Sales Forecasts for Existing Consumer Products and Services: Do Purchase Intentions Contribute to Accuracy?

J. Scott Armstrong
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104
Phone 215-898-5087
Fax 215-898-2534
e-mail: Armstrong@wharton.upenn.edu

Vicki G. Morwitz
Stern School of Business, New York University
44 W. 4th Street
New York, NY 10012
e-mail: vmorwitz@stern.nyu.edu

V. Kumar
Marketing Department
University of Houston
Houston, TX 77204-6283
e-mail: vkumar@uh.edu

April 2000 R

ABSTRACT

Purchase intentions are routinely used to forecast sales of existing products and services. While past studies have shown that intentions are predictive of sales, they have only examined the absolute accuracy of intentions, not their accuracy relative to other forecasting methods. For example, no research has been able to demonstrate that intentions-based forecasts can improve upon a simple extrapolation of past sales trends. We examined the relative accuracy of four methods that forecast sales from intentions. We tested these methods using four data sets involving different products and time horizons; one of French automobile sales, two of U.S. automobile sales, and one of U.S. wireless services. For all four products and time horizons, each of the four intentions-based forecasting methods was more accurate than an extrapolation of past sales. Combinations of these forecasting methods using equal weights lead to even greater accuracy, with error rates about one-third lower than extrapolations of past sales. Thus, it appears that purchase intentions can provide better forecasts than a simple extrapolation of past sales trends. While the evidence from the current study contradicts the findings of an earlier study, the consistency of the results in our study suggest that intentions are a valuable input to sales forecasts.

Keywords: combining, expectations, extrapolation, intentions, purchase probabilities

Assume that you need to forecast next year's sales for a consumer product. Also assume that you have time-series data on the sales of this product and that you have access to consumer survey data about whether the respondents intend to purchase it within the next year. To what extent, if any, would you expect the intentions data to improve the accuracy of forecasts based solely on past sales data?

Purchase intentions are often used to forecast sales of existing consumer durables and new consumer products (Lehman 1989, pg. 140; Morrison 1979). Purchase intentions are inexpensive to acquire and easily understood by managers, which may account for their widespread use. Organizations in many countries collect data about people's intentions to purchase consumer goods. For example, starting in the 1940's, the U.S. government conducted surveys to forecast consumer spending on durables such as new cars and houses (McNeil 1974). Such surveys are currently conducted by organizations in the U.S. and other countries. (IFO in Munich keeps track of these studies.[1]) The implicit assumption made by those who gather consumer intentions data is that they will aid in the prediction of future sales. Indeed, many studies have found a positive correlation between purchase intentions and purchase behavior (Adams 1974; Juster 1966; McNeil 1974; McNeil and Stoterau 1967; Morwitz and Schmittlein 1992; Morwitz, Steckel and Gupta 1996; Tobin 1959). However, most past studies have only examined the absolute predictive accuracy of purchase intentions, not their accuracy relative to other forecasting methods.

In practice, purchase intentions can be used to make a variety of managerial decisions (Morrison 1979). For example, consumer durable-goods producers can use purchase intention measures to help anticipate major shifts in consumer buying so that they can adjust their production and marketing plans accordingly. Buyer-intention surveys can also be useful in estimating demand for new products (Silk and Urban 1978).

While purchase intention measures have a wide range of applications, the focus of the present study is on product category sales for existing consumer durable products/services. Category-level sales forecasts are often used by product manufacturers as well as by consultants, trade associations, and the government (Kalwani and Silk 1982). For example, in the U.S. wireless industry, managers use intentions measures to assess category growth. Although the intentions are measured at the category level, the forecasts are often used by brand managers (Infosino 1986; Kalwani and Silk 1982). Product category-level sales forecasts are similarly used in auto, appliances and services industries.

We compare intentions-based forecasts to simple extrapolations of past sales data. While there are many different benchmark forecast methods one could choose, we decided to use extrapolations of past sales for several

reasons. First, we wanted a comparison model that was as easy to use and easy to understand. Second, managers typically have information on past sales and can use time-series methods. However, the use of sophisticated time-series methods is not always possible as these require longer time series than what is normally available. Prior comparative studies have concluded that there is no benefit to sophisticated extrapolations in such a situation (Armstrong 2001a). Finally, and of key importance, a research study compared intentions-based forecasts to simple extrapolations of past sales and found the latter to be more accurate (Lee, Elango, and Schnaars 1997). Our goal was to reexamine this issue with different methods and a greater variety of data sources.

The theoretical literature is equivocal about whether intentions-based forecasts or past sales trends should be more accurate. Received wisdom suggests that the best predictor of future behavior is past behavior. On the other hand, the social psychology literature states that a good predictor of what individuals will do is their stated intentions to perform the behavior (Fishbein and Ajzen 1975, p. 368-369). Other research suggests that intentions data are useful for predictions under certain conditions. Armstrong (1985, pp. 81-84) summarizes these conditions: (1) the event being predicted is important, (2) the respondent has a plan (at least the high intenders do), (3) the respondent can fulfill the plan, (4) new information is unlikely to change the plan over the forecast horizon, (5) responses can be obtained from the decision maker, and (6) the respondent reports correctly. Such conditions are likely to be met for short-term purchase intentions for expensive goods and services. This suggests that intentions data could potentially improve accuracy of forecasts based solely on past sales behavior for these products.

The study uses durable goods and a high technology service, where intentions are expected to contribute to accuracy. We examined the relative accuracy of four methods for forecasting sales from intentions as well as a combination of these four methods. The results have practical implications for marketing managers, economists, and government officials who rely on purchase intentions to forecast future demand for consumer products and services. The results also add to the existing academic body of literature on forecasting product sales using time-series data and using intentions.

PRIOR RESEARCH

The extant literature on intentions has focused primarily on (1) determining which questions should be used to measure intentions, and (2) assessing the strength of the intentions-behavior relationship. The literature has not focused on assessing the relative predictive validity of data obtained from various intentions measures.

A variety of survey questions have been used to measure consumers' purchase intentions and we use the term "intentions" for all. Among the most commonly used measures are Juster's 11-point purchase probability scale (Juster 1966) and a 5-point likelihood of purchase scale. Day, et al. (1991) concluded that Juster's 11-point purchase probability scale provides substantially better predictions of purchase behavior than other types of intentions scales. They based their conclusion on the evidence from two New Zealand studies and also on prior research by Juster (1966), Byrnes (1964), Stapel (1968), and Gabor and Granger (1972). This 11-point purchase probability scale (with verbal descriptions at each point) may provide better predictions than other scales because it captures more information than scales with fewer points (Pickering and Isherwood 1974).

Purchase probabilities (e.g., What is the probability that you will buy a car?) and expectations (e.g., How likely are you to buy a car? Do you expect to buy a car?) are broader than direct intentions questions (e.g., Do you intend to buy a car? Do you plan to buy a car?) because they refer to actions that might be unplanned as well as planned. Assessing purchase probabilities and expectations may be advantageous in situations where people realize that they may purchase an item even though they have no plans at the time of the survey (e.g., they realize that their old automobile might break down and that might lead them to purchase a car). Therefore, a smaller proportion of respondents reports "zero" on purchase probability scales than report "no" on intentions scales (Juster 1966). For many studies, most purchases are made by those who had reported no plans to buy (Juster 1966). This occurs because although non-intenders seldom purchase, they are often the largest group of respondents. Probability and expectations questions may also be a better measure of future purchasing for low-involvement goods than intentions questions because respondents may be less likely to have explicit purchase plans at the time of the survey, but they may expect to make a purchase. For example, consumers may not intend or plan to purchase a candy bar in the next month but may realize that it is likely they will make an unplanned purchase.

Models have been developed to describe how purchase intentions relate to purchase behavior for durable goods such as cars. Morrison (1979) developed a model that allows for random and systematic error in intent measurement, and for changes in true intentions over time. Morrison's model was proposed as a descriptive, rather than a predictive model, and it requires purchase data for estimation. Bemmaor (1995) extended Morrison's model to allow for differences between non-intenders and intenders in the probability that their true intentions change over time. In addition, Bemmaor's model can be used to develop *ex ante* forecast bounds on the proportion of respondents who will purchase. Bemmaor derives upper and lower bounds for the proportion of consumers who will purchase a

5

wide range of products. He demonstrates that, for many existing consumer products, these forecast bounds include the actual proportion purchased.

Despite much prior research, we know little about the value of consumers' reported purchase intentions for predicting purchases when we also have time series data on product sales. There is theoretical literature in psychology that suggests that a good predictor of an individual's future behavior is his or her stated intention (Fishbein and Ajzen 1975). However, the psychological literature also suggests that past behavior is an important predictor of future behavior (Bentler and Spechart 1979). The empirical evidence suggests that both intentions and past sales are useful for forecasting future sales. However, there is little research that examines the relative accuracy of intentions and past sales. We found only one study that addressed this issue. Lee, Elando and Schnaars (1997) analyzed intentions data (for cars and houses) from the Conference Board, and past sales data from different secondary data sources. They concluded (p. 127), "Very little support is found for using buying intentions as a forecasting tool for predicting the sales of [existing] durable goods."

INTEGRATING INTENTIONS AND SALES DATA

Two commonly used methods to forecast sales from intentions predict that the proportion of consumers who will purchase will equal 1) the mean intent (transformed to lie between zero and one to represent the mean probability of purchase), or 2) the proportion of respondents indicating a positive purchase intent (Morwitz et al. 1996). However, several studies have shown that these methods often provide biased estimates of sales, overstating or understating actual purchasing (Morrison 1979, Manski 1990). Thus, when possible, sales data should be used to adjust for the bias in intentions. The simplest way to do this is to relate an aggregate measure of purchase intentions to an aggregate measure of sales. For example, Morrison (1979) estimates the bias in intentions measures as the difference between the mean intent across respondents (transformed to a number between 0 and 1) and the proportion that purchased. For an existing product, this bias could be estimated in a previous period and used to adjust for the bias in a current intent score. This method should be accurate if the bias is stable over time.

If panel data are available, a second way to adjust for the bias in intentions data is to use sales data to estimate the proportion of respondents in each intender and non-intender group who will make a purchase (i.e., intent-behavior conversion rates). For example, for existing products, intention-behavior conversion rates can be obtained from past time periods (Morwitz and Schmittlein 1992). This method should lead to accurate sales forecasts if these conversion rates are stationary over time.

While our study deals with existing products, there is an analogy to new product forecasting. For new products, intent-behavior conversion rates can come from past observations of similar products or from established industry norms. Jamieson and Bass (1989) describe a set of different intent-behavior conversion norms that have been used by industry practitioners. The accuracy of this method for new product forecasting will depend on the extent to which intent-behavior conversion rates vary across similar products.

Given the uncertainty about the best way to use intentions, one strategy is to use different approaches to develop forecasts, then to combine the forecasts from these approaches. The evidence on combining forecasts is substantial and favorable (Batchelor and Dua 1995; Clemen 1989; Schmittlein, Kim, and Morrison 1990; Winkler 1989). Combined forecasts are expected to be especially effective when the component forecasts differ substantially from one another with respect to the data and methods used. Combining forecasts from different methods with equal weights is recommended unless a strong case can be made for differential weights (Armstrong 2001b). Therefore, in this study we develop several different intentions-based forecasts and combine them using equal weights.

DESIGN

This section describes the data examined in this study, the methods used to develop the intentions forecasts, and the criteria used to evaluate them.

Data

To test the value of intentions, we sought situations where they are expected to be useful. Automobile sales meet the key conditions where intentions are expected to be useful. The automobile purchase decision is an important one and the typical consumer plans this purchase. Of course, some consumers might encounter new conditions that would change their plans (such as the loss of a job, or loss of a car due to theft or accident), but, given an adequate sample, these effects are not expected to be large over a one-year forecast horizon. Similarly, we believed that the conditions were also favorable for using intentions to predict subscriptions to wireless telephone services, although less so because new information might change the intentions.

French Car Data (1977-1988). We used eleven years of data on French consumers' intentions to purchase automobiles, along with annual sales data. These data were originally collected by the National Statistics Institute in France, and were analyzed in Bemmaor (1995). The intentions were measured annually in October. Each year's intentions were obtained using a scale that asked for the intent to purchase in the next 14 months.

The question was stated as "Within the two to three years to come, do you (or some member of your household) plan to buy a ____?" The possible answers were "Yes, certainly," "Yes, perhaps," "Probably not," "Certainly not," and "Don't know." Those answering "Yes, certainly" and "Yes, perhaps" were then asked: "And more precisely, do you have a plan to buy a ____ between now (October) and the end of next year?" followed by the same scale for responses. If the respondent answered "Yes, certainly" or Yes, perhaps" he or she was then asked: "At what approximate date do you plan to make that purchase?" "Before the end of the year," "Between January 1 and April 1 of next year," "Between April 1 and the beginning of the summer vacation next year," "During or after the summer vacation of next year," and "Don't know." Bemmaor (1995, p. 180, Table 1) transformed these responses into a ten-point scale to reflect 14-month ahead purchase intentions.

Bemmaor reported the proportion of respondents in each intender and non-intender group (i.e., for each point on the 10-point scale). Purchase data were obtained by examining the proportion of respondents from the previous year who purchased a car one year after they expressed their intentions. Bemmaor reported both aggregate annual purchase rates and the purchase rate for each intender and non-intender group. The panel size varied, ranging from 2,190 to 2,890 respondents per year.

U.S. Car Data (1961-1965). Theil and Kosobud (1968) reported data from the United States Bureau of the Census' Current Population Survey on American consumers' intentions to purchase an automobile. They report intentions and purchases of respondents re-interviewed after twelve months over a four-year period 1961 to 1965. The sample size per year ranged from 4,000 to 12,000 respondents (quarterly data were pooled to increase the sample size). Respondents were asked about their expectations as follows: "In the next six months, does any member of this family expect to buy a car?" (with response alternatives "yes–definitely," "yes–probably," "maybe–other reasons," "maybe–depends on old car," "don't know," and "no"). Theil and Kosobud considered a respondent to be an intender if he or she marked any of the first four categories. They report the percentage of intenders and non-intenders, the overall twelve-month purchase rates, and the twelve-month purchase rates for intenders and non-intenders.[2]

U.S. Car Data (1987-1989). These data were originally collected by Allison Fisher, Inc.,[3] and a subset of these data was analyzed in Morwitz and Schmittlein (1992). Purchase intentions were measured twice a year from July 1987 through January 1989. Each period's intentions were obtained using a scale that asked respondents to indicate their expectations: "When will the next new (not used) car (not truck or van) be purchased by someone in

your household?" with the following six response alternatives: "6 months or less," "7-12 months," "13-24 months," "25-36 months," "Over 36 months," and "Never." We used a binary measure indicating whether a respondent expected to purchase a new automobile in the next six months. Purchase data indicated the proportion of respondents from the previous year who purchased a car six months after they expressed their expectations. We computed aggregate six-month purchase rates and six-month purchase rates for intenders and non-intenders. The panel size varied across the four six-month periods, ranging from 124,749 to 139,359 respondents per period.

U.S. Wireless Telephone Service (1993 - 1996). Compared with automobiles, the decision to purchase wireless service is less important and requires less planning. Also, changes in consumer acceptance for this relatively new high tech product are likely to be rapid. Overall, then, we expected intentions to be relevant, though less useful than for automobiles. We also expected intentions with shorter time horizons (e.g., the next 3 or 6 months) to be more useful than those for larger time horizons (e.g., 12 months), given the rapid change in this industry. These data were collected through primary research sponsored by a commercial firm. A sample of respondents who met the target market criteria were asked about their purchase expectations for wireless telephone service. The question was:

How likely are you to subscribe to wireless service in the next X months?

Definitely   Probably   Might   Probably Not   Definitely Not

The value for "X" was varied at 2, 3, 6, and 12 months. Respondents were presented with only one of these four time periods. The sample size for each period was 100. Thus, a total sample of 400 respondents provided responses each period. The study was repeated six times in a given year for the question that measured purchase intent in the next two months. Similarly, the study was repeated four, two, and one times in a given year to reflect quarterly, semi-annual, and annual purchase intentions. Thus, over a four-year horizon, twenty-four data points are available for evaluating bimonthly purchase intentions. Likewise, 12, 8, and 4 observations are available for evaluating purchase intentions across quarterly, semi-annual, and annual time periods. The respondents were contacted again after the set interval and were asked if they had subscribed to a wireless service.

Methods

We first describe the benchmark method, then provide alternative intentions forecasting methods. We use bold font to denote the names that we give to each method.

Benchmark. In the benchmark method, we forecast that sales are equal to last period's sales plus drift. This is referred to as the **sales extrapolation method**. Sales were expressed on a per capita basis by dividing by the

number of adults. We only do this to establish the past sales trend for periods preceding those for which we have intentions and purchasing data. For these periods, we know the percent that bought directly from our data. Our original intent for drift was to estimate the annual percentage change rate for each series by using a regression across time and setting the drift at ½ of this rate. This was adopted in line with the favorable evidence on damping forecasts (Gardner and McKenzie 1985), along with some uncertainty on our part as to the expected direction of trend. The procedure is equivalent to combining the no-change (random walk) and the trend model. As it turned out, the regression models for automobiles indicated that the trends were negligible (and statistically insignificant), so we used the naive, "no-change," model. This sales extrapolation method has been shown to perform well in previous studies. In particular, it was among the better models in a comparative study used to forecast annual sales of consumer durables by Schnaars (1984), and it is one of the two models used by Lee et al. (1997). For wireless service, there was a significant sales trend. For example, with the bimonthly and quarterly data, the trend varied from 10% to 20%. For these data, drift was set at 1/2 of the estimated historical trend in percentage growth terms. We also estimated the "no-change" model for these data.

Methods that Use Past Sales to Adjust for the Bias in Intentions. Intentions, by themselves, provide only a crude way to predict sales. For new products, intentions are sometimes used directly to forecast sales. However, when sales figures are available, it is sensible to calibrate intentions against them. In other words, we look at a category of intenders (e.g., "we are certain we will buy a car in the next six months") and determine what percent actually did purchase in that period. This relationship is then extended to the period to be forecast. We used four methods that relate intentions data to sales in the subsequent time period.

1. **Multiple intender groups:** In the multiple intender groups method, we assumed that the intent-behavior sales rate for respondents in each intender and non-intender group is equal to the corresponding rate for the previous time period. This method was used to forecast sales of durable goods in Morwitz and Schmittlein (1992).

2. **Intender/non-intender groups:** Morwitz and Schmittlein (1992) found that forecast accuracy could be improved over multiple intender groups by computing one intent-behavior conversion rate for all those who indicated that they intended to purchase and one for non-intenders. The intender/non-intender groups method could be better than multiple intender groups because pooling across the different intender groups increases the accuracy in the estimated intent-behavior conversion rate.

10

However the intender/non-intender groups method could be worse than multiple intender groups because it gives up some richness in distinguishing different intender groups.

3. **Average intentions:** Morrison (1979) developed a descriptive model of the relationship between purchase intentions and subsequent purchasing. Morrison proposed that there are three threats to the predictive validity of purchase intention measures. First, intentions are measured with error. Second, respondents' purchase intentions might change over time because of exogenous events (e.g., current car breaks down, sudden rise in income). Third, average stated purchase intentions might be a biased estimate of the proportion that actually buy the product because of systematic error (e.g., response style biases, promotional effects, changes in the economy, as noted in Kalwani and Silk 1982). Morrison developed a mathematical model that addresses these three threats.

   The average intentions method is based on his descriptive model of the purchase intention-purchase behavior relationship. Estimating the model requires both intentions and actual sales data, and thus it cannot be used in its original form as a predictive model. However, as suggested by Bemmaor (personal correspondence, August 21, 1995), we used Morrison's framework to formulate a predictive model:

$$\text{Percent of buyers (t)} = \text{Mean intent (t) - bias (t-1)}.$$

To estimate the predictive model based on the *average intent* method, we computed the mean intention (transformed to a number between 0 and 1) using the current year's intentions data. (The mean intention for the French car data was provided in Bemmaor's paper). We estimated "bias" by taking the difference between mean intentions and the percent who purchased during the previous year.

   The average intentions method could be better or worse than the multiple intender groups and the intender/non-intender groups methods. It could be worse because it uses less information in that it assumes the same process holds for all intender groups. The multiple intender groups method allows the bias estimates to vary across all intender groups, and the intender/non-intender groups method allows separate bias estimates for intenders and non-intenders, while the average intentions method has one bias estimate. Alternatively, the average intentions method may be better because it may eliminate noise in the data.

4. **Adjusted average intentions:** Bemmaor (1995) proposed a method based on a probabilistic model of the relationship between purchase intentions and behavior. Whereas the average intentions method assumes that all respondents have the same probability of changing true intentions over time, ρ, the adjusted average intentions method relaxes that assumption and assumes that intenders and non-intenders have different probabilities of changing true intentions (ρ_{int} and $\rho_{non-int}$).[4] Bemmaor also relaxes the assumption of the average intentions method that bias is constant across respondents.

    Based on equation (2) in Bemmaor's paper and his suggestions (personal correspondence, August 21, 1995), we used the following predictive model:

    Percent buyers (t+1)   =   Mean "true" intent (t) * [1 + ($\rho_{non-int}$(t) - ρ_{int}(t)) *intent dispersion (t) * % of "true" non-intenders (t)]

    The details for deriving the estimates for mean "true" intent, intent dispersion, % of "true" non-intenders, $\rho_{non-int,}$ and ρ_{int} are described in Bemmaor's paper.[5]

    Specifically, the first term, $\rho_{non-int}$, represents the probability that non-intenders switch their true intentions after reporting their stated purchase intent at time t. Estimation of this parameter requires stated intentions data at time t, and the intentions-behavior conversion rate for non-intenders at time t-1. The second term, ρ_{int}, represents the probability that intenders switch their true intentions. Estimation of this parameter requires stated intentions data at time t, and the intentions-behavior conversion rate for intenders at time t-1. For the French automobile data, Bemmaor has already included estimates of these parameters in his paper.

    We might expect the adjusted average intentions method to perform better than the average intentions method and to be comparable to intender/non-intender groups method because Bemmaor allows the intention-behavior conversion process to differ for intenders and non-intenders. However, the adjusted average intentions method might be better or worse than the multiple intender groups method. It might be worse because, like the intender/non-intender groups method, it only allows two different conversion processes: one for intenders and one for non-intenders, while the multiple intender groups method allows every intender group to have a different purchase rate. On the other hand, the adjusted average intentions method might perform more accurately because the conversion

rates computed for two larger groups are likely to be more stable than the rates computed using

multiple intender groups.

Combined Forecasts: One commonly used basis for combining forecasts is to use methods that make

different assumptions about the same data (Armstrong 2001b). Thus, it seemed sensible to combine forecasts from

the four intentions methods (multiple intender groups, intender/non-intender groups, average intentions and adjusted

average intentions methods). Following the findings of prior research (Clemen 1989), we used equal weights to

combine the forecasts.

Combined forecasts are expected to be especially effective when the component forecasts differ

substantially from one another (Armstrong 2001b). We had enough periods in the French data set to examine

agreement. There was a great deal of variation in the correlation among the errors in the different intentions methods

(they ranged from -0.78 to 0.99). Given the differences across these methods, we expect to see improvements from

combining.

Criteria

The Relative Absolute Error or RAE (Armstrong and Collopy 1992) is the mean absolute error in the forecast

divided by the mean absolute error of the naive forecast. This measure was designed for situations where different

series are being examined because it adjusts for scale and controls for the amount of change in the series.

RESULTS

All four methods that used survey data provided more accurate forecasts than the benchmark extrapolations

for all forecast horizons. The results are shown in Table 1. The adjusted average intentions method was most

accurate in each data set. As expected, combined forecasts were more accurate than the average of the component

methods. On average, the combination method's forecast errors were about one-third less than the benchmark

extrapolation and about 5.5% less than that for the typical intentions method.

TABLE 1

Relative Absolute Errors For
Intentions-Based Forecasting Methods
(F = number of Forecasts; M = horizon in Months)

	Automobile			U.S. Wireless Telephone Service, 1993-96			
Method	US 1987-89 M=6 F=3	US 1961-65 M=12 F=4	French 1978-87 M=14 F=10	M=2 F=24	M=4 F=12	M=6 F=8	M=12 F=4
Sales extrapolation	1.00	1.00	1.00	.96[2]	.96	.97	.97
Multiple intender groups	.91	.67	.92	.67	.65	.59	.72
Intender/non-intender groups	.91	.67	.79	.71	.68	.63	.78
Average intentions	.55	.67	.97	.73	.69	.66	.82
Adjusted average intentions	.48	.67	.73	.65	.63	.57	.71
Average of the intentions methods	.65	.67	.85	.69	.66	.61	.76
Combination of intentions[1]	.57	.67	.74	.66	.64	.59	.74

[1] For the French data, we averaged across the four different intentions methods. In the two U.S. data sets, the intentions measure was binary, so *multiple intender groups* and *intender / non-intender groups* forecasts are equal. Therefore, we averaged across the three different intention methods.

[2] The values are less than one for sales extrapolation because the forecasts with sales drift were compared to a naive model forecasts with no drift.

For different products, time horizons, countries, and types of intentions questions, the intentions surveys when combined with prior sales data, were more accurate than forecasts based solely on past sales. The improvements can be important. For example, in the French data, during any given year, about twenty percent of the sample purchased automobiles. The error for the sales extrapolation was about 0.66 points (mean absolute error). If one were using this to forecast automobile sales for the next year, it would imply an error in the percentage of purchasers of about 3.3%. The combined intentions method reduced the error by 31%, which would have produced an error of 2.3%.

In the U.S. Wireless data, the adjusted average intentions and multiple intender group methods performed consistently better than the sales extrapolation and other methods. The sales extrapolation method itself produced an average error in the range of 6 to 8 percent (of the actual number of people subscribing to the wireless service) depending on the purchase intention time period. The combined intentions method reduced the error by about 33%. This would result in an overall error of about 3 to 5% of the actual number of people subscribing to the wireless depending on the purchase intention time period. The error reduction is highest for the semi-annual purchase intention measure and lowest for the annual purchase intention measure.

DISCUSSION

Implications

Purchase intentions are often gathered and used to forecast sales of existing consumer durables (e.g., automobiles, appliances) in order to manage production and sales (Kalwani and Silk 1982). For these types of products, past sales histories are often available. Our findings provide support for the use of purchase intentions to forecast sales even when time series data are available.

Comparison to Previous Findings

Our findings are in conflict with those from the one prior study (Lee, Elango & Schnaars 1997). Since both the data and the forecasting methods used in their papers differed from ours, it is possible that either their results or ours are an artifact of the methods used to forecast sales from intentions. Therefore, to determine why our results differ, we reanalyzed their data using our methods and used their methods on our data. Unfortunately, this effort did not help to resolve the conflict. Using our data, intentions-based forecasts were always more accurate than past sales extrapolations, regardless of the methods used. In contrast, using Lee et al.'s data, past sales extrapolations were always more accurate than intentions-based forecasts.

Using our approach with the Lee et al. data

For each of the seven data series reported in this paper, intentions and sales come from the same samples. This allowed us to compute intentions-sales conversion rates for different intender groups. However, this is not possible with the data analyzed by Lee et al. Their intention measures come from samples of households gathered by the Conference Board. Their sales data do not come from the same samples. Rather, sales for new and used cars came from 1978-1993 issues of *Ward's Automotive Yearbook*, and sales of new and existing houses were obtained from publications of the National Association of Realtors and the U.S. Department of Commerce. Because intentions and sales data do not come from the same sample, we cannot compute intention-sales conversion rates for different intender groups, and therefore cannot use the *multiple intender groups, intender/non-intender groups* or the *adjusted average intentions* method. We are only able to analyze their data using the *average intentions* method.

Specifically, for each month, we related the percent of people who intended to buy each product (i.e., all cars, new cars, used cars, all houses, new houses, existing houses) in the next six months to the percent who purchased each product in that period and assumed the same relation would hold into the next six-month period. In the case where this led to a negative forecasts (because the difference between intentions and the percent purchased

in one period was greater than the percent of intenders in the next period) we set the *average intentions* forecast to zero. We divided Lee et al.'s monthly sales for each product by the total number of U.S. households each year (obtained from the U.S. Census Bureau), to obtain the percent of households that purchased each product for each month. Since the question measured intentions to purchase within six months, we summed across 6-month intervals to obtain 6-month sales. We compared the *average intentions* method to the *sales extrapolation* (i.e., sales for this six-month period = sales in the prior six-month period). We also considered that there could be seasonal variation in the intentions accuracy (different intention-sales conversion rates during the first and last six-months of the year). Therefore, rather than only calibrating intentions to sales during the previous 6 months, we also calibrated intentions to sales during the same 6-month period in the previous year. Since there could also be seasonal variations in sales, we also used these sales as the new *sales extrapolation* forecast.

The results of both analyses indicate that when we use the *average intentions* method with their data, we obtain the same findings as Lee et al; for all six time-series, forecasts based on past sales are more accurate than intentions-based forecasts. (Details can be obtained from the second author.)

Using Lee et al.'s approach with our data

Lee et al.'s approach involves comparing the percentage change in the intentions to the percent change in past sales. Specifically we compared both the percentage change intentions $((intent(t)-intent(t-1))/(intent(t-1))$ and the percentage change in past sales $((Sales(t)-Sales(t-1))/(Sales(t-1))$ to the percentage change in actual sales $((Sales(t+1)-Sales(t))/(Sales(t))$.

For the French car and wireless data we used two different measures of intent: (1) the percent of respondents with positive intent (e.g., 1-9 on Bemmaor's scale), and (2) the mean intent. For the two U.S. car sales data sets we used the percent that intended to buy as the intent measure. The results showed that intentions-based forecasts outperform forecasts based on sales extrapolations for the data analyzed in this paper, even when we use the approach of Lee et al. (Details can be obtained from the second author.)

Reconciling these differences

While we cannot know with any certainty why these discrepancies have occurred, we believe that Lee et al.'s results may be a function of the type of intentions scale used by the Conference Board or the type of sampling used. The Conference Board's intentions scale is a simple 3-point scale that reads "Does anyone in your household

plan to buy a car/house in the next six months? Yes __, No __, Maybe __." The intentions literature suggests that purchase probability scales and scales with more scale-points have greater validity and reliability.

In the seven data series examined in our paper, intentions and sales were provided from the same sample of respondents. In contrast in Lee et al.'s paper intentions and sales data come from different sources because the Conference Board does not do follow up surveys of their respondents to see if they later purchased. It is possible that their intentions sample is not representative of the population. In addition, the sample sizes used by the Conference Board (about 3,000 households per month) are relatively small considering the proportion of households that purchase automobiles and homes over a six-month period. This could result in more noise in their intentions data than in their sales data, which would tend to lower the relative accuracy of their intentions-based forecasts.

Finally, it is possible that the type of sampling used in the data we analyzed led to artificially high accuracy of intentions-based forecasts. In the seven time series analyzed in this paper, intentions and sales data were obtained from the same samples of respondents. It is therefore possible that the correspondence between intentions and sales are high in part simply because intentions were measured. Past research has shown that the simple act of measuring intentions changes behavior (Fitzsimons and Morwitz 1996; Morwitz, Johnson and Schmittlein 1993). However, we were able to rule out this explanation using data on U.S. Wireless subscriptions. In addition to the panel data used in this paper (i.e., the same respondents provided both intentions and sales data), we were also able to obtain subscription data from a different representative cross-section of households and from a firm that provides market level data. We find that intentions-based forecasts are still more accurate than extrapolations of past sales even when sales data come from these other sources. (Details can be obtained from the third author.) Thus we cannot conclude that it was the act of measuring intentions alone that lead to intentions accuracy. It seems most likely that Lee et al.'s result is a function of the type of sampling used or the type of measures used by the Conference Board.

Given that the only two studies on this topic have produced conflicting results, we recommend further research to help determine the conditions under which intentions might improve upon extrapolation.

Limitations and Opportunities for Future Research

In our paper as well as in the research by Lee et al. (1997), the forecasts were all at the product category level. However intentions are also used to forecast sales of specific products and brands. Although intentions have been shown to be a useful input to brand-level sales forecasts (Kalwani and Silk 1982), no one has compared their

accuracy to an extrapolation of past sales. While we have no reason to believe that our results would not hold at the product and brand level, further study would be needed.

The methods used to forecast sales from intentions in this paper depend on being able to calibrate intentions against prior behavior. Thus, these methods cannot be used in situations where prior sales data do not exist. We therefore cannot make any generalizations about the relative predictive accuracy of intentions for new product forecasting. Future research should compare intentions-based new product forecasting models to other commonly used methods, such as expert opinions, for forecasting new product sales.

We did not address the relative costs in terms of dollars and time of obtaining these data. Because it takes time to commission a study to gather intentions and calibrate them to sales (e.g., over a 6-month or 1-year period), some firms may have to trade off concerns of having an immediate forecast versus delaying to gather more data but having a more accurate forecast.

Finally, the relative accuracy of intentions-based forecasting methods to past sales extrapolations may depend upon various conditions. For example, if substantial changes are expected in the future, and if people have good information about their future behavior, it may make sense to put more emphasis on intentions. Future studies should examine factors that affect the relative accuracy of these different forecasting approaches.

SUMMARY

Previous studies have shown that consumer intentions data are predictive of sales of consumer durables (Bemmaor 1995). However, only one study has examined whether intentions data can contribute to the accuracy of sales extrapolations, and it concluded that intentions data were of no value (Lee et al. 1997). Using three sets of data on automobile sales and four sets of wireless telephone data, we found that methods that combine intentions with past sales data produced substantial improvements in accuracy. In general, all of the four purchase intention measures were more accurate than time-series extrapolations. Although we had no basis to expect this in advance, the adjusted average intentions method was most accurate of the four methods that we examined. This method used past sales to adjust for bias and allowed different bias for intenders and non-intenders. As we expected, combined forecasts were more accurate. An equally weighted combination of forecasts from four different intentions methods reduced the error from the typical intention method by 5.5% and by about one-third in comparison with a simple extrapolation of past sales.

REFERENCES

Adams, F. G. (1974), "Commentary on McNeil, *Federal Programs to Measure Consumer Purchase Expectations*," *Journal of Consumer Research*, 1, 11-12.

Armstrong, J. S. (1985), *Long-range Forecasting*. New York: John Wiley.

Armstrong, J. S. (2001a), "Extrapolation for time-series and cross-sectional data," in J. S. Armstrong (ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers: Norwell, MA.

Armstrong, J. S. (2001b), "Combining forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers: Norwell, MA.

Armstrong, J. S. & Collopy, F. (1992), "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting,* 8, 69-80. (Full text at hops.wharton.upenn.edu/forecast.)

Batchelor, Roy & Dua, P. (1995), "Forecaster diversity and the benefits of combining forecasts," *Management Science*, 41, 68-75.

Bemmaor, A. C. (1995), "Predicting behavior from intention-to-buy measures: The parametric case," *Journal of Marketing Research*, 32, 176-191.

Bentler, P. M. & G. Spechart (1979), "Models of Attitude-Behavior Relationships," *Psychological Review*, 86, 452-464.

Byrnes, J. C. (1964), "Consumer intentions to buy," *Journal of Advertising Research*, 4, 49-51.

Clemen, R, T. (1989), "Combining forecasts : A review and annotated bibliography," (with discussion) *International Journal of Forecasting*, 5, 559-588.

Day, D., Gan, B., Gendall, P. & Esslemont, D. (1991), "Predicting purchase behavior," *Marketing Bulletin*, 2, 18-30 (full text of this paper is provided at http:marketing-bulletin.massey.ac.nz/bulletin/index.html).

Fishbein, M. & Ajzen, I. (1975), *Belief, Attitude, Intention, and Behavior*. Reading, MA: Addison-Wesley.

Fitzsimons, G. J. &. Morwitz, V. G. (1996), "The effect of measuring intent on brand-level purchase behavior," *Journal of Consumer Research,* 23, 1-11.

Frankenberger, K. D. & Ruiming, L. (1994), "Does consumer knowledge affect consumer responses to advertised reference price claims?" *Psychology and Marketing,* 11, 235-251.

19

Gabor, A. & Granger, C. W. J. (1972), "Ownership and acquisition of consumer durables: Report on the Nottingham consumer durables project," *European Journal of Marketing*, 6, 234-248.

Gardner, E. S., Jr. & McKenzie, E. (1985), "Forecasting trends in time series," *Management Science*, 31, 1237-1246.

Holak, S. L. & Lehmann, D. R. (1990), "Purchase intentions and the dimensions of innovation : An exploratory model," *Journal of Product Innovation Management,* 7, 59-73.

Infosino, W. J. (1986), "Forecasting new product sales from likelihood or purchase ratings," *Marketing Science*, 5 (Fall), 372-384.

Jamieson, L. F. & Bass, F. M. (1989), "Adjusting stated intention measures to predict trial purchase of new products: A comparison of models and methods," *Journal of Marketing Research*, 26, 336-345.

Juster, F. T. (1966), "Consumer buying intentions and purchase probability: An experiment in survey design," *Journal of the American Statistical Association*, 61, 658-696.

Kalwani, M. U. &. Silk, A. J. (1982), "On the reliability and predictive validity of purchase intention measures," *Marketing Science*, 1, 243-286.

Lee, M., Elango, B. & Schnaars, S. P. (1997), "The accuracy of the conference board's buying plans index: A comparison of judgmental vs. extrapolation forecasting methods," *International Journal of Forecasting,* 13, 127-135.

Lehmann, D. R. (1989), *Marketing Research Analysis*, Homewood, IL: Liuin.

Manski, C. F. (1990), "The use of intentions data to predict behavior: A best case analysis," *Journal of the American Statistical Association*, 85, 934-940.

McNeil, J. M. (1974), "Federal Programs to Measure Consumer Purchase Expectations, 1946-73: A Post-Mortem," *Journal of Consumer Research*, 1 (December), 1-10.

McNeil, J. M. & Stoterau, T. L. (1967), "The Census Bureau's new survey of consumer buying expectations," in *Proceedings, 1967 Meeting of the American Statistical Association*, Business and Economic Statistics Section, 97-113.

Morrison, D. G. (1979), "Purchase intentions and purchase behavior," *Journal of Marketing*, 43, 65-74.

Morwitz, V. G., Johnson, E. & Schmittlein, D. (1993), "Does measuring intent change behavior," *Journal of Consumer Research,* 20, 46-61.

Morwitz, V. G. & Schmittlein, D. (1992), "Using segmentation to improve sales forecasts based on purchase intent: Which intenders actually buy?" *Journal of Marketing Research*, 29, 391-405.

Morwitz, V. G., Steckel, J. & Gupta, A. (1996), "When do purchase intentions predict sales?," Working Paper, Stern School of Business, New York University, NY 10012.

Mullet, G. M. & Karson, M. J. (1995), "Analysis of purchase intent scales weighted by probability of actual purchase," *Journal of Marketing Research,* 22, 93-96.

Pickering, J. F. & Isherwood, B. C. (1974), "Purchase probabilities and consumers durable buying behavior," *Journal of the Market Research Society,* 16, 203-226.

Schmittlein, D. C., Kim, J. & Morrison, D. G. (1990), "Combining forecasts: Operational adjustments to theoretically optimal rules," *Management Science*, 36 1044-1056.

Schnaars, S. P. (1984), "Situational factors affecting forecast accuracy," *Journal of Marketing Research*, 21, 290-297.

Sherman, S. J. (1980), "On the self-erasing nature of errors of prediction," *Journal of Personality and Social Psychology,* 39, 211-221.

Silk, S. J. & G. L. Urban (1978), "Pre-test market evaluation of new packaged goods: A model and measurement methodology, *Journal of Marketing Research*, 171-191.

Stapel, I. (1968), "Predictive attitudes," in Adler, L. & Crespi, I., *Attitude Research on the Rocks*, Chicago: American Marketing Association, 96-115

Theil, H. & Kosobud, R. F. (1968), "How informative are consumer buying intention surveys?" *Review of Economics and Statistics*, 50, 207-232.

Tobin, J. (1959), "On the predictive value of consumer intentions and attitudes," *Review of Economics and Statistics*, 41, 1-11.

Winkler, R. L. (1989), "Combining forecasts: A philosophical basis and some current issues," *International Journal of Forecasting*, 5, 605-609.

NOTES

1. The IFO address is Poschingerstrasse 5, 81678 Munich, Germany. Their website is www.ifo.de.

2. Note that respondents indicated their intentions to purchase in the next 6 months, but purchases were measured over a 12 month period.

3. The address for Allison Fisher, Inc. is 3000 Town Center #2300, Southfield, Michigan 48075-1192.

4. Bemmaor used δ' and δ to reflect these probabilities.

5. The parameters mean true intent, intent dispersion and % of true intenders can be estimated from intentions data if one assumes that the intentions data are distributed beta binomial (Morrison 1979; Kalwani and Silk 1982). These parameters can be estimated using the method of moments or maximum likelihood.