing
Willingness to pay
Discrete choice experiments
Bayesian estimation
Bucket pricing
Two-part pricing plans
Three-part pricing plans

A B S T R A C T

A bucket pricing plan charges a periodic (usually monthly) fixed price that allows consumers to use the service up to a set allowance. The determination of optimal plans requires knowledge about each consumer's simultaneous decision about service subscription, plan choice, and consumption, which are interrelated and difficult to predict. In addition to prices, service attributes also influence these three decisions, but how they do so depends on the particular service attribute. This article describes a novel method to predict consumers' reactions to bucket pricing plans with varying service attributes and develops an algorithm to optimize these plans. Methodologically, we show that the failure to model the influence of service attributes correctly leads to non-optimal prices and profits, which differ by up to 22.75% from the optimal solution. Substantially, we show that bucket pricing plans are approximately as profitable as other nonlinear pricing plans if at least three bucket pricing plans serve to segment the market. Bucket pricing plans therefore present an attractive alternative for service providers to differentiate consumers according to their WTP and consumption.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

With a bucket pricing plan, a service provider charges a periodic (usually monthly) fixed price, in exchange for which consumers may use the service without further charges up to a preset allowance during the period. Due to the great heterogeneity among consumers, most providers offer more than one bucket pricing plan to differentiate consumers with different demands. Thus, consumers can choose freely among several combinations of fixed prices and allowances. If they use up their allowance, they might change to another plan with a higher allowance or stop using the service for the remaining period. The purchase of single units of the service is not possible, but changes in the plan are allowed (typically once a month) without any additional charges.

Bucket pricing plans have become increasingly popular in various industry sectors. With the introduction of Apple’s iPad in 2010, the mobile broadband service provider AT&T waived its unlimited data plans and began charging $15 per month for 200 MB (this translates to 7.5 cents per MB of allowance), $25.00 per month for 2 GB (1.25 cent per MB), $45.00 per month for 4 GB (1.125 cent per MB). In the music industry, eMusic offered three monthly bucket pricing plans in 2007 for downloads: 30 songs per month for $12.99, 50 songs for $16.99, or 75 songs for $20.99. Cloud computing service vendors such as Epipacific also charge a monthly price that depends on the number of transactions (e.g., 2000 transactions for $99.95, 5000 for $199.95). Even car insurance companies such as Allianz offer bucket pricing plans with allowances of 6,000 km, 10,000 km, or more. The health care industry uses bucket pricing as well. Evolution to Wellbeing, a personal training provider, charges $160 per month for up to 8 training sessions ($20 per session), $200 monthly for up to 12 ($17 per session), or $220 for up to 24 sessions ($9 per session).

Despite the prevalence of bucket pricing, research on the topic remains scarce (see Table 1). Compared with two-part pricing plans (i.e., a usage-independent fixed price plus a marginal price per unit), bucket pricing plans enable consumers to benefit from a fixed allowance rather than pay for each unit of consumption. Thus, payment is separated from consumption, which enables consumers to enjoy their consumption more, according to the theory of mental accounting (Prelec & Loewenstein, 1998). In addition, the monthly

2 We acknowledge that the situations of payments and hedonic consumption, which are studied in Prelec & Loewenstein (1998), occur only once. In contrast, bucket pricing plans involve reoccurring charges at the beginning of each period. Clear evidence that pre-payment is also desired for reoccurring expenses and that the theory of mental accounting still holds is, however, missing, despite the promising findings of Lambrecht & Skiera (2006).

1 Tel.: +49 69 798 34649; fax: +49 69 79 35001.

* Corresponding author. Tel.: +49 69 798 34638; fax: +49 69 79 35001.
E-mail addresses: schlereth@wiwi.uni-frankfurt.de (C. Schlereth), skiera@skiera.de (B. Skiera).

0167-8116/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijresmar.2011.08.004
price stays the same and is known at the beginning of each month, which is often preferable for risk-averse consumers (Lambrecht & Skiera, 2006). Compared with three-part pricing plans (i.e., a usage-independent fixed price, an allowance, and a marginal price per unit), bucket pricing does not allow users to purchase single units of the service if they exceed the allowance. Most providers that use bucket pricing thus allow consumers to change and even cancel their plan each month such that consumers have only a short-term commitment. This ability represents a major difference from most three-part pricing plans, which require consumers to retain their selected plan over time, often for as long as 24 months (Lambrecht, Seim, & Skiera, 2007). Similar to most three-part pricing plans, unused allowance expires after the end of each period such that consumers might feel regret because they have paid for something that they did not use.

Fig. 1 illustrates the differences between bucket pricing plans as well as two-part and three-part pricing plans. The consumer can usually select among several pricing plans. In Fig. 1, the dotted line indicates the pricing plan, which is more expensive than the other plan for a given consumption. Bucket pricing plans lead to a stepwise function for the bill amount, which only increases if consumers move to the next-expensive pricing plan. Two-part pricing plans, in contrast, increase the bill amount with every unit that is consumed. Three-part pricing plans share characteristics with bucket pricing plans (in particular, the stepwise function) and two-part pricing plans (in particular, the increasing part). A bucket pricing plan is a special case of a three-part pricing plan because its marginal price is infinite, i.e., it is so high that no consumer will be willing to pay that price. A two-part pricing plan is also a special case of the three-part pricing plan because its allowance is zero.

The determination of optimal bucket pricing plans remains challenging. If service providers offer more than one plan, as shown in Fig. 1, a consumer has to decide simultaneously on subscription, plan choice, and consumption, which are interrelated and difficult to predict. The reason for this is the interdependency between prices and consumption (Iyengar, Jedidi, & Kohli, 2008; Lambrecht et al., 2007). For example, if AT&T decreases the price of their 250 MB plan, the number of consumers who choose that plan will likely increase. However, each consumer with the 250 GB plan generates less profit, and consumers switching from a higher plan will use the service less frequently, which is likely to reduce profit.

The simultaneous prediction of consumers’ decisions on subscription, plan choice, and consumption becomes even more complicated if service providers vary in service attributes to differentiate their service offerings further (Eggers & Sattler, 2009). The reason for this is that different types of service attributes may influence the three decisions in a different way. For example, Apple’s iPad comes in versions with various storage capacities (e.g., 16, 32, or 64 GB) and access to 3G networks. The differences in storage capacity might have little influence on consumption of 3G network capacity because most bandwidth-extensive downloads are performed over WiFi networks, and even the smallest storage capacity is still large enough for the bandwidth provided by current 3G network operators. Differences in the availability of 3G networks instead, which can vary strongly across operators, might significantly impact the consumption of capacity. Despite the practical importance of this distinction, all previous research has applied a “one-size-fits-all approach” (see Section 2.3); most of the studies link service attributes to the subscription decision and thereby neglect their potential influence on consumption.

This article aims to develop a novel method to predict consumers’ reactions to bucket pricing plans that also differ in their service attributes and to develop an algorithm to optimize a menu of bucket pricing plans. In particular, we attempt to examine whether service attributes affect plan choice with respect to the subscription decision or the consumption decision. This difference is important because the two mechanisms affect willingness to pay and the resulting optimal prices differently. Furthermore, we analyze the profit differences

---

**Figure 1:** Visualization of bucket pricing plans opposed to two-part and three-part pricing plans.
between bucket and other popular pricing plans. Because transaction data rarely are available for such innovative offerings, we use survey data elicited through discrete choice experiments (e.g., Moore, 2004; Wedel, Vriens, Bijmolt, Krijnen, & Leeflang, 1998).

In Section 2, we outline previous research on nonlinear pricing to develop our model of consumer decision making for bucket pricing plans in Section 3. We also outline how service attributes might vary in their influence on subscription and consumption decisions. Through an empirical study, in Section 4, we measure preferences for a music download service that uses bucket pricing and compare the validity of models that predict different influences of service attributes on consumers’ reactions to bucket pricing plans. In Section 5, we also compare bucket pricing plans against other popular plans, such as two- and three-part pricing plans. Finally, we conclude with a summary of our findings.

2. Literature review

Table 1 summarizes previous studies on consumers’ reactions to nonlinear pricing plans, which are plans in which the price per unit is not strictly proportional to the number of units purchased. It also outlines how our research differs from previous studies.

2.1. Nonlinear pricing plans

Previous research on nonlinear pricing has its origins in welfare economics and has primarily tackled consumer decision making for two-part pricing plans (i.e., fixed price and a marginal price per unit; e.g., Hui, Yoo, & Tam, 2007; Leland & Meyer, 1976; Murphy, 1977). The contributions include analytical models that reveal the interdependence between prices and consumption. Revealed preference data from field experiments (Danaher, 2002) and actual market transactions (Goettler & Clay, 2011; Narayanan, Chintagunta, & Miravete, 2007) also provide empirical evidence of such interdependence. Transaction data have been used to study three-part pricing plans (e.g., Iyengar, Ansari, & Gupta, 2007; Lambrecht et al., 2007) in studies that also account for consumers’ uncertainty about preferences, quantity during subsequent periods, and learning effects when making long-term decisions.

Unlike three-part pricing plans, bucket pricing plans include no single-usage price (i.e., no usage is possible beyond the allowance) but allow consumers to switch easily across different plans. With the exception of Iyengar (2010) and Iyengar & Jedidi (2011), who compare alternative specifications of the utility function and show that differences in internal validity are modest, previous research has ignored such pricing. Our work differs from theirs because neither Iyengar (2010) nor Iyengar & Jedidi (2011) consider the potential effects of service attributes on consumers’ choice and consumption decisions or the ability of bucket pricing plans to generate profit when compared to other nonlinear pricing plans.

2.2. Data sources

 Revealed preference data have high external validity, such that they support estimates of consumer learning over time (Iyengar et al., 2007; Lambrecht et al., 2007; Narayanan et al., 2007). However, prices in revealed preference data often vary only over a limited range, so estimations of reactions to price changes are difficult to make. Moreover, revealed preferences are unavailable for companies that enter new markets or sell new services not previously sold under real market conditions (Swait & Andrews, 2003; Wertenbroch & Skiera, 2002). Stated preference methods can offer assistance in such situations, which is why we focus on them in this study (Eggers & Sattler, 2009; Louviere, Hensher, & Swait, 2000). Iyengar et al. (2008) also explain how to use discrete choice experiments to estimate demand for three-part pricing plans in a cellular phone service context. This data source is inexpensive and provides good control over the experimental setting to enable tests of consumers’ reactions to new attribute ranges and pricing.

2.3. Service attributes

Prior consideration of service attributes in models that capture consumers’ responses to nonlinear pricing plans is very limited; most models focus on homogeneous services that do not differ in their attributes (e.g., Hui et al., 2007; Maskin & Riley, 1984; Narayanan et al., 2007). However, a few recent models have started to account for differences in service attributes, such as brands, roll-over minutes, Internet access (Iyengar et al., 2008), service quality, catalogue size (Iyengar, 2010), switching costs (Ascarza, Lambrecht, & Vlckassim, 2010), and pricing plan types (Iyengar et al., 2007; Lambrecht et al., 2007). These studies consistently assume that every service attribute affects the same behavioral process, such as consumers’ usage-independent utility, which then influences the likelihood of subscribing to a service. Only Iyengar & Jedidi (2011) anticipates that all service attributes affect the perceived utility of each service unit and thus consumption.

Unlike these prior studies, we account for influences in different behavioral processes and therefore expect that some service attributes affect the consumption decision, whereas others affect the subscription decision. We propose a flexible extension to a bucket pricing model that can distinguish among different influences of the varying types of service attributes. This extension could be included easily into other models of nonlinear pricing. We also outline the extent to which the failure to account for differences in service attributes can lead to non-optimal pricing recommendations.

2.4. Counterfactual simulations

Wilson (1993) introduces a rich framework of descriptive statistics to understand consumers’ demand behavior, which is rooted in statistics for uniform pricing plans. However, it cannot provide information regarding bucket pricing decisions because service providers still require a full understanding of the holistic system, including all relevant interactions across optional pricing plans and consumption rates. Counterfactual simulations can complement such understanding and enable an analysis of market conditions with different pricing plans. For example, prior research has studied consumption sensitivity and elasticity when prices change, such as shifts in choice probabilities in response to pricing changes (Iyengar et al., 2008; Lambrecht et al., 2007) or alterations in pricing plan choices, consumption, and profits if decision making uncertainty increases (Ascarza et al., 2010; Lambrecht et al., 2007).

Danaher (2002), Iyengar et al. (2008), and Iyengar (2010) use grid search techniques as a heuristic to determine profit-maximizing prices, though these solution spaces increase exponentially with the number of pricing plans. This exponential increase might explain why most studies consider only one or two pricing plans. Schlereth, Stepanchuk & Skiera (2010) show that simulated annealing reduces this shortcoming and can easily optimize four plans. Additionally, they show that two optional two-part pricing plans are frequently sufficient to skim the market. However, these authors do not reveal how to optimize bucket or three-part pricing plans. Consequently, we have little knowledge about how pricing plans compare to one another, despite the great importance of such understanding for service providers that must determine which type and number of pricing plans to use. We follow Schlereth et al. (2010) and develop a simulated annealing algorithm to optimize bucket pricing and three-part pricing plans. We then extend this application to compare their profitability.
### Table 1
Summary of studies that measure consumers’ reactions to nonlinear pricing plans.

<table>
<thead>
<tr>
<th>General Topic</th>
<th>Research Work</th>
<th>Nonlinear Pricing Plans</th>
<th>Empirical Data Sources</th>
<th>Influence of Service Attributes</th>
<th>Counterfactual Simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Two-part pricing plans</td>
<td>Three-part pricing plans</td>
<td>Bucket pricing plans</td>
<td>No influence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No empirical data</td>
<td>Stated preferences</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Revealed preferences</td>
<td>Subscription decision</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Consumption decision</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Depends on service attribute</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand under different pricing plans</td>
<td>Danaher (2002)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Usage uncertainty and/or learning</td>
<td>Narayanan et al. (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Iyengar et al. (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Lambrecht et al. (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Iyengar et al. (2008)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Goettler &amp; Clay (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Ascarza et al. (2010)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Comparison across different specification</td>
<td>Iyengar (2010)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
</tr>
<tr>
<td></td>
<td>Iyengar &amp; Jedidi (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Comparison across different data collection methods</td>
<td>Schlereth et al. (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Optimization of pricing plans</td>
<td>Hui et al. (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Service attributes and bucket pricing</td>
<td>Schlereth et al. (2010)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Our paper</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Notes: (X) indicates working paper.
3. Model Development

We develop a model for choice decisions about bucket pricing plans with different service attributes. To do so, we begin with a basic model that considers bucket pricing for homogeneous services that do not differ in their attributes, then specify prior assumptions about the distribution of the parameters in the willingness-to-pay (WTP) space (also known as the surplus model). Next, we introduce a flexible extension of the model for heterogeneous services that describes how differences in service attributes might influence plan choice through either the subscription or the consumption decision. Finally, we discuss the different implications that arise and outline the application of hierarchical Bayes techniques to estimate parameters.

3.1. Modeling plan choice for bucket pricing

Let \( J \) represent the set of bucket pricing plans that can be chosen by all consumers during each period. Each plan \( j \in J \) consists of an allowance \( q_j \) and a periodic, fixed price \( p_j \). We assume a utility-maximizing consumer \( i \) who does not choose more than one plan, and who consumes \( n_{i,j}(q_j) \) units of the service, depending on his or her preferences and the allowance \( q_j \) (\( 0 \leq n_{i,j}(q_j) \leq q_j \)). The amount of consumption \( n_{i,j}(q_j) \) in a period can occur all at once or be summed over different consumption phases within that period.

To analyze data with multiple units of a homogeneous service, we need a nonlinear utility specification of the utility function to capture the unique WTP for each quantity unit (Iyengar et al., 2008; Lambrecht et al., 2007). Let \( v_{i,j}(n_{i,j}(q_j), z_{i,j}) \) represent the deterministic part of the utility that consumer \( i \) obtains by choosing bucket pricing plan \( j \), consuming \( n_{i,j}(q_j) \) units of the service, and spending the remaining budget on \( z_{i,j} \) units of an unobservable outside good. The outside good provides a utility of \( \omega_i \cdot z_{i,j} \), where \( \omega_i \) represents the price parameter (Sonnier, Ainslie, & Otter, 2007). We assume that utility increases with quantity at a decreasing rate. We choose the price parameter (Sonnier, Ainslie, & Otter, 2007). We assume a utility-

\[
v_{i,j}(n_{i,j}(q_j), z_{i,j}) = a_{i,j} \cdot n_{i,j}(q_j) - \frac{b_{i,j}}{2} \cdot n_{i,j}(q_j)^2 + c_{i,j} + \omega_i \cdot z_{i,j} \quad (i \in I, j \in J),
\]

subject to the following budget constraint:

\[
Y_i \geq z_{i,j} \cdot p_j + p_j \quad (i \in I, j \in J),
\]

where \( Y_i \) represents the budget of consumer \( i \) and \( p_j \) is the price of the outside good. Furthermore, the parameter \( a_{i,j} \) reflects the increase in utility that accompanies an increase in consumption; \( b_{i,j} \) accounts for the decrease in the marginal utility; and \( c_{i,j} \) is the usage-independent utility, that is, the utility for zero quantity. This usage-independent utility equals 0 for many services, but it might be greater than 0 if the service provider includes services that are free of additional charge in a subscription, such as subsidized hardware that comes with telecommunication services or free website space bundled with Internet access.

The index \( j \) of the parameters \( a_{i,j}, b_{i,j}, \) and \( c_{i,j} \) reflects the differences in the attributes of a pricing plan \( j \), as we describe in more detail later. To ensure a semi-concave function, the parameters \( a_{i,j}, b_{i,j}, \) and \( c_{i,j} \) should be greater than or equal to 0. Without loss of generality, we normalize the price of the outside good \( p_j \) to 1. Assuming that a consumer exhausts his or her budget and substituting the rearranged term of the budget constraint in Eq. (2), \( z_{i,j} = Y_i - p_j \), into Eq. (1), we obtain an indirect utility function:

\[
v_{i,j}(n_{i,j}(q_j), p_j) = a_{i,j} \cdot n_{i,j}(q_j) - \frac{b_{i,j}}{2} \cdot n_{i,j}(q_j)^2 + c_{i,j} + \omega_i \cdot (Y_i - p_j) \quad (i \in I, j \in J).
\]

Thus, according to Eq. (3), \( u_{i,j} = \omega_i \cdot Y_i \) if consumers do not choose the service, such that they spend all their money on the unobservable outside good. By forming a Lagrange function, we derive the optimal consumption \( n_{i,j}^* \) of consumer \( i \) for bucket pricing plan \( j \) (see the Appendix):

\[
n_{i,j}^*(q_j) = \begin{cases} q_j, & \text{if } q_j \leq \frac{a_{i,j}}{b_{i,j}} \\ \frac{a_{i,j}}{b_{i,j}}, & \text{if } q_j > \frac{a_{i,j}}{b_{i,j}} \\ \end{cases} \quad (i \in I, j \in J).
\]

This model predicts that consumers either fully use the allowance \( q_j \) or leave some of it unused if his or her saturation level (i.e., \( a_{i,j}/b_{i,j} \)) is below the allowance. In this case, consuming an additional unit would not provide any additional benefit for the consumer but might cause some disutility (e.g., opportunity cost of time, as implicitly considered herein). Substituting Eq. (4) into Eq. (3), we can derive the indirect utility function as a function of the allowance \( q_j \). Eq. (5) then uses the indicator variable \( \text{Ind}_{i,j} \) to distinguish the increase in utility with every unit below the saturation level from the constant part of the utility function for every unit above the saturation level. We assume that the consumer does not consume the \( \left( \frac{a_{i,j}}{b_{i,j}} + 1 \right) \)th unit of the service, which provides no additional utility.

Thus, substituting Eq. (4) into Eq. (3) gives

\[
v_{i,j}(q_j, p_j) = \text{Ind}_{i,j} \cdot \left( a_{i,j} \cdot q_j - \frac{b_{i,j}}{2} \cdot q_j^2 \right) + \left(1 - \text{Ind}_{i,j}\right) \cdot \left( \frac{a_{i,j}^M}{2 \cdot b_{i,j}^M} \right)^2 + c_{i,j} + \omega_i \cdot (Y_i - p_j) \quad (i \in I, j \in J),
\]

where \( \text{Ind}_{i,j} = \begin{cases} 1, & \text{if } q_j \leq \frac{a_{i,j}}{b_{i,j}} \\ 0, & \text{otherwise} \end{cases} \)

Finally, we transform the indirect utility function in Eq. (5) into a surplus function (Sonner et al., 2007) that directly expresses the utility for zero quantity. This transformation is motivated by the belief that consumers typically prefer to think in monetary rather than utility terms, a preference that must be taken into account by appropriate prior distribution specifications in hierarchical Bayesian estimation methods. Similar to classical additive utility models (Sonner et al., 2007), we transfer the surplus model modification to the context of services and determine the parameters \( a_{i,j}, b_{i,j}, \) and \( c_{i,j} \) as monetary values (see the Appendix):

\[
a_{i,j}^M = \frac{a_{i,j}}{\omega_i}, \quad b_{i,j}^M = \frac{b_{i,j}}{\omega_i}, \quad c_{i,j}^M = \frac{c_{i,j}}{\omega_i} \quad (i \in I, j \in J).
\]

With this transformation, we can rewrite the indirect utility function in Eq. (5) as a surplus model:

\[
v_{i,j}(q_j, p_j) = \omega_i \cdot \left( \text{Ind}_{i,j} \cdot \left( a_{i,j}^M \cdot q_j - \frac{b_{i,j}^M}{2} \cdot q_j^2 \right) + \left(1 - \text{Ind}_{i,j}\right) \cdot \left( \frac{a_{i,j}^M}{2 \cdot b_{i,j}^M} \right)^2 + c_{i,j}^M - p_j \right) + \omega_i \cdot Y_i \quad (i \in I, j \in J).
\]
3.2. Service attributes

Few studies include service attributes in models of nonlinear pricing, and when they do, they model all service attributes in an identical way. For example, lyengar et al. (2008) feature brand names, Internet access, and roll-over minutes in their choice model for three-part pricing plans. It is reasonable to assume that the availability of Internet access for a cell phone provides additional utility but does not alter the number of calls made by a consumer. As in most other studies, these authors link differences in the utility of service attributes to usage-independent WTP, which we capture with the parameter \(c_{ij}^M\). This linkage implies that the attributes have no influence on the consumption decision (i.e., the number of units consumed) but make the general offering of the service more or less attractive, which affects the likelihood of subscription.

However, what is neglected in prior literature is the idea that other attributes might require a different linkage. Consider downloads for digital music and the effects of differences in digital rights management (DRM) restrictions. As an access control technology, DRM is frequently applied by the music industry to prevent the unauthorized copying and distribution of purchased songs. For example, DRM can restrict consumers’ ability to use downloaded songs on different players or computers or the number of songs that may be burned on a CD. These restrictions likely affect consumption, because the utility function only imagines a consumption (e.g., Sundararajan, 2004). However, the utility function only imagines a consumption increase on the consumption attributes of service plan \(j\) and that the in...
For this study, respondents make a choice among several offerings with different bucket pricing plans and service attributes as well as a no-choice option. We assume that consumer i chooses from each choice set s the alternative that yields the highest utility and is subject to the budget constraint. To account for any additional factors that influence utility and are known to the consumer but unobservable to the data analyst, we introduce a stochastic component \( \varepsilon_{i,s} \) (usually labeled the error term) and thus can make statements about the probability \( P_{r_{i,s}} \) that consumer i picks plan j in choice set s:

\[
P_{r_{i,s}} = \frac{\exp(v_{i,j} \mid q_i, p_j) + \varepsilon_{i,s} \mid q_i, p_j)}{\exp(v_{i,j} \mid q_i, p_j)}
\]

The error term \( \varepsilon_{i,s} \) has a mean of 0 and covariance matrix \( \Sigma \). We assume it is distributed independently and at extreme values (or type I extreme value, Gumbel). Let \( \theta_i \) summarize all parameters in the utility function of consumer i, that is, \( a_i^{b_i}, b_i^{b_i}, c_i^{b_i}, \) and the vector \( \beta_i^{b_i}, \) and let \( d_{i,s} \) represent an indicator variable equal to 1 if consumer i chooses plan j from choice set s and 0 otherwise. Thus, matrix \( \theta \) contains all vectors \( \theta_i \) and the following likelihood function of the mixed logit model enables us to estimate the distributions of \( \theta \):

\[
L(d \mid \theta, \Sigma) = \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{s=1}^{S} P_{r_{i,s}}(d_{i,s} = 1 \mid \theta, \Sigma)
\]

\[
= \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{s=1}^{S} \left( \frac{\exp(v_{i,j} \mid q_i, p_j))}{\exp(v_{i,0}) + \sum_{f=1}^{J} \exp(v_{i,j} \mid q_i, p_j)\right)^{d_{i,s}}
\]

Note that the income term \( \eta_i \cdot Y_i \) is canceled out in the likelihood function because it has no effect on the differences across the utilities of all alternatives, including the alternative of not using the service at all.

We specify the density of the parameters \( \theta \) to be normally distributed, with mean \( \theta \) and covariance matrix \( \Omega \), as denoted by \( g(\theta \mid \theta, \Omega) \). The conditional posterior on \( \theta_i, \forall i \), given \( \theta \) and \( \Omega \), is:

\[
\Lambda(\theta \mid \theta, \Omega) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{s=1}^{S} L(d_{i,s} \mid \theta, \Sigma) \cdot g(\theta_i \mid \theta, \Omega).
\]

4. Empirical study

Through our empirical study, we illustrate our model’s ability to reflect consumers’ chosen subscriptions, plans, and consumption in a bucket pricing scenario. It also outlines how the linkages of service attributes affect the validity of the results. We use an online survey to study consumers’ preferences for digital music offerings with different bucket pricing plans, which are alligned to the business model of the company eMusic introduced to the European market six months prior to this study.

4.1. Digital music context

New technical possibilities for distributing digital contents and services through the Internet have not been beneficial to all industry sectors. In particular, the music industry has suffered mightily from digital music piracy. After enjoying annual revenue growth rates of 10% or more in the 1990s (Rob & Waldfogel, 2006), it experienced a sudden 42% decline in revenue between 2000 and 2008 (from $15 to $8 billion; RIAA, 2009). Technologies such as MP3 compression of audio and peer-to-peer platforms have enabled consumers to encode purchased CDs and distribute or consume them for free through the Internet. The music industry has reacted with both sticks and carrots (Sinha & Mandel, 2008); it discourages illegal download activities through lawsuits but also provides legal alternatives to consume digital music. Such alternatives include services that sell individual songs or bundles on a download-to-own basis (e.g., iTunes) and those that stream music through the Internet, for which customers usually pay a monthly price (e.g., Napster).

These legal alternatives complement traditional CD sales but as of yet have not completely compensated for losses (RIAA, 2009). They have also triggered a growing debate about the extent to which digital music piracy harms the music industry (Bhattacharjee, Gopal, Lertwachara, Marsden, & Telang, 2007; Rob & Waldfogel, 2006) and the factors that will encourage people to stop illegally downloading music (Dilmperi, King, & Dennis, 2011; Sinha & Mandel, 2008). The discussion also features new business models such as advertisement-sponsored streaming services (e.g., the Swedish DRM-based music streaming service Spotify; Papi, Eggers, & Wolpert, 2010) and the abandonment of single-song sales in favor of album-only sales (Elberse, 2010). To assess new business models, a key metric is the value of digital music to consumers (Papi et al., 2010; Rob & Waldfogel, 2006; Sinha & Mandel, 2008; Sinha, Machado, & Sellman, 2010). However, most prior studies attempt to capture consumers’ WTP for just one particular – e.g., a favorite or previously purchased – song or album. Papi et al. (2010) underline the strong need for models that consider changes in consumption by individual consumers to assess price recommendations for new business models in the music industry.

In this context, the business model of eMusic is interesting for two reasons. First, it does not distinguish between songs for which consumers have varying WTP; instead, it treats them as a commodity good, for which consumers have a recurring demand, which can be influenced by pricing. Second, eMusic offers consumers ownership in the form of DRM-free MP3 downloads, which constituted an innovative service attribute at the time of the study. This strategy conflicts with almost all other services that use DRM restrictions to decrease the incidence of digital piracy (e.g., Sundararajan, 2004). However, DRM also restricts consumers’ ability to listen to music they have purchased on different technical devices such as MP3 players, car stereos, or various PCs. Thus, in 2007, a month after we completed our study, Apple’s chief executive officer Steve Jobs called for the elimination of DRM from legally sold digital songs and introduced DRM-free songs on iTunes for a surcharge of 0.39€ per song (i.e., 0.99€ for DRM-restricted and 1.29€ for DRM-free songs).

4.2. Study design

4.2.1. Questionnaire

The questionnaire consists of three parts: self-stated preferences and experiences with legally purchasing digital music, choice sets that ask respondents to select the most attractive alternative, and demographic information items. The bucket pricing plans in the choice sets differ in their DRM restrictions (DRM-free and DRM-restricted), brand names (Musicload, Napster, iTunes, and a fictive name, Loadasong), monthly subscription prices (4.99€, 7.99€, 14.99€, 24.99€, and 34.99€), and allowance levels (5, 10, 20, 40, and 60 songs). Consistent with the eMusic business model, we informed respondents that they could switch their plan once a month. Even though illegal downloading is pervasive in the music industry, we did not explicitly consider it in the choice of attributes and levels (e.g., by offering an illegal option at the price of 0€). The reason

---

3 We informed respondents that they could burn DRM-restricted songs up to five times on a CD and were unrestricted in copying them to a mobile player or playing them on a single PC.
for this is two-fold. First, the resulting decision model would also have to account for respondents’ perceived probability of being exposed to the risks of a lawsuit and possibly moral consequences. Therefore, illegally downloaded songs may also provide some negative utility, which is difficult to elicit from discrete choice experiments. Second, including an option that depicts an act as being against the law might cause social desirability bias, which would influence the parameter estimates.

4.2.2. Choice set design
We created two D-optimal versions of the choice designs, each consisting of 15 choice sets, and added two additional choice sets to each version for holdout predictions (see the Appendix). Each choice set consists of three bucket pricing plans and a no-choice option. Because illegal downloading has been such a big challenge for the music industry (Rob & Waldfogel, 2006), we incorporated music piracy implicitly in the no-choice option. That is, if a respondent prefers illegal sources of digital music, he or she should select the no-choice option because none of the pricings provide sufficient utility.

4.2.3. Data collection
We conducted an online survey of undergraduate and graduate students of a major European university and received 123 completed questionnaires. We considered students suitable respondents for this study because they represent one of the main target groups for music download providers. Of the respondents, 26.83% stated that they had legally purchased digital music, mainly using iTunes, Musicload, or Napster (64.71%, 41.18%, and 26.47%, respectively; multiple responses were possible).

4.2.4. Estimation
We use the 1845 discrete choice decisions (=123×15) that constitute the data set for the estimation and use Eq. (7) as the indirect utility function of choice in the estimation. Our model links DRM restrictions to the parameter $c_{i,j}^M$, brand name to $c_{i,j}^b$, and the name of the platform to the usage-independence parameter $c_{i,j}^{\text{ci}}$ (Model 1). Therefore, Model 1 reflects our expectation that DRM restrictions affect plan choice though the consumption decision, whereas the name of the platform is connected to brand awareness (Agarwal & Rao, 1996) and influences plan choice through the subscription decision. In addition, we estimate three possible combinations of attribute linkages to either $a_{i,j}^M$ or $a_{i,j}^b$ (Models 2–4). For example, linking the brand to parameter $a_{i,j}^b$ indicates that the brand influences consumption, perhaps because brands indicate differences in platform usability or the availability of a recommender system (Senecal & Nantel, 2004). The model with the best fit should reveal the best match of attribute linkages to respondents’ decision making.

The base levels of the parameters $a_{i,j}^M$, $b_{i,j}^M$, $c_{i,j}^M$, and $c_{i,j}^{\text{ci}}$, as well as the parameter for the DRM restrictions, are reparameterized (e.g., $a_{i,j}^M = \exp(a_{i,j}^M)$) to ensure positive values and the desired concave form of the utility function. Dummy variables for the service attributes in the matrix $x$ are effect coded. Estimations rely on hierarchical Bayes techniques and use standard diffuse priors. The reported results are determined from 20,000 iterations that we retain after discarding the initial 40,000 iterations (=60,000 iterations total). We assess convergence according to the trace plot of the likelihood function.

### Table 4
Parameter estimates.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level</th>
<th>Mean</th>
<th>Median</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>willingness-to-pay function</td>
<td>$a_{i,j}^M$</td>
<td>0.61</td>
<td>0.41</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.04)</td>
<td>(0.27)</td>
<td></td>
</tr>
<tr>
<td>Price parameter</td>
<td>$b_{i,j}^M$</td>
<td>0.18</td>
<td>0.03</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.00)</td>
<td>(1.94)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{i,j}^M$</td>
<td>0.03</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.00)</td>
<td>(0.26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{i,j}^{\text{ci}}$</td>
<td>1.57</td>
<td>1.81</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{M}$</td>
<td>0.23</td>
<td>0.15</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.23)</td>
<td>(0.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{n}$</td>
<td>0.41</td>
<td>0.34</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.22)</td>
<td>(0.18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{i}$</td>
<td>0.25</td>
<td>0.21</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.28)</td>
<td>(0.19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{B}$</td>
<td>−0.38</td>
<td>−0.43</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.32)</td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{DRM}$</td>
<td>0.20</td>
<td>0.09</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.28)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{B}$</td>
<td>−0.29</td>
<td>−0.09</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.28)</td>
<td></td>
</tr>
</tbody>
</table>

4 We also tested a model in which consumers’ experience with legal music download platforms was entered as an observable covariate for explanation of the parameters but found no improvement in the model fit (see the Appendix).
We calculate the WTP for various allowances and show its median and 5%, 25%, 75%, and 95% quantiles in the boxplots in Fig. 2. The left- and right-hand sides present the willingness to pay estimates for DRM-free and DRM-restricted songs, respectively. We used iTunes as the brand name, but the differences are similar for the other brand names. On average, respondents are willing to pay 4.94€ for 10 DRM-free songs but 46.03% less (i.e., 2.67€) if they are DRM-restricted. For 60 songs, the WTP difference increases to 49.54% (7.59€ = 15.31€ for DRM-free – 7.72€ for DRM-restricted songs). Furthermore, the median WTP continues to increase for 50 to 60 DRM-free songs; for 30 songs or more, it stays nearly constant with that for DRM-restricted songs. This finding supports the claim that offering DRM-free songs increases consumption.

5. Counterfactual simulations

Our discrete choice model attempts to capture decision making processes for bucket pricing plans and emphasizes the importance of choosing the correct linkage of service attributes to the parameters of the utility function. However, service providers are more interested in recommendations of optimal (profit-maximizing) prices and allowances in bucket pricing plans (e.g., Danaher, 2002; Iyengar et al., 2008). Therefore, in a counterfactual simulation, we first analyze the extent to which price recommendations vary across service attribute linkages. Considering the increasing popularity of bucket pricing, in a second counterfactual simulation, we tackle the question of how their profitability compares with that of other pricing plans (e.g., pay-per-use, two-part pricing, and three-part pricing).

5.1. Comparison of pricing recommendations across different service attribute linkages

To quantify the extent to which managerial implications vary when different types of behavioral processes are associated with each service attribute, we compare the recommendations of optimal (profit-maximizing) bucket pricing plans obtained from the four models in our empirical study. The individual draws of the Metropolis Hastings step in the hierarchical Bayesian estimation indicate respondents’ simultaneous decisions about subscriptions, plan choices, and consumption; these predictions then provide a means to estimate the profits of various menus of bucket pricing plans denoted by $\pi^*(p_j, q_j)$:

$$\pi(p_j, q_j) = \sum_{j} \sum_{k} \left( p_j - k_v \cdot n_i(q_j) \right) \cdot \text{Pr}_{ij}(p_j, q_j) \rightarrow \max$$  \hspace{1cm} (13)$$

Profit equals the sum of the profit contributions of all consumers, which consists of two components: the probability of a consumer choosing plan $j$ and the margin, or the monthly price $p_j$ minus the product of variable costs $k_v$ and consumption $n_i(q_j)$, as specified in Eq. (4). Although they are not considered here, it is straightforward to incorporate other types of costs, such as fixed costs for each subscriber served.

From the MCMC sampler, we consider a subsample of the posterior distributions of the individual parameters to determine optimal pricing recommendations. We use a subsample rather than posterior means to describe respondents according to the distribution of their individual parameters, which can be asymmetric. This subsample helps us calculate profits for a given set of bucket pricing plans $J$. The challenge then is to determine the optimal bucket pricing plans that maximize profit because the objective function in Eq. (13) is nonlinear and nonconvex, with many local maxima. We implemented and extensively tested various heuristic search methods; simulated annealing provides the best search performance (see also Schlereth et al., 2010). It randomly accepts solutions with a decreasing objective functional value and can thus overcome local maxima. To our knowledge, ours is the first study to determine optimal bucket and three-part pricing plans while investigating more than just one or two options.

We apply simulated annealing to menus of one, two, and three bucket pricing plans using the individual parameter distributions of each model, which results in 12 optimization runs (=3 menus x 4 models). We neglect competition to clearly illustrate the effects of linking attributes on either consumption or subscription decisions. It is straightforward to account for static competition by including the pricing schemes of the most important competitors. We use the fictive brand name “Loadasong”; this scenario reflects the situation when eMusic entered the European market and was not as popular or familiar as iTunes, Musicload, or Napster. We assume that variable costs, such as those for licensing, taxes, and technical infrastructure, are 0.22€, which accounts for approximately 80% of the price per song in eMusic’s largest plan. We use the parameter estimates of the best-fitting Model 1 to simulate the effects of the optimal bucket pricing plans on profits across all models. The differences in Table 5, compared with the optimal profit in Model 1, indicate the degree to which the recommendations and strategies for segmenting the market vary if one uses the wrong model to obtain price recommendations.

---

We depict the full model of the profit maximization problem in the Appendix.

---

The price histories of eMusic and market leaders such as iTunes indicate no observable dependencies in how prices are set. For other service fields, it might be possible to include Bertrand competition or Stackelberg leader–follower price competition (Kadiyali, Suddhir, & Rao, 2001) in the simulation.
According to Table 5, applying bucket pricing recommendations from Models 2–4 decreases profits for the three bucket pricing plans from 8.96% to 22.75% (see last column). The recommendations of Model 2 yield the smallest deviations because the DRM restrictions influence the parameter $a_i^M$. The greatest deviation appears in Model 4, which links DRM restrictions on parameter $c_i^M$ and brand on $a_i^M$ (i.e., the reverse of Model 1). With fewer plans, this difference increases up to 30.76%.

The recommended prices and allowances in the bucket pricing plans thus indicate differences in segmenting the market. Model 1 suggests three bucket pricing plans with allowances of 8, 31, and 80 songs. The prices per song then range between 0.75€ (=6.006/8) and 0.60€ (=47.82€/80), which is lower than the prices currently charged by iTunes or Musicload (i.e., 0.99€–1.39€ per song). Moreover, eMusic’s monthly bucket pricing plans (i.e., as of 2007 in Europe, 30 songs for 12.99€, 50 for 16.99€, or 75 for 20.99€) provide songs at substantially lower prices than its competitors. Thus, the results have high face validity.

In contrast with this pricing plan, the results of Model 1 suggest targeting consumers with a demand of less than 30 songs and charging higher prices per song for consumers with higher demand quantities. Both suggestions predict actual changes in eMusic’s subscription plans in the years 2008–2010; it reduced the number of songs in its lowest plan and increased the average price per song for all bucket pricing plans (i.e., in 2010, 24 songs for 11.99€ per month, 35 for 15.99€, and 50 for 20.79€).

The model that deviates most in terms of profit (i.e., Model 4) recommends skimming usage-independent WTP, as reflected by the parameter $c_i^M$. Therefore, for the three bucket pricing plans, the model recommends offering very few songs for a high average price (i.e., 4.94€ for two songs or 2.47€ per song). In addition, the price for 80 songs is unreasonably high and exceeds 73€, or 0.92€ per song. Thus, we conclude a lack of face validity for the recommendations of Model 4.

If the menu consists of just one bucket pricing plan, Models 1 and 2 recommend offering 30–48 songs for about 14.00–18.57€, whereas Models 3 and 4 recommend fewer songs for a lower price. Both Models 3 and 4 include DRM restrictions in the usage-independent WTP, so their price recommendations target consumers with lower demand. Similarly, when offering two bucket pricing plans, the recommendations in Models 3 and 4 differ from those in Models 1 and 2 because they emphasize the outer limits of the price range. Therefore, Models 3 and 4 recommend one plan with few songs (about 10) and one plan with 80 songs for a high price (i.e., at least 59.70€). In contrast, the models that link the service characteristic to the parameter $a_i^M$ suggest targeting consumers whose demand is approximately 25 songs and asking a lower price for 80 songs.

5.2. Comparison with other pricing plans

Bucket pricing plans differ from pay-per-use, two-part, and three-part pricing plans in that they do not contain a marginal price (see also Fig. 1). A two-part pricing plan for music downloads is a plan that charges a monthly fixed fee plus a marginal price per song, which is below the marginal price per song of common pay-per-use operators like iTunes. A three-part pricing plan offers an allowance in exchange for the fixed fee; however, it differs from bucket pricing plans in that consumers can purchase single songs for a certain marginal price if they exceed that allowance. Thus, the bucket pricing plans are less flexible, but they encourage consumers to use a preset number of units.

In a second counterfactual simulation, we analyze the extent to which profits differ across the alternative plans. Furthermore, service providers frequently must choose a particular number of plans, which involves a trade-off between offering more plans to attain better market segmentation (Maskin & Riley, 1984; Murphy, 1977) and offering fewer plans to avoid confusing the consumers and minimize the administrative costs for managing the plans (Hui et al., 2007). Extant recommendations only apply to two-part pricing plans (e.g., Murphy, 1977; Schlereth et al., 2010).

Using the subsample from the posterior distribution of Model 1, we determine profit-maximizing plans, analogous to what was performed in the first counterfactual simulation. We vary the type and number of pricing plans, such that they range between one and four, except for the pay-per-use plan, which can feature only one. We also vary the variable costs (i.e., 0.02€–0.42€ per song in two increments of 0.20€) because previous research shows a strong dependence for some pricing plans (i.e., flat rates) between variable costs and profitability (e.g., Schlereth et al., 2010). For each of the 39 conditions (=3 types of nonlinear pricing plans×4 plans×3 variable costs×3 pay-per-use plans), we determine the optimal pricing plans and report the results in Table 6.

The second counterfactual simulation requires some important assumptions. First, we assume that the individual parameters $a_i^M$, $b_i^M$, and $c_i^M$ remain the same for different types of pricing plans and neglect potential pricing plan effects (e.g., Lambrecht & Skiera, 2010).
While increasing the number of two-part pricing plans enables service providers to better differentiate among consumers because of the already differentiated market, when compared to bucket pricing plans, the profit potential of both pricing plans eventually seems to be about the same. For three-part pricing plans, we expect the explanation to be somewhere in between, given the fact that three-part pricing plans consist of a combination of allowance and marginal price after exceeding the allowance.

In summary, service providers can freely choose among multiple two-part, three-part, and bucket pricing plans because their differences in profit are small. They must only realize that they will likely need to offer more bucket pricing plans, especially if variable costs are high. Therefore, digital music providers such as eMusic could benefit from offering even more bucket pricing plans than the three it offered between 2007 and 2010.

6. Conclusions

Learning about the heterogeneous demand of consumers and using that information to define optimal pricing plans is an important challenge. Modern companies increasingly adopt bucket pricing, an alternative to flat-rate, pay-per-use, two-part, and three-part pricing plans. Thus, in this work, we develop a discrete choice model of consumers’ preferences for bucket pricing plans that accounts for the interdependencies between consumers’ decisions about their subscription, plan, and consumption. These preferences in turn provide a means to derive bucket pricing recommendations that differ between consumers.

We propose a model to capture consumers’ decisions and provide a flexible extension of our model that allows for different considerations of service attributes with respect to the utility function. For example, eMusic has traditionally offered DRM-free songs and thereby built a unique brand image. These attributes potentially affect consumers’ subscriptions, plan choices, and consumption decisions, but the specific nature of the effects depends strictly on the attributes. An analyst might have an intuitive understanding of the influence of some attributes, but without utter certainty about the most appropriate linkage that reflects consumers’ common decision making processes, analysts must undertake a comparison of the fit of alternative modeling approaches. In the case of digital music, for example, DRM restrictions influence consumption decisions, but the choice of service provider only influences the decision about service subscriptions. Our results also show that a well-specified linkage of attributes has a strong impact on bucket pricing recommendations, so failures to account for such linkages correctly may possibly cause profit losses of up to 22.75%.

This study also has several important implications for marketers. We show that profits under bucket pricing plans are almost the same as those under two-part and three-part pricing plans but substantially higher than those under pay-per-use plans. Unlike two-part and three-part pricing plans, it is beneficial to increase the number of bucket pricing plans, which substantially increases profit (in our study, by an average of up to 29.28%). The greater number of bucket pricing plans enables service providers to better differentiate among heterogeneous consumer demands because the charges for consumption reflect specific allowances. We therefore conclude that bucket pricing plans present an attractive alternative for service providers.

Our study also has some limitations that warrant attention. For example, we do not incorporate a potential learning effect of consumers over time, despite the likely influence on consumers’ plan choices and consumption (e.g., Iyengar et al., 2007; Narayanan et al., 2007). However, when using discrete choice experiments, these effects are not observable because respondents do not receive feedback about their decisions and therefore cannot mentally process the outcome of previous decision making for future decisions. Further research therefore might combine discrete choice experiment data and transaction data.

### Table 6
Comparison of optimal profits of alternative pricing plans relative to four bucket pricing plans.

<table>
<thead>
<tr>
<th>Variable Costs</th>
<th>Pricing Plan</th>
<th>1 Plan</th>
<th>2 Plans</th>
<th>3 Plans</th>
<th>4 Plans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (i.e., 0.02€)</td>
<td>Bucket Pricing</td>
<td>90.42%</td>
<td>96.34%</td>
<td>99.21%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>Pay-Per-Use Plan</td>
<td>86.61%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Two-Part Pricing</td>
<td>98.32%</td>
<td>99.94%</td>
<td>100.51%</td>
<td>100.65%</td>
</tr>
<tr>
<td></td>
<td>Three-Part Pricing</td>
<td>98.35%</td>
<td>100.33%</td>
<td>100.62%</td>
<td>100.89%</td>
</tr>
<tr>
<td>Medium (i.e., 0.22€)</td>
<td>Bucket Pricing</td>
<td>66.58%</td>
<td>89.97%</td>
<td>97.14%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>Pay-Per-Use Plan</td>
<td>89.11%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Two-Part Pricing</td>
<td>98.91%</td>
<td>100.60%</td>
<td>100.60%</td>
<td>100.95%</td>
</tr>
<tr>
<td></td>
<td>Three-Part Pricing</td>
<td>98.91%</td>
<td>100.66%</td>
<td>100.66%</td>
<td>100.95%</td>
</tr>
<tr>
<td>High (i.e., 0.42€)</td>
<td>Bucket Pricing</td>
<td>55.15%</td>
<td>88.55%</td>
<td>95.70%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>Pay-Per-Use Plan</td>
<td>86.57%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Two-Part Pricing</td>
<td>100.87%</td>
<td>101.28%</td>
<td>101.62%</td>
<td>102.76%</td>
</tr>
<tr>
<td></td>
<td>Three-Part Pricing</td>
<td>100.87%</td>
<td>101.88%</td>
<td>101.64%</td>
<td>102.86%</td>
</tr>
<tr>
<td>Average (over the three variable costs)</td>
<td>Bucket Pricing</td>
<td>70.72%</td>
<td>91.62%</td>
<td>97.35%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>Pay-Per-Use Plan</td>
<td>87.42%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Two-Part Pricing</td>
<td>99.37%</td>
<td>100.61%</td>
<td>100.91%</td>
<td>101.46%</td>
</tr>
<tr>
<td></td>
<td>Three-Part Pricing</td>
<td>90.38%</td>
<td>100.96%</td>
<td>100.97%</td>
<td>101.37%</td>
</tr>
</tbody>
</table>

2006). Second, we ignore the length of the commitment to a pricing plan, which tends to be longer for two- and three-part than for bucket pricing plans. Third, we neglect the strategic role of the type of pricing plans in oligopolies. Yang & Ye (2008) argue that instead of offering cheaper prices, service providers should employ pricing plans strategically to differentiate themselves from competitors. We welcome research that eliminates at least some of these assumptions.

In Table 6, we provide the optimal profits relative to the profits of four optimal bucket pricing plans. The latter are always lower than the profits from the four two-part or three-part pricing plans. However, the differences range between 0.65% and 2.86%, which we consider modest. Variable costs have no observable impact on differences in profit.

Bucket pricing plans substantially outperform pay-per-use plans by at least 10.89% (i.e., = 101.86%). Unfortunately, we find only marginal differences in profit between two-part and three-part pricing plans. The largest difference occurs for high variable costs, which only reaches as high as 2.86% (102.86%). However, the differences range between 0.65% and 2.86%, which we consider modest. Variable costs have no observable impact on differences in profit.

Bucket pricing plans provide a strong impact on bucket pricing recommendations, so failures to account for such linkages correctly may possibly cause profit losses of up to 22.75%.

In summary, service providers can freely choose among multiple two-part, three-part, and bucket pricing plans because their differences in profit are small. They must only realize that they will likely need to offer more bucket pricing plans, especially if variable costs are high. Therefore, digital music providers such as eMusic could benefit from offering even more bucket pricing plans than the three it offered between 2007 and 2010.

6. Conclusions

Learning about the heterogeneous demand of consumers and using that information to define optimal pricing plans is an important challenge. Modern companies increasingly adopt bucket pricing, an alternative to flat-rate, pay-per-use, two-part, and three-part pricing plans. Thus, in this work, we develop a discrete choice model of consumers’ preferences for bucket pricing plans that accounts for the interdependencies between consumers’ decisions about their subscription, plan, and consumption. These preferences in turn provide a means to derive bucket pricing recommendations that differ between consumers.

We propose a model to capture consumers’ decisions and provide a flexible extension of our model that allows for different considerations of service attributes with respect to the utility function. For example, eMusic has traditionally offered DRM-free songs and thereby built a unique brand image. These attributes potentially affect consumers’ subscriptions, plan choices, and consumption decisions, but the specific nature of the effects depends strictly on the attributes. An analyst might have an intuitive understanding of the influence of some attributes, but without utter certainty about the most appropriate linkage that reflects consumers’ common decision making processes, analysts must undertake a comparison of the fit of alternative modeling approaches. In the case of digital music, for example, DRM restrictions influence consumption decisions, but the choice of service provider only influences the decision about service subscriptions. Our results also show that a well-specified linkage of attributes has a strong impact on bucket pricing recommendations, so failures to account for such linkages correctly may possibly cause profit losses of up to 22.75%.

This study also has several important implications for marketers. We show that profits under bucket pricing plans are almost the same as those under two-part and three-part pricing plans but substantially higher than those under pay-per-use plans. Unlike two-part and three-part pricing plans, it is beneficial to increase the number of bucket pricing plans, which substantially increases profit (in our study, by an average of up to 29.28%). The greater number of bucket pricing plans enables service providers to better differentiate among heterogeneous consumer demands because the charges for consumption reflect specific allowances. We therefore conclude that bucket pricing plans present an attractive alternative for service providers.

Our study also has some limitations that warrant attention. For example, we do not incorporate a potential learning effect of consumers over time, despite the likely influence on consumers’ plan choices and consumption (e.g., Iyengar et al., 2007; Narayanan et al., 2007). However, when using discrete choice experiments, these effects are not observable because respondents do not receive feedback about their decisions and therefore cannot mentally process the outcome of previous decision making for future decisions. Further research therefore might combine discrete choice experiment data and transaction data.
to obtain the advantages of both data sources (e.g., Swait & Andrews, 2003). We consider the application of this approach to nonlinear pricing valuable because such a model could capture such (e.g., learning) effects and strategically control for different pricing offers, which are not observed in real markets. Another valuable extension of our model is that of quality-differentiated subscription plans. We acknowledge that service providers can differentiate themselves from their competitors not only through their prices but also through their choices of service attributes. For example, the Swedish streaming service Spotify offers a free streaming service, which has restrictions such as lower music quality and radio-like advertisements, as well as a premium subscription for a monthly fee with no restrictions. How to link preferences to differentiated pricing plans is another research topic of high practical importance.

Acknowledgments

The authors gratefully acknowledge the help of Johanna Burkhardt during the data collection and the excellent suggestions of Thomas Otter. In addition, the editor Marnik Dekimpe, the associate editor David Soberman, and the two anonymous referees provided many very helpful and excellent suggestions. We wrote large parts of this manuscript during stays at the Centre for the Study of Choice (CenSoC) at the University of Technology Sydney (Australia), for which we thank Jordan Louviere for his support and valuable comments. This research was funded by the German Ministry for Education and Research (BMBF – FKZ: 01 IA 08003C), by the German Research Foundation (DFG – GZ: SCHL 1942/1-1), and by the E Finance Lab Frankfurt.

Appendix A. Proof of optimal consumption

We derive optimal consumption $n_{ij}^*$, subject to $0 \leq n_{ij}(q_j) \leq q_j$, by forming a Lagrange function for the problem outlined in Eq. (3):

$$L(n_{ij}(q_j), \lambda) = v_{ij}(n_{ij}(q_j), p_j) - \lambda \left( q_j - n_{ij}(q_j) \right) \rightarrow \max \quad (i=1, j=1).$$

(A.1)

Differentiating the Lagrange function in Eq. (A.1) yields:

$$\frac{\partial L(n_{ij}(q_j), \lambda)}{\partial n_{ij}(q_j)} = a_{ij} - b_{ij} n_{ij}(q_j) + \lambda \quad (i=1, j=1).$$

(A.2)

We obtain two optimal solutions based on the Kuhn-Tucker conditions $\frac{\partial L(n_{ij}(q_j), \lambda)}{\partial n_{ij}(q_j)} = 0, \lambda \geq 0$, and $\lambda \cdot (q_j - n_{ij}(q_j)) = 0$. A consumer uses the service up to the saturation level if this saturation level is smaller than the allowance; otherwise, the consumer completely exhausts the available allowance:

$$n_{ij}^*(q_j) = \begin{cases} q_j, & \text{if } q_j \leq \frac{a_{ij}}{b_{ij}} \; n_{ij}(q_j) \\ \frac{a_{ij}}{b_{ij}}, & \text{if } q_j > \frac{a_{ij}}{b_{ij}} \; n_{ij}(q_j) \end{cases} \quad (i=1, j=1).$$

(A.3)

Appendix B. Estimation space transformation

We transform the indirect utility function in Eq. (5) into the surplus function. Thus, we move the preference space, also known as the utility model, to the WTP space, which is also known as the surplus model (Sonnier et al., 2007). For this purpose, we define WTP$_{ij}(q_j)$ as the price for an allowance $q_j$, for which the consumer is indifferent between purchasing and not purchasing ($v_{0i} = v_{ij}(q_j)$; e.g., Moorthy, Ratchford, & Talukdar, 1997). Thus, we can write

$$\text{WTP}_{ij}(q_j) = \frac{1}{\varpi_i} \left( \text{Ind}_{ij} \left( a_{ij} q_j - b_{ij} q_j^2 \right) + \left( 1 - \text{Ind}_{ij} \right) \left( \frac{a_{ij}^2}{2 b_{ij}} \right) + c_{ij} \right) \quad (i=1, j=1).$$

(B.1)

Transforming Eq. (B.1) so that WTP$_{ij}(q_j)$ appears on the right-hand side yields:

$$WTP_{ij}(q_j) = \frac{1}{\varpi_i} \left( \text{Ind}_{ij} \left( a_{ij} q_j - b_{ij} q_j^2 \right) + \left( 1 - \text{Ind}_{ij} \right) \left( \frac{a_{ij}^2}{2 b_{ij}} \right) + c_{ij} \right) \quad (i=1, j=1).$$

(B.2)

and the consumer surplus, given the monthly package price $p_j$, then becomes:

$$CS_{ij}(q_j, p_j) = WTP_{ij}(q_j) - p_j \quad (i=1, j=1).$$

(B.3)

If we compare Eqs. (5) and (B.2), we find that the relationship between the parameters that express preferences and the monetary parameters (indicated by the subscript M) yields statements about the WTP:

$$a_{ij} = \frac{a_{ij}}{\varpi_i}; \; b_{ij} = \frac{b_{ij}}{\varpi_i}; \; c_{ij} = \frac{c_{ij}}{\varpi_i} \quad (i=1, j=1).$$

(B.4)

which enables us to transform Eq. (5) into Eq. (7).

Eqs. (5) and (7) are behaviorally equivalent, and their usage theoretically leads to the same parameter estimates. This equivalence occurs if we use estimation methods that do not incorporate prior information, such as the maximum likelihood estimator. However, when incorporating prior information, as required by hierarchical Bayes methods, the distributions of these priors typically are not equivalently included in both estimation spaces. In particular, the commonly applied standard diffuse priors of monetary parameters derived from the utility model are normal and usually divided by normal or lognormal distributions. These priors may emphasize the tails of the distribution and place greater weight on outliers, which may result in unreasonably high parameter estimates. Sonnier et al. (2007) thus argue that, in the surplus model, the prior distribution for WTP expressions is simply normal or lognormal. This model thus yields more face-valid estimates than the utility model, especially if data are scarce, as is the case in most discrete-choice experiments. We account for their considerations by using Eq. (7) as an indirect utility function of choice in our empirical estimation.

Appendix C. Choice design

In the empirical study, 2 designs each consisted of 15 choice sets for the estimation and 2 additional holdouts, which were the same for both designs. Thus, respondents answered 17 choice sets in total. With the exception of the brand attribute, the attributes in the discrete choice experiment contain levels with associated benefits or costs that monotonically increase or decrease. Therefore, it is possible for choice sets to include bucket pricing plans

---

5 The topic of how many choice sets to use per person yields little consensus. On the one hand, researchers want to collect as much information as possible, but on the other hand, they should avoid straining respondents’ cognitive effort to prevent answers that have little value or are subject to high error variance. Marketing research frequently employs 16–32 choice sets per respondent (e.g., Hensher, Stopher, & Louviere, 2001; Iyengar et al., 2008; Parker & Schrift, 2011), and Hensher et al. (2001) recommend around 16 treatments.
that strongly dominate other bucket pricing plans in the same choice set. A dominant plan has more of every benefit attribute (allowance and DRM restrictions) and less of the monthly package price attribute than any other plan. When such plans are present, the decision is trivial, and thus, such a choice set provides limited information.

There are several possibilities for creating the choice design, including the methods suggested by Street, Burgess, & Louviere (2005) or the use of software (e.g., NGene or Sawtooth) to maximize D-efficiency. To make the choices more realistic, we excluded two unlikely bucket pricing plans (i.e., 5 songs for 34.99€ and 60 songs for 3.99€) and generated a D-efficient starting design. We avoid the dominance of any alternative by creating a set of 80 additional candidates. Next, we employed the analog to Strategy 5 from Street et al. (2005), which is an iterative approach, and carefully tested each choice set for dominating or dominated occurrences. We replaced each such occurrence with five candidates and selected the one that maximized the resulting D-optimal design efficiency. The final design appears in Table A1. Column 1 features the allowance level, column 2 the price, column 3 the brand, and column 4 DRM restrictions. Holdouts are indicated by “H” in the left-side column. Using online software provided by Street and Burgess (http://crsu.science.uts.edu.au/choice/choice.html), we find that each of the designs obtains an efficiency of at least 76% compared with the optimal design (which does not account for dominant alternatives) or 86% when combined.

To test the potential gain in model fitting, we extend the Bayesian algorithm and account for differences pertaining to whether a respondent already has legally purchased music using an effect-coded covariate variable. The basic implementation for linear utility functions has been described by Lenk, DeSarbo, Green, & Young (1996), Rossi, McCulloch, & Allenby (1996), and Renken (1997). The major difference between the models with and without covariates appears in the upper level of the Bayes sampler. The hierarchical Bayes model with upper-level covariates assumes that respondents’ part-worths relate to the covariates through a multivariate regression model of the following form:

$$\theta_i = \theta \cdot z_i + \xi_i, \quad \text{where} \quad \xi_i \sim \text{Normal}(0, \Omega) \quad (i \in \Omega).$$  \tag{D.1}$$

If we assume that individual preferences are modeled by m part-worths and n covariates (including a constant), then $\theta$ is an m × n matrix of regression parameters, $z_i$ is a vector of n elements, and $\xi_i$ is a vector of random error terms. The partworths are drawn from a normal distribution with means $\theta \cdot z_i$.

Comparing the results of Table A2 with those of Table 3, we find only marginal gains in internal validity. For example, the log marginal density increases for the first model from $-597.24$ to $-587.64$. However, the predictive validity decreases slightly. The changes in internal and predictive validity are small and inconsistent for the other models. Orme & Howell (2009) observe similar results, which indicate only modest or no improvements when they include covariates in their models. In summary, we find no significant influence of experience with legal purchases of music downloads on any of the parameters. Therefore, experience provides no additional information for improving parameter estimates or predictions.

### Appendix D. Influence of experience with the download platform on decision making

In our study, 26.83% of the respondents reported that they already had purchased songs from legal music download platforms. Their experience might cause their decision process to differ from that of the remaining 73.17% of respondents without any experience. Our hierarchical Bayes estimation in the empirical study assumes only one common normal distribution on the population level for both types of respondents, but if the groups behave differently, it seems reasonable to shrink their individual estimates not to one common population mean $\theta$ but to a conditional mean $\theta \cdot z_i$, given experience with online music downloads.

### Appendix E. Profit maximization problem

**Table E.1**

<table>
<thead>
<tr>
<th>Version 1</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pricing Plan 1</td>
<td>Pricing Plan 2</td>
</tr>
</tbody>
</table>
| 3 2 3 1 2 4 0 0 0 0 0 2 0 3 1 1 0 0 0 0 1 4 4 2 1 1 0 1 1 4 2 2 1 2 3 3 0 1 0 3 1 4 2 1 1 2 4 2 0 0 0 3 1 1 3 1 3 4 2 1 1 0 0 3 0 3 3 1 1 2 0 0 1 3 4 2 1 1 3 1 0 0 0 0 0 3 3 0 0 0 2 1 0 4 4 2 1 2 0 0 0 4 2 3 0 0 0 2 1 3 4 3 0 4 3 2 1 2 1 0 1 1 4 0 0 3 3 2 1 1 0 1 1 1 0 0 2 0 2 0 2 3 1 1 1 0 3 1 2 1 0 2 0 2 0 0 1 1 1 0 0 2 1 4 3 0 1 0 2 3 0 0 0 1 0 2 0 4 2 0 0 0 1 3 4 2 1 1 3 1 0 0 0 0 0 3 1 3 1 2 0 4 4 1 1 4 4 0 4 1 1 3 2 1 0 1 0 0 3 4 1 0 1 3 2 1 0 1 0 0 4 4 0 3 1 2 2 1 0 1 0 2 0 1 0 3 0 3 1 1 1 1 4 2 2 0 3 1 3 1 0 0 2 0 4 2 0 1 3 1 2 0 4 23 1 0 1 1 0 2 3 2 2 3 1 3 4 3 1 1 2 1 0 2 1 1 1 1 0 0 1 0 2 3 0 2 4 3 0 0 2 2 0 4 3 1 1 2 2 1 1 1 0 3 0 4 4 2 0 0 1 1 3 2 1 2 0 1 4 3 0 1 0 4 3 0 0 1 2 2 0 3 1 1 1 1

Holdouts are indicated by “H” in the left-side column.

### Internal and predictive validity (covariate model)

<table>
<thead>
<tr>
<th>Internal Validity</th>
<th>Predictive Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMD</td>
<td>HR</td>
</tr>
<tr>
<td>Model 1: DRM restrictions linked to $a_{i}^{M}$, brand name to $c_{i}^{M}$</td>
<td>$-563.99$</td>
</tr>
<tr>
<td>Model 2: All attributes linked to $a_{i}^{M}$</td>
<td>$-623.99$</td>
</tr>
<tr>
<td>Model 3: All attributes linked to $c_{i}^{M}$</td>
<td>$-671.54$</td>
</tr>
<tr>
<td>Model 4: DRM restrictions linked to $c_{i}^{M}$, brand name to $a_{i}^{M}$</td>
<td>$-767.24$</td>
</tr>
</tbody>
</table>

Notes: LMD: log marginal density; HR: hit rate; MAD: mean absolute deviations; DRM: digital rights management.

We adapt the model provided by Schlereth et al. (2010) to formulate the profit maximization problem for a given number of bucket pricing plans:

$$\pi(p_j, q_j) = \sum_{j=1}^{m} \left[ \left( p_j - k_i \cdot q_j \right) \cdot p_{r_j} \right] \rightarrow \max,$$  \tag{E.1}$$

$$q_j \leq q_f,$$  \tag{E.2}$$

$$p_j \leq p_f,$$  \tag{E.3}$$

and

$$p_j \in \mathbb{R}_+,$$  \tag{E.4}$$

The objective function in Eq. (E.1) searches for the best set of prices and allowances ($p_j, q_j$), which maximizes the profit, calculated as the monthly price minus variable costs times individual consumption, as in Eq. (4). Furthermore, we specify $p_j$ to be positive and continuous and $q_j$ to be positive but discrete.
Assuming the parameters do not change for other types of pricing plans, it is straightforward to formulate the profit maximization problem for three-part pricing plans (see Eqs. (E.5)–(E.6)). Each three-part pricing plan consists of a monthly fixed price \( p_j \), an allowance \( q_j \), and a marginal price \( m_j \). The objective function is then

\[
\Pi(p_j, q_j, m_j) = \sum_{j=1}^{m} \left( p_j + m_j \cdot \max(0, q_j - m_j) - k_j \cdot n_j(q_j, m_j) \right) \times Pr_j(p_j, q_j, m_j) \rightarrow \max.
\]

with consumption analogous to that described in Iyengar et al. (2008) or Lambrecht et al. (2007), namely,

\[
n_j(q_j, m_j) = \left\{ \begin{array}{ll}
\frac{a_j - \sigma_j - m_j}{a_j} & \text{if } q_j < \frac{a_j}{b_j} \\
\frac{a_j - \sigma_j - m_j}{b_j} & \text{if } q_j > \frac{a_j}{b_j} \\
\frac{a_j - \sigma_j - m_j}{q_j} & \text{if otherwise}
\end{array} \right.
\]

using the preference model

\[
\text{using the surplus model}
\]

\[
(6.6)
\]

References


10 This very strong assumption should be omitted in future research. Lambrecht & Skiera (2006) analyze why flat rates are perceived differently from pay-per-use plans; Ascarza et al. (2010) incorporate differences between two-part and three-part pricing plans in their choice model. A potential extension of our study would be to present respondents with multiple types of pricing plans and capture the differences in decision making with additional parameters.
Effects of formal sales control systems: A combinatory perspective

C. Fred Miao a,⁎, Kenneth R. Evans b,1

a 362 New Snell Hall, School of Business, Clarkson University, Potsdam, NY 13699-5765, United States
b Michael F. Price College of Business, University of Oklahoma, Norman, OK 73019-4007, United States

ARTICLE INFO

Article history:
First received in 7, October 2010 and was under review for 4 months
Available online 16 March 2012
Area Editor: Stefan Wuyts

Keywords:
Sales control systems
Salesperson knowledge
Role ambiguity
Intrinsic motivation
Salesperson performance

ABSTRACT

This research paper investigates the combinatory effects of three well-established formal sales control styles: outcome, capability, and activity control. Drawing on Expectancy Theory and Cognitive Evaluation Theory, the authors theorize that sales control combinations have differential impacts on three key intermediary variables (salesperson knowledge, role ambiguity, and intrinsic motivation), which subsequently affect salesperson performance. A Partial Least Squares (PLS) analysis using data from industrial salespeople and their managers indicates that (1) the capability and outcome control styles have positive combinatory effects that enhance salesperson knowledge and intrinsic motivation while mitigating role ambiguity, (2) combining outcome control and activity control can increase role ambiguity, and (3) a combination of activity control and capability control can enhance role clarity but dampen intrinsic motivation. Finally, our results indicate that the effects on manager-rated salesperson performance of capability control and activity control are fully mediated while the effects of outcome control are partially mediated by salesperson knowledge and role ambiguity.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Effectively managing a sales force is more challenging today than ever before. As the nature of many buyer–seller relationships has shifted from arms-length transactions to more long-term, focused relational exchanges, the sales force must be more effective in forging cooperative partnerships, providing integrative solutions, and adapting to the changing needs of relational customers (Dwyer, Schurr, & Oh, 1987; Hunter & Perreault, 2007; Krafft, Albers, & Lal, 2004; Langerak, 2001; Morgan & Hunt, 1994; Panagopoulos & Avlonitis, 2010). As Hunter and Perreault (2007, p.16) note, “[T]he strategic importance of the sales force to the organization’s success may be at an all-time high”. These relational issues are further complicated by the challenges of the unstable economic environment and the ever-present issues surrounding sales force productivity.

Among the tools at a sales manager’s disposal, perhaps the most obvious means of shaping salespeople’s attitudes and behaviors is the sales control system (Anderson & Oliver, 1987). Following the seminal work of Anderson and Oliver (1987), most empirical research has focused on the direct/indirect effects of behavior- and outcome-based control. The great majority of empirical evidence seems to favor behavior-based control because of its stronger positive associations with various salespersons’ psychological, behavioral and performance outcomes (see Baldauf, Cravens, & Piercy, 2005 for a comprehensive review). However, given that most sales organizations typically integrate the two types of sales control philosophies with varying degrees of emphasis (Cravens, Lask, Low, Marshall, & Moncrief, 2004; Jaworski, Stathakopoulos, & Krishnan, 1993; Oliver & Anderson, 1995; Onyemah & Anderson, 2009), this tradition of focusing research on the main effects of sales control systems seems unnecessarily restrictive.

Notable studies exploring this issue in detail include Jaworski et al. (1993) and Cravens et al. (2004). These studies investigated four combinations of control philosophies (high, bureaucratic, clan, and low control) employing both formal and informal sales control systems. While these studies made significant contributions to the sales control literature, important gaps remain. First, some results reported by these studies are not consistent. For example, in contrast to Jaworski et al. (1993), who found no differences in salesperson performance across different control combinations, Cravens et al. (2004) reported the highest mean value of salesperson performance in the high control group. This discrepancy begs the question: do sales control combinations always produce positive synergies? Second, both studies linked control combinations directly to salesperson performance. Walker, Churchill, and Ford’s (1977) classic framework for industrial sales force management and a later meta-analysis on the determinants of salesperson performance (Churchill, Ford, Hartley, & Walker, 1985) suggest that the effects of supervisor styles (e.g., sales control systems) on salesperson performance may be indirectly determined by three key intermediary variables: salesperson knowledge, role ambiguity, and intrinsic motivation.
development and empirical support found in the recent sales literature provides further evidence for these indirect relationships (e.g., Leigh & McGraw, 1989; Miao, Evans, & Zou, 2007; Oliver & Anderson, 1994; Piercy, Cravens, & Lane, 2001; Weitz, Sujan, & Sujan, 1986). Third, Jaworski et al. (1993) and Cravens et al. (2004) examined behavior control as a unidimensional construct. Behavior control can be further disaggregated into activity and capability control (Challagalla & Shervani, 1996). Activity and capability controls have been demonstrated to have differential and even opposite psychological and behavioral consequences in the sales literature (Fang, Evans, & Landry, 2005; Miao et al., 2007). Therefore, investigating activity and capability controls as distinct constructs may provide additional and important insights into the deployment of sales control combinations. For example, under Expectancy Theory (Vroom, 1964), outcome–capability and outcome–activity control combinations may have opposite effects on role ambiguity. The outcome–capability control combination may enhance role clarity by highlighting the importance (valence) and effectiveness of skills and knowledge (expectancy) in attaining outcome rewards (instrumentality). In contrast, the outcome–activity control combination may reduce role clarity because some of the requirements (e.g., call or expense reports) may come at the expense of sales activity, thereby compromising outcome performance expectancy. As such, the outcome–activity control combination can send contradictory messages to salespeople due to non-compatible role expectations.

This study advances the model depicted in Fig. 1 and subjects it to an empirical analysis. A Partial Least Squares (PLS) analysis is employed, which uses dyadic data from industrial salespeople and their managers to find substantive support for sales control combinatory effects on key intermediary variables, the key hypothesis of this study. Specifically, positive combinatory effects are found for outcome and capability controls. Outcome control effectively enhances salesperson knowledge and intrinsic motivation while reducing role ambiguity only when capability control is high. In contrast, the outcome–activity control combination can actually increase role ambiguity. Moreover, the activity–capability control combination can enhance role clarity while concurrently dampening intrinsic motivation. Finally, our results indicate that the effects of capability control and activity control on manager-rated salesperson performance are fully mediated while the effect of outcome control is partially mediated by salesperson knowledge and role ambiguity.

The remainder of this paper is organized as follows: first, a review of the pertinent theories is provided, and the hypotheses are developed; next, the research method is described, and the data analysis procedure and results are presented; then, the research findings are discussed, and their implications are provided; finally, the paper concludes with a discussion of its limitations and future research directions.

2. Theory and hypotheses

2.1. Conceptual framework and theoretical foundation

The model developed in the study is based on Walker et al.’s (1977) classic conceptualization of industrial sales force management. Their framework suggests that organizational variables, such as supervisory styles, affect salesperson performance indirectly via salesperson motivation, aptitude, and role perceptions. Moreover, these intermediary variables may have interactional effects on salesperson performance. A subsequent meta-analysis by Churchill et al. (1985) demonstrates that role perception, skill, and motivation are the top three determinants of salesperson performance and that their associations with performance are much stronger than the direct effects of organizational variables.

Because aptitude or skill can be considered a manifestation of one’s knowledge, salesperson knowledge is examined as one of the key intermediary variables in this study. This is consistent with Vargo and Lusch (2004, p. 6) that “knowledge is the fundamental unit of exchange” in the logic of the new dominant paradigm. Regarding role perception, this study focuses on role ambiguity because research has shown that role ambiguity is especially prevalent in industrial selling contexts and that it has a particularly detrimental impact on salesperson performance (Miao & Evans, 2007; Singh, 1993). Moreover, because intrinsic motivation plays a prominent role in the context of sales control combinations (Mallin & Pullins, 2009; Oliver & Anderson, 1995), salespeople’s intrinsic motivation is examined, and the effect of extrinsic motivation is treated as a control. The main hypotheses of this paper examine the combinatory effects of sales control on these intermediary variables and are drawn from Expectancy Theory and Cognitive Evaluation Theory, both of which will be discussed in detail in the following sections.

While Jaworski et al. (1993) and Cravens et al. (2004) investigate both formal and informal control combinations, this study focuses on...
formal control combinations. Formal control combinations are consistent with Anderson and Oliver’s (1987) original conceptualization of this topic and are much more likely (than informal control) to be the dominant control that is used in larger sales units (Jaworski et al., 1993), which is consistent with this study's sample size (mean employee number = 1079).

2.2. Sales control systems

In their seminal work on sales control systems, Anderson and Oliver (1987, p. 76) define a control system as “[A]n organization’s set of procedures for monitoring, directing, evaluating, and compensating its employees”. In this framework, a formal sales control system is considered to be a continuum ranging from behavior-based control to outcome-based control. Behavior control often entails intense managerial involvement in directing, training, evaluating and rewarding salespeople according to their inputs (e.g., selling activities and/or strategies), rather than simply focusing on immediate sales output. In contrast, outcome control approximates a market contract arrangement that uses incentives (e.g., commission or bonus) to reward salespeople on the basis of their sales outcomes (e.g., sales volume) with very limited involvement of management in guiding the selling process. Behavior control can be further disaggregated into activity and capability controls (Challagalla & Shervani, 1996). Under activity control, salespeople are required to perform a prescribed combination of selling activities that are deemed important for achieving desirable levels of performance. However, not all required activities directly contribute to sales outcomes, and these activities can sometimes reduce the time devoted to selling (e.g., filling out expense or call reports). Sales managers typically monitor salespeople’s actual behavior and reward them on the basis of their performance of these required activities. In contrast, capability control does not pre-specify a set of required selling/non-selling activities for the salesperson to follow. Instead, managers employing capability control set goals for the skill level and abilities of their salespeople, monitor those skills and abilities, coach salespeople, and reward them on the basis of their skill level and abilities. The sales manager can, for instance, illustrate why a particular course of action in a sales encounter is more effective in one situation but less so in another. Prior research has shown that activity control and capability control have differential and even opposite effects (Fang et al., 2005; Miao et al., 2007). Therefore, these concepts should be treated as distinct entities, not part of a global behavior control construct.

While the majority of empirical research on sales control systems has focused on comparing the effects of behavior and outcome control, some researchers have explored the synergistic nature of behavior and outcome control combinations. For example, using a sample of independent sales representatives in the electronic component industry, Oliver and Anderson (1995) demonstrated that a hybrid control strategy can be more effective than employing a singular control strategy based on outcome or behavior control. They showed that this hybrid strategy was effective at enhancing the sales reps’ intrinsic motivation and work planning and helped to achieve company sales/profit goals. A more recent study by Onyemah and Anderson (2009) also illustrated the synergistic nature of sales control combinations and showed a combinatory approach to be effective as long as the behavior and outcome control elements are internally congruent. More research, however, is needed to uncover the pathways through which sales control combinations deliver their differential impacts on salesperson performance. It is less clear whether activity, capability, and outcome controls have differential combinatory effects because previous research on hybrid control strategies has not distinguished between the two different types of behavior control (as this study does).

2.3. Role ambiguity

Singh (1998, p.70) defines role ambiguity as “the perceived lack of information a salesperson needs to perform his or her role adequately (e.g., effort instrumentalities) and his or her uncertainty about the expectations of different role set members”. Role ambiguity has been shown to be a critical intermediary variable in the sales context because salespeople who experience a high level of role ambiguity may be unsure whether their selling behaviors are appropriate in a given sales situation (Roman & Iacobucci, 2010).

Outcome control, activity control, and capability control can reduce role ambiguity because these sales control systems specify detailed behavior or outcome goals and provide necessary feedback (Challagalla & Shervani, 1996); however, their combinatory effects on role ambiguity are not well understood. Through utilizing Expectancy Theory (Cron, Dubinsky, & Michaels, 1988; Vroom, 1964; Walker, Churchill, & Ford, 1979), this study examines the combined effects of outcome–capability control, outcome–activity control, and activity–capability control. Expectancy Theory is based on three key components: valence, expectancy, and instrumentality. Valence refers to the perceived desirability of attaining a particular type of reward (e.g., commission) that is available to a salesperson. Expectancy is a salesperson’s estimate of the probability that expending a given amount of effort on a task will lead to an improved level of performance. Instrumentality alludes to a salesperson’s perception of the linkage between job performance (e.g., sales volume) and the attainment of rewards (e.g., monetary compensation). Because outcome control enhances the valence and instrumentality of goal achievement, it enables salespeople to infer role expectations within their sales organization, thereby reducing role ambiguity. The effect of outcome control on salespeople’s estimated probability of achieving sales goals, however, is less clear. This is because, in the context of outcome control, salespeople are left to devise their own best course of action. As a result, salespeople may expend greater effort without necessarily improving performance. For example, salespeople may believe that simply working harder can lead to better sales performance, when a change of strategies across selling situations would be more effective (Spirio & Weitz, 1990).

Capability control dictates that salespeople can best achieve sales goals through superior skills and abilities. That is, capability coaching and feedback enhance outcome performance expectancy by directing salespeople’s attention to the valence and instrumentality of working smarter, as opposed to simply working harder, in attaining outcome rewards. Importantly, capability control and outcome control are cognitively consistent because they provide role information in a complementary and holistic fashion (Simon, Snow, & Read, 2004). In contrast, activity control is not cognitively consistent with outcome control and can reduce salespeople’s role clarity when used concurrently with outcome control. The valence and instrumentality of certain activity goals may be at odds with those identified by outcome control techniques. Recall that some activities required by a job (e.g., call/expense reports) may reduce the amount of salespeople’s time devoted to selling, which can lead to lower outcome performance expectancy and role uncertainty because activity goals are achieved at the expense of outcome goal attainment. While previous literature has not considered the interaction between activity and capability control, Expectancy Theory would suggest that capability control may amplify the negative effect of activity control on role ambiguity. This amplification occurs because the valence and instrumentality of skills identified by capability coaching and feedback can increase salespeople’s expectancy to successfully perform requisite activity tasks. For example, a salesperson following basic selling techniques, as identified by activity control, may lose a sales opportunity even though the product seems to be ideal for the new account or client. In this situation, the sales manager may be able to identify the cause of this outcome; for example, it could be that the salesperson...
lacks assertiveness in his communication with the new account. Such diagnostic feedback can help the salesperson understand new customers’ expectations in future sales encounters, thereby providing important role information for the salesperson to fulfill activity goals. The above discussion leads to the following hypotheses:

H1a. The outcome–capability control combination is negatively associated with the salesperson’s role ambiguity.

H1b. The outcome–activity control combination is positively associated with the salesperson’s role ambiguity.

H1c. The activity–capability control combination is negatively associated with the salesperson’s role ambiguity.

2.4. Salesperson knowledge

Prior research suggests that sales control systems may have a significant impact on the development of the salesperson’s knowledge (Cravens, Ingram, LaForge, & Young, 1993; Kohli, Shervani, & Challagalla, 1998). In sales force management literature, selling effectiveness is believed to be a function of the salesperson’s knowledge structure, among other variables (e.g., Evans, Kleine, Landry, & Crosby, 2000; Leigh & McGraw, 1989; Rapp, Ahearne, Mathieu, & Schillewaert, 2006; Sharma, Levy, & Evanschitzky, 2000; Sujan, Weitz, & Sujan, 1988; Wilken, Cornelissen, Backhaus, & Schmitz, 2010). Previous research in this area has found that greater knowledge allows salespeople to handle different selling situations more effectively without experiencing undue cognitive stress (Evans et al., 2000). In addition, more effective salespeople were found to have more elaborate scripts (Leong, Busch, & John, 1989), show more variations in selling behaviors across customer types (Leigh & McGraw, 1989), and demonstrate more complete and complex selling strategies (Sharma et al., 2000). While an effective knowledge structure is critical for sales force productivity, the extent to which sales control combinations may facilitate or detract from the development of salesperson knowledge is less clear. Since most prior research on salesperson knowledge was conducted in a context outside business-to-business interactions, the data for this study targets this knowledge gap and addresses the industrial sales environment.

While the primary interest of this study lies in the combinatorial effects of sales control systems, the effects of outcome, capability, and activity control on salesperson knowledge are briefly discussed first. According to Expectancy Theory, outcome control elevates the valence and instrumentality of achieving sales goals, and this may have a positive effect on a salesperson’s learning orientation (Kohli et al., 1998), leading them to obtain relevant marketing intelligence and higher levels of knowledge. Similarly, capability control may directly improve salesperson knowledge because it emphasizes the salience and utility of selling capabilities and provides coaching and diagnostic feedback necessary for capability improvement. While some of the required routine activities (e.g., expense reports) do not directly contribute to sales-related knowledge, a high level of activity control nevertheless provides knowledge of the basic steps and procedures that a salesperson should use in the selling process (Fang et al., 2005). Therefore, a positive association between activity control and salesperson knowledge is also expected.

Capability control is likely to enhance the positive effect of outcome control on salesperson knowledge. To achieve outcome goals, salespeople will seek, process, and integrate diverse sources of information concerning their capabilities in order to regulate their effort and behavioral strategies (Carver & Scheier, 1982; Klein, 1989). Capability control can provide an important source of diagnostic information regarding the aspects of their skills/abilities that need improvement and/or reinforcement (i.e., valence) in order to attain outcome rewards (i.e., instrumentality). This type of capability assessment and feedback can help salespeople recalibrate their selling strategies during the sales process, thereby increasing performance expectancy via enhanced knowledge. Activity control, in contrast, is expected to dampen the positive effect of outcome control on salesperson knowledge. Although activity control elevates the valence and instrumentality of fulfilling activity goals, it may conflict with the valence and instrumentality of outcome goals. Non-sales related activities (e.g., filling out expense reports) may detract from salespeople’s time devoted to sales, thereby reducing their chance to learn during the selling process. Moreover, activity control does not provide diagnostic feedback, and thus, salespeople are unable to recalibrate their strategies in different sales encounters effectively (Mallin & Pullins, 2009). For example, a salesperson may be required to visit a certain number of new accounts every month. Activity control may not inform the salesperson as to why some new accounts declined the salesperson’s visit. Consequently, the salesperson may simply try to contact more new accounts without necessarily improving their knowledge and performance expectancy. When activity control is combined with capability control, Expectancy Theory would suggest that they have a positive effect on salesperson knowledge. Capability control can highlight the importance (i.e., valence) and utility (i.e., instrumentality) of understanding the basic steps and procedures required by activity control. Consequently, salespeople will likely have enhanced expectancy when performing activity tasks due to their improved knowledge. We hypothesize the following:

H2a. The outcome–capability control combination is positively associated with salesperson knowledge.

H2b. The outcome–activity control combination is positively associated with salesperson knowledge.

H2c. The activity–capability control combination is positively associated with salesperson knowledge.

2.5. Intrinsic motivation

Motivation is typically defined as a psychological state that initiates and guides a person’s behavior or conscious choices (Brown & Peterson, 1994). Whereas extrinsic motivation reflects the extent to which salespeople treat work as a means for obtaining external rewards (e.g., money, recognition, and promotion), intrinsic motivation measures the extent to which salespeople are driven by a passionate interest and deep level of enjoyment in what they do (Amabile, Hill, Hennessey, & Tighe, 1994; Weitz et al., 1986). Intrinsic motivation is particularly relevant to sales research because of its implications for selling effectiveness (Roman & Iacobucci, 2010) and the well-being of salespeople (Ryan & Deci, 2000). In the sales control context, the hybrid control system seems to be more strongly linked to salespeople’s intrinsic motivation compared with the singular use of behavioral or outcome control (Mallin & Pullins, 2009; Oliver & Anderson, 1995). Previous research suggests that capability control may have a positive effect on intrinsic motivation because it emphasizes selling skills and knowledge that can increase salespeople’s intrinsic reward orientation (Weitz et al., 1986). The main effect of activity control on intrinsic motivation, however, is somewhat ambiguous. On the one hand, feedback relating to sales strategies embedded in activity control may increase intrinsic motivation because of the perception of enhanced competence (Anderson & Oliver, 1987); on the other hand, top-down determination of routine activities may lead to an external locus of control and loss of self-determination, which can dampen intrinsic motivation. Similarly, outcome control may dampen intrinsic motivation, depending on whether such outcome feedback and rewards are construed as “controlling” or “informational” in nature (Anderson & Oliver, 1987). Cognitive Evaluation Theory (CET) suggests that if outcome feedback is considered relevant to improving
one’s competence and abilities, it can enhance an internal locus of control and intrinsic motivation; however, if salespeople consider outcome feedback and rewards to be measures of control, intrinsic motivation will decrease (Deci & Ryan, 1985).

In accordance with CET, capability control and outcome control may have a positive combinatory effect on intrinsic motivation because capability feedback potentially alters salespeople’s perception of the nature of outcome rewards. Specifically, capability evaluation provides diagnostic feedback and adds informational value by showing salespeople why they were or were not successful in particular selling situations (Mallin & Pullins, 2009). Therefore, outcome rewards will more likely be perceived by the salesperson as an indicator of how well she/he has mastered adequate knowledge and professional competence. In contrast, when activity control is combined with outcome control, it may dampen intrinsic motivation. Unlike capability control, salespeople may perceive activity control as manipulating his/her behavior, and this can occur because activity control pre-specifies what and how tasks should be performed (Mallin & Pullins, 2009). Therefore, in this context, outcome rewards are more likely to be construed by salespeople as controlling rather than informational in nature, thereby leading to external locus ascription and reduced intrinsic motivation. Furthermore, activity control and capability control may have a negative combinatory effect on intrinsic motivation, because activity control may restrict salespeople’s perceived autonomy and hinder their ability to utilize new strategies learned from capability coaching and feedback, thereby dampening the positive effect of capability coaching on intrinsic motivation. The following hypotheses are proposed:

H3a. The outcome–capability control combination is positively associated with intrinsic motivation.

H3b. The outcome–activity control combination is negatively associated with intrinsic motivation.

H3c. The activity–capability control combination is negatively associated with intrinsic motivation.

2.6. Impact on salesperson performance

Fig. 1 suggests that sales control systems indirectly affect salesperson performance via salesperson knowledge, role ambiguity, and intrinsic motivation. The main effects of these intermediary variables on salesperson performance have been established in the previous literature. Research suggests that salesperson knowledge is one of the primary determinants of salesperson performance because extensive knowledge enhances a salesperson’s effectiveness across many different selling situations (e.g., Leigh & McGraw, 1989; Sharma et al., 2000; Sujan et al., 1988). The detrimental effect of role ambiguity on salesperson performance has been demonstrated extensively throughout previous research on this subject. When salespeople experience a high level of role ambiguity, they have a difficult time understanding what is expected of them and what courses of action are needed to complete their required tasks, thereby resulting in lower levels of performance (Jaworski & Kohli, 1991; Ramaswami, 1996). Moreover, salespeople who have high levels of intrinsic motivation are more likely to practice adaptive selling and to accept failure as a learning experience, leading to enhanced performance (Roman & Iacobucci, 2010; Weitz et al., 1986).

Beyond these main effects, Fig. 1 also suggests that there are interactive effects of these intermediary variables on salesperson performance, and this is consistent with Walker et al.’s (1977) argument that these variables likely possess multiplicative effects. Specifically, it is suggested that (1) intrinsic motivation and (2) salesperson knowledge mitigate the deleterious effect of role ambiguity on salesperson performance. Role ambiguity is unavoidable in the sales environment because of the wide range of tasks and contexts that employees must address. Accordingly, intrinsically motivated salespeople should be better able to tolerate higher levels of role ambiguity because they possess an inherent need to learn more about selling and are more willing to accept failure in ambiguous environments if it provides a learning opportunity (Roman & Iacobucci, 2010; Weitz et al., 1986). Therefore, the negative effect of role ambiguity on salesperson performance may be less pronounced when intrinsic motivation is high. Similarly, knowledgeable salespeople may be able to outweigh the negative effects of role ambiguity on performance because their superior knowledge allows them to adapt to an otherwise ambiguous selling environment (Evans et al., 2000). The following hypotheses are developed:

H4. Intrinsic motivation mitigates the negative effect of role ambiguity on salesperson performance.

H5. Salesperson knowledge alleviates the negative effect of role ambiguity on salesperson performance.

3. Research method

3.1. Data collection

This study was conducted within the US manufacturing sector (SIC codes 20–39), using industrial salespeople and their managers as respondents. This empirical approach is appropriate for the current study because its theoretical framework requires an adequate variability of behavior and outcome control combinations across selling contexts (e.g., Fang et al., 2005; Ramaswami, 1996). A random name list of 1561 industrial sales managers within SIC codes 20–39 was obtained from a leading list broker. A marketing research firm pre-qualified sales managers from the list, and these managers agreed to provide access to their salespeople to participate in the study. The pre-qualification procedure ensured that (1) the respondent was a sales manager, (2) he or she directly supervised salespeople, and (3) the sales manager was interested and willing to fill out the sales manager survey and to distribute salesperson surveys to up to three salespeople that he or she directly supervised. To encourage participation, sales managers were promised a copy of the summary of the study findings, and the participating sales managers and their salespeople were entered into a raffle of ten $25 gift certificates. This pre-qualification procedure generated a total of 471 qualified sales managers (for a 30.2% response rate) who subsequently provided a pool of 1371 salespeople for this study.

Data collection began three days after the completion of the telephone pre-qualification. Specifically, an envelope enclosing a cover letter providing instructions to the sales manager, a copy of the sales manager survey, and a postage-paid return envelope were sent to each of the 471 sales managers. Also enclosed in each envelope were cover letters and surveys for salespeople and postage-paid return envelopes. Three weeks after the first mailing, a reminder package enclosing the same materials was sent to each of the 471 sales managers. This two-wave mailing effort generated 223 completed salesperson surveys (for a response rate of 16.3%) and 100 completed sales manager surveys (for a response rate of 21.2%), providing a total of 282 individual salesperson-performance evaluations. Of these responses, 195 salesperson-sales manager dyads were obtained and used for testing our conceptual framework.

To assess potential nonresponse bias, several methods were used. First, comparisons of early and late responses (first 20% versus last 20%) across study variables resulted in nonsignificant differences (p > .10). Second, the matched (n = 195) and unmatched salespeople samples (n = 28) were compared. No significant differences (p > .10) were found across study variables. Finally, the matched (n = 195) and unmatched (n = 87) data reported by sales managers were compared.
Again, no significant differences were found (p > .10) in company characteristics (i.e., company size, market share and sales growth) or salesperson performance. These results collectively suggest that nonresponse bias is not a serious concern.

3.2. Measurement scales

Whenever possible, validated multi-item measures in the existing literature were used; all are reflective scales except for salesperson performance and selling effort, which are treated as formative scales² (see Appendix A). Sales managers rated salespeople’s performance and provided information regarding the sales organization’s size, market share of its primary product line, as well as its annual sales growth rate. Salespeople reported the remaining variables depicted in Fig. 1.

The three sales control constructs were measured with scales from Kohli et al. (1998). Specifically, activity control (α = .86), capability control (α = .92), and outcome control (α = .94) were each measured with five items. Because of the lengthy and complex nature of measures in previous studies focusing on salesperson knowledge (e.g., Leigh & McGraw, 1989; Sujan et al., 1988) and space constraints in the survey, a five-item scale of salesperson knowledge (α = .89) was developed for this study, allowing us to determine the salesperson’s knowledge structure. Role ambiguity (α = .76) was measured with three items from Netemeyer, Alejandro, and Boles (2004). Intrinsic motivation (α = .90) was measured with four items adapted from Miao et al.’s (2007) study in a similar sales control context. Manager-rated salesperson performance was measured with five items from Cravens et al. (1993). Three control variables were included—extrinsic motivation, salesperson experience, and selling effort—as they may influence salesperson performance. In addition, because intrinsic motivation, salesperson knowledge, and role ambiguity may change as the salesperson gains more experience (Cron et al., 1988), the effect of salesperson experience on these intermediary variables was also taken into account. Extrinsic motivation (α = .80) was measured with two items from Miao et al. (2007), and sales experience was assessed with a single-item measure asking for the respondent’s number of years of full-time experience. Selling effort was measured with three items from Sujan, Weitz, and Kumar (1994).

3.3. Measurement model

Confirmatory factor analysis (CFA) was performed using EQS 6.1 to evaluate the measurement properties of the seven latent reflective constructs in this study. Each item was restricted to its a priori factor and each factor was allowed to correlate with all of the other factors. The measurement model demonstrated an acceptable fit (χ²(1565) = 725.470, p < .01; NFI = .935, NNFI = .961, CFI = .966, SRMR = .061, RMSEA = .074). All a priori factor loadings of the latent constructs were large, positive, and significant (p < .01), demonstrating convergent validity (Bagazzi, Yi, & Phillips, 1991). Discriminant validity was established in two ways. Each pair of constructs was assessed using nested CFA models in which a one-factor model was compared with a two-factor model using chi-square difference tests; in each case, the two-factor model had a significantly better fit (p < .01). In addition, the average variance extracted (AVE) by each construct was greater than its shared variance with all other constructs (Fornell & Larcker, 1981). Table 1 summarizes the descriptive statistics of the data.

3.4. Common method variance

Because salespeople provided information about the sales control systems and the intermediary variables, there is some concern regarding the presence of potential common method variance (CMV). To assess the extent to which CMV may exist in the data, we employed a procedure used by Carson (2007) to estimate a combined congeneric measurement model with an additional common method factor. The common method factor loads on all items and controls for variance and covariance among the items introduced by responses from a single informant. To achieve model convergence, all loadings of the method factor were specified as equal in size, reflecting the assumption that CMV affects all items equally. In addition, the common method factor is not correlated with other trait constructs, reflecting the assumption that the degree of CMV is independent of the true magnitude of trait constructs (cf., Homburg, Muller, & Klamann, 2011). Variance decompositions and attenuated AVE for each construct are given in Table 2. Trait variance (average 84.1%) was found to significantly exceed both method (average 3.9%) and error variance (average 12.0%) for all reflective scales. Because the method factor accounts for a very small percentage of the variance explained and because this study’s framework involves complex interaction effects, common method variance is not likely a serious threat in this study.

4. Results

4.1. Hypotheses testing

A PLS analysis was performed for hypotheses testing because PLS analysis is capable of accommodating both reflective and

---

² According to Rossiter’s (2002) article in International Journal of Research in Marketing, many existing scales previously considered to be reflective in nature should actually be formative. Consistent with Rossiter, salesperson performance and selling effort were measured as formative scales because each item measures a distinct aspect of the construct, which may not be interchangeable. Thanks to one of the anonymous reviewers for identifying this issue.
formative scales. In addition, PLS analysis allows the multiplication of standardized item scores to create latent interaction factors (Ahearne, MacKenzie, Podsakoff, Mathieu, & Lam, 2010). Outcome control was the only sales control style to have a significant effect on salesperson performance. Therefore, the influence of outcome control is partially mediated, and the effects of capability and activity controls on salesperson performance are fully mediated. The results of the PLS tests are reported in Table 3. In addition, significant interactions were plotted using simple slope analysis. These plots depict how the effects of independent variables differ when high and low levels of the moderating variables are present. The results appear in Fig. 2.

Table 2

<table>
<thead>
<tr>
<th>Scale</th>
<th>Trait†</th>
<th>Method‡</th>
<th>Error§</th>
<th>Attenuated AVE¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome control</td>
<td>0.888</td>
<td>0.031</td>
<td>0.081</td>
<td>0.690</td>
</tr>
<tr>
<td>Capability control</td>
<td>0.866</td>
<td>0.031</td>
<td>0.103</td>
<td>0.630</td>
</tr>
<tr>
<td>Activity control</td>
<td>0.824</td>
<td>0.038</td>
<td>0.138</td>
<td>0.548</td>
</tr>
<tr>
<td>Salesperson Knowledge</td>
<td>0.854</td>
<td>0.061</td>
<td>0.085</td>
<td>0.669</td>
</tr>
<tr>
<td>Role ambiguity</td>
<td>0.799</td>
<td>0.036</td>
<td>0.165</td>
<td>0.625</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>0.852</td>
<td>0.039</td>
<td>0.109</td>
<td>0.662</td>
</tr>
<tr>
<td>Extrinsic motivation</td>
<td>0.804</td>
<td>0.035</td>
<td>0.161</td>
<td>0.718</td>
</tr>
</tbody>
</table>

† Percentage of variance explained by trait (construct).
‡ Percentage of variance explained by common method factor.
§ Percentage of variance explained by error.
¶ Attenuated for common method variance/covariance in the items.

H1a predicts that the outcome–capability control combination is negatively related to role ambiguity. As shown in Table 3, H1a is supported (β = −.32, p < .05). The simple slope analysis (Fig. 2, panel A) shows that, at a high level of capability control (one standard deviation above the mean), the effect of outcome control on role ambiguity is negative (β = −.32, p < .05). At a low level of capability control (one standard deviation below the mean), outcome control has no effect (β = −.12, ns). H1b posits that activity control and outcome control possess a positive combinatorial effect on role ambiguity, and this is supported (β = .24, p < .05). The simple slope analysis (Fig. 2, panel A) shows that, at a high level of activity control, outcome control is not related to role ambiguity (β = −.14, ns); at a low level of activity control, outcome control is negatively related to role ambiguity (β = −.31, p < .05). As predicted, the activity–capability control combination is negatively related to role ambiguity (β = −.20, p < .05), thereby supporting H1c. The simple slope analysis (Fig. 2, panel A) shows that, at a high level of capability control, activity control is negatively related to role ambiguity (β = −.31, p < .01); at a low level of capability control, activity control is not related to role ambiguity (β = −.09, ns). With respect to main effects of sales control systems on role ambiguity, the effect of activity control is negative and significant (β = −.30, p < .01); capability control, however, has a positive effect (β = .34, p < .01). Such an effect must be interpreted cautiously considering this variable’s significant interaction effects with outcome control and activity control. Finally, outcome control has a negative main effect on role ambiguity (β = −.42, p < .01). H2a states that the outcome–capability control combination is positively related to salesperson knowledge, which is supported by the data (β = .19, p < .10). The simple slope analysis (Fig. 2, panel B) shows that, at a high level of capability control, outcome control has a positive effect on salesperson knowledge (β = .55, p < .01); at a low level of capability control, outcome control has no effect on salesperson knowledge (β = .27, ns). H2b and H2c are rejected because neither the outcome–activity control combination (β = .07, ns) nor the activity–capability control combination (β = −.02, ns) is significantly related to salesperson knowledge. In addition, activity control has a positive main effect on salesperson knowledge (β = .20, p < .05); capability control, however, has a negative main effect (β = −.20, p < .01). This latter result countered our expectations and will be discussed in more detail later in this paper. Outcome control has a positive effect (β = .47, p < .01) on salesperson knowledge, which is consistent with what was expected. H3a posits that the outcome–capability control combination is positively associated with intrinsic motivation, and it is supported (β = .33, p < .05). The simple slope analysis (Fig. 2, panel C) shows that, at a high level of capability control, the effect of outcome control on intrinsic motivation is significant and positive (β = .49, p < .01); at a low level of capability control, however, outcome control has a much lower positive effect (β = .28, p < .05). H3b is rejected because the activity–outcome control combination is not associated with intrinsic motivation (β = −.04, ns). H3c is supported because the activity–capability control combination is negatively related to intrinsic motivation (β = −.21, p < .05). Fig. 2 (panel C) demonstrates that, at a high level of activity control,
capability control has no effect on intrinsic motivation ($\beta = .15$, ns); at a low level of activity control, capability control has a positive effect on intrinsic motivation ($\beta = .29$, $p < .05$). Moreover, outcome control was found to have a positive main effect on intrinsic motivation ($\beta = .35$, $p < .01$), whereas activity control ($\beta = .11$, ns) and capability control ($\beta = .10$, ns) have no direct effects.
In terms of impact on salesperson performance, salesperson knowledge has a positive effect ($\beta = .17, p < .05$), role ambiguity has a negative effect ($\beta = -.20, p < .05$), and intrinsic motivation has no effect ($\beta = -.03, ns$). Beyond these direct effects, intrinsic motivation was found to mitigate the negative effect of role ambiguity on salesperson performance ($\beta = .26, p < .01$), thereby supporting H4. The simple slope analysis (Fig. 2, panel D) shows that, at a high level of intrinsic motivation, role ambiguity has a negative but non-significant effect on salesperson performance ($\beta = -.29, ns$); at a low level of intrinsic motivation, however, role ambiguity has a significant negative effect on salesperson performance ($\beta = -.52, p < .01$). In contrast to H5, however, salesperson knowledge was found to amplify the negative effect of role ambiguity on performance ($\beta = -.15, p < .05$), and thus, H5 is rejected. The simple slope analysis (Fig. 2, panel D) demonstrates that, at a high level of salesperson knowledge, role ambiguity has a negative effect on salesperson performance ($\beta = -.39, p < .01$); at a low level of salesperson knowledge, however, role ambiguity has no effect on salesperson performance ($\beta = -.18, ns$). It appears that role ambiguity has an especially deleterious effect on performance when the salesperson has a high level, as opposed to a low level, of knowledge. We explore a possible explanation of this finding in the Discussion section. Finally, the PLS analysis reported a total combined effect of sales control systems (activity, capability, and outcome control) on salesperson performance via salesperson knowledge and role ambiguity, which is measured at .25 ($p < .05$).

4.2. Robustness check

In the data used for this study, salespeople (level one) are clustered under their sales manager (level two). That is, within-cluster errors may not be independent, and this may affect the results derived from the PLS analysis. To test the potential of this within-cluster effect, we followed the recommendations of Mizik and Jacobson (2009) by allowing for correlated error terms within the clusters. Specifically, in the Generalized Estimating Equations (GEE) procedure, manager-rated salesperson performance was regressed on all salesperson-reported constructs (i.e., activity control, capability control, outcome control, salesperson knowledge, role ambiguity, intrinsic motivation, salesperson experience, extrinsic motivation, and selling effort). In the GEE procedure, the working correlation matrix accounts for the within-cluster error dependencies using cluster-robust standard errors estimation. It displays Quasi-likelihood under independence model criterion (QIC) and Corrected Quasi-likelihood under independence model criterion (QICC) for choosing the best correlation structure (with a “smaller-is-better” criterion). It was found that (1) the results of the correlated-errors GEE model were consistent with those derived from the PLS analysis and that (2) the independence model has a better goodness-of-fit (QIC = 2557.792) than does the correlated-errors model (QICC = 2559.827). Therefore, it is more parsimonious to employ the independence model, and the results based on the PLS analysis are reliable.

5. Discussion

This study investigates the combinatory effects of sales control systems above and beyond their singular effects. Substantive support from the empirical evidence provides several important theoretical and managerial contributions to the existing sales management literature. First, in contrast to previous research linking sales control combinations directly to salesperson performance (e.g., Cravens et al., 2004; Jaworski et al., 1993; Oliver & Anderson, 1995; Onyemah & Anderson, 2009), this research shows that the differential effects of sales control combinations are mediated by salesperson knowledge and role ambiguity. Second, this research extends themes present in the literature on sales control combinations by distinguishing between activity and capability controls within the global behavior control construct. When outcome control is combined with capability control, salespeople are more likely to have a lower level of role ambiguity and enhanced levels of knowledge and intrinsic motivation. In contrast, when outcome control is coupled with activity control, undesirable consequences may result. Outcome goals and activity goals may not be perceived to be consistent by salespeople, thereby leading to role confusion. Perhaps more interestingly, the activity-capability control combination was found to simultaneously enhance role clarity and reduce intrinsic motivation. Therefore, these findings suggest that mixing apples (i.e., activity control) with oranges (i.e., capability control) in sales control combinations can have unexpected consequences, and thus, researchers should distinguish between these two types of behavior control when investigating hybrid control systems.

To investigate the extent to which managers have combined the “correct” sales control styles, the construct of outcome control was median split into high (>5.6) versus low groups (<5.6). The analysis examined how many salespeople in the high outcome control group also reported a comparable level of capability control versus activity control (i.e., >5.6). Managers who have combined a high level of outcome control with a high level of capability control in only 6.59% of cases, whereas as many as 25.27% of cases reported a high outcome–high activity control combination (which should be avoided as it may increase role ambiguity). Moreover, only 27.84% of cases were characterized by high activity–high capability control combination, which may reduce role ambiguity. These results are particularly informative, and managers are encouraged to integrate capability control into outcome goals because of the beneficial combinatory effects found on role clarity, salesperson knowledge, and intrinsic motivation, which can subsequently have a positive effect on salesperson performance.

The results of this study also show that sales control styles have significant main effects on role ambiguity, salesperson knowledge, and intrinsic motivation. While the results of outcome control and activity control are consistent with our expectations, the main effects of capability control are somewhat surprising. Capability control has a negative effect on salesperson knowledge and a positive effect on role ambiguity. These findings indicate that the oft-touted benefits of capability control as a standalone control philosophy need to be qualified (Kohli et al., 1998). Capability control appears to effectively enhance salesperson knowledge and role clarity only when it is accompanied with a high level of outcome control. When outcome control is not employed concurrently, capability control does not seem to be very effective. These findings confirm Challagalla and Shervani’s (1996) observation that the subjective appraisal of salesperson capability may be viewed as suspicious and biased. Because capability control is prone to variability in interpretation and reference standards, pure reliance on capability control without a more objective reference point could inhibit the role clarity and knowledge accumulation of the salesperson. These results suggest that capability control, when employed as a standalone supervisory tool, may not have the presumed effect on performance as sales control theory, to date, has suggested (e.g., Challagalla & Shervani, 1996; Fang et al., 2005; Kohli et al., 1998; Miao et al., 2007). Therefore, to attain a more comprehensive understanding of the various sales control systems, knowledge of their combinatory effects is required.

This study also illustrates the interactive effects of the intermediatory variables on salesperson performance. While salespeople with a high level of intrinsic motivation may reduce the negative effect of

---

\(^{1}\) Thanks to the editor for his suggestion for performing this post hoc analysis.
role ambiguity, as anticipated, the moderating effect of salesperson knowledge is contrary to what was hypothesized. Specifically, role ambiguity’s deleterious effect on performance was especially acute when the salesperson is very knowledgeable. A plausible explanation for this counterintuitive finding might be that knowledgeable salespeople learn to typify selling situations which, in turn, provide role information and guide behaviors. However, in sales tasks that are characterized by a high level of role ambiguity (e.g., atypical selling situations), their knowledge structure becomes less useful, thereby leading to lower levels of performance. In contrast, less knowledgeable salespeople have a much lower base level of performance. As such, the drop in their performance due to role ambiguity is much less noticeable compared with their knowledgeable and normally high-performing counterparts. From a managerial perspective, sales managers should always strive to reduce salespeople’s role ambiguity. It does not appear that knowledgeable salespeople are less likely to suffer from the negative consequences of role ambiguity.

Overall, our findings highlight the critical intermediary roles of salesperson knowledge and role ambiguity, as well as the moderating role of intrinsic motivation, in the sales control context. Managers should continuously update and strengthen the knowledge structure of their salespeople, reduce role ambiguity in the selling environment, and maintain a higher level of intrinsic motivation in their sales force with appropriately designed sales control combinations.

6. Limitations and future research directions

As with any research project, this study is subject to some limitations. First, although salesperson performance was evaluated by the sales manager, the remaining variables came from the same source, since they were reported by the salesperson. Nevertheless, post hoc analysis indicated that common method bias was not a serious threat in this study. In addition, the significant interaction effects observed also alleviated the concern of common method bias because it would be very difficult, if not impossible, for the respondents to have correctly guessed both the positive and negative interaction hypotheses. Second, all respondents came from manufacturing industries involving business-to-business selling in the United States. Therefore, the generalizability of the results to other settings such as consumer marketing or sales force management outside the US (e.g., within the European Union) could not be assumed without further empirical testing.

This study also provides inviting avenues for future research. Activity, capability, and outcome controls were measured in a global scale; however, it has been suggested that each type of control includes dimensions of feedback, rewards, and punishments. The extent to which those dimensions of sales control may affect salesperson knowledge, role ambiguity, and intrinsic motivation in a combinatorial fashion is not clear. In the same vein, the measures of activity control are global in nature without distinguishing between repetitive routine activities (e.g., sales report) and strategic selling procedures (e.g., steps to overcome customer objections). To the extent that activity control provides feedback on an appropriate combination of selling strategies that salespeople should utilize (Fang et al., 2005), developing separate measures to capture the more repetitive (versus the more strategic) aspects of activity control is warranted. Perhaps more interestingly, salespeople’s attribution ascriptions may change depending on success or failure conditions (Fang et al., 2005). The model developed in this paper may behave differently when salespeople receive positive feedback (i.e., success condition) as opposed to negative feedback (i.e., failure condition) from their sales managers. These intriguing questions can only be answered in future studies.

Appendix A. Study measures and loadings

<table>
<thead>
<tr>
<th>Activity control (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My manager informs me about the sales activities I am expected to perform</td>
<td>0.731</td>
</tr>
<tr>
<td>2. My manager monitors how I perform required sales activities</td>
<td>0.845</td>
</tr>
<tr>
<td>3. My manager informs me on whether I meet his/her expectations on sales activities</td>
<td>0.867</td>
</tr>
<tr>
<td>4. My manager readjusts my sales activities when necessary</td>
<td>0.697</td>
</tr>
<tr>
<td>5. I would be recognized by my manager if I perform sales activities well</td>
<td>0.671</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capability control (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My manager periodically evaluates the selling skills I use to accomplish a task (e.g., how I negotiate)</td>
<td>0.807</td>
</tr>
<tr>
<td>2. My manager provides guidance on ways to improve my selling skills and abilities</td>
<td>0.887</td>
</tr>
<tr>
<td>3. My manager evaluates how I make sales presentations and communicate with customers</td>
<td>0.880</td>
</tr>
<tr>
<td>4. My manager assists me by illustrating why using a particular sales approach may be effective</td>
<td>0.869</td>
</tr>
<tr>
<td>5. I would be commended if I improve my selling skills</td>
<td>0.733</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome control (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My manager tells me about the expected level of achievement on sales volume or market share targets</td>
<td>0.830</td>
</tr>
<tr>
<td>2. My manager monitors my performance on achieving sales volume or market share targets</td>
<td>0.867</td>
</tr>
<tr>
<td>3. I receive frequent feedback on whether I am meeting expected achievement on sales volume or market share targets</td>
<td>0.927</td>
</tr>
<tr>
<td>4. My manager ensures that I am aware of the extent to which I attain sales volume or market share targets</td>
<td>0.916</td>
</tr>
<tr>
<td>5. I would be recognized by my manager if I perform well on sales volume or market share targets</td>
<td>0.793</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Salesperson knowledge (1 = poor, 7 = outstanding)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My knowledge of identifying distinct customer categories in terms of their characteristics and preferences is…</td>
<td>0.78</td>
</tr>
<tr>
<td>2. My knowledge of identifying distinct customer preferences is…</td>
<td>0.783</td>
</tr>
<tr>
<td>3. My knowledge of matching appropriate selling strategies with distinct customer categories is…</td>
<td>0.831</td>
</tr>
<tr>
<td>4. My knowledge of matching my solutions with distinct customer needs is…</td>
<td>0.790</td>
</tr>
<tr>
<td>5. My knowledge of adjusting my selling approaches when new customer information becomes available is…</td>
<td>0.755</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Role ambiguity (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I feel certain about how much authority I have (reverse-coded)</td>
<td>0.520</td>
</tr>
<tr>
<td>2. I know what my responsibilities are (reverse-coded)</td>
<td>0.852</td>
</tr>
<tr>
<td>3. I know exactly what is expected of me (reverse-coded)</td>
<td>0.901</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intrinsic motivation (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What matters most to me is enjoying my selling job</td>
<td>0.825</td>
</tr>
<tr>
<td>2. It is important for me to be able to enjoy my selling job</td>
<td>0.808</td>
</tr>
<tr>
<td>3. I enjoy selling for the pleasure of it</td>
<td>0.808</td>
</tr>
<tr>
<td>4. It is the experience of selling that gives me the most pleasure</td>
<td>0.866</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extrinsic motivation (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am strongly motivated by the money I can earn through my sales job</td>
<td>0.753</td>
</tr>
<tr>
<td>2. I am keenly aware of the income goals I have for myself</td>
<td>0.908</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selling effort (1 = strongly disagree, 7 = strongly agree)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I work long hours to meet my sales objectives</td>
<td>–</td>
</tr>
<tr>
<td>2. I do not give up easily when I encounter a difficult customer</td>
<td>–</td>
</tr>
<tr>
<td>3. I work unceasingly at selling a customer until I get an order</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Salesperson performance (1 = poor, 7 = outstanding)</th>
<th>Standardized loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Listening attentively to identify and understand the real concerns of customers</td>
<td>–</td>
</tr>
<tr>
<td>2. Convincing customers that s/he understands their unique problems and concerns</td>
<td>–</td>
</tr>
<tr>
<td>3. Using established contacts to develop new customers</td>
<td>–</td>
</tr>
<tr>
<td>4. Communicating his/her sales presentation clearly and concisely</td>
<td>–</td>
</tr>
<tr>
<td>5. Providing satisfying solutions to customers’ problems</td>
<td>–</td>
</tr>
</tbody>
</table>

*Formative scale.
Impact of online channel use on customer revenues and costs to serve: Considering product portfolios and self-selection

Sonja Gensler, a,⁎, Peter Leeﬂang, a,b,1, Bernd Skiera c,2

a Department of Marketing, University of Groningen, Post Box 800, 9700 AV Groningen, The Netherlands
b LUISS Guido Carli, Rome, Italy
c Department of Marketing, University of Frankfurt, Grüneburgplatz 1, 60323 Frankfurt am Main, Germany

A R T I C L E   I N F O

Article history:
First received in 5, March 2010 and was under review for 9 months
Available online 8 March 2012

Area Editor: Koen H. Pauwels

Keywords:
Channel management
Self-selection effect
Product portfolio
Matching method
Online marketing
Distribution

A B S T R A C T

Developing a strategy for online channels requires knowledge of the effects of customers’ online use on their revenue and cost to serve, which ultimately influence customer profitability. The authors theoretically discuss and empirically examine these effects. An empirical study of retail banking customers reveals that online use improves customer profitability by increasing customer revenue and decreasing cost to serve. Moreover, the revenue effects of online use are substantially larger than the cost-to-serve effects, although the effects of online use on customer revenue and cost to serve vary by product portfolio. Self-selection effects also emerge and can be even greater than online use effects. Ignoring self-selection effects thus can lead to poor managerial decision-making.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Channels are an integral part of any firm’s customer management strategy, and many firms have established online channels, often in an attempt to improve customer profitability (Gensler, Dekimpe, & Skiera, 2007; Hitt & Frei, 2002). Popular press reports imply that there is a positive association between customers’ online use and customer profitability; an e-commerce executive at Bank of America asserted, for example, that the firm’s 12.6 million online banking customers are 27% more profitable than its offline customers (Tedeschi, 2005). Such statements likely encourage managers to push customers to online channels. There are several reasons to believe, however, that simple comparisons of online and offline customer profitability can provide only limited insight into the effects of online use.

First, differences in profitability between online and offline customers may reﬂect self-selection effects, which exist if online and offline customers differ in characteristics such as age. Differences in customer profitability thus would be driven not by customers’ online use but mainly by differences in these customer characteristics. Accordingly, ﬁrms must account for self-selection effects when investigating whether customers’ online use affects key metrics such as customer profitability and whether managing customers’ use of channels is even useful.

Second, ﬁrms need to identify the primary function of their online channels, and they need to separate the effects of customers’ online use into the speciﬁc ways in which it inﬂuences customer revenue and cost to serve. Online channels can develop and enhance customer relationships if their use improves customer revenue (Blattberg, Kim, & Neslin, 2008). If online channels allow for decreasing the cost to serving customers, they are particularly suited to managing customer relationships that have limited cross- or up-selling opportunities. Thus understanding the revenue and cost-to-serve effects of online use suggests which strategy to pursue to ensure effective customer relationships.

Third, ﬁrms must determine whether the effects of customers’ online use vary depending on the mix of products used by the customers (i.e. product portfolio). If the effects of online use vary by customer product portfolio, a better understanding of the differential effects of customers’ online use would offer key insights into which customers the ﬁrm should target when aiming to actively manage customers’ use of channels.

Accordingly, we aim to determine the effects of customers’ online use on customer revenue and cost to serve by taking into account self-selection effects and the moderating impact of the customer’s product portfolio. We further study the effect of customers’ online use on customer profitability empirically. To determine the extent of self-selection effects, we use data pertaining to approximately 87,000 customers of a large European retail bank. Our results offer three main contributions.

First, our study reveals both the revenue and cost effects of customers’ online use, enabling us to assess which effect is the primary driver of differences in customer profitability. In contrast, most previous studies (see Table 1) have focused on either customer profitability or...
customer revenue in relation to online use. Campbell and Frei (2010) provide insights into the profitability and cost effects of online use, but the way in which they present their results does not allow any inference regarding the revenue effects of online use. Consequently, our knowledge about how to employ online channels to manage customer relationships remains quite limited, as also noted by Neslin and Shankar (2009).

Second, we investigate for the first time the moderating effects of the customer’s product portfolio on revenue effects and cost effects of customers’ online use. Although some studies note product-specific effects of online use (see Table 1; Hitt & Frei, 2002; Pauwels, Leeflang, Teerling, & Huizingh, 2011; Thomas & Sullivan, 2005; Xue, Hitt, & Chen, 2011), research has yet to examine the effects of online use when customers combine products with different characteristics. Such knowledge is necessary if firms are to manage their customers’ use of channels effectively.

Third, we demonstrate the importance of accounting for self-selection effects. Only a few previous studies have considered self-selection effects at all (Table 1). We control for self-selection effects through hybrid matching. By introducing the method of hybrid matching into the marketing literature, we also demonstrate the viability of this approach.

The remainder of this article proceeds as follows. First, we discuss the expected effects of online use on customer revenue and on cost to serve. Next, we explain the importance of controlling for self-selection effects and describe different methods for doing so. We then describe our data, detail our methodology and present the empirical results. We end with a discussion of our findings, managerial implications, research limitations, and suggestions for further research.

2. Conceptual development

By employing an online channel, managers aim to increase customer profitability by improving customer revenue and decreasing cost to serve. Online channels can increase customer revenue and decrease cost to serve if the use of online channels leads customers to alter their behavior (Degeratu, Rangaswamy, & Wu, 2000). Customer behavior is in this study reflected by customers’ demand for one or more products and the number of transactions that customers undertake in each channel. Product demand influences customer revenue, which is defined as the customer’s product demand multiplied by the contribution margin earned by the firm on a particular product. Meanwhile, the cost to serve a customer is the product of the number of transactions a customer undertakes in a channel multiplied by the channel-specific costs.

In the following section, we consider the incremental effect of online use on customer behavior, in line with previous studies, because most online customers use traditional offline channels as well (e.g., Chu, Chintagunta, & Cebollada, 2008). Online customers are thus defined as customers who use an online channel for at least some transactions, even if they do not use that channel exclusively.

2.1. Effect of customers’ online use on customer revenue

Online use can increase the customer’s product demand for several reasons. When searching for product information, customers who use online channels generally perceive that they have greater information control than do customers using offline channels. These perceived measures of control include the selection of which information is presented, for how long, and which information follows from it (Ariely, 2000). Greater information control online thus improves the customer’s ability to understand information relevant to the customer’s choices, which is a critical determinant of decision making (Weathers, Sharma, & Wood, 2007). Furthermore, online channels often provide interactive decision tools that help customers to picture themselves using the product, a feature that should increase their purchase likelihoods (Huang, Lurie, & Mitra, 2009; Schlosser, 2003). Online channels also offer greater convenience and accessibility than do offline channels (Brynjolfsson, Hu, & Smith, 2003; Montoya-Weiss, Voss, & Grewal, 2003); customers do not have to consider opening hours or wait in checkout lines. Because online use should increase the customer’s product demand, it should also enhance customer revenue (Hitt & Frei, 2002). We expect, in turn, online customer revenue, which combines online and offline revenue, to be higher than revenue earned from customers who use offline channels exclusively.

H1. Online use increases customer revenue.

The verification of this hypothesis is particularly important because previous research into the association between online use and customer revenue frequently has ignored self-selection effects (Table 1).

2.2. Effect of customers’ online use on cost to serve

Beyond increasing customer revenue, firms invest in online channels because the costs per transaction are much lower than with offline channels, decreasing the cost to serve customers (Campbell & Frei, 2010). Online channels also reduce the cost from a customer’s perspective (e.g., no travel, no waits) and should improve overall customer efficiency by lowering the marginal cost of transactions (Bitner, Brown, & Meuter, 2000), assuming customers substitute offline transactions with online transactions (Xue et al., 2011).

The reduction in marginal costs from the customer’s perspective may, however, increase the total number of transactions. That is, online customers may engage in more transactions because they expend fewer resources on any single transaction (Xue et al., 2011). Whether the firm’s total cost to serve customers increases or decreases, therefore, depends on how customers allocate their transactions across channels. More transactions can reduce costs only if customers substitute some of their costly offline transactions with less costly online

### Table 1

Studies that associate online use with customer profitability, customer revenue, and cost to serve.

<table>
<thead>
<tr>
<th>Study</th>
<th>Variables of interest</th>
<th>Moderating effect of product portfolio</th>
<th>Control for self-selection effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ansari et al. (2008)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Campbell and Frei (2010)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hitt and Frei (2002)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Kumar and Venkatesan (2005)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pauwels et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Thomas and Sullivan (2005)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Venkatesan, Kumar, and Ravishanker (2007)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Xue et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Xue et al. (2007)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>This study</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

OLS = ordinary least squares.

* Investigates the effect of multichannel use, which usually indicates that customers use both offline and online channels.
transactions; without such a substitution effect, the cost to serve increases. Because online use reduces the marginal cost of a transaction to the customer and increases the customer’s efficiency, we expect customers to substitute online transactions for offline transactions. The substitution effect, combined with the lower cost of online transactions from the firm’s perspective, may reduce the cost to serve an online customer. We propose the following:

H2. Online use lowers the cost to serve a customer.

Conventional wisdom suggests that online use decreases the cost to serve, yet empirical evidence is scarce (Table 1). Campbell and Frei (2010) provide the only investigation of online cost to serve, and they find a positive effect: online use increases cost to serve. Our test of H2 thus adds another data point to the discussion of how online use affects the cost to serve a customer.

2.3. Moderating effects of the customer’s product portfolio

The size of the effects of online use may depend on a customer’s product portfolio (i.e., the products of a firm that the customer uses) because some products tend to be more closely related to a particular channel than others (Inman, Shankar, & Ferraro, 2004; Liang & Huang, 1998). In this context, frequency of use and perceived product risk represent important product characteristics for matching products to channels (Gutiérrez, Izquierdo, & Cabezudo, 2010; Inman et al., 2004). Specifically, frequently used products are likely to benefit more from online use (Danaher, Wilson, & Davis, 2003; Xue, Hitt, & Harker, 2007) because customers search for more information about frequently used products than they do for less frequently used products (Bhatnagar & Ghose, 2004). The increase in search behavior online may result in greater product demand. Furthermore, frequently used products may benefit more from the accessibility and convenience of online channels—that is, online channels should make it easier to use frequently used products, which also should increase the customer’s product demand.

Yet many customers associate greater risk with online channels rather than offline channels (Forsythe & Shi, 2003). Research further shows that risk in the channel, in turn, positively influences product-related risk (Gutiérrez et al., 2010). For example, customers may be more afraid of making a mistake when using an online channel (Forsythe & Shi, 2003), often because personalized advice (e.g., from a sales representative) — which can reduce consumers’ perceptions of product-related risk (Alba et al., 1997) — is not available online. Additional information on a website (e.g., third-party evaluations, consumer feedback) may help reduce product-related risk, but such information tends to be regarded as less valuable than personalized advice (Weathers et al., 2007). High-risk products thus may appear more closely related to offline than online channels, and high levels of risk could attenuate the positive effect of online use on the customer’s product demand (Smith & Sivakumar, 2004). However, we do not expect a negative effect of online use on the customer’s product demand for high-risk products because online customers continue to use offline channels and have access to personalized advice. With all these arguments, we propose that the effect of online use on the customer’s product demand, and ultimately on customer revenue, depends on the composition of a customer’s product portfolio:

H3. The composition of the customer’s product portfolio moderates the positive effects of online use on customer revenue.

2.4. Effect of customers’ online use on customer profitability

Customer profitability, in a service setting such as retail banking, is a function of customer revenue, cost to serve, and risk costs — the latter being a cost measure determined by the bank that reflects the customer’s risk of default in payment (Skiera, Bermes, & Horn, 2011). To compute a customer’s risk cost, the bank takes into account customer characteristics such as age and history with the bank.

The influence of online use on customer profitability should depend on its revenue (H1) and cost-to-serve (H2) effects. We predict a positive effect of online channel use on customer profitability because it should improve customer revenue and reduce cost to serve. However, we refrain from formulating a separate hypothesis for this effect on customer profitability because if it exists, it stems from the effects detailed in H1 and H2. Instead, we summarize our conceptual framework and proposed relationships in Fig. 1.

3. Main effects and self-selection controls

3.1. Formal definition of the effects of online use

To determine the effects of online use on customer revenue and cost to serve, we examine the specific effects of online use on product demand and number of transactions that customers undertake, i.e., factors that can translate into changes in customer revenue and cost to serve. The basic problem in identifying these effects is that we can observe both product demand and transactions in either the online transactions for offline transactions to improve their own efficiency (Campbell & Frei, 2010), with the added effect of decreasing firms’ cost to serve.

Again though, this improved efficiency might lead customers to undertake more transactions, especially if they are making decisions about high-risk products. For customers who use high-risk products, an online channel provides continuous access to detailed product information, which may result in increased information monitoring and more active product management (Campbell & Frei, 2010). For example, online brokerage accounts encourage more transactions because customers perceive an opportunity to take personal control of their investments and improve their returns by actively managing their investments (Strader & RamaSwami, 2004). An increasing number of transactions limit the firm’s opportunity to reduce cost to serve. We therefore propose that the composition of a customer’s product portfolio moderates the effect of online channel use on cost to serve.

H4. The composition of the customer’s product portfolio moderates the negative effects of online use on cost to serve.

Note: This figure only shows the relationships of focal interest. In our empirical application, we control for additional effects not depicted in this figure.

Fig. 1. Conceptual framework.
online or the offline scenario. For example, we can observe the number of online transactions by an online customer only if he or she uses the online channel.

Specifically, we use the average treatment on the treated effect (ATT) to evaluate the effects of online use on product demand and number of transactions (Heckman & Navarro-Lozano, 2004; Zhao, 2004). The ATT provides a basis for assessing whether online use influences product demand and the number of transactions completed among customers who actually use the online channel:

$$ \text{ATT}_k = E(y^1_k | d_k = 1) - E(y^0_k | d_k = 1) \quad \forall k \in K, \quad (1) $$

where $d_k$ represents a binary variable that equals 1 if customer $i$ is online and 0 otherwise. Meanwhile, $E(y^1_k | d_k = 1)$ is the value of the outcome variable $k$ (e.g., number of transactions completed) for online customer $i$, and $E(\ldots)$ is the expected value of $k$ for all online customers. The counterfactual outcome in Eq. (1) that requires estimation is $E(y^0_k | d_k = 1)$ or the expected value of outcome variable $k$ for all online customers if they were offline.

To determine this counterfactual outcome in Eq. (1), we could use the observed mean values for all offline customers, $E(y^0_k | d_k = 0)$. Yet online customers also may differ in their characteristics from offline customers, and the same characteristics could influence the customer’s product demand as well as the number of transactions. Simple mean comparisons, then, cannot represent the effects of online use because self-selection effects (SE) are confounded with the effects of online use. More formally,

$$ \left[ E(y^1_k | d_k = 1) - E(y^0_k | d_k = 0) \right] = \text{ATT}_k + \text{SE}_k \quad \forall k \in K. \quad (2) $$

In Eq. (2), the left-hand side represents the observed difference in means between online and offline customers. SE is the self-selection effect for outcome variable $k$.

### 3.2. Methods to control for self-selection effects

The most popular approaches to control for self-selection effects are matching, instrumental variables (IV) and control function methods (Heckman & Navarro-Lozano, 2004). The latter two methods control for selection on unobserved characteristics and rely on instrumental variables, which must satisfy two conditions: They must be exogenous, and they must be highly correlated with the binary treatment variable (i.e., customers’ online use) (Stock, Wright, & Yogo, 2002). If we do not observe variables that meet these conditions, the estimated effects can be highly biased; the biases induced by such weak instrumental variables can be so great that even a simple ordinary least squares (OLS) regression that used observed customer characteristics to control for selection could perform better (Woglom, 2001).

Matching methods do not rely on instrumental variables but instead eliminate self-selection effects by comparing online and offline customers with similar observed characteristics. Such methods thus rely on the selection of observed customer characteristics to build matched samples; the selection should be based on theory and previous empirical findings. If researchers fail to simultaneously include all observed customer characteristics that affect the outcome (e.g., checking account balances) and treatment decision (i.e., to use the online channel), the estimated effects will be biased. Matching methods thus have high data demands. Moreover, the observed customer characteristics must be sufficient to ensure that the outcome variables are independent of the treatment and conditional on customer characteristics – that is, the observed customer characteristics must address the conditional independence assumption (Rosenbaum & Rubin, 1983). Furthermore, the methods fail if the observed customer characteristics are perfect predictors of the treatment decision because no matching partners can then be identified. Matching methods thus assume, given a set of observed characteristics, that some unspecified randomization is capable of allocating customers to online use (Heckman & Navarro-Lozano, 2004).

There are a variety of matching methods available to build matched samples. **Covariate matching** builds samples on the basis of observed characteristics (Zhao, 2004). Its limitation is that if there are many characteristics that could drive self-selection effects, it becomes infeasible to match customers directly. The Mahalanobis distance can map multiple characteristics into a single measure that expresses the gap between any two customers (Rosenbaum, Ross, & Silber, 2007). However, this approach requires customer characteristics measured on metric scales. Such difficulties have limited the use of covariate matching in the past.

An alternative approach is **propensity score matching** (Dehejia, 2005; Mithas, Krishnan, & Fornell, 2005). The propensity score is in our case the conditional probability that a customer with a specific vector of observed characteristics uses the online channel. This probability can be estimated with a logit or probit model. With the propensity score, we ensure that the distribution of characteristics in the two groups (i.e., online versus offline) is the same (Rosenbaum & Rubin, 1983). Yet propensity score matching cannot guarantee that the matched online and offline customers are directly comparable in all of their characteristics (e.g., age).

**Hybrid matching** compares online and offline customers according to both the propensity score and certain selected customer characteristics (Rosenbaum & Rubin, 1985), ensuring that matched customers are directly comparable with respect to these characteristics.

In our empirical study, we use customer transaction data and customer characteristics that are typically stored in customer databases (e.g., age, length of relationship). These customer characteristics are not exogenous and thus are not strong instrumental variables. We therefore focus on the use of matching methods and show that IV and control function methods are less appropriate in such a setting (Section 5.4). Specifically, we use hybrid matching to determine the counterfactual outcomes in Eq. (1). We expect that in our setting, hybrid matching may improve the accuracy of the estimated effects because hybrid matching ensures that online and offline customers correspond on selected characteristics in addition to their probability of using the online channel. We subsequently compare the predictive performance of propensity score and hybrid matching and find that the latter yields slightly higher predictive performance (Table 6).

### 4. Empirical study

#### 4.1. Data

We used transaction data from a random sample of approximately 87,000 private clients of a large European retail bank over a three-month period. Of these clients, 38.4% were enrolled in online banking, although only 1.9% actively used it. For our purposes, a customer actively uses the online channel if he or she conducts at least two transactions with the firm through the online channel during our three-month observation period.

In Table 2, we provide an overview of the variables. For each customer, we collected information about the monthly number of transactions completed with the retail bank. We also determined the number of transactions each customer completed in each channel. Furthermore, we had information about each customer’s product
To assess the predictive performance of the hybrid matching method, we collected information about online customers during a three-month period that started 6 months before our focal observation period. We thus were able to obtain information about online customers at two different points in time: the observation period \((t=0)\), and an earlier period \((t=-6 \text{ months})\). At \(t=-6\), there were 108 current online customers who were still offline; their use of the online channel started later. These 108 customers represent the holdout sample; we assume that no effects other than online use cause any differences in each person's product demand or number of transactions completed between \(t=0\) and \(t=-6\). This assumption is defendable, considering the short time frame. We do not include the customers in the holdout sample in our estimates of the effects of online use.

### 4.2. Implementation of the hybrid matching method

We estimate each customer’s propensity to use the online channel as a function of the following: the customer’s age \((\text{AGE})\); the length of the customer’s relationship with the firm \((\text{LOR})\); the customer’s ownership of checking accounts \((\text{CHECK})\), savings accounts \((\text{SAV})\), brokerage accounts \((\text{BROK})\) and credit cards \((\text{CREDIT})\); and whether the checking account is a joint account \((\text{JOINT})\). The ownership variables indicate whether a customer uses a particular product. We use these variables because previous studies show that online customers tend to be younger than offline customers and have shorter relationships with the firm (Campbell & Frei, 2010; Degeratu et al., 2000). Moreover, when a customer uses a specific product, she or he also may be more or less likely to use an online channel. Owners of checking accounts, for example, may find it more convenient to move online than owners of savings accounts, because the online channel offers greater convenience for managing checking accounts than for managing savings accounts, which require less upkeep. Owners of joint accounts may realize greater cost savings from online banking, making them more prone to use it than other account holders would be (Campbell & Frei, 2010; Hitt & Frei, 2002).

In Table 3, we list the estimated effects of customer characteristics on propensity to use the online channel. As expected, customers who are younger, have shorter relationships with the firm, and have a joint checking account are more likely to use online banking. Customers who own a checking account or credit card are also more likely to use the online channel, whereas owners of savings and brokerage accounts are less likely. However, the latter two effects are not significant.

Next, we use the propensity score to build matched samples of online and offline customers who are comparable in their characteristics. The percentage reduction in bias is an important metric with which to evaluate whether the two groups are more comparable after matching (Table 3). To compute the percentage reduction in bias, we compare the difference in means for characteristics measured on a metric scale (or relative frequencies for characteristics measured on a nominal scale), after matching versus before matching (Rosenbaum & Rubin, 1985):

\[
\text{reduction\_bias}_m = \left(1 - \frac{X_{d=1,m}^{after} - X_{d=0,m}^{after}}{X_{d=1,m}^{before} - X_{d=0,m}^{before}}\right) \cdot 100 \quad \forall m \in \mathbb{M}
\]

where \(X_{d=1,m}^{after}\) and \(X_{d=0,m}^{after}\) are the mean values (relative frequencies) for the customer characteristic \(m\) for online and offline customers.

---

5. Information about credit cards is available only if customers also own a checking account. Moreover, the credit card balance is affected by the checking account balance because balances on the credit card are paid from the checking account. To address this interdependence in further analyses, we consider the relative credit card balance by dividing the credit card balance by the checking account balance.

6. We thank an anonymous reviewer for providing this argument.
Table 3
Parameters of customer characteristics in propensity score model (logit model, dependent variable: online use).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Parameter</th>
<th>t-Value</th>
<th>Reduction in bias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>−0.04</td>
<td>−8.51</td>
<td>55.36</td>
</tr>
<tr>
<td>Length of relationship (month)</td>
<td>−0.02</td>
<td>−19.94</td>
<td>43.94</td>
</tr>
<tr>
<td>Joint account</td>
<td>0.32</td>
<td>3.31</td>
<td>−82.14</td>
</tr>
<tr>
<td>Checking account ownership</td>
<td>5.35</td>
<td>5.34</td>
<td>74.20</td>
</tr>
<tr>
<td>Credit card ownership</td>
<td>4.48</td>
<td>5.74</td>
<td>34.59</td>
</tr>
<tr>
<td>Brokerage account ownership</td>
<td>0.00</td>
<td>0.02</td>
<td>6.94</td>
</tr>
<tr>
<td>Savings account ownership</td>
<td>−0.03</td>
<td>−0.59</td>
<td>35.14</td>
</tr>
<tr>
<td>Intercept</td>
<td>−7.22</td>
<td>−7.16</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>86,754</td>
<td>log-likelihood = −7,023.04, pseudo R² = 0.131</td>
<td></td>
</tr>
</tbody>
</table>

Bold values are significant at the 5% level.

* Computation based on Eq. (3).

After matching, and $x_{k_m}^{before}$ and $x_{k_m}^{before}$ are the mean values (relative frequencies) for the same customer characteristic for online and offline customers before matching. The percentage reduction in bias thus indicates whether the comparability of the two groups improves after matching (Rosenbaum & Rubin, 1984).

Overall, the reduction in bias is substantial for most characteristics (Table 3), such that online and offline customers become more similar with respect to their observed characteristics after matching. For example, the percentage reduction in bias is 55.36% for the customer characteristic ‘age’. Only for the joint account characteristic does the metric decrease, such that the two groups become less comparable. However, we retain this characteristic in estimation of the propensity score because the joint account characteristic has a significant effect and because the difference in the relative frequency before and after matching is rather small (0.010 versus 0.017).

To ensure that the matched online and offline customers are actually comparable, we employ a common support restriction. This restriction eliminates all customers who do not lie within the region of common support (Heckman, Ichimura, & Todd, 1997), a construct defined as the overlap between the propensity score distributions of online and offline customers. This approach excludes customers with a propensity score smaller (larger) than the minimum (maximum) value of the propensity score of the region of common support. In our study, 77.6% of the offline customers and 100% of the online customers were within the region of common support, providing us a sufficient number of matching partners for each online customer (Zhao, 2004).

Studies in a variety of settings indicate that online and offline customers differ substantially in age and length of relationship with the firm (e.g., Hitt & Frei, 2002; Shankar, Smith, & Kangaswamy, 2003). Therefore, we included these characteristics as well as the propensity score when estimating the effects of online use. In other words, matched customers are explicitly comparable with respect to age and length of relationship with the firm. To identify matching partners, we use the Mahalanobis distance (M) (Rosenbaum et al., 2007) and compute the effects of interest with the following specified version of Eq. (1):

$$\text{ATT}_{k} = E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right) - E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 0\right) = E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right) - E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 0\right) = E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right) - E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 0\right) \forall k \in K$$

where $E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right)$ equals the mean value of outcome variable $k$ across all online customers, and $E\left(\left(y_{ik}^{0}\right)\mid d_{i} = 0\right)$ equals the mean value of outcome variable $k$ across all matched offline customers (counterfactual outcome).

To estimate the effects of online use on the customer’s product demand and number of transactions completed, we implement a Gaussian kernel algorithm that matches each offline customer with a specific online customer and assigns a weight to each match that reflects the distance between members of the pair (for details, see the Appendix). This algorithm is effective for samples with large numbers of offline customers who possess characteristics similar to those of the online customer (Smith & Todd, 2005), as we find to be the case in our data.7

4.3. Assessment of the hybrid matching method

To test the robustness of the estimated effects, we use propensity score stratification and determine if the effects of online use vary across customers. We designate five strata using the estimated propensity scores; the strata differ with respect to the estimated likelihood of using the online channel (Dehejia & Wahba, 2002; Mithas & Krishnan, 2009). After estimating the effects of online use on the customer’s product demand and number of transactions completed for each stratum, we apply an analysis of variance (ANOVA) to evaluate if the effects of online use differ across strata. We use each online customer’s observed value on an outcome variable $k$ and its corresponding counterfactual outcome; if we find significant differences across strata, heterogeneity exists across customers, and the average effects of online use measured across all customers would be biased.

We also assess the predictive performance of the hybrid matching method using information from the holdout sample. That is, we compare actual and predicted values for the customers in the holdout sample in terms of number and balance of checking accounts, number of transactions completed, and cost to serve. We focus on product demand for checking accounts because too few customers in our holdout sample use other products. To assess the predictive performance, we calculate the absolute percentage error (APE):

$$\text{APE}_k = \frac{\left| E_i\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right) - \left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right) + \text{ATT}_{k}\right|}{E_i\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right)} \forall k \in K,$$

where $E_i\left(\left(y_{ik}^{0}\right)\mid d_{i} = 1\right)$ represents the predicted value of variable $k$ (e.g., balance of checking accounts) for the holdout sample, equal to the sum of the observed mean value for $k$ in the holdout sample at $t = 6$ (i.e., when customers were still offline) and the estimated effect of online use on $k$. The actual value of $k$ is the observed mean value in the holdout sample at $t = 0$.

With this information, we compare the hybrid matching results against the results of propensity score matching, OLS regression, and the IV and control function methods to evaluate predictive performance. For the IV and control function methods, we use customer age and length of relationship as instruments. Although customer age and length of relationship with the bank are important drivers of online use, we would expect both characteristics to be weak instrumental variables because they are correlated with the customer’s product demand and number of transactions completed. As a result, their predictive performance is expected to be poor.

5. Results

Customers’ online use first affects the customer’s product demand and number of transactions completed, and then it affects customer revenue and cost to serve. We will therefore detail the effects of online use on product demand and number of transactions completed before

---

7 We also implemented the one-nearest and four-nearest neighbor algorithm and compared their predictive performance. The Gaussian kernel algorithm offers the highest predictive performance. These results are available from the first author on request. We discuss how we assess predictive performance in the subsequent section.
extending our discussion to the effects on customer revenue and cost to serve and to the moderating effect of product portfolios.

5.1. Effect of customers’ online use on product demand and number of transactions completed

We employ Eq. (4) to assess the treatment effects and provide the results in Table 4. Online use has a significant positive effect on the number of checking accounts used by a customer (ATT = 0.01, p = 0.02).8 However, we find no significant effects of online use on the number of brokerage accounts (ATT = 0.00, p = 0.24) or on the number of credit cards (ATT = 0.00, p = 0.68) that the customer holds. The effects of online use on the customer’s balances are not significant at the 5% level. We also find that online use affects the number of transactions completed positively (ATT = 2.50, p = 0.00). On average, online customers conduct 2.50 more transactions per month than do their offline counterparts.

If we ignore self-selection effects and simply compare product demand and number of transactions completed for online and offline customers, we find that online customers use significantly more checking accounts (Δ = 0.18, p = 0.00) and credit cards (Δ = 0.02, p = 0.00) but have fewer brokerage accounts (Δ = −0.02, p = 0.00) (Table 4).9 Moreover, online customers have significantly less money in their checking accounts (Δ = −201.38 EUR, p = 0.02) and savings accounts (Δ = −883.33 EUR, p = 0.00) than offline customers. Finally, online customers conduct more transactions with the bank (Δ = 3.68, p = 0.00). The differences between the observed mean difference and ATT thus indicate substantial self-selection effects. For example, the 0.18 difference in number of checking accounts held by online versus offline customers appears to be caused mainly by self-selection. With Eq. (2), we find the self-selection effect to be 0.17 (p = 0.00), while the effect of online use is only 0.01. The observed difference in means is thus driven by self-selection more than by online use.

5.2. Effect of customers’ online use on customer revenue and cost to serve

To assess the effect of online use on customer revenue, we multiply the individual observed values of the online customer’s product demand and estimated counterfactual outcome by the corresponding net contribution margins for each product. We sum these outputs across products to ascertain actual and estimated customer revenues. Finally, we compute the mean difference between actual and estimated customer revenues to derive the effect of online use (Eq. (4)).

The information on cost to serve is based on the number of transactions conducted in a particular channel multiplied by the channel-specific cost. We use the mean difference in actual and estimated costs to serve to derive the effect of online use.

In support of H1, online use improves customer revenue significantly (ATT = 0.18, p = 0.03). Online customers generate, on average, 0.18 EUR in additional revenue per month (Table 4). If we ignore self-selection effects, the difference in means is −0.99 EUR per month, implying there exists a negative self-selection effect of −1.17 EUR per month (Eq. (2)). The size and sign of the self-selection effect underscore the importance of controlling for such effects.

Cost to serve decreases by 0.12 EUR per month (p = 0.00), as we predicted in H2. This result suggests that online customers substitute offline transactions with online transactions; the number of transactions completed increases by 2.50 transactions per month on average, but cost to serve decreases. When comparing only the costs to serve of online and offline customers, we find a difference in means of

8 To evaluate whether the differences in means for the matched samples are significant, we use a paired sample t-test (Rosenbaum & Rubin, 1985).
9 The results are based on simple mean comparison tests using t-tests for independent samples (Rosenbaum & Rubin, 1985).

5.3. Moderating effect of the customer’s product portfolio

Theory suggests that the impact of customers’ online use on revenue and cost to serve may vary depending on customers’ product portfolio. We employ an ANOVA to test whether product portfolios moderate the effect of online use on customer revenue; in support of H3, they do (F-value = 7.60, p = 0.00).10

In Table 5, we show the effects of online use on customer revenue across product portfolios, based on the portfolios actually used by online customers. The overall effect of online use on customer revenue reported in Table 4 is equal to the weighted average of the second column of Table 5 with the number of online customers using a product portfolio as weights. The overall effect of online use on cost to serve reported in Table 4 is equal to the weighted average of the fourth column of Table 5 with the number of online customers using a product portfolio as weights. The revenue effect is largest for a product portfolio that contains checking and brokerage accounts (2.84 EUR per month) — a surprising result, considering that brokerage accounts represent high-risk products (Wang, Keller, & Siegrist, 2011). The effect is significantly larger than for product portfolios consisting of the following: checking accounts alone (0.18 EUR per month, p = 0.00); checking and savings accounts (−0.30 EUR per month, p = 0.00); or the combination of checking accounts, savings accounts, and credit cards (−0.12 EUR per month, p = 0.00). The second largest effect appears in the portfolio with checking accounts and credit cards (1.50 EUR per month). This effect is larger compared to the revenue effect on checking accounts only (p = 0.00).

10 The ANOVA is based on online customers and considers portfolios only of products they actually used. We employ the estimated individual effect of online use on customer revenue (observed value — estimated counterfactual outcome) as the dependent variable and product portfolio as the independent variable. A similar approach indicates whether the product portfolio moderates the effect of online use on cost to serve.
As we expected, the IV and control function estimates are heavily about online customers for the three-month period at \( t = \) the IV and control function methods (Table 6), we use information method with that of propensity score matching, OLS regression, and effects of online use.

ences between the estimated effects of online use on the customer's differ across customers. Across strata, we account for such heterogeneity. We thus employ propensity score effects in our previous analyses would be biased if we failed to actions completed could vary across customers. In this case, the ef- customers with savings accounts in their product portfolios react less favorably to on- accounts. Online use actually increases cost to serve by 0.17 EUR per month). Although online use increases the number of transactions that the bank realizes signi- cant cost savings.

Cost to serve is not reduced as much for customers with savings accounts. Online use actually increases cost to serve by 0.17 EUR per month when customers own checking and savings accounts and by 0.59 EUR per month when they additionally have credit cards. Cust- imers with savings accounts substitute online transactions for off- line transactions to a far lesser degree. Overall, customers with savings accounts in their product portfolio react less favorably to on- line use, and the substitution effect is significantly smaller for these customers than for customers without savings accounts in their prod- uct portfolios. Customers with savings accounts tend to make exten- sive use offline channels and seem hesitant about migrating to the online channel.

5.4. Further evaluation of the estimated effects

The effects of online use on product demand and number of trans- actions completed could vary across customers. In this case, the ef- fects in our previous analyses would be biased if we failed to account for such heterogeneity. We thus employ propensity score stratification and an ANOVA to test whether the effects of online use differ across customers. Across strata, we find no significant differ- ences between the estimated effects of online use on the customer's product demand or number of transactions completed. Overall, there is no substantial customer heterogeneity with respect to the ef- fects of online use.

To compare the predictive performance of the hybrid matching method with that of propensity score matching, OLS regression, and the IV and control function methods (Table 6), we use information about online customers for the three-month period at \( t = -6 \) months and turn again to our holdout sample of 108 customers. As we expected, the IV and control function estimates are heavily biased because we lack strong, valid instrumental variables. The OLS regression (mean APE [MAPE] = 16.14%) performs much better than the IV (MAPE = 313.55% and 197.01%) or control function (MAPE = 158.25% and 101.36%) methods with weak instrumental variables. The control function method seems less affected by weak instrumental variables than the IV method, but its predictive perform- ance is still very low.

Propensity score matching exhibits a MAPE of 14.97%. It performs very well with respect to the number and balance of checking ac- counts and presents the lowest APE for these two outcome variables. However, hybrid matching attains slightly better results with respect to predictive performance (MAPE = 13.45%).

It is comforting to find that hybrid matching performs best in the holdout exercise. However, we acknowledge that we have only demon- strated adequate predictive performance. Further research is need- ed to examine the validity of the estimated effects.

6. Discussion and implications

This study deepens our understanding of the effects of customers’ online use on customer revenue, cost to serve, and profitability as well as the moderating effect of the customer's product portfolio. Al- though we investigate a retail bank, the need to develop a strategy for online channels is not limited to this sector, and our study's insights have value in other settings as well.

We argue in particular that it is important to determine the reve- nue and cost effects of customers' online use when developing an online channel strategy. Previous studies have focused on either revenue or cost effects; we address both. Our results show that online use af- fects customer revenue positively, on average, as other authors have found (Kumar & Venkatesan, 2005; Thomas & Sullivan, 2005). Online use also decreases cost to serve, in contrast with the findings of Campbell and Frei (2010). The revenue effect of online use is approxi- mately 50% greater than the effect on cost to serve (0.18 versus 0.12 EUR per month); in other words, increased revenue accounts for approxi- mately 60% of the profitability effect. Managers may hope to in- crease customer profitability by reducing cost to serve, but our results indicate that moving customers online can also lift profitability by

\[ \text{Hybrid matching} = 4.92, p = 0.00 \]

11 We also estimated the IV and control function models using both instrumental vari- ables (age and length of relationship) at the same time, but the performance of the models declined.

12 A detailed comparison of the results of hybrid matching versus propensity score matching indicated very similar estimated effects of online use. Only for the balance in the checking account do we find a significant positive effect of online use when propen- sity score matching is used and no significant effect with hybrid matching.
increasing customer revenue. Therefore, the online channel seems well suited for encouraging customer relationships, and managers need to develop effective online channel strategies to exploit these revenue benefits.

The effects of online use on customer revenue and cost to serve also vary across customer product portfolios, a finding that could provide a good basis for customer segmentation. Managers need differentiated strategies to motivate customers to use online channels, yet most evidence from the popular press suggests that firms use generic strategies that ignore differences in customer product portfolios. For example, Catherine Palmieri, managing director at Citibank, noted, “We’ve given away cash, we’ve had $25,000 sweepstakes and other contests... We do a lot of things in the branch where we get the sales force excited about getting the customers online” (Gurlaci, 2005).

Our results show, however, that customers with savings accounts are less attractive targets for moving online because their revenues are likely to decrease, and their costs to serve are likely to increase. Savings accounts are used relatively less frequently than other products in our study, and the frequency of use may be an important product characteristic for segmentation efforts. The convenience and accessibility of the online channel proves especially beneficial for frequently used products; managers should work to motivate customers to move to the online channel when those customers own frequently used products. On a firm level, firms that offer high-usage products may benefit more from actively managing customer channel use than do firms selling products that are used less frequently.

We also provide evidence of substantial self-selection effects. Ignoring these effects can lead to poor managerial decision making. For example, the observed mean difference for customer revenue is −0.99 EUR per month, but the estimated revenue effect of online use is 0.18 EUR per month, indicating a negative self-selection effect of −1.17 EUR per month. In other words, online customers appear to produce less revenue than offline customers. Yet online use improves customer revenue, so supporting the online channel can prove beneficial for the bank. Because self-selection effects account for a critical portion of observed differences in means between online and offline customers, managers must take these effects into account when assessing the impact of online use on customer metrics such as customer revenue and cost to serve.

Finally, we demonstrate an effective application of hybrid matching and propensity score matching that provided much higher predictive performance than the IV and control function methods. The latter two tactics seem inappropriate in the face of weak instrumental variables. Instead, matching methods are viable when controlling for self-selection effects, especially if only weak instrumental variables are available. These insights should transfer meaningfully to other situations in which managers attempt to quantify the effectiveness of online use on customer behavior.

7. Limitations and further research directions

This study suffers from several limitations that suggest avenues for further research. First, researchers should replicate our study with firms in other industries to investigate the generalizability of our findings, particularly our finding that the revenue effect of online use is substantially larger than the cost effect. Researchers also may wish to examine whether customer product portfolios provide actionable bases of segmentation when developing customer migration strategies in other settings, and which product characteristics are most important in determining whether moving customer to the online channel will increase profitability. Of further interest is whether product characteristics, such as frequency of use, moderate the effects of online use on customer revenue and cost to serve.

In addition, our empirical data set provides only limited information about individual customers. Additional characteristics that are likely to be correlated with online use (e.g., attitudes toward Internet use) could be influential, although the good predictive performance of our results provides some support for their adequacy. Further studies could examine, in detail, which method of controlling for self-selection effects is most appropriate in certain settings.

A further limitation of our data is that they are cross-sectional; consequently, we cannot address the long-term effects of online use by examining the effects on customer retention. Some previous studies provide initial evidence that online use improves customer retention (e.g., Hitt & Frei, 2002; Shankar et al., 2003). Time-series data could reveal whether the effects of online use on the customer’s product demand and number of transactions completed change over time and whether the extent of the customer’s online channel use moderates the effects of online use. Investigating the dynamic effects of online use on customer behavior would be particularly interesting. Furthermore, it would be worthwhile to investigate competition and whether use of the online channel increases the firm’s share of the customer’s budget. If customers were routed to an online channel when buying or using a product, causality would be reversed. Using panel data and formal causality testing could address this issue. Reverse causality is not a problem in our study—the bank managers we consulted assured us that customers were not routed to the online channel—but it may be of concern in other applications.

Online marketing activities are also part of the incremental effect of online use on customer behavior. Additional research should disentangle the effects of online use and online marketing activities to provide more detailed insights in this realm.

Finally, moving customers across channels requires appropriate marketing instruments. An additional step for researchers would be to identify ways of encouraging customers to use online channels. Reinders, Dabhoklar, and Frambach (2008), Ansari et al. (2008), and Thomas and Sullivan (2005) provide some insights, but more research is needed to reveal how firms can manage customers’ channel use actively and effectively.

Acknowledgment

We thank the E-Finance Lab at the House of Finance at Goethe-University and its partners for supporting this research.

Appendix A. Hybrid matching approach

We briefly outline the implementation of the hybrid matching method for estimating ATT using the Mahalanobis distance and a Gaussian kernel algorithm:

1. Randomly order online customers and offline customers.
2. For online customer i, determine similarity to all offline customers in the sample, according to the Mahalanobis distance (M), by considering the propensity score, customer’s age (AGE), and length of relationship (LOR).
3. Estimate the counterfactual outcome for every outcome variable for online customer i using a weighted average of the outcome variable of all offline customers. The weights assigned to each offline customer j depend on the similarity between online and offline customers, as determined by the Gaussian kernel algorithm. The bandwidth parameter τ is a function of the standard deviation σ of the similarity measure and sample size N, such that 

$$
\tau = 1.06 \cdot \sigma \cdot N^{-0.2} \quad \text{(Silverman, 1986)}.
$$

The weight of an offline customer j for an online customer i equals:

The standard deviation σ is the sample standard deviation after taking the common support restriction into account. Thus, it is estimated across all customers.
\[
W_{ij} = \frac{K\left(\frac{H_{ij}}{\tau}\right)}{\sum_{i,j=1}^{n} K\left(\frac{H_{ij}}{\tau}\right)}, \quad \forall i \neq j, j = 1, 2, \ldots, n,
\]

where \(K(\cdot)\) is a normally distributed kernel function.

4. Remove online customer \(i\) from the list of online customers.

5. Does \(i\) equal \(L_i\)? If no, set \(i = i + 1\) and go to Step 2. If yes, go to Step 6.

6. Compute the ATT for every outcome variable, using the average outcome for the online customers and the average estimated counterfactual outcome.

References


The joint effects of choice assortment and regulatory focus on choice behavior

Anirban Som a,⁎, Yih Hwai Lee b,1

a Department of Marketing, Faculty of Business, Building 2, Bond University, Gold Coast, Queensland, 4229, Australia
b Department of Marketing, NUS Business School, Mochtar Riady Building, BIZ 1, 8-27, National University of Singapore, 15 Kent Ridge Drive, 119245, Singapore

A R T I C L E   I N F O

Article history:
First received in 10, June 2010 and was under review for 18 months
Available online 23 March 2012

Area editor: Harald Van Heerde

Keywords:
Regulatory focus
Promotion
Prevention
Alignable assortment
Non-alignable assortment
Confidence

A B S T R A C T

Past research presents contrasting views regarding the effect of assortment size on consumer decision making. Research has suggested that large assortments provide a diverse range of choices to consumers and thus increase their choice confidence levels and likelihood of making choices. Other research, however, has suggested that large assortments may have a negative impact on consumers’ choice confidence and choice likelihood because of cognitive load and anticipatory post-purchase regret. The current research aims to address these contrasting conclusions by examining the issue from the motivational goal perspective. Specifi

1 Tel.: +65 6516 3168, +65 6516 3058.

© 2012 Elsevier B.V. All rights reserved.

1 Tel.: +61 410628682 (mobile), +61 7 55955588 (office).
E-mail addresses: somaniran81@yahoo.co.in, asom@bond.edu.au (A. Som), bizleeyh@nus.edu.sg (Y.H. Lee).

This research is based on a Master’s degree dissertation at the NUS Business School, National University of Singapore.

⁎ Corresponding author. Tel.: +61 410628682 (mobile), +61 7 55955588 (office).
E-mail addresses: somaniran81@yahoo.co.in, asom@bond.edu.au (A. Som), bizleeyh@nus.edu.sg (Y.H. Lee).

1 Tel.: +65 6516 3168, +65 6516 3058.

How does an increase in assortment size, whether alignable or non-alignable, affect consumer choice behavior? Extant literature suggests that there may be both positive and negative influences. Larger assortments generally provide greater diversity of choices and therefore have greater ability to satisfy consumers (Anderson, 2006). Accordingly, retailers who offer diverse choices are able to generate greater sales volumes compared to competitors who offer a narrower range of choices (Bown, Read, & Summer, 2003). Further, Berger, Dragnaska, and Simonson (2007) demonstrate that providing a high level of variety within a product category can lead to increased market share of a brand. There are other benefits of having many options to choose from. For instance, an assortment consisting of a range of options with unique characteristics enables consumers to engage in more direct comparisons, thus heightening their choice confidence (Hutchinson, 2005). Choosing from a wide range of choices also satisfies the desire for novelty and reduces the level of choice uncertainty (Ariely & Levav, 2000).

Gourville and Soman (2005) suggest that consumers experience higher cognitive load and higher anticipatory post-purchase regret if the assortment from which they need to make a choice is non-alignable (vs. alignable). Specifically, for a given assortment size, more mental steps are needed to solve a choice conflict in a non-alignable assortment [e.g., “Do I value a sunroof or a leather interior or an alarm system?” (Gourville & Soman, 2005, p. 389)] than in an alignable assortment (e.g., “Do I need an engine with a higher or lower mileage?” (Gourville & Soman, 2005, p. 389)). As assortments become bigger, the difference in choice effort between a non-
alignable and an alignable assortment increases. In addition, unlike choosing from an alignable assortment, choosing from a non-
alignable assortment requires a consumer to forgo one attribute and
select another. Thus, for a given assortment size, a non-alignable as-
sortment, compared to an alignable assortment, produces signifi-
cantly greater anticipatory post-purchase regret for consumers. As the non-alignable assortment becomes bigger, this regret is magnified.
Thus, it has been suggested that an increase in the size of a non-
alignable assortment negatively affects consumer purchase incidence, whereas an increase in the size of an alignable assortment positively
affects consumer purchase incidence.

Given the divergent views in the literature on the effect of assort-
ment size on consumer choice decisions, it is important to identify
and understand the conditions under which the beneficial and ad-
verse consequences from an increase in assortment size are likely to
occur. The current research contributes to this effort by studying the
influence of consumers’ motivational orientations or goals on con-
sumer choice behavior with increases in assortment size. Specifically, we propose that a consumer’s regulatory focus (promotion or preven-
tion) interacts with assortment alignability (alignable or non-alignable) to
determine whether an assortment size increase would have a posi-
tive or negative impact on choice behavior. By doing so, our paper
also extends Gourville and Soman’s (2005) paper. While their article
introduced assortment alignability as a moderator that influences the
relationship between assortment size and consumer choice processes,
the current research moves forward and examines consumers’ regula-
tory focus as a moderator that affects the relationship between assort-
ment alignability, assortment size and consumers’ choice processes.

2. Literature review

Regulatory focus theory (Higgins, 1997) suggests that when mak-
ing a purchase decision, consumers may display one of two basic mo-
tivational orientations: a promotion-focused or a prevention-focused
orientation. Individuals with a promotion-focused inclination are
likely to focus on achievement and on maximizing their gains. In con-
trast, individuals with a prevention-focused inclination are likely to
focus on safety and on minimizing losses. Furthermore, a focus on
promotion makes the presence or absence of positive outcomes sa-
lent, while a focus on prevention makes the presence or absence of
negative outcomes salient (Crowe & Higgins, 1997).

Researchers have estimated that approximately half of all con-
sumers in a market are relatively more promotion-focused, while the
other half are relatively more prevention-focused (Zhao & Pechmann, 2007). Aside from being a chronic disposition, one’s ten-
dency to be promotion-focused or prevention-focused may also be
temporarily heightened by external stimuli. For example, Zhou and
Pham (2004) showed that exposure to information about investment
products such as common stocks (government bonds) can moment-
arily induce a promotion (prevention)-focus. It should be noted
that dominant promotion-focus and dominant prevention-focus pro-
duce similar effects on the actions of individuals, regardless of wheth-
er they are chronically salient or have been made temporarily salient
by administering a promotion or a prevention prime on individuals.
For the purpose of our research, we take into account the fact that a
dominant promotion-focus or a dominant prevention-focus is chron-
ically salient in individuals.

The choice behavior we examine involves consumers making choices
from assortments that are either alignable or non-alignable with the
choices in the assortments consisting of positive as well as negative attri-
butes. Including both positive and negative attributes may be more re-
ective of assortments in the marketplace that consist of common
positive feature(s) and unique negative feature(s), as well as common
negative feature(s) and unique positive feature(s) (Dhar & Nowlis, 1999).

We propose two key differences between promotion-focused and
prevention-focused individuals in their choice behavior when faced
with assortments of different alignability. First, we suggest that promotion-focused individuals place greater consideration on the
presence or absence of positive attributes when making a choice. In
contrast, prevention-focused individuals place greater consideration
on the presence or absence of negative attributes when making a
choice. This difference is supported by Chernev (2004), who found
that promotion-focused individuals are likely to give more weight
to hedonic and performance-related attributes, or attributes that are
more directly related to enhancing the positive outcomes of a choice.
On the contrary, the same research found that prevention-focused in-
dividuals are likely to give more weight to utilitarian and reliability-
related attributes, or attributes that are related to minimizing the
negative outcomes of a product choice.

The second difference between promotion-focused and prevention-
focused individuals is that the former adopt a selection approach
(choosing the best available alternative from a choice set), while the lat-
ter adopt a rejection approach (choosing by rejecting the less desirable alternatives in a choice set). The basis of this assertion comes from pre-
vious studies (e.g., Chernev, 2009) that have suggested that promotion-
oriented individuals are likely to receive higher utility from approach
means, whereas prevention-oriented individuals derive greater utility
from avoidance means. As such, it may be construed that selecting the
best possible match with the goal of promotion is more compatible
with the approach means and thus would be considered an appropriate
choice strategy for promotion-focused consumers. On the contrary,
rejecting mismatches with the goal of prevention is more compatible
with avoidance means and thus would be considered an appropriate
choice strategy for prevention-focused consumers.

We employ the two differences stated above to make predictions
concerning the interplay between regulatory focus, assortment
alignability, and assortment size on choice decision as well as on
choice confidence. In the following section, we explicate our hy-
potheses under alignable followed by non-alignable assortment
contexts.

3. Alignable assortment, variation in assortment size and
consumers’ regulatory focus

Suppose an energy drink brand has five brand variants: E1, E2, E3,
E4 and E5. The alignable features in an assortment of size two are as
follows:

3.1. Promotion-focused consumers

While making a choice from the assortment depicted in Table 1a,
promotion-focused consumers are likely to place more importance on
the positive attribute thiamin compared to the negative attribute
sulfonamide. The brand variant that contains the maximum propor-
tion of thiamin is E2. Thus, the promotion-focused consumers are
likely to prefer E2 over E1.

Let the size of the alignable assortment be increased to three as
reflected in Table 1b.

In this case, promotion-focused consumers are likely to prefer the
brand variant E3 over the brand variants E1 and E2. For promotion-
focused consumers, the greater the perceived success of promotion
goal fulfillment, the higher the level of confidence they should have
in their choice decision. Thus, compared to making a choice that is su-
perior in terms of promotion goal fulfillment relative to one

<table>
<thead>
<tr>
<th>Table 1a</th>
<th>Alignable assortment of size two.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand variant</td>
<td>Proportion of thiamin that increase the strength of muscles</td>
</tr>
<tr>
<td>E1</td>
<td>5%</td>
</tr>
<tr>
<td>E2</td>
<td>7.5%</td>
</tr>
</tbody>
</table>
alternative (assortment of size two), making a choice that is superior in terms of promotion goal fulfillment relative to two alternatives (assortment of size three) should increase the promotion-focused consumers’ confidence in the relative superiority of the selected alternative in terms of promotion goal fulfillment. Thus, with an increase in the size of an alignable assortment, the promotion-focused consumers’ choice confidence as well as choice incidence (i.e., the likelihood of making a choice) should also increase.

3.2. Prevention-focused consumers

While making a choice from the alignable assortment of size two, prevention-focused consumers are likely to focus on minimizing the presence of the negative attribute sulfonamide in their selected item to fulfill their prevention goal. The brand variant that contains the minimum proportion of sulfonamide is E1. Therefore, they are likely to reject E2 and select E1.

Following similar logic, when prevention-focused consumers select an item from the alignable assortment of size three, they are likely to reject the brand variants E2 and E3 and select the brand variant E1. Prevention is centered on maximizing correct rejections while pursuing an activity (Chernev, 2009). Thus, the greater their perceived successes at executing this act of rejection, the greater their prevention goal fulfillment. Therefore, for someone with a prevention orientation, correctly rejecting two potential mismatches would potentially be construed as bringing them closer to fulfilling their prevention goal compared to when only one prevention goal mismatch is rejected. As such, an increase in the size of the alignable assortment should lead to an increase in the prevention-focused consumers’ choice confidence as well as choice incidence.

The discussion thus far leads us to propose the following:

In a within-brand choice context, when the assortment type is alignable and when the choices in the assortment consist of positive as well as negative attributes,

H1(a). the perceived confidence of consumers about the correctness of their choice increases as the size of the alignable assortment increases (regardless of their level of regulatory focus).

H2(a). the preference of consumers for making a choice from the alignable assortment increases as the size of the assortment increases (regardless of their level of regulatory focus).

4. Non-alignable assortment, variation in assortment size and consumers’ regulatory focus

Consider a non-alignable assortment of size two of an energy drink as depicted in Table 1c.

### Table 1c
Non-alignable assortment of size two.

<table>
<thead>
<tr>
<th>Component</th>
<th>Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotin</td>
<td>Present</td>
</tr>
<tr>
<td>Guarana</td>
<td>Present</td>
</tr>
<tr>
<td>Niacin</td>
<td>Present</td>
</tr>
<tr>
<td>Sulfonamide</td>
<td>Present</td>
</tr>
</tbody>
</table>

4.1. Promotion-focused consumers

When selecting from a non-alignable assortment of size two, promotion-focused consumers can gain one positive attribute that is present in the alternative they select while forgoing another positive attribute that is present in the alternative that they do not select. If the non-alignable assortment extends to a set size of three as shown in Table 1d, the promotion-focused consumers gain one positive attribute while failing to gain two other positive attributes. As suggested earlier, the criterion for judging promotion goal fulfillment is the presence or absence of positive attributes in a choice option. In this regard, as the size of the non-alignable assortment increases, the number of positive attributes that the promotion-focused consumers fail to gain increases while the number of positive attributes that they gain does not increase. As a result, as assortment size increases, promotion-focused consumers will increasingly perceive themselves as being unable to fulfill their promotion goal, leading to a reduction in their choice confidence as well as choice incidence.

4.2. Prevention-focused consumers

In contrast, when making a choice from the same non-alignable assortment of size two, prevention-focused consumers are able to avoid the presence of one negative attribute while having to accept another in their chosen item. When the selection is made from the non-alignable assortment of size three, prevention-focused consumers are able to avoid the presence of two negative attributes while having to accept only one negative attribute in their chosen item.

For prevention-focused consumers, as the size of the assortment increases, the number of negative attributes that they are able to avoid in their choice increases while the number of negative attributes that they need to accept in their choice remains constant. Prevention goal fulfillment is attained by avoiding the presence of negative outcomes. Hence, as the size of the assortment increases, there should be a greater fulfillment of the prevention goal. This should then lead to an increase in choice confidence and choice incidence. The following hypotheses are thus proposed:

In a within-brand choice context, when the assortment type is non-alignable, and when the choices in the assortment consists of positive as well as negative attributes,

H1(b). the perceived confidence of promotion-focused (prevention-focused) consumers about the correctness of their choice from the non-alignable assortment decreases (increases) as the size of the assortment increases.

H2(b). the preference for making a choice from the non-alignable assortment for promotion-focused (prevention-focused) consumers decreases (increases) as the size of the assortment increases.

5. Method

We employed a 2 (Consumer’s self-regulatory focus: Promotion vs. Prevention) × 2 (Assortment type: Alignable vs. Non-alignable) × 4 (Size of assortment: 2 vs. 3 vs. 4 vs. 5) between-subjects design. The participants were 275 undergraduates from an Asian business school.
5.1. Stimuli development

Alignable and non-alignable assortments of different sizes containing brand variants of an energy drink brand were created (refer to Appendix A). A pretest was conducted with 45 undergraduate students to test whether their brand attitudes toward the variants that were used to construct the non-alignable assortments were comparable (7-point items: bad-good; unfavorable-favorable; dislike-like). Cronbach's alpha computed separately for the items measuring the attitudes toward each of the five brand variants ranged from 0.84 to 0.91. The responses for the three brand attitude measures were then averaged, and an ANOVA suggested that the participants possessed similar attitudes toward the brand variants (t=1, p>.05 in all pairwise comparisons).

5.2. Procedure

Participants first completed an 18-item regulatory focus scale, which has been validated by Lockwood, Jordan, and Kunda (2002). Next, each participant was given information about a hypothetical brand assortment of an energy drink. The size and alignability of the assortment presented to each participant were based on the experimental condition to which the participant was randomly assigned. After having read the information, each participant was asked to select a brand variant of their choice from the assortment presented to them.

They then responded to the confidence and choice-likelihood dependent measures. The participants' prevailing mood states were also assessed at this time. We administered cognitive load and anticipatory post-purchase regret measures to assess their potential influence on our dependent variables. Measures were also administered to assess whether, in accordance with our stated assumption, promotion-focused consumers adopted a “selection” approach whereas prevention-focused consumers adopted a “rejection” approach in their decisional task.

5.3. Dependent measures and covariate measures

Participants indicated how confident (1=Not at all confident to 9=Very confident) they were about making the correct choice of an energy drink from the assortments provided to them. They could also indicate instead why they preferred not to make any choice from the assortments that were presented to them. The following question was posed to them in this regard: “Had there been a ‘I would prefer not to choose’ option, how likely is it that you would have gone for it?” (1=Extremely unlikely to go for the no choice option, and 9=Extremely likely to go for the no choice option). We then reverse recoded the responses and labeled this item the “choice likelihood” measure, i.e., a higher value indicates a greater likelihood of making a choice from the assortment. The mood measure consisted of 4 items (happy, worried, calm, sad) on a seven-point scale (1=not at all happy/worried/calm/sad, and 7=Extremely happy/worried/calm/sad) as adapted from Crowe and Higgins (1997). The cognitive load assessment consisted of two questions: “How difficult or how easy is it to make the choice decision?” (1=Easy to understand, and 7=Difficult to understand) and “How much attention do you think is needed to make the choice decision?” (1=Requires little attention, and 7=Requires a lot of attention) as adapted from Keller and McGill (1994). Anticipatory post-purchase regret was gathered when participants responded to the question, “If I do not choose the other available options, I will regret it later” (1=Very unlikely, 7=Very likely). This regret measure was adapted from Tsiros (2008).

5.4. Process measures

After the participants rated the dependent measures and the covariate measures, they were further asked to rate their level of agreement with the following statements on a scale of 1 (not at all agree) to 7 (very likely).

Statement A: While making my choice, I searched for the option that I thought was the most favorable.
Statement B: While making my choice, I looked to reject the options that I thought were unfavorable.

6. Analysis of findings

There were 252 completed responses. To classify the participants as being promotion- or prevention-focused, we adopted a procedure that is similar to what has been used in past research (Zhao & Pechmann, 2007). Specifically, each participant's responses on the promotion and prevention measures (α=0.83 for promotion-focus and 0.77 for prevention-focus) were averaged. A measure of dominant regulatory focus was then created by subtracting the prevention score from the promotion score. Subsequently, participants were classified as being either promotion- or prevention-focused on the basis of a median split (Mdn=0.78). With this procedure, 123 participants were deemed to be prevention-focused, and 129 participants were classified as being promotion-focused.

6.1. Tests of Hypotheses 1(a) and 1(b)

We first conducted a full-factorial ANCOVA with self-regulatory focus, assortment type, and assortment size as the explanatory factors impacting the “choice confidence” response. The mood, cognitive load, and regret measures were treated as covariates. As predicted, the three-way interaction was significant after controlling for the covariates (F (3,229)=4.71, p=.003). The covariate main effects were not significant (p>.05 in all cases). Furthermore, no significant interaction between the covariates and the fixed factors was observed (p>.1 in all cases). Hence, we excluded the covariates from subsequent analyses.

Follow-up analyses were performed via a method adapted from Gourville and Soman (2005). Specifically, assortment sizes 2 and 3 were grouped to form a “small” assortment size condition, while assortment sizes 4 and 5 were grouped to form a “large” assortment size condition. Consistent with H1(a), when the assortment is alignable [Fig. 1A], confidence to make a correct choice was significantly higher when the assortment size was “large” as opposed to when it was “small” for both promotion- and prevention-focused participants (Promotion: Maligne=7.65 vs. Msmall=5.79, t(236)=3.61, p=.0001; Prevention: Maligne=6.8 vs. Msmall=5.65, t(236)=1.96, p=.05).

Under the non-alignable assortment type (Fig. 1B), the promotion-focused condition showed a significantly lower “choice confidence” when the assortment size was “large” (Maligne=3.93 vs. Msmall=5.53, t(236)=−3.61, p=.0004). In contrast, for the prevention-focused condition, i.e., a higher value indicates a greater likelihood of making a choice from the assortment.
condition, “choice confidence” was significantly higher when the assortment size was “large” ($M_{\text{Large}} = 6.78$ vs. $M_{\text{Small}} = 5.85$, $t(236) = 2.38$, $p = .01$). Hence, H1(b) was also supported.

6.2. Tests of Hypotheses 2(a) and 2(b)

A similar full-factorial ANCOVA was performed on the “choice likelihood” dependent measure. The three-way interaction (illustrated in Fig. 2A and B) was significant after controlling for the mood measures, perceived task difficulty measures and the regret measure ($F(3,229) = 2.84$, $p = .03$). The main and interaction effects involving the covariates were non-significant ($p > .05$ in all cases). Once again, we excluded the covariates for subsequent analyses.

For the alignable assortment (Fig. 2A), the “choice likelihood” was significantly higher for promotion-focused participants when the assortment size was “large” as opposed to when it was “small” ($M_{\text{Large}} = 5.07$ vs. $M_{\text{Small}} = 3.71$, $t(236) = 1.94$, $p = .05$). For prevention-focused participants, the “choice likelihood” was only higher when the assortment size was “large” as opposed to when it was “small” with the difference between the means not achieving statistical significance ($M_{\text{Large}} = 5.95$ vs. $M_{\text{Small}} = 5.1$, $t(236) = 1.06$, $p = .29$). Hence, H2(a) was partially supported.

For the non-alignable assortment (Fig. 2B), “choice likelihood” was only directionally lower for promotion-focused participants when the assortment size was “large” as opposed to when it was “small”, although statistical significance was not achieved ($M_{\text{Large}} = 6.92$ vs. $M_{\text{Small}} = 7.54$, $t(236) = -1.04$, $p = .28$). However, for the prevention-focused participants, in keeping with our predictions in H2(b), “choice likelihood” was significantly higher when the assortment size was “large” compared to when it was “small” ($M_{\text{Large}} = 6.7$ vs. $M_{\text{Small}} = 4.91$, $t(236) = 3.24$, $p = .001$). Thus, H2(b) was partially supported.

6.3. Process measures to test the choice strategies of promotion- and prevention-focused consumers

We expected promotion-focused consumers to display a greater level of agreement with Statement A compared to Statement B (see Section 5.2 and to Section 5.4 for details on the process measures), and we expected the opposite for prevention-focused consumers.

We compared the “selection” scores and the “rejection” scores separately for the promotion-focused participants and the prevention-focused participants using t-tests. The results were in accordance with our expectations. Promotion-focused participants showed greater levels of agreement with Statement A than Statement B ($M = 7.36$ vs. $M = 6.17$, $t(128) = 3.94$, $p = .001$). On the contrary, prevention-focused consumers showed agreement with Statement B compared to Statement A ($M = 7.06$ vs. $M = 6.05$, $t(122) = 2.98$, $p = .003$).

6.4. “Spotlight analysis” to test the strengths of our interaction

A spotlight analysis was conducted to supplement the median split approach of dichotomizing the regulatory focus scores (Fitzsimons, 2008). We regressed the “choice confidence” measure on the dominant self-regulatory scores, assortment alignability, and size of the assortment. The three-way interaction was significant ($β = -2.31$, $p = .006$).

To clarify the nature of this interaction, we performed regressions for each assortment alignability condition. In the alignable assortment condition, as expected, there was no interaction between the size of the assortment and the dominant self-regulatory scores ($β = .134$, $p = .64$). Participants with a high prevention-focus (one standard deviation below the mean dominant regulatory scores) exhibited significantly greater “choice confidence” when the assortment size was “large” compared to when it was “small” ($β = .54$, $p = .06$). Participants with a high promotion-focus (one standard deviation above the mean dominant regulatory scores) exhibited significantly greater “choice confidence” when the assortment size was “large” compared to when it was “small” ($β = 0.46$, $p = .01$).

In the non-alignable assortment condition, as expected, there was a significant interaction between size of the assortment and the dominant self-regulatory scores ($β = -0.94$, $p = .0002$). At one standard deviation below the mean level of the dominant regulatory scores (a high prevention-focus), participants exhibited significantly greater “choice confidence” when the assortment size was “large” compared to when it was “small” ($β = 0.39$, $p = .05$). In contrast, at one standard deviation above the mean level of dominant regulatory scores (a high promotion-focus), participants exhibited significantly lower “choice confidence.”

![Fig. 1. Three-way interaction of consumers’ self regulatory focus, assortment type and assortment size with “choice confidence” DV.](image-url)
confidence” when the assortment size was “large” compared to when it was “small” ($\beta = -0.46, p = .003$).

The results of the spotlight analysis provide further support for our proposed hypotheses and lessen the possibility of methodological bias in our findings.

6.5. “Choice confidence” as a mediator of the relationship between the independent variables and the “no-choice preference”

Because decision confidence plays a major role in positively predicting buying intentions (Chernev, 2009), we construed that “choice confidence” may mediate the relationship between the regulatory focus × alignability × size interaction and “choice likelihood”. We performed the mediation analysis using the bootstrapping procedure of Preacher and Hayes (2004). The independent variables in our mediation test are regulatory focus (dummy coded with 0 for promotion and 1 for prevention), alignability (0 for alignable and 1 for non-alignable), and size (0 for sizes 2 and 3 and 1 for sizes 4 and 5). All possible main and interaction effects were included as explanatory factors. The mean indirect effect of the three-way interaction from the bootstrap analysis was found to be negative ($a \times b = -1.92$) and significant [a 95% confidence interval excluding zero (−2.71 to −1.13)], providing support for the mediating role of “choice confidence”. In the indirect path, a unit increase in the three-way interaction term decreased “choice confidence” by $a = 3.19$ units on a 1 to 9 scale ($p = .0001$).

Further, it can be interpreted that by holding the three-way interaction constant, a unit change in “choice confidence” increased “choice likelihood” by $b = .61$ units on a 1 to 9 scale ($p = .001$). The direct effect of the three-way interaction $c = -.42$ on “choice likelihood” was not significant ($p = .51$). Because the direct effect was not significant, it was an indirect-only mediation model (Zhao, Lynch, & Chen, 2010), i.e., the mediator identified was consistent with our proposition. The mediation model is illustrated in Fig. 3.

7. General discussion

The current article makes a significant contribution to the literature on assortment size and consumer behavior. A comparison can be drawn between our paper and that of Gourville and Soman (2005) in which the main intention was to test the interaction effects of assortment size, assortment alignability and consumer post-purchase regret/cognitive load in making decisions on consumer choice likelihood. This article demonstrates that something other than regret and choice difficulty, specifically consumers’ regulatory focus, can interact with assortment alignability and assortment size and have interesting implications for consumers’ choice confidence and choice likelihood.

The current paper contributes also to the divided literature on whether increases in assortment size are favorable for a brand. Studies (e.g., Ariely & Levav, 2000) have suggested that large assortments...
can be beneficial for a brand. However, Gourville and Soman (2005) have clearly suggested otherwise. The current study tries to reconcile the two views by taking into account their regulatory focus or their goals. Along this line, the current research demonstrates the boundary point at which an increase in assortment size is beneficial for a brand (alignable assortment type-promotion and prevention-focused consumers as well as non-alignable assortment type-prevention-focused consumers) and where it becomes detrimental (non-alignable assortment type-promotion-focus).

8. Managerial implications and future research

Past research suggests that promotion-focus (prevention-focus) is generally nurtured in individualistic (collectivistic) countries (Aaker & Lee, 2001; Zhao & Peckmann, 2007). Our findings imply that it might be useful to tailor assortment types and sizes to different cultures and countries. Companies operating in a predominantly individualistic consumer culture might benefit more from highlighting alignable positive differences among products than from highlighting non-alignable positive differences between them (for example, when selling a nutritional drink, it might be useful to highlight the extra calcium content and hence the extra bone strength offered by the different variants). Alternatively, they might think of limiting the number of non-alignable options available to promotion-focused consumers (e.g., they can limit the number of available options to n = 2 or n = 3). For instance, Dell enables consumers to narrow their choice (e.g., they can limit the number of available options to n = 2 or n = 3). For instance, Dell enables consumers to narrow their choice of variants. Alternatively, they might think of limiting the number of non-alignable options available to promotion-focused consumers (e.g., they can limit the number of available options to n = 2 or n = 3). For instance, Dell enables consumers to narrow their choice of products at the initial stages of their search by ticking checkboxes related to the attributes they deem essential in their desired product. The product list that is subsequently generated is a short one. This can be a smart way to limit the number of non-alignable options that are available for promotion-focused consumers.

Marketers could benefit from providing consumers with collectivistic orientations (who may be predominantly prevention-focused) with assortments of larger sizes (whether alignable or non-alignable). Furthermore, marketers could help to highlight potential mismatches with prevention goals. For instance, when shopping for clothes, encouraging consumers to try on more designs to highlight the number of successful rejections might be a good idea. Along the same vein, highlighting the product’s target market as Dell did in grouping PC offerings into “gaming”, “home office”, and “multimedia tasks” desktops can be viewed as an effort to inform prevention-focused consumers of the needs that the products might not be able to serve. This may help heighten their choice confidence through mismatch avoidances.

In addition to the managerial implications, our article has interesting implications for future research. Novemsky and Dhar (2005) demonstrated that in a sequential choice setting, goal fulfillment in an initial choice heightens the risk-taking behavior of individuals in subsequent choices. Heightened risk-taking behavior may be reflected by a preference for untested choices that may have the potential to provide the desired level of goal fulfillment. In our large, non-alignable choice context, prevention goal fulfillment may result in a possible increase in risk-taking among prevention-focused consumers (who in general are risk averse) (Pham & Avnet, 2004) as they move from their initial choice to subsequent choices.

Taking a more dynamic view of goal orientation, Levav, Kivetz, and Cho (2010) have proposed that having too many means of attaining a self-regulatory goal may produce counter-normative behaviors. The proposition may have interesting implications toward the effects we found. For instance, beyond a certain set size limit, non-alignable assortments may cause promotion-focused consumers to make counter-normative choices due to the presence of too many options that fit with their promotion goal. In contrast, it is possible that such malleability in choice might not be observed for prevention-focused consumers as the options in the choice set would be perceived as misfits that need to be rejected correctly to make a choice. Future research could potentially explore these possibilities.

Acknowledgments

The authors thank the editor, the area editor and the two reviewers for their helpful comments and suggestions. The authors also thank Professor Mark Spence for his helpful suggestions. The authors would also like to thank Amrit Kaur for her assistance with proofreading.

Appendix A Example of the assortments used in the Experiment

A.1 Alignable assortment of size 3

<table>
<thead>
<tr>
<th>Features of brand variants</th>
<th>Proportion of riboflavin that improves the hemoglobin content in blood</th>
<th>Proportion of glutamine that may cause stomach discomfort</th>
<th>Proportion of thiamin that may increase the strength of muscles</th>
<th>Proportion of sulfonamide that may cause sleep disturbances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand variant</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Variant 1</td>
<td>5%</td>
<td>5%</td>
<td>2.5%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Variant 2</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Variant 3</td>
<td>5%</td>
<td>5%</td>
<td>7.5%</td>
<td>7.5%</td>
</tr>
</tbody>
</table>

A.2 Non-alignable assortment of size 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand variant</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Variant 2</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Variant 3</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
</tbody>
</table>
References


Offensive versus defensive marketing: What is the optimal spending allocation?

Guisomar Martín-Herrán a,b, Shaun McQuitty c, Simon Pierre Sigüé c,*

a Depto. Economía Aplicada (Matemáticas), Universidad de Valladolid, Spain
b GERAD, Montréal, Canada
c Faculty of Business, Athabasca University, Canada

A R T I C L E   I N F O
Article history:
First received in November 9, 2010 and was under review for 8 months
Available online 22 March 2012
Area Editor: David Soberman

Keywords:
Customer acquisition
Customer retention
Defensive marketing
Offensive marketing
Lanchester model

A B S T R A C T
This article investigates the optimal spending allocation between offensive and defensive marketing in a dynamic, mature market when two firms are competing for market share. A modified Lanchester model is used to determine Nash stationary feedback strategies that allow the competitors to adjust their marketing expenditures as their market shares evolve over time. The interaction between offensive and defensive marketing activities is an important component of the model. Previous studies have not considered this variable. Our findings suggest that a cost differential between offensive and defensive marketing cannot fully explain resource allocation in a competitive market. Instead, optimal allocation largely depends on the firms’ relative positions in the market, their competitive advantages in offensive and defensive marketing, and the costs and effectiveness of these two classes of marketing activities. This article discusses the theoretical and managerial implications.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Marketing managers are routinely challenged to rationalize their marketing budgets, improve the effectiveness of their marketing activities, and strengthen their firms’ competitive positions. One of the fundamental challenges that marketers face is how to divide their budgets between offensive and defensive marketing activities. We address the problem by studying a pure market share rivalry in a mature market. We classify marketing activities aimed at attracting a rival firm’s customers as offensive marketing, and we classify marketing activities focused on retaining a firm’s current customers or fostering brand loyalty as defensive marketing. We do not consider growing markets where competitors can attract new customers from outside the duopoly.

Traditional marketing focuses on a firm’s competitors and its customers (Fornell, 1992; Rust & Zahorik, 1993). Each firm monitors its competitors’ activities and reacts to significant moves that may change the dynamics of competition. The impact of marketing activities thus largely depends on how a firm’s activities compare with those of its competitors (Bridges & Freytag, 2009; Reibstein & Wittink, 2005; Shugan, 2005). However, the growing literature on relationship marketing proposes an alternative paradigm that focuses on customer satisfaction and retention (Day & Montgomery, 1999; Fornell, 1992; Fornell & Wernerfelt, 1987, 1988; Fruchter & Sigüé, 2004; Morgan & Hunt, 1994), with the goal of building and maintaining lasting relationships with existing customers. This customer-centric view of marketing is based on the belief that individualized marketing programs foster customer loyalty and retention, reduce marketing costs, and improve firm profitability (Fruchter & Sigüé, 2009; Kumar, Shah, & Venkatesan, 2006; Palmatier, Scheer, Houston, Evans, & Gopalakrishna, 2007). As a consequence, firms develop and implement relationship-based marketing tools such as loyalty programs, special treatment programs, and direct marketing activities that target their own customers (Leenheer, Heerde, Bijmolt, & Smidts, 2007; Vargo & Lusch, 2004). These defensive marketing initiatives prevent customer churn and increase consumption by current customers.

Although a few voices advocate the supremacy of defensive over offensive marketing (e.g., Berry, 1995; Parvatiyar & Sheth, 2001), intuition, marketplace evidence, and empirical research suggest that offensive marketing still plays a significant role in the ways that firms conduct business (Fruchter & Sigüé, 2005). Indeed, a study of how competitors react to each other’s promotion and advertising attacks reveals that the dominant form of competitive response is passive in nature (Steenkamp, Nijs, Hassens, & Dekimpe, 2005). However, when a reaction does occur, it typically is retaliatory in the same instrument, i.e., sales promotion and advertising attacks are countered with sales promotions and advertising, respectively. At the same time, the merit of defensive marketing cannot be underestimated. A firm that enjoys a high level of customer loyalty is less vulnerable to competitors’ attacks. Defensive marketing creates psychological, economic, emotional, and affective switching costs that restrain customers from switching to competitors. Note that these costs can be described in other ways, e.g., Chiu, Hsieh, Li, and Lee (2005) discuss the financial, social, and structural bonds that create utilitarian and hedonic value for loyal customers.
The use of both offensive and defensive marketing instruments gives competitors various strategic options and allows them to compete in different ways. Consequently, in some cases the best response to a competitor’s offensive marketing may be a defensive activity. This defensive reaction runs contrary to recommendations from literature that focuses exclusively on offensive marketing and suggests that the optimal response to a competitor’s attack is a retaliatory attack (e.g., Erickson, 1985; Jarrar, Martín-Herrán, & Zaccour, 2004). Defensive marketing also can be used preemptively to discourage future competitors’ assaults (Roberts, 2005). For example, if a firm can secure a captive market through a well-designed customer retention program, then competitors may find it difficult or costly to change consumer behavior in such a market. However, in other cases, a firm may have no alternative but to attack its competitors. New entrants to a market typically have to attack the incumbents to appeal to the incumbents’ customers and gain market share (Hauser & Shugan, 1983).

Unfortunately, despite the popularity of both offensive and defensive marketing strategies in the literature, formal knowledge about the conditions of their implementation is sparse. A cost argument often is used to advocate for the use of defensive marketing due to the belief that it is far cheaper to retain current customers than to attract new ones (e.g., Fruchter & Sigle, 2009; Reichheld & Sasser, 1990). This idea pervades the marketing literature but suffers from serious shortcomings. First, the idea assumes that offensive and defensive marketing are independent activities with no interaction. This assumption is unrealistic, particularly given the empirical evidence for interactions between marketing activities (Naik, Raman, & Winer, 2005). Advertisers who experience interaction or synergy effects between two marketing activities should increase the total budget and allocate more funds to the less effective activity (Naik & Raman, 2003). Second, the focus on reducing marketing expenditures ignores overall marketing effectiveness and profitability. For example, it is unproductive to focus on defensive marketing activities solely because they are less expensive when the strategic goal of a new entrant in a market should be building and expanding its customer base (Hauser & Shugan, 1983). Lastly, the competitive advantages a firm has over its rivals affect the firm’s resource allocations across business activities (Porter, 1985), including marketing. For example, if a firm’s defensive (offensive) marketing programs are more effective than a rival’s or if the firm has more loyal customers than its rivals, then marketing resource allocation should rely on that firm’s competitive marketing strengths and weaknesses.

The purpose of this article is to explore the implementation of offensive and defensive marketing strategies and offer an alternative resource allocation perspective that goes beyond the cost differentials of these strategies. We integrate critical factors such as the effects of offensive and defensive marketing on market shares, the firms’ competitive advantage/disadvantage in performing marketing activities, and the firms’ relative positions in the market. We develop an extended Lanchester model in a duopoly to investigate optimal resource allocation between defensive and offensive marketing in a dynamic competitive setting. We assume that two firms in a mature market battle over time for market shares through defensive and offensive marketing activities to maximize their respective profits. Each firm’s marketing activities have positive effects on its own market share and negative effects on the other’s market share. Considering the interactions between marketing activities in dynamic competitive markets is encouraged (Naik et al., 2005), and our model specification acknowledges that offensive and defensive marketing activities interact to improve the appeal of a firm’s offering to its rival’s customers. We assume that the interaction effect of a firm’s offensive and defensive marketing is added to the main effect of its offensive marketing to determine the full effect of its raid on the rival’s market share, while its defensive marketing only limits the rival’s attack on market share. We use a differential game methodology (Jørgensen & Zaccour, 2004) to derive Nash stationary feedback strategies that allow the competitors to adjust both the offensive and defensive marketing investments as their market shares change over time.

The remainder of this article is organized as follows. Section 2 reviews related literature. Section 3 presents our extended Lanchester model. Section 4 derives a Nash feedback equilibrium. Section 5 presents numerical illustrations. Section 6 provides conclusions and discusses the implications of our findings.

2. Overview of the literature

An extensive amount of analytical work has been completed on dynamic competition (e.g., Chintagunta & Villasim, 1992; Erickson, 1985, 1992, 1997; Fruchter, 1999; Fruchter & Kalish, 1997; Jarrar et al., 2004). Most of these works focus exclusively on offensive marketing activities and use the classical Lanchester model or some of its modified versions (see Jørgensen and Zaccour (2004) for a review). In a Lanchester model, competitive firms typically battle for market share, and a firm’s marketing activities target its competitors’ customers, who are assumed to be vulnerable to the rival’s marketing. Under this scenario, a firm’s best strategic response to a competitor’s marketing assaults is to organize its own attacks on the rival’s customer base. Most recent works prefer closed-loop or feedback strategies—which allow competitors to adjust their marketing efforts as their market shares evolve over time—to open-loop strategies (e.g., Breton, Jarrar, & Zaccour, 2006; Jarrar et al., 2004). This research establishes that offensive marketing strategies increase and decrease with a firm’s own market share when the discount rate is, respectively, zero and positive (see Erickson, 1991; Jarrar et al., 2004). However, one of these works’ main drawbacks is that the competitors are left without any defensive capability to protect their current customers and must rely entirely on offense to strengthen or maintain their competitive position.

Although the importance of focusing on customers in competitive markets has been extensively underlined, there remains a paucity of analytical dynamic competitive work that focuses on defensive marketing or a combination of defensive and offensive marketing. Retaining customers is important because customer defections reduce firm profitability and because the cost of getting new customers is much higher than keeping existing customers (Keaveney & Parthasarathy, 2001; Reichheld & Sasser, 1990). Customer complaints and service recovery strategies are critical for firm success (Hart, Heskett, & Sasser, 1990; Patterson, Cowley, & Prasongsukarn, 2006). Complaints give firms an opportunity to make service recoveries, avoid lost sales and negative word-of-mouth, and identify problems with service delivery (de Ruyter & Brack, 1993). Service quality and customer relationship management initiatives reflect efforts to develop and leverage customer loyalty through retention programs that increase firm profitability (Anderson, Fornell, & Lehmann, 1994; Anderson & Sullivan, 1993; Rust, Zahorik, & Keiningham, 1995).

An early analytical competitive framework of defensive marketing is offered by Hauser and Shugan (1983), who propose a model of how an incumbent firm should defend its position against the marketing assault of a new entrant to an industry. They find that a defensive marketing strategy largely depends on the distribution of buyer preferences and the position of the new entrant relative to the position of the incumbent in a multidimensional attribute space. A few extensions to this model have been proposed (see Hauser and Shugan (2008) for a review). A defensive marketing strategy is defined in this research stream as the reaction of an incumbent brand to the launch of a new competitive brand. The new entrant attacks while the incumbent firm defends, and neither firm in this model is able to use a full arsenal of defensive and offensive marketing tools.

In what may be the first study of resource allocation between defensive and offensive marketing in a dynamic competitive setting, Fornell and Wernerfelt (1987) develop an analytical model based on Hirschman’s (1970) exit–voice conceptualization. They find that investing in complaint management (a form of defensive marketing) should yield positive income for business organizations due to the costs associated with losing dissatisfied customers and then attempting to replace
them through offensive marketing. Complaint management also offers long term benefits, such as lowering overall marketing expenditures, by retaining existing customers. In a subsequent study, Forrell and Wernerfelt (1988) examine the characteristics of optimal complaint management systems and the market structures under which they are most valuable. The study's findings support the view that firms should encourage dissatisfied customers to complain because complaint management systems increase both customers' expected utility and a firm's profitability. Thus, an investment in complaint management to avoid customer defection reduces the need to invest in offensive marketing to acquire new customers (Forrell & Wernerfelt, 1988).

Because we use an extended Lanchester model of the offensive and defensive marketing tradeoffs in a dynamic duopoly, our article is closely related to Erickson (1993), who finds that offensive (defensive) marketing monotonically decreases (increases) with a firm's market share. However, the critical conceptual innovation of our model, when compared to Erickson's model, arises from our acknowledgment of the interaction effect between offensive and defensive marketing for determining the full power of a firm's raid on the rival's customers. In their study of advertising and promotion strategies in an exclusively offensive Lanchester-type model, Naik et al. (2005) encourage the consideration of interaction terms among marketing activities when these activities are necessary to compete effectively in dynamic markets. The rationale for the interaction is that different marketing activities can create a synergy that increases a firm's overall marketing effectiveness. Although defensive marketing activities such as loyalty programs primarily target existing customers, new customers also may be attracted by such programs and may decide to switch from one firm to another. Offensive marketing may be required, however, to make new customers aware of a firm's loyalty programs.

An increasing amount of customer relationship management (CRM) literature investigates resource allocation between customer acquisition and retention (see Jain and Singh (2002) for a review). For example, approaches for determining the optimal promotional allocation between customer acquisition and retention activities to maximize customer equity are proposed by Berger and Nasr-Bechwati (2001) and Blattberg and Deighton (1996). Similarly, the optimal level of contacts with customers across various channels to maximize customer lifetime value (CLV) is explored by Venkatesan and Kumar (2004). Resource allocation between social relationship and transactional marketing when a firm maximizes CLV is studied by Fruchter and Sigüé (2009). However, these models disregard competition and focus on a single firm. The first attempt in the CRM literature to investigate how two firms compete to retain or acquire customers is offered by Musalem and Joshi (2009) but is limited to two periods. Their work reveals that the optimal resource allocation between customer acquisition and retention should consider customers' preferences toward each firm, their contribution margins for each firm, and their responsiveness to each firm's acquisition and retention effort. Depending on these aspects, the competitors should battle to acquire or retain some customers more than others.

3. Extended Lanchester model

Consider a duopoly in which two competing firms face two different market segments: a segment of a firm's current customers and a segment of the rival's customers. Examples include any two natural gas, cable, or electricity providers that serve a community, although any competitive market where one firm competes against an amalgam of rival firms could serve as an example. Households that require such services must purchase from one of the two rival service providers. To grow its market share, each firm must try to keep its own customers while attracting the rival's customers. We assume that the overall market is mature and not increasing in size, so there are no new customers to target. We also assume that the competitors use offensive marketing to attract the rival's customers and defensive marketing to retain the firm's own customers. Let \( a(t) \) and \( b(t) \), respectively, be Firm 1's, \( i \in \{1, 2\} \), offensive and defensive marketing efforts at time \( t \). Let \( M \) be the market share of Firm 1 and \( 1 - M \) be the market share of Firm 2. The dynamics of Firm 1's market share at time \( t \) is given by the following modified Lanchester model:

\[
M(t) = f_1(a_1(t), b_1(t))(1 - M(t)) + f_2(b_1(t))M(t) - f_3(a_2(t), b_2(t))M(t) - f_4(b_2(t))(1 - M(t)),
\]

where \( M_0 \) is the initial market share of Firm 1 and the functions \( f_1, f_2, f_3, f_4 \) satisfy:

\[
\frac{\partial f_i}{\partial a_1}(a_1, b_1) \geq 0, \quad \frac{\partial f_i}{\partial b_1}(a_1, b_1) \geq 0, \quad \frac{\partial f_i}{\partial a_2}(a_2, b_2) \geq 0, \quad \frac{\partial f_i}{\partial b_2}(a_2, b_2) \geq 0.
\]

The two firms' market shares are not static and change continuously over time. The instantaneous change of Firm 1's market share depends on the two firms' offensive and defensive marketing activities and has four components. The first component in the right hand side (RHS) of the first equation above indicates that Firm 1's market share changes positively with the impact of its offensive and defensive marketing on the rival's customers. The second component signifies that Firm 1's defensive marketing activities contribute to increasing its market share. On the other hand, from the third and fourth components, the rival's defensive and offensive marketing activities negatively impact Firm 1's existing customers, while the rival's defensive marketing limits the increase of Firm 1's market share.

As in previous Lanchester-type models, we consider the competitive nature of the market in the evolution of market shares. However, a distinctive feature of our model's specification is that defensive marketing activities also may contribute to attracting customers from the rival, while offensive marketing is assumed not to affect customer retention. To better illustrate this specification, consider the case of two competing cable providers in a community. Defensive marketing targeted at existing customers could include activities like complaint management, customer support services, and discounts for service packages, although such activities also could appeal to the rival's customers who might consider switching to the alternative service provider. Conversely, offensive marketing that encourages the rival's customers to defect could include activities such as time-limited offers for new subscribers. We assume that current customers are ineligible for the incentives offered to new customers (as typically is the case). In other words, customers can decide to switch from one firm to another based not only on the competitor’s special offers to attract them (offensive marketing) but also on the competitor's treatment of its current customers (defensive marketing).

The dynamics of Firm 1's market share can be presented as follows:

\[
M(t) = [f_1(a_1(t), b_1(t)) - f_4(b_2(t))](1 - M(t)) + [f_2(b_1(t)) - f_3(a_2(t), b_2(t))](M(t)),
\]

\[
M(0) = M_0.
\]

We follow Erickson (1993) and assume that all of a firm's customers are exposed to the competitor's offensive marketing and therefore need to be protected. Consequently, the effects of Firm 1's offensive and defensive marketing on the rival's existing customers is countered by the competitor's defensive marketing activities, while Firm 1's own defensive marketing diminishes the effect of the competitor's marketing programs on its current customers.

For simplicity, we choose the following linear and linear-quadratic specifications for functions \( f_1(a_1, b_1), f_2(b_1), f_3(a_2, b_2), \) and \( f_4(b_2) \):

\[
f_1(a_1, b_1) = \alpha_1 a_1 + \beta_1 a_1 b_1 + \gamma_1 b_1,
\]

\[
f_3(a_2, b_2) = \alpha_2 a_2 + \beta_2 a_2 b_2 + \gamma_2 b_2.
\]
where $\alpha_i$, $\beta_i$, and $\gamma_i$, $i \in \{1, 2\}$, are non negative parameters denoting the effectiveness of offensive marketing, the interaction between offensive and defensive marketing, and defensive marketing, respectively. Note that with this specification, a firm’s defensive marketing alone does not impact the rival’s customer base. The interaction effect of a firm’s offensive and defensive marketing is added to the exclusive or main effect of offensive marketing to determine the full strength of its attack on the competitor’s market share. The rationale is that, for a firm’s defensive marketing to reach its rival’s customers, the offensive marketing program must generate some degree of awareness of both the brand and the defensive marketing program. The benefits of a firm’s defensive marketing can then be evaluated by the rival’s customers, who may consequently switch firms.

Finally, we assume that offensive and defensive marketing costs are quadratic and specified as follows:

$$C_{fi}(a_i) = \frac{C_{1f_i}}{2} a_i^2, \quad C_{fi}(b_i) = \frac{C_{2f_i}}{2} b_i^2, \quad i \in \{1, 2\},$$

where $C_{1f_i}$ and $C_{2f_i}$ are positive cost parameters for Firm $i$’s offensive and defensive marketing. The two firms strive to select the best allocation of their marketing budgets across offensive and defensive marketing activities that allow them to compete effectively and to maximize their respective discounted profits over an infinite horizon. The objective functions of the two firms are formalized as follows:

$$\max_{a_1,b_1} \int_0^\infty e^{-rt} \left[ C_{1f_1}(a_1(t) - \frac{C_{1f_1}}{2} a_1^2(t)) - C_{2f_1}(b_1(t) - \frac{C_{2f_1}}{2} b_1^2(t)) \right] dt,$$

$$\max_{a_2,b_2} \int_0^\infty e^{-rt} \left[ C_{1f_2}(a_2(t) - \frac{C_{1f_2}}{2} a_2^2(t)) - C_{2f_2}(b_2(t) - \frac{C_{2f_2}}{2} b_2^2(t)) \right] dt,$$

both subject to:

$$\dot{M}(t) = \left[ f_1(a_1(t), b_1(t)) - f_4(b_1(t)) \right] (1 - M(t)),$$

$$\dot{M}(0) = M_0,$$

where $r \geq 0$ is the common discount rate and $g$ are market share gross profit rates. It is assumed that $r$ is equal across firms and that the $g$ do not change over time regardless of sales volume. In other words, our model disregards economies of scale that might improve the return as a firm increases its sales.

4. Analysis

Our model’s specification requires us to use the methodology of differential games to effectively handle continuous strategic interactions. In a differential game, two key concepts typically must be considered. We must first consider the solution concept used for deriving the equilibrium (Jørgensen & Zaccour, 2004). In this article, we use the Nash equilibrium to account for the fact that the two competitors do not cooperate and make their marketing decisions simultaneously. Each firm optimizes its objective function and takes its competitor’s decisions as given. The second key concept to consider in a differential game is the information available to the players when they make their decisions. We use the feedback information structure (Breton et al., 2006; Jarrar et al., 2004; Jørgensen & Zaccour, 2004), which assumes that the players base their decisions on the current market shares and time. However, because the planning horizon is infinite and the problem is autonomous, i.e., all model parameters are time-invariant, we can disregard the time argument and confine our interest to stationary feedback strategies. In other words, we assume that, at any time, the competitors observe the current state of their market shares and then make their respective offensive and defensive marketing decisions. The feedback strategies are subgame perfect and are known to appeal to managers because they prescribe decision rules based on the current state of the dynamics (the market shares).

The standard condition for determining a stationary feedback Nash equilibrium requires that we find bounded and continuously differentiable functions $V_i(M), i \in \{1, 2\}$, satisfying for all $M, 0 \leq M \leq 1$, the following Hamilton–Jacobi–Bellman (HJB) equations (the time argument is omitted hereafter):

$$rV_1(M) = \max_{a_1,b_1} \left\{ g_1(M) - \frac{C_{1f_1}}{2} a_1^2 - \frac{C_{2f_1}}{2} b_1^2 + V_1'(M)M \right\},$$

(1)

$$rV_2(M) = \max_{a_2,b_2} \left\{ g_2(1-M) - \frac{C_{1f_2}}{2} a_2^2 - \frac{C_{2f_2}}{2} b_2^2 + V_2'(M)M \right\},$$

(2)

where $V_1(M)$ and $V_2(M)$, respectively, denote the value functions of Firms 1 and 2, and $V_1'(M)$ and $V_2'(M)$ are their derivatives.\footnote{Note that the model does not fit neither in the linear-quadratic structure nor the linear state structure. For that reason, we cannot use the usual conjecture method on the functional form of the value function. In this case, the solutions are genuinely non-linear and we have to resort to different techniques.}

1. The first-order optimality conditions for the maximization of the RHS of the HJB equations are:

$$-c_{1f_1}a_1 + V_1'(M)(\alpha_1 + \beta_1 b_1) (1-M) = 0,$$

(3)

$$-c_{1f_2}b_1 + V_1'(M)(\beta_1 a_1(1-M) + \gamma_1 M) = 0,$$

(4)

$$-c_{2f_2}a_2 - V_2'(M)(\alpha_2 + \beta_2 b_2) M = 0,$$

(5)

$$-c_{2f_1}b_2 - V_2'(M)(\beta_2 a_2 M + \gamma_2 (1-M)) = 0.$$

(6)

2. The second-order conditions that assure a maximum are given by:

$$c_{1f_1} c_{1f_2} - (\beta_1(1-M)V_1''(M))^2 \geq 0 \text{ and } c_{2f_1} c_{2f_2} - (\beta_2 MV_2''(M))^2 \geq 0.$$

(7)

These conditions are assumed to be satisfied and are fulfilled in all numerical illustrations presented in Section 5.

From Eqs. (3)–(6), we obtain the derivatives of the value functions as functions of the market share of Firm 1, $M$, the state variable in our model.

$$V_1'(M) = \frac{c_{1f_1} a_1}{\alpha_1 + \beta_1 b_1 (1-M)} V_1(M) = \frac{c_{1f_2} b_1}{\beta_1 a_1 (1-M) + \gamma_1 M},$$

(8)

$$V_2'(M) = \frac{c_{2f_1} a_2}{\alpha_2 + \beta_2 b_2 M} V_2(M) = \frac{c_{2f_2} b_2}{\beta_2 a_2 M + \gamma_2 (1-M)}.$$

(9)

In Lancaster models, the value function of Firm 1 typically is positively related to its market share, while the value function of its rival, Firm 2, decreases as the market share of Firm 1 increases.

When we rearrange the expressions of the derivatives of the firms’ value functions, the following two equalities must hold:

$$c_{1f_1} a_1 (\beta_1 a_1 (1-M) + \gamma_1 M) = c_{1f_2} b_1 (1-M)(\alpha_1 + \beta_1 b_1),$$

(10)

$$c_{2f_2} b_2 (\beta_2 a_2 M + \gamma_2 (1-M)) = c_{2f_1} a_2 (\alpha_2 + \beta_2 b_2).$$

(11)

Using these equalities, we can summarize the relationship between each firm’s two marketing instruments in the following proposition.

**Proposition 1.** The optimal offensive and defensive marketing investments for each firm are related as follows for $M \in (0,1)$:

\[
a_1(M) = -\frac{\gamma_1}{2\beta_1} M \left( \frac{M}{1-M} \right)^{\frac{\gamma_1}{\alpha_1}} + \sqrt{\left( \frac{\gamma_1}{2\beta_1} \right)^{\gamma_1} + \frac{c_{1f_2}}{c_{1f_1} \beta_1} (\alpha_1 + \beta_1 b_1 M) b_1(M)}.
\]
\[ a_1(M) \delta \gamma_1 \delta M = - \frac{\gamma_2}{2\gamma_2} M \left[ 1 - \frac{M}{1 - \gamma_2} \right \{ \left( \frac{\gamma_1}{\gamma_2} M \right)^2 + \frac{c_{12} b_2 b_1(M) b_2(M)}{c_{21} b_1(M)} \} \right] \]

\[ a_2(M) = - \frac{\gamma_2}{2\gamma_2} M \left[ 1 - \frac{M}{1 - \gamma_2} \right \{ \left( \frac{\gamma_1}{\gamma_2} M \right)^2 + \frac{c_{22} b_2 b_1(M) b_2(M)}{c_{21} b_1(M)} \} \right] \]  

\[ (a) \]

**Proof.** Eq. (10) can be rewritten as a second-order polynomial equation for function \( a_1 \). One of the quadratic equation’s two solutions can be disregarded because that solution implies a negative value of \( a_1 \) for any positive value of \( b_1 \) and any value of the market share, \( M = 0,1 \). The other root of the quadratic equation, as shown in expression (12), always positive for any positive value of \( b_1 \) and any value of the market share, \( M = 0,1 \). The same procedure can be applied to Eq. (11) to obtain the expression of \( a_2 \) in (13).

Eqs. (12) and (13) specify the optimal offensive marketing strategies for the two competitors. They establish that when market shares are held constant, Firm 1 will increase its offensive marketing as the effectiveness of these activities and the cost of defensive marketing increase. On the other hand, increases in the effectiveness of defensive marketing and the cost of offensive marketing will lead to a decrease in offensive marketing activities. This result derives from the following equations:

\[ \frac{\partial a_1}{\partial M}(M) = - \frac{1}{2\gamma_2} M \left[ 1 - \frac{M}{1 - \gamma_2} \right \{ \left( \frac{\gamma_1}{\gamma_2} M \right)^2 + \frac{c_{12} b_2 b_1(M) b_2(M)}{c_{21} b_1(M)} \} \right] \]

\[ \frac{\partial a_2}{\partial M}(M) = - \frac{1}{2\gamma_2} M \left[ 1 - \frac{M}{1 - \gamma_2} \right \{ \left( \frac{\gamma_1}{\gamma_2} M \right)^2 + \frac{c_{22} b_2 b_1(M) b_2(M)}{c_{21} b_1(M)} \} \right] \]

where the market share \( M \) of Firm 1 is given and constant.

Interestingly, there exists a positive non-linear relationship between the firms’ two marketing instruments for a given and fixed value of Firm 1’s market share \( M \), i.e., an increase in a firm’s defensive marketing effort leads to an increase in its offensive marketing expenditures. This relationship exists because the two marketing instruments generate a positive synergy between them that strengthens a firm’s attractiveness to its rival’s customers.

Note that if the market share \( M \) of Firm 1 is given and constant, then

\[ \frac{\partial a_1}{\partial M}(M) = \frac{c_{12} a_1 V_1(0)}{c_{12} b_1 V_1(0) - \beta^2 V_1(0)^2} > 0, a_1(1) = 0, \]

\[ \frac{\partial a_2}{\partial M}(M) = \frac{c_{22} a_2 V_2(0)}{c_{22} b_2 V_2(0) - \beta^2 V_2(0)^2} > 0, a_2(1) = 0, \]

The analysis of the relationship between the two marketing instruments and the state variable (Firm 1’s market share \( M \)) does not lead to a clear conclusion. Thus, it is difficult to predict how a firm should adjust its investments in offensive and defensive marketing given changes in the market share. We summarize in the following proposition a partial result that can be analytically obtained at this stage.

**Proposition 2.** The following implications hold when the two competitors undertake both offensive and defensive marketing activities at the equilibrium and \( M = 0,1 \).

- If \( b_1(M) < 0 \), then \( a_1(M) < 0 \), or equivalently, if \( a_1(M) \geq 0 \), then \( b_1(M) \geq 0 \).
- If \( b_2(M) > 0 \), then \( a_2(M) > 0 \), or equivalently, if \( a_2(M) \leq 0 \), then \( b_2(M) \leq 0 \).

**Proof.** The derivatives of expressions (12) and (13) with respect to the state variable \( M \) are given by the following equations:

\[ a_1(M) = \frac{c_{12} a_1 V_1(0)}{c_{12} b_1 V_1(0) - \beta^2 V_1(0)^2} > 0, a_1(1) = 0, \]

\[ a_2(M) = \frac{c_{22} a_2 V_2(0)}{c_{22} b_2 V_2(0) - \beta^2 V_2(0)^2} > 0, a_2(1) = 0, \]

The proposition’s implications resulted because the term in the square bracket in the expression of \( a_1(M) \) (\( a_2(M) \)) is negative (positive).

**Proposition 2** suggests that both offensive and defensive marketing activities can simultaneously decrease or increase with Firm 1’s market share. That is, if Firm 1’s market share grows, the firm may optimally benefit by either reducing or increasing the investment in both offensive and defensive marketing activities. Alternatively, the proof reveals that for some intervals of the market share and in some areas of the parameter space, we could obtain \( a_1(M) < 0 \) and \( a_1(M) > 0 \) for Firm 1 and \( a_2(M) < 0 \) and \( a_2(M) > 0 \) for Firm 2. In this context, a firm would benefit most from increasing one marketing activity and decreasing the other as its market share changes over time. For example, Firm 1 may invest heavily in offensive marketing to grow its market share and subsequently invest more in defensive marketing to protect the share already captured. Thus, the relative importance of both offensive and defensive marketing changes as the market share evolves over time. Due to the complexity of the model, however, it is impossible to obtain a precise characterization of the conditions that consistently produce \( b_1(M) > 0 \) and \( a_1(M) < 0 \) for Firm 1 and \( b_2(M) < 0 \) and \( a_2(M) > 0 \) for Firm 2. The following section’s numerical simulations and graphical analysis provide some illustrations.

The following proposition summarizes the players’ strategies when one of the firms controls the entire market.

**Proposition 3.** The offensive and defensive marketing activities when a single firm controls the entire market are given by:

\[ a_1(0) = \frac{c_{12} a_1 V_1(0)}{c_{12} b_1 V_1(0) - \beta^2 V_1(0)^2} > 0, a_1(1) = 0, \]

\[ a_2(0) = 0, a_2(1) = \frac{c_{22} a_2 V_2(0)}{c_{22} b_2 V_2(0) - \beta^2 V_2(0)^2} > 0, \]

\[ b_1(0) = \frac{\beta^2 a_1 V_1(0)}{c_{12} b_1 V_1(0) - \beta^2 V_1(0)^2} > 0, b_1(1) = 0, \]

\[ b_2(0) = \frac{\beta^2 a_2 V_2(0)}{c_{22} b_2 V_2(0) - \beta^2 V_2(0)^2} > 0, b_2(1) = 0, \]

The evaluation of Eqs. (3)–(6) at \( M = 0 \) and \( M = 1 \) with straightforward manipulations allows us to characterize the offensive and defensive marketing strategies when one of the firms controls the entire market. On one hand, it is apparent that \( V_1(0) > 0 \) and \( V_2(1) > 0 \). In other words, if a firm’s market share is zero, then any increase to its share leads to an increase in its value function. On the other hand, the second-order conditions for the maximization in the HJB equations in (7) ensure that the denominators


\[ c_{12} \mathbf{c}_{11} - \beta \mathbf{g}_{12} \mathbf{V}(0) \] and \[ c_{22} \mathbf{c}_{21} - \beta \mathbf{g}_{22} \mathbf{V}(0) \] are positive. These conditions guarantee the positivity of Firm 1's offensive and defensive marketing activities when its market share is zero.

Proposition 3 suggests that a firm should start investing in defensive marketing activities even when its market share is zero and keep defending even when it controls the entire market. This nonintuitive result follows from the positive interaction between offensive and defensive marketing. If no interaction takes place, then the firm should not invest in defensive marketing when its market share is zero. When customers anticipate the future benefits that a firm's defensive marketing can offer them, defensive marketing programs act as critical parts of the firm's full offensive arsenal, and firms can market the defensive marketing program benefits to attract new customers. Aaker (2003) provides examples of customer loyalty programs that are used to attract new customers. Conversely, although a firm should invest in offensive marketing when its market share is zero, it is not optimal to continue investing in offensive marketing when a firm controls the entire market.

As in Chintagunta and Vilcassim (1992) and Erickson (1992), we set the discount rate to zero (\( \tau = 0 \)). This assumption is restrictive but necessary to make the HJB equations analytically tractable. We could use numerical approaches to overcome this difficulty, but these approaches also have limitations (e.g., Jarar et al., 2004). The strategies for offensive and defensive marketing obtained when \( \tau = 0 \) are good approximations for all small values of the discount rate, which is especially important when long-term profit maximization in a dynamic setting is considered (Erickson, 1993). Because the firms' profit integrals are evaluated over an infinite horizon with a zero discount rate, the integral cannot converge. However, the steady state profit converges to zero faster than time tends to infinity, ensuring that the improper integral is convergent. Erickson (1993) also finds that steady state profit is zero when the firm follows the feedback equilibrium strategy.

Considering the assumption that \( \tau = 0 \), we replace the expressions of the derivatives of the value functions given by Eqs. (8) and (9) in HJB Eqs. (1)–(2) and obtain the following simplified system of four non-linear algebraic equations with four unknowns, \( ai(M), bi(M), i \in \{1,2\} \):

\[
\begin{align*}
g_1(M) + c_{11} \frac{ai}{Z} - c_{12} \frac{bi}{Z} &= \frac{1}{\alpha_i} \left[ \gamma_i(1-M) - (\gamma_i - a_i)(\gamma_i + b_i) \right], \\
g_2(M) - c_{21} \frac{ai}{Z} + c_{22} \frac{bi}{Z} &= \frac{1}{\beta_i} \left[ (a_i - b_i)(1-M) + (a_i + b_i) \right], \\
g_1(1-M) + c_{11} \frac{ai}{Z} - c_{12} \frac{bi}{Z} &= \frac{1}{\alpha_i} \left[ \gamma_i(1-M) - (\gamma_i - a_i)(\gamma_i + b_i) \right], \\
g_2(1-M) - c_{21} \frac{ai}{Z} + c_{22} \frac{bi}{Z} &= \frac{1}{\beta_i} \left[ (a_i - b_i)(1-M) + (a_i + b_i) \right].
\end{align*}
\]

Replacign the expressions of \( ai \) and \( bi \) as functions of \( bi \) and \( b_2 \), respectively, as given in Eqs. (12) and (13), these equations can be transformed as a two non-linear equation system in variables \( bi \) and \( b_2 \). We carry out numerical simulations using MATLAB to obtain the variables \( bi \) and \( b_2 \) as functions of Firm 1’s market share. Once these functions are known, we can easily derive the behavior of variables \( a_1 \) and \( a_2 \) as functions of Firm 1’s market share with Eqs. (12) and (13).

5. Numerical illustrations

We employ numerical analyses to better evaluate and understand the impact some model parameters have on the firms’ strategies. Specifically, we analyze the effects of both offensive and defensive marketing costs and effectiveness. Our numerical analysis is based on two elements: the intensity of both offensive and defensive marketing when market share is zero or one, and the evolution of both types of marketing at different market shares over time.

5.1. The effects of costs on firms’ strategies

This subsection evaluates how the costs of the two marketing activities affect competitors’ strategies. We first assume \( \alpha_i = \beta_i = \gamma_i = g_i = 1, i \in \{1,2\} \), and then we vary the values of the marketing costs. In the first example (Fig. 1, upper left), we assume that all marketing costs are identical and take the values \( c_1 = c_2 = 0.25, i \in \{1,2\} \). We next assume that the cost of offensive marketing is higher than the cost of defensive marketing with, respectively, \( c_1 = 1 \) and \( c_2 = 0.25 \) (Fig. 1, upper right) and \( c_1 = 1 \) and \( c_2 = 0.75 \) (Fig. 1, lower).

In Fig. 1 (upper left), Firm 1’s initial investments in both offensive and defensive marketing are \( a_1(0) = 1.9 \) and \( b_1(0) = 1.4 \). The investments in the two instruments are identical when the two firms hold the same market share, i.e., \( M = 0.5 \). In all other areas where the market shares of the two firms differ, the weaker firm is predominantly offensive while the leader is largely defensive. As expected from Proposition 3, \( a_1(1) = 0 \) and \( b_1(1) = 1.0 \). Under these conditions, the investment in offensive marketing decreases with increasing market share, but the investment in defensive marketing may either decrease or increase with market share. Most importantly, this figure shows that when the value of the market share of Firm 1 (M) is in the approximate interval 0.15–0.70, \( a_1(M) < 0 \) and \( b_1(M) > 0 \). Outside of this interval, \( a_1(M) > 0 \) and \( b_1(M) < 0 \). In other words, with an extremely small or large market share, it is optimal for Firm 1 to reduce both offensive and defensive marketing; when Firm 1’s market share takes intermediate values, the firm should emphasize defensive marketing at the expense of offensive marketing activities.

Fig. 1 (upper right) shows a significant change in the players’ strategies compared to the strategies used in the first scenario. Defensive marketing activities now are significantly cheaper than offensive marketing (with \( c_1 = 1 \) and \( c_2 = 0.25, i \in \{1,2\} \)), and Firm 1 should increase its initial investment in defensive marketing so that \( b_1(0) = 2.6 \). The firm should reduce its initial investment in offensive marketing to \( a_1(0) = 1.5 \). Regardless of Firm 1’s relative position in the market, more resources are allocated to defensive marketing. Additionally, Firm 1’s (Firm 2’s) investments in both offensive and defensive marketing decrease (increase) with the increase of Firm 1’s market share (\( a_1(M) < 0 \) and \( b_1(M) > 0 \); \( a_1(M) > 0 \) and \( b_1(M) < 0 \)).

In Fig. 1 (lower), defensive marketing activities remain cheaper than offensive marketing, but the difference between the two costs is significantly reduced (\( c_1 = 1 \) and \( c_2 = 0.75 \)). Firm 1’s initial investments in offensive and defensive marketing are \( a_1(0) = 1.2 \) and \( b_1(0) = 1.0 \). The investment in offensive marketing decreases monotonically with increasing market share. Depending on the competitive position of the firm, the investment in defensive marketing may not change with small variations in market share, or such investment may either increase or decrease with market share, but the investment in defensive marketing does not change significantly with market share. When Firm 1’s market share is within the interval \( M = 0.15–0.70 \), \( b_1(0) > 0 \) and \( a_1(M) < 0 \). Most importantly, although it still is cheaper to undertake defensive marketing, Firm 1 finds it optimal to allocate more resources to offensive marketing when \( M < 0.75 \).
Fig. 1 shows that when offensive and defensive marketing are equally effective, the two firms’ strategies are driven by their market shares and the relative costs of offensive and defensive marketing activities. Regardless of the firms’ market shares, the widely held belief that firms should invest more resources in defensive marketing than in offensive marketing holds when customer acquisition costs are substantially higher than customer retention costs (e.g., Fruchter & Sigué, 2009; Reichheld & Sasser, 1990). On the other hand, despite moderately higher customer acquisition costs, investing more in offensive marketing can be effective when the firm’s market share is small. When a firm’s market share is small, the increase in its market share due to an increase in offensive marketing is greater than a similar increase in defensive marketing and overcomes the gain in costs. Thus, when customer acquisition costs are only slightly higher than customer retention costs, the firm with less market share can increase its share by investing heavily in offensive marketing to the leader’s customer base. The focus should shift to primarily protecting a firm’s own customer base only when its market share becomes relatively large.

5.2. Dominant defensive marketing effects

We next consider a scenario in which defensive marketing’s role in strengthening customer retention is a key success factor in the industry. This scenario applies to many industries, such as airlines, wireless phones, health clubs, and grocery stores. In this scenario, defensive marketing has a larger effect on market share than offensive marketing and the interaction between offensive and defensive marketing, i.e., $\gamma_i = 1$ and $\alpha_i = \beta_i = 0.25$, $i \in \{1,2\}$. The marketing cost values are normalized to one, $c_1 = c_2 = 1$, $i \in \{1,2\}$. Fig. 2 displays the firms’ marketing strategies with respect to Firm 1’s market share.

Observe from Fig. 2 that if Firm 1’s market share initially is zero, then the firm launches the greatest possible effort in offensive marketing.
Offensive marketing remains higher than defensive marketing until Firm 1 should allocate more resources to offensive than to defensive activities. More broadly, the dominant offensive marketing effect is more effective than defensive marketing in this scenario, offensive marketing will attract a few additional customers from a smaller rival who still can gain by investing in offensive activities.

5.3. Dominant offensive marketing effect

Under the third scenario, we assume that offensive marketing has a larger effect on market shares than does defensive marketing and that the interaction between offensive and defensive marketing, i.e., \(a_i = 1\) and \(\beta_i = \gamma_i = 0.25, i = \{1, 2\}\). This scenario typifies industries that produce household appliances, soft drinks, and electronic products where customers typically are more receptive to offensive marketing activities. The marketing cost values are normalized to one, \(c_i = c_2 = 1, i = \{1, 2\}\).

Our findings for this case are displayed in Fig. 3. In Fig. 3, Firm 1’s offensive marketing effort is relatively low when its market share is zero \((a_1(0) = 0.4)\). As market share increases, the intensity of offensive marketing initially decreases slightly, quickly increases to reach a maximum when Firm 1’s market share exceeds 0.75, and then decreases to zero as Firm 1 controls the entire market. Firm 1 begins with only a minimal investment in defensive marketing when its market share is zero. The investment in defensive marketing increases to reach a maximum \((b_1 = 1.35)\) when the market share is approximately 0.95, and, from this maximum, defensive marketing decreases to just over 0.80 as Firm 1 controls the entire market. Thus, for market shares in the interval \(M = 0.80–0.95\), the investment in defensive marketing decreases to just over 0.80 as Firm 1 controls the entire market. For this reason, the investment in defensive marketing is not always the optimal tool for increasing market share for the dominant firm in the market. The player with the larger market share in an environment where customers are sensitive to offensive marketing activities is more vulnerable to its rival’s attacks. In this context, it is more effective for the dominant leader to defend its own customers than to try attracting a few additional customers from a smaller rival who still can gain by investing in offensive activities.

5.4. Asymmetric firms

Lastly, we consider two scenarios where Firm 1 is either more effective at offensive marketing or more effective at defensive marketing than Firm 2. We assume that a firm will invest primarily in the activity that it performs better than its rival because the firm has a competitive advantage in this activity. In the two scenarios, the marketing cost values are normalized to one, \(c_1 = c_2 = 1, i = \{1, 2\}\), while the values of the remaining parameters are, respectively, \(\alpha_1 = 1.25, \alpha_2 = \beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 1\) and \(\gamma_1 = 1.25, \alpha_1 = \alpha_2 = \beta_1 = \beta_2 = \gamma_2 = 1\). The findings of these two cases are displayed in Fig. 4 (left) and (right).

Fig. 4 (left) suggests that, despite Firm 1 being more effective at offensive marketing than Firm 2, Firm 1’s market share still can be optimized by investing mainly in defensive marketing if Firm 1 is the dominant player in the market. On the other hand, Firm 1 should invest primarily in offensive marketing when its market share is small, even though it is more effective at defensive marketing (Fig. 4 (right)). Thus, adding asymmetry to the model does not change the qualitative finding that the dominant firm mostly defends, while the smaller firm primarily targets the rival’s customer base through offensive marketing to expand its market share. However, observe that in Fig. 4 (left), \(a_1(0) \simeq 1, b_1(0) \simeq 0.6, \alpha_2(0) \simeq 0, \) and \(b_2(0) \simeq 0.65\); while in Fig. 4 (right), \(a_1(0) \simeq 1.2, b_1(0) \simeq 0.8, \alpha_2(0) \simeq 0, \) and \(b_2(0) \simeq 0.6\). These numbers suggest that Firm 1’s competitive advantage in defensive marketing generates greater initial investments in both offensive and defensive marketing than a similar asymmetry in offensive marketing. Knowing that the rival’s customers are less loyal than its own customers, Firm 1 has additional incentive to attack its rival and transform newly acquired customers to loyal customers. The reverse also holds, and the incumbent initially defends slightly more to counter the new entrant’s competitive advantage in offensive marketing, which limits the investments in both offensive and defensive marketing. Such an analysis can be conducted at any level of the firms’ market shares. Any competitive advantage in a given type of marketing activity of course will impact the two firms’ strategies.

The intent of these numerical simulations is to demonstrate that the qualitative tradeoffs between offensive and defensive marketing activities identified in the literature can indeed arise at reasonable values of the various model parameters, and the findings in Propositions 2 and 3 hold. We hope that these tradeoffs form the basis for further empirical analysis.

6. Discussion and conclusions

This article aims to develop a comprehensive decision-making framework that allows marketing managers to optimally allocate their scarce resources between offensive and defensive marketing activities in dynamic competitive markets. We now discuss our main findings, their implications for marketing researchers and managers, the limitations of our study, and suggestions for future research.

6.1. Results

We find that the suggestion that firms should allocate more resources to defensive marketing than to offensive marketing because it is cheaper to retain current customers than to attract new ones applies only to the particular case where the cost differential between customer acquisition and customer retention is extremely large. Otherwise, allocating resources across offensive and defensive marketing is not straightforward and largely depends on the firms’ relative positions in the market, the...
firms’ competitive advantage/disadvantage in performing marketing activities, and the costs and effectiveness of the two classes of marketing activities. For example, in a context where retaining one’s own customers does not have significant costs, the firm with the smallest market share may find it optimal to predominantly allocate resources to offensive marketing. We also find that the challenger in a duopoly should invest mostly in offensive marketing to grow its market share, even when defensive marketing has a larger effect on market share. Meanwhile, the dominant firm should invest primarily in defensive marketing. However, even when the impact of defensive marketing on market share is weak relative to offensive marketing, focusing solely on offensive activities does not serve the dominant firm’s interest. Defending its customer base through defensive marketing is a firm’s best strategy to limit the impact of a challenger’s assaults. Importantly, adding asymmetry to our model to take into account competitors’ relative strengths in performing marketing activities does not change the previously discussed qualitative results.

6.2. Discussion

This research has several theoretical and managerial implications. First, because our model considers the interaction or synergy between offensive and defensive marketing, we are able to demonstrate that this interaction is necessary for a thorough appraisal of the relative effects on market share of these two classes of marketing activities (e.g., Naik & Raman, 2003; Naik et al., 2005). Models that do not consider the interaction disregard the benefits of a firm’s advertising on its customer retention programs. This interaction makes offensive marketing activities more important than expected by those who ignore the interaction. Thus, our findings show that modeling this interaction gives the competitors a full range of realistic strategic options that cannot be obtained otherwise.

Second, our model and results integrate some of the apparently conflicting results from previous research. For example, some studies recommend that a firm’s offensive marketing increases with its market share (Chintagunta & Vilcassim, 1992; Erickson, 1991; Jarrar et al., 2004), whereas Erickson (1993) finds that offensive and defensive marketing strategies monotonically decrease and increase with a firm’s own market share. Our findings show that, depending on the area in the parameter space, both of these behaviors are possible, i.e., a firm’s offensive marketing can either increase or decrease with its own market share.

Third, our findings support the view that a firm with zero market share entering a market should combine both offensive and defensive marketing, whereas the incumbent firms with large market shares primarily should use defensive marketing against new entrants. In this regard, our work extends Hauser and Shugan (1983), a study that analyzes how an incumbent firm defends itself from the assault of a new entrant. We are able to characterize how intensely the incumbent should respond to the new entrant’s attack, and we show that it is optimal for the incumbent to add offensive marketing to its arsenal once the new entrant starts building its customer base.

Lastly, we extend the CRM literature that investigates a single firm’s resource allocation decisions between customer acquisition and retention (e.g., Musalem & Joshi, 2009), and we show that the firms’ current positions in the market and their relative strengths in performing offensive and defensive marketing activities should be considered in marketing resource allocation decisions in a competitive market.

6.3. Limitations and future research

The findings derived here are based on several simplifying assumptions. This research makes the pivotal assumption that marketing resources can be divided between customer acquisition and retention activities. This assumption is implicitly or explicitly used in several other resource allocation articles (e.g., Blattberg & Deighton, 1996; Reinartz, Thomas, & Kumar, 2005). The separation can be straightforward when some marketing tools such as direct marketing are used, but significant overlap can also exist across offensive and defensive marketing activities. For simplicity, we also assume that the evolution of a firm’s market share depends exclusively on its own and the rival’s investments in offensive and defensive marketing activities. This assumption can be relaxed if a firm’s marketing activities generate positive or negative word-of-mouth, which can affect the firm’s market share. We also assume a mature market, but Erickson (1985) shows that the Lanchester model can be modified to study a growing market (at the expense of analytical tractability). Our analytical findings could hold in a growing market; although the market may increase over time, market shares may not change. Offensive marketing also may become more important due to an increased opportunity to obtain new customers (and defensive marketing and their interaction may become relatively less important).

We confined our interest to an analytical investigation that, with the support of numerical simulations, shows that various tradeoffs between offensive and defensive marketing can be considered in a dynamic competitive environment. This work could be extended by conducting empirical studies to examine whether some of the tradeoffs derived from our proposed framework are observed in the business world. In this perspective, at least two different types of research can be conducted. A longitudinal study could examine how firms change the allocation of their spending across customer acquisition and retention activities.
as their market shares change over time. A cross-sectional study could examine firms with different market shares and determine whether the observed firms are investing in defensive and offensive marketing as predicted by our analysis. This research also suggests a related question regarding whether firms in different industries vary their spending in both defensive and offensive marketing according to the costs and effectiveness of these activities and their competitive positions (Kim & Hahn, 2004).

Acknowledgments

The first author’s research was partly supported by MICINN and JCYL under projects ECO2008-01551 and VA001A10-1 and was co-financed by FEDER funds. The last two authors’ research was supported by Athabasca University. We thank two anonymous reviewers, an associate editor, and the editor, Professor M.G. Dekimpe, for their helpful comments.

References