

Principles for Examining Predictive Validity: The Case of Information Systems Spending Forecasts

Fred Collopy

The Weatherhead School
Case Western Reserve University
Cleveland, Ohio 44106

Monica Adya

The Weatherhead School
Case Western Reserve University
Cleveland, Ohio 44106

J. Scott Armstrong

The Wharton School
University of Pennsylvania
Philadelphia, PA 19104

Research over two decades has advanced the knowledge of how to assess predictive validity. We believe this has value to information systems (IS) researchers. To demonstrate, we used a widely cited study of IS spending. In that study, price-adjusted diffusion models were proposed to explain and to forecast aggregate U.S. information systems spending. That study concluded that such models would produce more accurate forecasts than would simple linear trend extrapolation. However, one can argue that the validation procedure provided an advantage to the diffusion models. We reexamined the results using an alternative validation procedure based on three principles extracted from forecasting research: (1) use *ex ante* (out-of-sample) performance rather than the fit to the historical data, (2) use well-accepted models as a basis for comparison, and (3) use an adequate sample of forecasts. Validation using this alternative procedure did confirm the importance of the price-adjustment, but simple trend extrapolations were found to be more accurate than the price-adjusted diffusion models.

Introduction

In the past two decades, much has been learned about testing the predictive validity of models. We believe that utilization of the findings from the forecasting literature *will* lead to improved assessments of predictive validity in information systems (IS) research. To demonstrate this, we reanalyzed the models proposed in an oft-cited study about information systems spending by Gurbaxani and Mendelson (1990), which we refer to as "G&M". The G&M paper presents models that, in our opinion, pass tests of face validity. In this paper we examine the models' predictive validity.

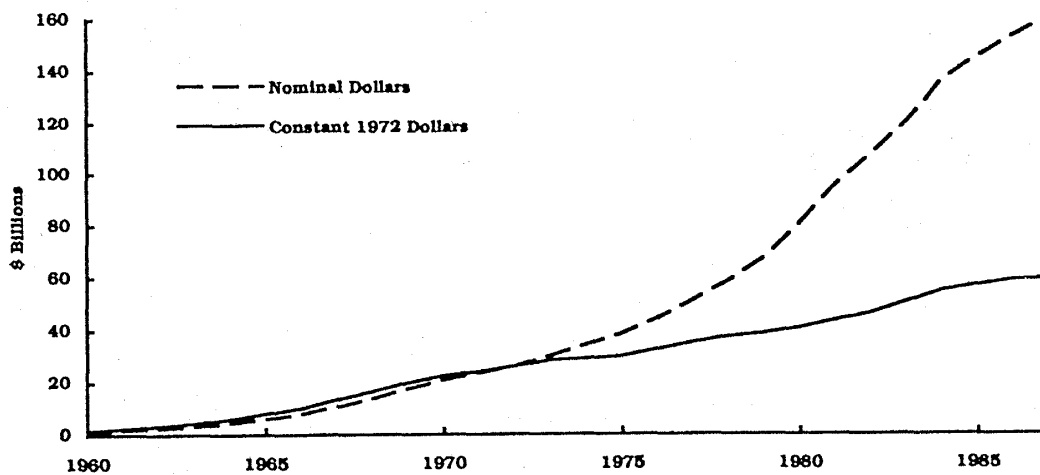
In their paper, G&M modified some traditional diffusion models to analyze aggregate spending on information systems in the United States. Their primary objective was to establish the importance of price as a determinant of IS spending. G&M is an important study because it provides empirical comparisons of alternative approaches. In our opinion, such studies represent the best in management science. Also, G&M is the most comprehensive attempt yet to quantify and to explain the growth of aggregate U.S. spending on IS. Its analysis of the role of price provides insights about changes in the historical rates of IS spending growth. Finally, given the role of IS in the economy, a model that improves our understanding of IS expenditures would seem to be of substantial importance.

In addition to using their price-adjusted diffusion models to provide causal explanations, G&M proposed that the models be used to forecast aggregate spending on information systems. They stated that their model ". . . has implications for forecasting the magnitude of IS expenditures into the future (p. 42)." In comparisons of their model with other prediction techniques, they concluded that their model provided more accurate forecasts. This study examines G&M's conclusions about the utility of their model for forecasting IS expenditures by providing an alternative analysis, using data provided by G&M and techniques drawn from the field of forecasting.

G&M's Data, Models, and Validation Procedure

G&M tested their models using a time series based on the responses of user organizations to surveys conducted by International Data Corporation (IDC). The respondents' estimated their IS expenditures, which included spending on mainframes, minicomputers, microcomputers, data entry equipment, data communications equipment, line charges, software, services, supplies, and overhead. This also included replacement, supplies, labor, and other recurring costs. The costs of users' time (outside of IS organizations) were not included. The IDC database contains information on more than 80% of the general purpose computers installed in the U.S. These data represent consumption, not adoption. Figure 1 shows this series in current dollars and in 1972 constant dollars, and Appendix A presents the data.

Figure 1
Information Systems Spending in the U.S.



G&M proposed an adaptation to diffusion models. Diffusion (or S-shaped) models posit a changing rate of growth that can depend on two opposing factors. First, as potential customers are brought into contact with the product, the extent to which the product has already been accepted exerts a positive influence on growth. Second, increases in the level of use can exert a negative influence on the growth rate because the number of potential customers decreases.

Can diffusion models be expected to perform well in forecasting the G&M data? Potential threats include a lack of empirical support for diffusion models, difficulty in estimating saturation levels, and the risks associated with their complex functional forms. There is little empirical evidence on the conditions under which diffusion models are most appropriate for forecasting. Meade's (1984) review of the empirical evidence for diffusion curves contains six studies that involve forecasting, but these studies were based on small samples and none of them provided benchmarks for comparing the performance of the diffusion models. Rao (1985) examined the forecast validity of several diffusion methods and found them to be less accurate than simpler extrapolation methods. However, the empirical evidence on this matter remains sparse.

The functional form of diffusion models is complex (Mahajan and Muller 1979, Meade 1984). Initially, the series grows slowly, then increases its growth rate at some point, and, finally, it decreases the growth rate as it approaches a saturation level. Standard diffusion models assume a constant total population of potential customers. But for many technologies there is a third segment of the diffusion process: the entry of individuals into the pool of potential customers (see Meade 1984, Mahajan et al. 1990). The population of potential customers can be expanded through a reduction in price, and G&M proposed an extension to incorporate price, arguing that "a considerable share of IS spending growth is due to price effects (p. 28)." They refer to their extended versions of the three basic diffusion models as price-adjusted diffusion models.

Diffusion models also require estimates of the saturation level. This is either provided as an input to the model or is estimated from the historical data, as part of the fitting process. An historical series often contains little evidence as to the eventual saturation level, so Martino (1983) recommends against estimating the saturation level in this way. In G&M, saturation levels are estimated from the historical data. Furthermore, because IS spending measures consumption rather than adoption, it is particularly difficult to establish a saturation level.

As we described above, diffusion models introduce complexity by their shape. Models with complex functional forms seldom improve forecast accuracy (Armstrong 1985, pp. 200-202, reviews the evidence on this issue). This apparently occurs because complex models increase the risk of extending false patterns.

G&M provide empirical support to show that their price-adjusted diffusion models yield accurate forecasts. However, one can argue that their test provided an advantage to the price-adjusted diffusion model. For each of the price-adjusted diffusion models, they used all of the data from 1960 to 1987 to produce a fitted model. They then compared the last value of the fitted model with the actual observed value for 1987 (one of the values that was used in fitting the model). The difference between the fitted value and the observed value represented the forecasting error for the price-adjusted diffusion model. For each of the alternative models (which they refer to as linear and exponential extrapolation), they obtained their parameter estimates using only two observations—the actual values for 1970 and 1980. Then they extrapolated to produce a forecast for 1987. The difference between the extrapolated value and the actual 1987 observation was used to assess the accuracy of the alternative model. The two sets of forecast errors were then compared. G&M concluded that the price-adjusted Gompertz curve provided the best forecast.

An Alternative Validation Procedure

We reexamined the results in G&M, using the same series they analyzed. We also examined the same diffusion models, including their price-adjusted diffusion models. We based the formulations on those presented in G&M. To ensure that we were using them as G&M did, we first replicated the results given in that paper. The details of the diffusion models and our application of them to produce forecasts are provided in Appendix B. Application of the models was a complex process. We think that practitioners would, in general, experience difficulty in using price-adjusted diffusion models correctly. Following the recommendation in Armstrong and Collopy (1992), we used the Median Absolute Percentage Error (MdAPE) to assess accuracy.

Empirical research has established principles that are useful in ensuring the predictive validity of time series models. To assess the forecasting validity of a proposed model: (1) use *ex ante* (out-of-sample) performance rather than the fit to the historical data, (2) use well-accepted models as a basis for comparison, and (3) use an adequate sample of forecasts. In this section, we describe how each of these principles effected the design of our validation procedure.

Use of ex ante Performance for Comparisons

The G&M models produced good fits to the historical time series, with R^2 s between .95 and .999. However, forecasting researchers have learned that the fit of a model to historical data is not a reliable indicator of the ability of that model to forecast future values of the series (for a summary, see Armstrong 1985, pp. 338-339). It is common in the forecasting field to validate models using *ex ante* (out-of-sample) forecasts, but in G&M, the diffusion models' "forecasts" are from within sample (they used the actual value for 1987 when fitting the model). The alternative extrapolation forecasts were seven-year-ahead *ex ante* forecasts (they did not use the 1987 value when fitting the model). A comparison of models should be based on having the same data available for fitting each model.

In this study we relied upon comparisons of *ex ante* forecasts from all of the methods. That is, models were fit using some of the historical data, forecasts were produced, and then the models' forecasts were compared on hold-out data (data which are not used at all in fitting the models).

Use of Well-accepted Models as a Reference

G&M recognized the need for comparisons with alternative models when they proposed comparing their price-adjusted diffusion models with linear and exponential trend models. The alternative models they chose, though, were estimated using only two observations, while their price-adjusted diffusion models were estimated using 28 observations.

The forecasting literature expresses a preference for simpler models unless a strong case has been made for complexity. The research findings indicate that relatively simple extrapolation models are robust (Armstrong 1984). We compared the diffusion models examined by G&M with simple extrapolation methods, including linear regression against time, Brown's linear exponential smoothing, damped trend exponential smoothing, and an average of the three (referred to as equal weights). These methods are relatively simple, widely used, well-documented in the literature, and have performed well in previous comparative studies. We chose them a priori. Below, we briefly describe why each of these models was chosen.

As a functional form for the simple linear regression against time, we chose an additive form. The additive form represents a more conservative choice when there is uncertainty about the proper form. In the case of IS spending, uncertainty arises because of changes in the basic (long-term) trend, and because the net effect of the underlying causal forces is not clear: price has a downward force on dollar sales while quality has an upward force.

The linear regression places equal weight upon all available historical data. For an alternative model that weights recent data more heavily, we selected exponential smoothing (Brown 1959). This approach is useful when the series is changing rapidly, as seems to be the case with IS spending.

We also used damped trend smoothing which offers another opportunity to improve the accuracy of extrapolations by reducing the magnitude of the trend component based on the variation in the historical series. This introduces some complexity, but the value of this procedure has been supported empirically (Gardner and McKenzie 1985).¹

Finally, we combined the three forecasts from simple methods by weighing them equally. Many studies have been conducted to identify the most effective combining strategies (Clemen 1989). A simple average has proven to be robust, in the sense that it is typically more accurate than the average error of the individual forecasts, and it is usually as accurate as other sets of a priori weights.

Use of an Adequate Sample of Forecasts

G&M based their conclusion about the forecast utility of the price-adjusted diffusion models on a single forecast. The use of successive updating provides a larger sample size for comparing the performance of the alternative methods. We first estimated the parameters of each model using the data from 1960 through 1974, then produced forecasts for ten years (1975-1984) and calculated forecast errors. The next observation, 1975, was then added to the historical data, the model parameters were re-estimated, forecasts were made, and errors were calculated. The procedure was repeated until the estimation data included all but the last observation. This produced 13 one-year-ahead forecasts, 12 two-year-ahead forecasts, and so forth, up to 4 ten-year-ahead forecasts. In all, there were 85 forecasts and 13 starting points for each method.

Results

The random walk (which simply extends the latest observation and assumes no change) was used as a benchmark for comparisons. The average MdAPE across all ten horizons for the three basic diffusion models was 21.3. This is about 20% less than the error from benchmark random walk forecasts. The logistic, which Martino's (1983) guidelines suggest would be the most appropriate of the diffusion models, did not perform well. Its average MdAPE was 31.9, which was 18% larger than that from the random walk.

¹ The damped trend forecasts were made using the Autocast II program.

The price adjustment improved the accuracy of each of the three diffusion models. Overall the MdAPE was reduced from 21.3 to 9.6. The price adjustment was particularly effective when applied to the logistic, resulting in a MdAPE of 6.0. For 194 of the 255 comparisons, the price-adjusted forecast had a lower error than the standard model. A paired sign test was significant at $p < 0.0001$. (There is likely some correlation among the errors, so this result overstates the significance). This supports G&M's position that a price parameter is an important addition to standard diffusion models not only for explanatory purposes, but also for prediction.

G&M concluded that their price-adjusted diffusion models produced better forecasts than did a simple linear trend extrapolation. However, when subjected to the above-described validation procedure, they did not. The linear regression against time had an overall MdAPE of 4.5, lower than that for any of the price-adjusted diffusion models.

Table 1 presents the MdAPEs for each method for each of the ten years of the forecast horizon. It also provides the unweighted averages across the ten years.

Table 1
Median Absolute Percentage Errors of Alternative Extrapolation Strategies

	Forecast Horizon										Average 1 to10 (85)
	1 (13)	2 (12)	3 (11)	4 (10)	5 (9)	6 (8)	7 (7)	8 (6)	9 (5)	10 (4)	
Random Walk	5.6	11.2	16.4	21.8	25.4	30.2	34.0	37.5	42.4	45.6	27.0
Logistic	19.4	21.6	23.7	28.0	32.1	34.6	37.1	39.3	41.3	41.6	31.9
Gumpertz	8.4	9.0	12.1	12.0	13.5	14.7	16.7	16.9	17.0	16.0	13.6
Modified Exponential	7.9	9.7	12.0	14.6	18.2	20.2	22.4	24.5	26.1	26.9	18.3
Avg Diffusion	11.9	13.4	15.9	18.2	21.3	23.2	25.4	26.9	28.2	28.1	21.3
Price-Adjusted Logistic	5.3	5.3	5.2	5.4	5.3	5.1	5.5	6.7	7.9	7.9	6.0
Price-Adjusted Gompertz	6.1	6.5	8.5	12.0	13.6	14.8	16.7	16.9	17.0	16.0	12.8
Price-Adjusted Mod Exp	3.2	4.9	4.8	6.2	8.2	10.6	11.8	13.7	17.4	18.6	9.9
Avg Price-Adjusted	4.9	5.6	6.2	7.9	9.1	10.2	11.3	12.4	14.1	14.2	9.6
Regression	2.9	2.9	3.2	3.9	4.5	5.3	4.7	5.5	6.4	5.3	4.5
Brown's Lin Exp Smooth	1.8	4.3	6.5	7.6	5.1	7.7	6.0	11.9	2.3	18.1	7.1
Damped Trend Exp Smooth	2.0	4.1	7.0	7.9	5.3	10.3	9.5	15.6	7.0	19.3	8.8
Avg Simple Extrapolation	2.2	3.7	5.6	6.5	5.0	7.8	6.8	11.0	5.2	14.3	6.8
Equal Weights	1.4	2.9	5.9	6.3	4.3	8.5	7.5	11.0	4.4	13.6	6.6

Number of forecasts in parentheses.

Brown's exponential smoothing and damped-trend exponential smoothing were relatively accurate. Their MdAPEs, averaged across the ten horizons, were 7.1 and 8.8, respectively. The simple trend extrapolation methods, with their average MdAPE of 6.8, were more accurate than the price-adjusted diffusion models, with their average MdAPE of 9.6.

An equally-weighted combination of the three simple trend extrapolations was slightly more accurate than the average of the components (MdAPE of 6.6 versus 6.8). Across all horizons, the errors from combining forecasts were about 3% less than the errors of the average component. The equally-weighted combination tended to be more useful for the short-range forecasts; for forecast horizons of up through five years, it reduced the error by about 10%, whereas for horizons of six through ten years, there was no error reduction.

The accuracy of the typical price-adjusted diffusion was not superior to that of the equally-weighted combined forecast models. On average, its error was about 45% larger.

The price adjusted logistic was the most accurate of the diffusion models, but there was no a priori basis for selecting this model in preference to the other price-adjusted models. Its error was slightly less than that for the equally-weighted combination of the simple methods.

Conclusions

We believe that G&M's models shed light on why simple S-shaped curves provide inadequate fits to the historical IS spending data and provide insight into the effects of price. G&M's price-adjusted models proved superior to traditional diffusion models, supporting their claim for the importance of price as a determinant of IS spending growth. However, their claim that such models would prove more accurate than simple and widely-used extrapolation methods was not supported.

It appears to be a commonly held view that spending on information systems is growing in an explosive fashion. The same view has been held at other times. If experience is any indication, one of the implications of the analysis presented here is that we are more likely to be close to true future values of spending on information systems if we assume that they will continue to grow in a simple linear fashion.

We propose that IS researchers who wish to validate their models for making forecasts should rely upon *ex ante* comparisons, make comparisons with well-established, reasonable alternatives, and use adequate samples. Given the evidence, we recommend that IS researchers and practitioners not use diffusion models for forecasting IS spending.

Acknowledgements. Gadi Ariav, Katherine Auer, Dick Boland, Chris Chatfield, Joshua Eliashberg, Katrine Kirk, Henry Lucas, Joseph Martino, Nigel Meade, Betty Vandenbosch, the editors, and several anonymous reviewers provided useful comments on earlier drafts of this article. Suzanne Berman and Vanessa Lacoss provided editorial assistance.

Appendix A
Information Systems Spending in the U.S. (\$ billions)

Year	Nominal	Constant 1972
1960	1.25	1.77
1961	1.81	2.53
1962	2.47	3.41
1963	3.35	4.57
1964	4.58	6.16
1965	6.11	8.11
1966	7.96	10.27
1967	10.81	13.51
1968	13.88	16.64
1969	17.54	19.99
1970	20.95	22.57
1971	23.54	24.23
1972	26.28	26.28
1973	30.15	28.38
1974	34.41	29.10
1975	38.68	30.01
1976	44.38	32.64
1977	51.38	35.32
1978	58.72	37.56
1979	67.56	38.75
1980	80.54	40.86
1981	95.22	43.76
1982	106.87	46.36
1983	120.42	50.47
1984	136.31	54.62
1985	145.45	56.68
1986	153.42	58.66
1987	160.76	59.10

Appendix B Description of Application of the Diffusion Models

We replicated and validated the six diffusion models considered by G&M. These were:

Gompertz	$B_t = KA^b$
Logistic	$B_t = 1/(K + Ab^t)$
Modified-Exponential	$B_t = e^{a-(b/t)}$
Price-Adjusted Gompertz	$B_t = KA^b e^{\lambda t}$
Price-Adjusted Logistic	$B_t = (1/(K + Ab^t))e^{\lambda t}$
Price-Adjusted Modified-Exponential	$B_t = e^{a-(b/t)+\lambda t}$

where

t = year,

B_t = real level of total IS spending in year t ,

$K, A, a,$ and b = parameters to be estimated,

$e^{\lambda t}$ = price adjustment to the basic diffusion models.

The price-adjusted models were the basic diffusion models multiplied by a function to account for the price adjustment. Because the price-adjusted models would have been difficult to estimate in their existing form, GRcM transformed them to the estimation equations:

Price-Adjusted Gompertz	$\log(B_t) = \log K + \lambda t + b^t \log A$
Price-Adjusted Logistic	$1/B_t = (K + Ab^t)e^{\lambda t}$
Price-Adjusted Modified-Exponential	$\log(B_t) = a - (b/t) + \lambda t$

All six models were estimated using the nonlinear regression procedure of SYSTAT 5.0. Specifically, we used the Simplex approach to estimate for all but the price-adjusted logistic model. In performing these analyses, our effort was to keep the estimation process as consistent as possible across all six methods. Another concern was to introduce as little external information as possible to the process. Details of estimating the six models follow.

Modified-Exponential Model

Both the original and a log-transformed version of this model were estimated. However, the transformed model did not yield reasonable parameter estimates or forecasts. Hence, we used the original model in the analysis.

Gompertz Model

The original version of the Gompertz model did not converge to an optimal solution. The log-transformed version of the model converged and produced reasonable parameters and forecasts.

Logistic Model

There were no convergence problems with either the original or log-transformed versions of the logistic model. However, the forecasts generated using the original model did not appear reasonable. Results from using the transformed model were much more reasonable. Consequently, our analysis was based on this version.

Price-Adjusted Modified-Exponential Model

G&M estimated parameter values using the log-transformed version of the price-adjusted modified-exponential model. When we used this version, we were able to replicate their results. However, some of the forecasts generated using this model looked unreasonable. We were able to produce much more reasonable forecasts using the original version of the model.

Price-Adjusted Gompertz Model

The original formulation of the price-adjusted Gompertz model did not converge even when given initial starting values. We were able to use the log-transformed version of the model to produce reasonable forecasts when we added the constraint that g be nonnegative.

Price-Adjusted Logistic Model

The original version of the price-adjusted logistic model failed to converge. We were unable to replicate G&M without providing their results as initial values. Even then some of the forecasts did not look reasonable. We changed the estimation technique to use the Quasi-Newton technique. This permitted us to replication of G&M's results and yielded forecasts that look reasonable.

References

- Armstrong, J. S. (1984), "Forecasting by extrapolation: Conclusions from 25 years of research," *Interfaces*, 14 (Nov.-Dec.), 52-66.
- Armstrong, J. S. (1985), *Long-Range Forecasting*. John Wiley, New York.
- Armstrong, J. S., and F. Collopy (1992), "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, 8, 69-80.
- Brown, R. G. (1959), *Statistical Forecasting for Inventory Control*, McGraw-Hill, New York.
- Clemen, R. T. (1989), "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5, 559-583.
- Gardner, E. S., Jr. and E. McKenzie (1985), "Forecasting trends in time series," *Management Science*. 31, 1237-1246.
- Gurbaxani, V. and H. Mendelson (1990), "An integrative model of information systems spending growth," *Information Systems Research*, 1, 23-46.
- Mahajan, V. and E. Muller (1979), "Innovation diffusion and new-product growth models in marketing," *Journal of Marketing*, 43, 55-68.
- Mahajan, V., E. Muller and F. Bass (1990), "New product diffusion models in marketing: A review and directions for research," *Journal of Marketing*, 54, 1-26.
- Martino, J. (1983), *Technological Forecasting for Decision Making*. North Holland, New York.
- Meade, N. (1984), "The use of growth curves in forecasting market development: A review and appraisal," *Journal of Forecasting*, 3, 429-451.
- Rao, S. (1985), "An empirical comparison of sales forecasting models," *Journal of Product Innovation Management*, 4, 232-242.