Taylor & Francis
Taylor & Francis Group

Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: A Business Value Approach

Author(s): Arun Bansal, Robert J. Kauffman and Rob R. Weitz

Source: *Journal of Management Information Systems*, Summer, 1993, Vol. 10, No. 1 (Summer, 1993), pp. 11-32

Published by: Taylor & Francis, Ltd.

Stable URL: https://www.jstor.org/stable/40398029

JSTOR

Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: A Business Value Approach

ARUN BANSAL, ROBERT J. KAUFFMAN, AND ROB R. WEITZ

ARUN BANSAL is currently Senior Analyst at Bear Stearns and Company, New York, where he is involved in management science modeling and applications of information technologies to support the firm's operations in the financial markets. He holds a Ph.D. from the Stern School of Business, New York University, where he specialized in research at the intersection of finance and information systems. Portions of his doctoral dissertation, which focused on data quality and systems design issues in the context of mortgage-backed securities portfolio management, have been published in *Information and Management*, and here. The work was sponsored by the Capital Markets Sector, Manufacturers Hanover Trust, New York.

ROBERT J. KAUFFMAN. See the Guest Editors' Introduction for biographical information.

ROB R. WEITZ is Associate Professor in the Department of Computing and Decision Sciences, Stillman School of Business, Seton Hall University. He has previously served on the faculty of INSEAD, the European Institute of Business Administration, and has been a Visiting Professor of Information Systems at the Stern School of Business, New York University. He received his Ph.D. in operations research from the University of Massachusetts at Amherst. His research interests focus on the areas of forecasting, applied artificial intelligence, and decision support. His publications have appeared in *Decision Sciences*, *Information and Management*, *AI Magazine*, and *The International Journal of Forecasting*.

ABSTRACT: Under circumstances where data quality may vary (due to inaccuracies or lack of timeliness, for example), knowledge about the potential performance of alternate predictive models can help a decision maker to design a business-value-maximizing information system. This paper examines a real-world example from the field of finance to illustrate a comparison of alternative modeling tools. Two modeling alternatives are used in this example: regression analysis and neural network analysis.

There are two main results: (1) Linear regression outperformed neural nets in terms of forecasting accuracy, but the opposite was true when we considered the business value of the forecast. (2) Neural net–based forecasts tended to be more robust than linear regression forecasts as data accuracy degraded. Managerial implications for financial risk management of mortgage-backed security portfolios are drawn from the results.

KEY WORDS AND PHRASES: business value of information technology, data quality, decision support systems, forecasting, information economics, neural networks, mortgage-backed securities, prepayment forecasting, risk management forecasting systems.

1. Introduction

WHEN DATA QUALITY VARIES, KNOWLEDGE ABOUT THE POTENTIAL PERFORMANCE of alternate predictive models can help a decision maker to design an appropriate information system in terms of predictive accuracy and payoff considerations. Model performance comparisons often presume perfect data—a presumption arguably more appropriate to textbook examples than to real-world problems. This research empirically examines the effects of data inaccuracy on the performance of two alternate forecasting frameworks: regression analysis and neural network analysis. A comparison can enable a decision maker to select the model that is least sensitive to predictive degradation in the range of observed data quality variation.

The application selected is a risk management problem [15] associated with the forecasting of prepayment rates[1] in *mortgage-backed securities* (MBS) portfolio management. Forecasting prepayments requires large data sets, which are available via commercial sources. In this work, model performance is evaluated based on a traditional accuracy metric, R^2, and on a payoff measure developed as part of this research. The payoff measure enables evaluation of trade-offs between the business value of improved decisions resulting from the use of more accurate data, and of the cost of obtaining such data.

This paper is organized as follows. Section 2 examines how empirical research in a variety of disciplines has dealt with data accuracy, and the kinds of conclusions that have been reached regarding the relationship between data accuracy and forecast performance. We also review research that compares the performance of regression and neural network analysis. Section 3 introduces the application, MBS portfolio management prepayment analysis. Section 4 presents linear regression and neural net models for forecasting prepayments. Econometric results provide a basis for comparing their forecasting performance. Section 5 discusses the forecasting systems' accuracy and payoffs, and then develops a specific metric for payoff in the context of MBS prepayment forecasting, based on hedge position creation. Section 6 presents the simulation and evaluative methods that are used to examine how forecast performance and payoff change as data become inaccurate. Section 7 presents and discusses the main results of the paper. Section 8 concludes with a discussion of the managerial significance of the results.

2. Data Quality and Forecasting Alternatives

DATA QUALITY CAN BE DESCRIBED IN TERMS OF A NUMBER OF DIMENSIONS, including frequency, accuracy, and response time. In general, a model performs better the more frequent and the more accurate the data, and the shorter the time to obtain it.

2.1. Information Economics, Accounting and Finance Research

A number of researchers have explored the effects of data quality in different decision-making situations, using modeling approaches that are based on information economics [1, 9, 14, 19]. An information economics model values information by comparing the business value or payoff from decisions made in the presence of information produced by an information system, less the business value of the same decisions made in the absence of the information, and then further subtracting the costs of the system and for obtaining the data [14]. Clearly, the higher this *payoff* that a user can derive from use of an information system, the better is the performance of the system. If payoff is based on forecast performance, an important potential determinant of the business value of the forecasting system is the quality of the data on which it relies.

Ballou and Pazer [6] showed how different aspects of data quality can be combined to yield a simple measure, representing the overall inaccuracy in the data. Other insights into the problem of optimizing data quality in various decision settings have been developed in the accounting literature, too. For example, Cushing [13] examined the effects of propagation of errors in internal control systems. And Ijiri and Itami [20] studied the effects of inaccuracies in demand estimation on a firm's profitability.

A stream of research in finance emphasizes the importance of high-quality data to make accurate predictions about future financial performance. For example, Beaver [10] emphasized the importance of data frequency in prediction of corporate bankruptcy. Later, Meyer and Pifer [26] learned that prediction of commercial bank failures was highly sensitive to the time frame from which the forecast data were drawn. Martin [25], whose work examined alternative types of early-warning models for predicting bank failures, recognized the importance of adequate sample size to ensure useful predictions [3]. Altman's [2] savings and loan industry bankruptcy forecasting system, though widely noted by the industry and the regulators, has been criticized for being developed from a data set representing too narrow a spectrum of institutions.

Although the above-cited work recognizes that data quality is important, the nature of the relationship between data quality degradation and the performance of the forecasting systems remains largely uncharted territory. In recent work we attempted to examine how such data quality variations (especially frequency and accuracy) affect the predictive accuracy and business value of a financial risk management forecasting system [7].

2.2. Alternative Forecasting Models: Regression and Neural Networks

Linear and nonlinear regression are well-known, widely used forecasting approaches. Given a suitably large historical data set upon which to base a forecast, the normal approach is to split the sample into a "training set" and a "test set" or "holdout" sample. The training set is used to develop the model, while the test set is used to examine the performance of the model in forecasting.

Neural networks can closely approximate linear and nonlinear regression. Comparison of neural nets with more traditional statistical techniques (various forms of regression, discriminant analysis, etc.) has been the focus of many recent studies [16, 31, 36, 40]. Some of the main differences between regression and neural networks are evident from these studies. First, neural nets consistently improve during training, if the neuron connection weights are adjusted each time mistakes occur. Regression techniques, on the other hand, process all training data simultaneously, before using new data.[2] Second, Lippmann [24] has suggested that, in theory, neural nets may be more robust than nonlinear classification models. Third, while regression models make it difficult to determine the right set of independent variables [16], the absence of direct links in many practical neural nets makes it impossible to determine which inputs affect the outputs directly. Therefore, unlike regression analysis, where the estimated coefficients enable the analyst readily to assess the effects of incremental changes in the *value* of an independent variable, with neural nets the focus shifts to the effects of incremental changes in the *number* of variables. Fourth, regression equations require model specification in advance. In nonlinear regression, specifying the exact nature of the nonlinearity may be a burdensome task [39]. Modeling with neural networks avoids model specification in the regression sense entirely.

Although they require no assumptions regarding distributions or dispersion of the data, however, neural nets do require specification of a network architecture; while guidance is available in the literature, the process still requires judgment, and trial and error. Neural nets typically have been shown to produce more accurate predictions with good-quality data than regression models, but the literature does not indicate empirically how this comparison might be altered if the data were inaccurate.

3. MBS Prepayment Forecasting

THERE ARE THREE MARKETS FOR MORTGAGE-RELATED FINANCIAL INSTRUMENTS: an origination market, a primary market and a secondary market. Lending institutions, such as banks, S&Ls, and mortgage lenders, deal directly with home buyers and borrowers in the *origination market*. Borrowers are obliged to repay principal and accrued interest to the lender. The lender, in turn, is interested in freeing up capital and spreading risk. This goal can be accomplished by selling the loans to the General National Mortgage Association (GNMA), the Federal National Mortgage Association (FNMA), and the Federal Home Loan Mortgage Corporation (FHLMC). The agencies purchase bank mortgage portfolios or individual loans through a process called *securitization*. The securities created in this way are called *mortgage-backed securities*

(MBS), and are sold by the agencies to investors in the *primary market*. Afterwards, they trade freely in the secondary market just like any other financial instrument. The lenders or the servicing agencies collect payments from customers and, after deducting fees, pass them to the current holder of the MBS [8, 18].

Although MBS are fixed-income securities, the home buyer or borrower retains an option to prepay the loan at any time. This makes MBS portfolios risky. Prepayments occur as interest rates fall; because the holder of an MBS can only invest at a lower interest rate, this will lead to a loss. Knowledge of the percentage of total customers that will prepay at a given time and an ability to predict future interest rates are essential for assessing MBS risk [29].

Information technology specialists in the financial services industry refer to the information systems that monitor and measure risk associated with holding financial instruments as *financial risk management systems* [35]. The cost of building these systems is high. In large money center banks, a typical financial risk management system may cost on the order of $10 million or more. Moreover, the required maintenance and periodic enhancements to keep up with changes in the industry can add a significant amount to this figure each year [34]. Data represent an important and recurring additional cost.

Examples of digital data sources relevant to a money market bank's capital markets functions include the Chicago Board Options Exchange, and the New York Stock Exchange, which are provided as services from firms such as Reuters, Telerate, ILX, Nikkei, and Dow Jones [5, 33]. These "quote vendors" consolidate data from the exchanges, from central banks worldwide, and from other governmental or private sector sources, and repackage it for digital transmission. In other cases, financial firms purchase large databases for infrequent or customized analyses. Selection of data vendor services, along with the models that use the data, usually involves a clear trade-off between decision quality and the costs associated with obtaining high-quality data. Thus, senior management should treat the design of an *effective* financial risk management system in the same manner as it crafts policies to secure firm profitability in risky markets: from the perspective of expected cost and benefit analysis. Practitioners emphasize the importance of making decisions about how to select a level of data quality that is optimal at the time that overall trading platform design decisions are made [23].

4. Forecasting Model, Data, and Base Case Results

FORECASTS OF MBS PREPAYMENT RATES often are made using regression models. These models estimate future prepayments based on current and past values of macroeconomic and investment-specific variables. The current standards are the models of the Public Securities Association and the Federal Housing Administration. The approach is simple: the prepayment rate growth during the first two and a half years of the mortgage is assumed to be constant, until the 6 percent prepayment level is reached; thereafter it is assumed to be fixed. The model is also static: it omits consideration of the interest rate environment [12], and other time-variant factors

considered in academic research, such as economic growth, the age of the mortgage, and so on [8, 17, 18, 27, 30], that have been shown to be useful predictors of prepayments. Claims for MBS prepayment forecast accuracy in terms of model R^2s as high as 99 percent have been reported [21], but observers recognize that incorporation of the primary factors that drive prepayments normally yields values of R^2 in the range of 60 percent to 80 percent. Only in cases where the model is "tailored" to a data set will it deliver such high explanatory power.

The mortgage prepayment rate in such models typically is expressed in terms of the *constant annual prepayment percentage* (CAPP), *which is a single-year prepayment rate for a given mortgage pool. A mortgage pool* is a group of mortgage loans that have similar characteristics, for example, the same maturity, the same coupon, and the same average prepayment rates.

4.1. Forecasting Model and Data

The model that we will employ as a basis for comparing the forecasting performance of linear regression and neural nets in the presence of data quality degradation is shown below. Managers at a major bank suggested that the model is representative of models that MBS portfolio analysts use to predict prepayment rates [4, 18], even though it is not as sophisticated as the proprietary models that firms utilize in portfolio management today.

$$CAPP_{pt} = \beta_0 + \beta_1\, MAT_p + \beta_2\, SPREAD_{pt} + \beta_3\, GXNP_t + \beta_4\, CONSENT_t + \beta_5\, TYPE_p$$
$$+ \beta_6\, SMALL_{pt} + \beta_7\, AGE_{pt} + \beta_8\, RATIO_t + \varepsilon_{pt}.$$

The investment-specific variables in the model are as follows:

$CAPP_{pt}$ = Prepayment rate expressed in annual percent for pool p at time t.

MAT_p = Maturity class, a qualitative variable with the value 1 if the MBS matures in 30 years, and 0 otherwise.

$SPREAD_{pt}$ = Difference between the coupon of pool p and the prevailing market rates for a similar mortgage at time t.

$TYPE_p$ = Type of security, a qualitative variable with the value 1 if the security type is GNMA, and 0 if it is FHLMC. (Note: our data set does not include any observations from FNMA, so this category need not be specified as a variable.)

$SMALL_{pt}$ = Small spread size, a qualitative variable with the value 1 if the spread exceeds 2 percent for pool p at time t, and 0 otherwise.

AGE_{pt} = Age of mortgage pool p at time t in years.

The macroeconomic variables included in the model are:

$GXNP_t$ = Annualized percentage change in GNP at time t.

$CONSENT_t$ = Consumer sentiment at time t on a scale of 0 to 100.

$RATIO_t$ = Ratio of personal income to expenditure for the United States at time t.

The βs represent linear regression coefficients for independent variable i in the prepayment forecasting model, and ε_{pt} represents normally distributed, zero mean residuals for pool p at time t.

The data for the study were gathered from two principal sources: the Capital Markets Sector of Manufacturers Hanover Trust Company, a large money center bank in New York City; CITIBASE, a widely available electronic database that provides data on macroeconomic indicators; and other published sources, such as the Salomon Bros. MBS Prepayment Profile reports [32]. The data set contained 1,170 monthly observations of numerous variables on thirty-eight MBS, including the nine variables included in our models. The time span of the data was from April 1987 to February 1990.[3]

4.2. Base Case Prepayment Estimation Results

We now report on the estimation of two prepayment models using this data set: a linear regression model and a neural net. The results were developed using a reference set of data, representing the highest-quality data available. We randomly divided the complete data set of 1,170 observations into two parts, each containing 585 observations. The *training data set* was used for evaluating alternative model specifications. The *test data set* was held out from our model specification process, and later used to test the performance of the model. Predictive performance, in terms of R^2 and "payoff," was evaluated using the test data.

Linear Regression Model Results

The following estimates for the coefficients were obtained using a statistical package (SAS) and the training data set.[4]

$$CAPP_{pt} = 186.655 \;-\; 1.041\,MAT_p + 2.363\,SPREAD_{pt} - 0.664\,GXNP_t + 0.115\,CONSENT_t$$
$$(9.03; 0.001)\;(-1.96; 0.051)\;(14.72; 0.001)\quad(-3.99; 0.001)\quad(1.70; 0.090)$$

$$+ -1.912\,TYPE_p + 6.109\,SMALL_{pt} + 0.542\,TERM_{pt} - 171.077\,RATIO_T.$$
$$(3.66; 0.001)\quad(6.80; 0.001)\quad(8.60; 0.001)\quad(-9.85; 0.001)$$

Note: The (fitted) R^2 for the model is 70.7 percent; corrected R^2 is 70.3 percent; the number of observations is 585. The numbers in parentheses under the coefficients are the coefficients' t statistics and their corresponding significance levels.

This model provides estimates of prepayment rates given the best available data, and acts as a base case for further comparisons we will make. Senior managers at our research site indicated that they expected variations in both the predictive accuracy and payoffs from the use of increasingly inaccurate data.

Neural Network Results

The input data employed for the neural net were the same as those used for linear regression. A back propagation neural net model with three layers—an input layer, an

output layer, and a single hidden layer—was used. Eight input nodes, corresponding to the eight independent variables, were created. One output node was required for predicting values of the prepayment rate. One hidden layer was appropriate, given the empirically established relationships between the input data and the prepayment rate, evidence from prior research [16, 24] and the results of trials with this data. The number of nodes in the hidden layer was set at five, using guidance from the literature, and trial and error.[5] We used a commercially available package called Neuralware Professional.[6] The predictive accuracy of the forecast, stated in terms of R^2, was 67.9 percent. The structure of the neural net is depicted in figure 1.

5. Alternate Evaluation Approaches for MBS Prepayment Forecasting Systems

WE NEXT CONSIDER TWO MECHANISMS FOR MEASURING THE PERFORMANCE of an MBS prepayment information system, as a means to discover the impact of data quality degradation. After briefly considering why two kinds of measures are appropriate, we proceed to develop a "payoff" measure, representing the business value to an MBS portfolio manager from use of a forecasting system.

5.1. Measuring Accuracy and Business Value

The *predictive accuracy* of a forecasting system can be measured by several different statistical indicators. Here we use a common metric, R^2, which reports the percentage of the variation of the dependent variable "explained by" the independent variables.

Although R^2 is a good measure of fit between the predicted value and the actual value of the dependent variable in a forecasting model, this summary value alone does not fully describe the relationship between data quality and the *business value* of an MBS prepayment forecasting system to a portfolio manager. Changes in R^2 accompanying degradation of data quality may not correspond very well with the standard units used to measure risk and reward. It is possible, for example, that a small change in the R^2 estimate of a financial indicator may lead to large portfolio losses. More importantly, however, the units are different from those that portfolio managers worry about. As a result, we construct a *payoff measure* in more appropriate units.

5.2. Payoff Measure

The key risk that the MBS portfolio manager faces is the extent to which interest income on investable capital will decrease if future mortgage prepayments are unexpectedly large. This frees up cash, but not at a time that is beneficial to a fund manager; when interest rates fall, triggering prepayments, the cash can only be invested at the prevailing lower rates.

With knowledge of the risks associated with a specific MBS, the portfolio manager can devise combinations of financial instruments to reduce the overall risk of the position to an acceptable level. The standard technique of reducing such risk is called
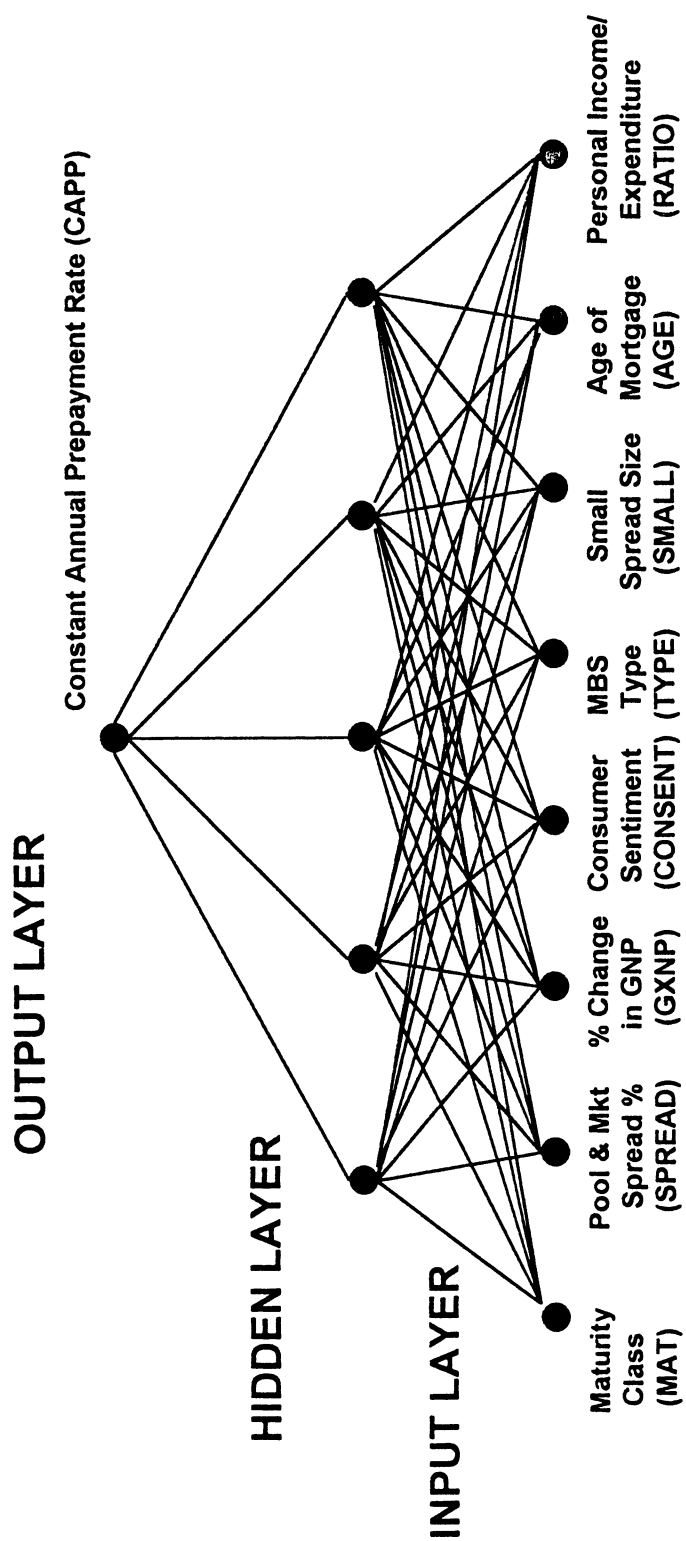
19

Figure 1. Neural Network Architecture and Training Data Results (Note: This architecture, with 8 input and 5 hidden nodes, produced an R^2 of 67.9 percent with training data)

"hedging." In hedging, the portfolio manager buys (or is "long" in) some positions and simultaneously sells (or is "short" in) others. To understand how such decisions are made based on information from an MBS prepayment forecasting system, we must consider finance concepts that apply to fixed-income investments: effective duration, an efficient hedge, and the hedge ratio.

Effective Duration

To monitor the performance of MBS, portfolio managers use a measure called *effective duration*. Effective duration calculates the risk associated with investments in MBSs. Its units are stated in terms of the percent elasticity of the instrument's price with respect to a 1 percent change in yield, and it is given by:

$$EFFECTIVE\ DURATION = \frac{PRICE_{UP} - PRICE_{DOWN}}{(PRICE_{INITIAL})\ (YIELD_{PLUS} - YIELD_{MINUS})}.$$

The variables are defined as follows:

$PRICE_{INITIAL}$ = Initial price of the financial instrument.

$PRICE_{DOWN}$ = Price if yield percentage goes down by x basis points (one basis point equals 0.01 percent).

$PRICE_{UP}$ = Price if yield percentage goes up by x basis points.

$YIELD_{MINUS}$ = Initial yield percentage minus x basis points.

$YIELD_{PLUS}$ = Initial yield percentage plus x basis points.

The prices in this equation are the present values of cash flows generated by the MBS during its lifetime. Because prepayments affect the cash flows, they will affect the prices and the yield, and hence, the effective duration [37]. In general, as the market interest rate decreases, the rate of prepayment increases: more and more borrowers find refinancing economically advantageous. Thus, MBS carry *interest rate risk*, the risk arising from interest rate fluctuations.

Efficient Hedge Design for Interest Rate Risk

To maximize return on invested capital, MBS portfolio managers attempt to create positions that yield the highest return for a specified level of risk. An *efficient hedge* enables a portfolio manager to make an investment to maximize expected return in the event that interest rates fall dramatically.

Two factors are important in designing efficient hedges in this context:

1. The types of instruments selected for the hedge; and,
2. The number of hedge securities that should be sold short for each MBS that is held.

Consider an MBS that is being hedged against a thirty-year U.S. Treasury bond—

whose price movements are inversely correlated with the MBS—with an effective duration of 8 percent.[7] Thus, the problem is to determine the ratio of the number of MBS held in the portfolio for each Treasury bond that is to be sold short. This is called the *hedge ratio*, and is defined as follows:

$$HEDGE\ RATIO = \frac{RISK_{PERMISSIBLE} + DUR_{TREASURY}}{DUR_{MBS}}.$$

The variables are defined as:

$RISK_{PERMISSIBLE}$ = Total allowable risk for the hedge, stated as expected percent loss in dollars per hundred at risk.

DUR_{MBS} = The effective duration of the MBS in percent.

$DUR_{TREASURY}$ = The effective duration of the thirty-year treasury bond in percent.

Based on discussions with senior managers at the field study site, values of $RISK_{PERMISSIBLE}$ in the area of 6 percent are reasonable in this application. This indicates that the maximum percentage loss of the total investment that the portfolio manager is willing to accept is 6 percent. Assuming that the value of the effective duration of a thirty-year Treasury bond can be estimated with some measure of confidence, this equation shows that the hedge ratio is dependent on an estimate of the effective duration for the MBS. Hence, predictions from the MBS prepayment forecasting system will affect the overall hedge ratio and the manner in which the portfolio manager creates positions to avoid unnecessary risk. Still, there is always a chance that the hedge ratio will be misspecified because the effective duration of the MBS has been estimated incorrectly.

Deviation In Hedge Ratio (DIHR)

The percent *deviation in hedge ratio* from the hedge ratio that would result in selection of an efficient hedge—what we will hereafter call DIHR—is given by:

$$DIHR = \frac{|(HEDGE\ RATIO_{INACCURATE} - HEDGE\ RATIO_{ACCURATE})|}{HEDGE\ RATIO_{ACCURATE}} * 100.$$

The variables are defined as follows:

$HEDGE\ RATIO_{INACCURATE}$ = Hedge ratio based on the inaccurate data set k used for predictions.

$HEDGE\ RATIO_{ACCURATE}$ = Hedge ratio based on accurate data.

Application of DIHR provides a mechanism by which to measure the business value of an MBS prepayment forecasting system. The investment strategy in this case would

be to sell Treasury bonds short, and continue to be long in the MBS. Thus, an effective risk management strategy here involves balancing the losses from the MBS portfolio that occur when interest rates move in an unfavorable direction and the gains from the securities that form the hedge. We provide a numerical illustration in figure 2 to assist the reader's understanding.

6. Experimental Design of the Simulation

MANY SOURCES OF INACCURACIES ARE POSSIBLE IN A DATA SET. A few examples are: an operator's typing mistake, the imprecise measurement of subjective data (e.g., consumer sentiment in our current example) or the lack of proper updates (e.g., data on most macroeconomic indicators are difficult to update on a day-to-day or month-to-month basis and thus often are interpolated). For this study we assume that most types of inaccuracies only affect a part, and not the complete data set. For example, typing errors usually appear in just a fraction of a data set. Similarly, imprecision in subjective data and lack of proper updates may also affect only a part of the data set. The inaccuracies are not persistent, and are best described as a zero mean "white noise" process.

To reflect this assumption in an empirical test, we simulated different levels of data inaccuracy by varying a *fraction* of all actual data values by some *amount* of error compared to the original values. We selected three different *fractions* of the observations in the data set: 4 percent, 8 percent, and 12 percent. We also selected four different *amounts* of error: 5 percent, 10 percent, 15 percent, and 20 percent. As white noise errors, the simulated *amount-errors* are equally likely to be +4 percent or –4 percent, or +8 percent or –8 percent, or +12 percent or –12 percent, depending upon the cell of the experimental design. Our selection of these ranges is meant to cover two types of errors that often occur in portfolio management for MBS: errors in subjective measures and errors due to lack of updates. The ranges of the potential errors are based on interviews conducted in our field study. (The reader should bear in mind that the larger the potential error, the more likely the bank's auditors will be to discontinue use of the digital data source altogether.)

We now describe the elements of the simulation experiment, focusing on the operational definitions of *fraction-error* and *amount-error*, the random number-generation process that enabled us to create our data sets, and the results of the simulation experiments.

6.1. Fraction-error and Amount-error

Results were obtained via an experiment [22], using data sets in which data inaccuracy is simulated. This was accomplished in the following way:

1. Data items in the test data set were randomly selected to be perturbed by a fixed amount; and

A Four-Step Business Valuation Procedure for an MBS Prepayment Forecast System

Measuring business value of an MBS prepayment forecasting system involves four steps:

Step 1.   *Estimate effective duration of the MBS, based on information from the MBS prepayment forecasting system.*

Step 2.   *Apply estimated value of effective duration to determine the hedge ratio in the target securities for the MBS position held.*

Step 3.   *Sell short the targeted hedge securities to form the efficient hedge.*

Step 4.   *Compare combined hedge position created based on inaccurate data with a second position based on accurate or the "best available" data after the fact.*

Illustration with Inaccurate Data

Inaccurate Estimation of Effective Duration.   Assume 30-year Treasury bonds with an effective duration of 8% are used as a hedge vehicle, and the amount of permissible risk specified by senior management is 6% (or $6 per $100 at risk). Inaccurate data are obtained from digital feeds, and used to forecast MBS cash flows. The resulting estimate of effective duration of the MBS position is 6%.

Application of an Estimated Hedge Ratio.   This results in an estimated hedge ratio of (6% + 8%)/6% = 2.33. Had accurate data been used, the estimate of effective duration would have been 5%. An effective duration of 5% results in a hedge ratio of (6% + 8%)/5% = 2.80.

Purchase of Targeted Securities to Form an Inefficient Hedge.   The estimated effective duration and the calculated hedge ratio together imply that 100 30-year treasury bonds should be sold short to form a hedge for every 2.33 of these MBSs that are held. This hedge is supposed to limit permissible risk -- at a minimum cost.

Use of DIHR to Measure Hedge Inefficiency.   But it turns out that cost was not minimized. Had the more accurate data been used to forecast prepayments, the resulting effective duration of 5% would have resulted in a true hedge ratio of 2.80. Selling short 100 Treasury bonds would have hedged the risk of 2.80 MBSs; only 82.1 bonds were needed to hedge 2.33 MBSs. The hedge ratio estimated using inaccurate forecast data is around 21.74% higher than it should be, in this case. The result is an imbalanced hedge position.

DIHR as an MBS Prepayment Forecasting System Value Measure.   DIHR tells the MBS portfolio manager the extent to which a lower cost, equally effective hedge position could have been created.

Figure 2.   An Illustration of the Application of *Deviation In Hedge Ratio*

2. For each cell of the experimental design, the size of the disturbance that creates the data inaccuracy and the number of inaccurate data items were fixed.

Thus, inaccuracy involves the *amount* of the error and the *fraction* of the observations in each data set involved. For amount, the inaccuracies were equally likely to be positive or negative. Test data (our "holdout" sample) from the original data set and the randomly selected data items were used as input to both the linear regression and neural net models. The forecasts for each combination of fraction and amount were

evaluated in terms of two different metrics: predictive accuracy and payoff. Hereafter, we refer to the two sources of simulated data inaccuracy as *fraction-error* and *amount-error*.

6.2. Antithetic Random Numbers

For each combination of *fraction-error* and *amount-error*, five pairs of antithetic random numbers were generated [22]. A pair of antithetic random numbers consists of a uniformly distributed random number in the interval [0,1], along with a second random number given by one minus the original random number. These were applied to the test data in the usual manner to yield the simulated errors, and then two separate estimations were run and their results were averaged. This procedure helped to reduce the variance of our simulation results.

6.3. Simulation Results

This process yielded five simulated estimation results per *fraction-error/amount-error combination* (an experimental cell) for each of two forecasting methods employed: neural nets and linear regression. With three *fraction-errors* and four *amount-errors* in all, we had twelve cells for each model, excluding the cell for the original data set (which had 0 percent for both *fraction-error* and *amount-error*).

7. Simulation Results

PREDICTIVE ACCURACY RESULTS IN TERMS OF R^2 for the forecasts, based on the simulated inaccuracies in the data (in terms of *amount-error* and *fraction-error*), are given in Table 1. R^2s in the table reflect comparisons between results obtained with the perturbed data, and the results obtained with the test data in the "holdout" sample. Table 2 presents similar results for payoff. Five simulations per *fraction error/amount-error cell* were produced; only the cell averages are reported.

7.1. ANOVA Tests

To determine the effects, if any, of varying data quality on predictive performance, we tested for the effects of *amount-error* and *fraction-error*, first in terms of predictive accuracy (R^2), and then on the basis of payoff (DIHR) for each model. We performed four two-factor analysis of variance (ANOVA) tests [28], using the data presented in Tables 1 and 2. One ANOVA run was conducted for each performance measure, for each model. For each run, the factors are *fraction-error* (with 4 percent, 8 percent, and 12 percent of the observations in error), and *amount-error* (with plus or minus 5 percent, 10 percent, 15 percent, and 20 percent inaccuracies).

Table 3 gives the calculated F values in each instance; critical values are given in the left-hand column under the factor.

Each significant result (where the calculated F value is greater than the critical F

Table 1   Simulated Variations in Predictive Accuracy When
Data Accuracy Varies (in Percent)

Simulated fraction-errors	Simulated amount-errors									
	0%	5%	10%	15%	20%	0%	5%	10%	15%	20%
		Linear Regression:					Neural Net:			
0%	69.9	Values of R^2				67.2	Values of R^2			
4%		69.34	68.32	67.20	66.62		64.64	63.36	62.98	62.90
8%		68.94	67.52	66.52	65.90		64.48	63.40	63.12	62.90
12%		68.62	66.88	65.24	65.26		64.48	63.32	63.02	63.84

Note: The data used to obtain these results were "test data," a sample that we held out from the complete data set, which also included "training data." The "training data" enabled us to determine and specify the linear regression model, and to determine the architecture of the neural network. Each cell reflects averaged results of 5 simulated estimations that were run for the test data sets involving different simulated inaccuracies. We observed that the coefficients of the linear regression models were very stable over the range of simulated data inaccuracies, as were the neural network architectures.

Table 2   Simulated Variations in Payoff When Data Accuracy Varies (inPercent)

Simulated fraction-errors	Simulated amount-errors									
	0%	5%	10%	15%	20%	0%	5%	10%	15%	20%
		Linear Regression:					Neural Net:			
0%	0.948	Deviation In Hedge Ratio				0.939	Deviation In Hedge Ratio			
4%		0.9480	0.9490	0.9506	0.9514		0.9430	0.9436	0.9432	0.9432
8%		0.9480	0.9504	0.9514	0.9524		0.9424	0.9434	0.9434	0.9426
12%		0.9486	0.9506	0.9524	0.9526		0.9432	0.9434	0.9432	0.9430

Note: The values shown represent "deviation in hedge ratio" (DIHR). Although the numbers are small (all are less than 1 percent), these minor deviations reflect a substantial amount of money when Treasury bond positions are created to hedge for MBS portfolios worth tens of millions of dollars.

value) is marked with an asterisk; these indicate when a factor has a significant effect on a predictive measure, when using a particular model. The table indicates, for example, that varying the amount of inaccuracy in the data reduced the payoff associated with the forecast from the linear regression model, but not the payoff associated with the forecast from the neural net model. For regression, in statistical terms, we have rejected the hypothesis that payoff is unaffected by increases in data

Table 3   Significance of Varying Amount-Error on Predictive Performance of Linear Regression and Neural Net Models —ANOVA Results

Factor/significance criterion	Linear regression		Neural Net	
	Predictive accuracy (R^2)	Payoff (DIHR)	Predictive accuracy (R^2)	Payoff (DIHR)
Amount-error F (0.05; 3,48) = 2.82	245.43*	88.14*	108.95*	1.44
Fraction errors F (0.05; 2,48) = 3.22	66.48*	13.42*	1.32	0.25
Amount-fraction error interaction F (0.05; 6,48) = 2.32	1.29	1.05	0.50	0.11

Significant results are marked with an asterisk.

inaccuracy; the alternate appears to be true. However, we cannot reject the hypothesis that payoff is unaffected by the amount of data inaccuracy when a neural net is the forecast tool.

The ANOVA results indicate for each model which factor is significant in terms of predictive accuracy. For example, the results indicate that as *fraction-error* increases from 4 percent to 8 percent to 12 percent, the decrease in R^2 for the linear regression model is significant. In other words, the mean value for R^2 is not equal for all three values of *fraction-error*. The ANOVA results do not indicate, however, where the significant difference, or differences, occur. That is, the significant difference in R^2 may result from a significant difference as *fraction-error* changed from 4 percent to 8 percent, or it changed from 8 percent to 12 percent. It could also have come from the change in R^2 as *fraction-error* took a larger jump from 4 percent to 12 percent. This result makes intuitive sense: as data quality degrades, the variance of the error term of the model increases, resulting in "looser" model fit.

7.2. Tukey Test

We performed a Tukey Studentized Range Test [28] in order to determine, for the five significant cases in Table 3, exactly where significant differences occurred. These figures indicate that significant differences in R^2 and payoff are evident at each change in *amount-error* and each change in *fraction-error* in almost every case. (The only exception was for the R^2 levels at AMT = 15 percent and AMT = 20 percent for the neural net model; no significant change in R^2 was evident.)

From the ANOVA and Tukey tests, we can draw the following conclusions:

1. For linear regression, increasing levels of *fraction-error* and *amount-error* result in statistically significant degradations in predictive performance as measured by both R^2 and payoff.

2. For the neural net model, changes in *fraction-error* seem to have no statistically significant effect on predictive performance, for both R^2 and payoff.

3. For the neural net model, increasing levels of *amount-error* result in statistically significant degradation in performance as measured by R^2; however, no performance degradation is apparent in terms of payoff.

We also note that the degradation in R^2 with increases in amount-error appears to be more pronounced for linear regression than for the neural net model.

Besides separately examining the robustness of the individual models with respect to data accuracy, it also is of interest to compare the relative performance of the linear regression and neural network models. A cell-by-cell comparison of Table 1 indicates that linear regression consistently outperforms the neural net model. Similarly, analysis of Table 2 indicates that the results are reversed with respect to payoff: with payoff as the criterion, the neural net appears to outperform linear regression. The differences in performance between the models in each case are statistically significant as determined by paired *t*-tests [28]. Table 4 presents the test results.

8. Conclusion

TYPICALLY, RESEARCHERS AND PRACTITIONERS INTERESTED in optimizing predictive performance in quantitative applications focus on selecting an appropriate model for their problem. An implicit assumption is that the data used for modeling and prediction are accurate, and that the data, irrespective of quality (i.e., frequency, accuracy, timeliness), are cost-free.

8.1. Discussion and Managerial Guidelines

The first assumption was relaxed in this research: the effects of both forecast method and data accuracy were examined in terms of predictive performance. Predictive performance was gauged by a traditional metric, R^2, and by a payoff measure called *deviation in hedge ratio* (DIHR). The latter provides a realistic yardstick of the benefit of improved predictive accuracy for forecasting prepayments in MBS portfolio management.

Predictive Accuracy versus Payoff

The results of the previous section indicate that for this application, the predictive performance of linear regression, as measured by *both* predictive accuracy and payoff, suffers as data accuracy degrades. A similar effect is observed for neural nets when predictive accuracy is measured by R^2, but only with respect to changes in *amount-error*, the amount of inaccuracy in a given piece of data. Of particular interest is that the predictive performance of neural nets, as measured by payoff, is unaffected by changes in data quality over the range of degradation examined in our experiment. This research therefore illustrates an instance where less accurate, and typically less expensive data may provide results equivalent to those obtained with more accurate and expensive data.

Table 4   Two t-tests for Paired Comparisons of Cell Means

Variable	Number of observations	Mean	Standard error	t-statistic	Significance level
Accuracy (R^2)	61	0.0378	0.00104	36.50	0.0001
Payoff (DIHR)	61	−0.0072	0.00025	−29.14	0.0001

Note: The paired t-tests compare the overall performance of the linear regression and neural net models as data quality, in terms of fraction-error and amount-error, varies. A positive sign for the mean indicates that linear regression performed better than the neural net; a negative sign indicates that the neural net performed better than the linear regression.

Linear Regression versus Neural Nets

One important overall conclusion for the MBS application is that linear regression outperforms neural nets when R^2 is the performance criterion. It is not surprising that linear regression performs well under a criterion which reflects the objective of the least squares algorithm. One circumstance where a neural net model would tend to outperform linear regression in terms of R^2 is when nonlinear relationships are present in the data.[8] We can probably conclude from these R^2 results that the linearity assumptions of the regression model employed were well satisfied.

However, when payoff is the performance measure, the neural net consistently outperforms linear regression. This suggests a second important result: payoff expressed as DIHR is arguably the more crucial measure of business value, a useful result for financial risk managers. For a realistic MBS application, then, where the practitioner is interested in the payoffs that result from creating hedged positions to guard against prepayments, and in circumstances where data costs vary with accuracy, we can conclude that the better approach is to use a neural net model—at least so long as the data are potentially inaccurate to the limited extent examined in this experiment. More generally, this research suggests that, from a real-world standpoint, well-designed portfolio management forecasting systems involve a balance of data and model qualities. However, when data inaccuracies may be present, it makes sense to bear in mind that "white noise" inaccuracies, such as were simulated in this paper, should be of less concern than those with systematic bias.

Management Guidelines

The following guidelines for management are suggested by our work:

1. Managers should define a meaningful payoff function for their application when comparing different forecasting approaches.

2. If possible, a simulation study should be performed using each forecasting method under the expected range of data quality variation, and with information about the associated data costs included.

3. Managers should select for use that data-model combination that maximizes cost-benefit, taking into account two things—the payoff associated with the forecasts and the cost of the data.

We believe that these guidelines have broad relevance: for example, they apply equally well in inventory management, sales forecasting, and many other settings.

8.2. Caveats and Continuing Research

Although this research provides a new line of thinking for forecasting system design, we should remind the reader of three important caveats. Even though the data set we used was large and rich, the results of this research were derived in a highly specialized domain of business, MBS prepayment predictions. It is possible that in other applications, and for other payoff functions, linear regression may perform better than neural nets with inaccurate input data. Interested readers are encouraged to apply our methodology to ascertain the generalizability of the results. A second caveat related to our results stems from our assumption of white noise, zero mean simulated data inaccuracies in the forecasting data. If the simulation allowed for biased or skewed errors, it is possible that different results would have been obtained. A third caveat concerns the potential sensitivity of the results to the settings we chose for the neural network simulations, as discussed in note 6.

We are working to extend this line of research in several other ways. First, we plan to examine whether these results hold when alternate proprietary, firm-specific models for MBS prepayments are used. This is a natural next step: the primary research question will be whether we can validate our results in a different context—one where payoff is a crucial day-to-day concern. Second, we will examine whether our results generalize across forecasting domains. One other such application that may be of interest is risk management forecasting for foreign exchange trading. In prior research, fluctuations in foreign exchange rates have been predicted successfully using both the neural net and the linear regression approaches, among others [11]. An open question motivated by this research remains: which approach would be the more effective one if the quality of the forecast data cannot be guaranteed? Finally, an obvious extension is to consider larger errors, and to assess the performance of additional models. This would help to focus the analysis on forecast performance in the face of massive data quality degradation, and to identify the sensitivity of the results to the underlying forecasting model.

Simulation and empirically driven analysis based on information economics approaches, such as the one proposed in this paper, can enable managers to make more rational decisions about the business value of forecasting systems that involve data quality design decisions. As Voltaire once remarked, "[d]oubt is not a pleasant situation, but certainty is an absurd one." With prior knowledge about the expected performance of forecasting systems that are subject to data quality degradation, the doubt becomes a little more bearable—even for a financial risk manager.

NOTES

1. A mortgage includes an "embedded option," which gives a loan borrower an opportunity to prepay a mortgage loan if interest rates make prepayment favorable.

2. Others might argue that a regression model also can be made adaptive to new information, by rerunning it at frequent intervals with completely new data sets. However, the process presumes that the analyst has a sense of how to change the estimation form of the underlying model.

3. The time frame for which we collected data spanned thirty-five months. With thirty-eight MBS data points per month, the data set should have included 1,330 observations total. But a small amount of data was missing for several of the months, reducing our data set to 1,170 observations.

4. We also carried out the following diagnostic tests: the Durbin–Watson statistic to check for autocorrelation; scatter plots to examine regression residuals for heteroscedasticity; and the Belsey, Kuh, Welch test for multicollinearity.

5. One rule of thumb is that the number of nodes in the hidden layer should be approximately 75 percent of the number of nodes in the input layer. Trials using numbers above and below this figure were used to settle on the exact number to use.

6. We set the *learning rate* of the hidden layer at 0.3, and that of the output layer at 0.15; learning rate controls the average size of the weight changes used in the net. We set the *epoch* at 16 iterations; epoch measures the interval of time between changes in the outputs of the neurons. We set the *momentum* at 0.4; it balances network training time improvement and forecast stability. Convergence occurred after 18,000 iterations; network weights gradually adjusted such that the network's output gradually approached the desired output. See Wasserstein [38] for additional details.

7. Our assumption about correlation is a simplification. In typical financial market operations, securities that are selected to form a hedge rarely exhibit a perfect inverse correlated with the position a portfolio manager holds. Including this additional detail would unnecessarily complicate the analysis. Although the mathematics are more complex, it can be shown that relaxing this assumption boils down to the insertion of a "correlation constant" in the present hedge ratio.

8. While nonlinear regression is one option, model specification is onerous, requiring the analyst to make assumptions about the nonlinearities. This is known to be potentially problematic [39].

REFERENCES

1. Ahituv, N. Assessing the value of information: problems and approaches. *Proceedings of the Tenth International Conference on Information Systems*, Boston, 1989, pp. 315–325.

2. Altman, E.I.; Haldeman, R.; and Narayanan, P. Zeta analysis: a new model for identifying bankruptcy risk. *Journal of Banking and Finance*, 1, 1 (1977), 29–54.

3. Altman, E.I.; Avery, R.B.; Eisenbeis, R.A.; and Sinkey, J.F., Jr. *Applications of Classification Techniques in Business, Banking and Finance*. Greenwich, CT: JAI Press, 1981.

4. Arak, M., and Goodman, L. S. Prepayment risk in ginnie mae pools. *Secondary Mortgage Markets* (Spring 1985), 20–25.

5. Arend, M. Financial databases flag overseas opportunities. *Wall Street Computer Review* (October 1989).

6. Ballou, D.P., and Pazer, H.L. Modeling data and process quality in multi-input, multi-output information systems. *Management Science*, 31, 2 (February 1985), 152–162.

7. Bansal, A., and Kauffman, R.J. Risk management and data quality selection: an information economics approach. Working Paper, Center for Research on Information Systems, Stern School of Business, New York University, September 1992.

8. Bartlett, W.W. *Mortgage-backed Securities: Products, Analysis and Trading*. New York: New York Institute of Finance, 1989.

9. Barua, A.; Kriebel, C.H.; and Mukhopadhyay, T. MIS and information economics: augmenting rich description with analytical rigor in information systems design. *Proceedings*

of the Tenth International Conference on Information Systems, Boston, 1989, pp. 327–339.

10. Beaver, W.H. Financial ratios as predictors of failure. In *Empirical Research in Accounting: Selected Studies 1966. Journal of Accounting Research*, 4 (supplement) (1967), 71–111.

11. Colin, A. Machine learning techniques for foreign exchange trading. *Proceedings of the Third Annual International Conference on Advanced Trading Technologies*, New York, July 1992.

12. Curley, C., and Guttentag, J. The yield on insured residential mortgages. *Explorations in Financial Research* (Summer 1974), 114–161.

13. Cushing, B.E. A mathematic approach to the analysis and design of internal control systems. *The Accounting Review*, 49, 1 (January 1974), 24–41.

14. Demski, J.S. *Information Analysis*, 2d ed. Reading, MA: Addison-Wesley, 1985.

15. Doherty, N.A. *Corporate Risk Management*. New York: McGraw-Hill, 1985.

16. Dutta, S., and Shekhar, S. Bond rating: a non-conservative application of neural networks. *Proceedings of the Second International Conference on Neural Networks*, San Diego, 1988, pp. 443–450.

17. Green, J., and Shoven, J.B. The effects of interest rates on mortgage prepayments. *Journal of Money, Credit and Banking* (February 1986), 41–59.

18. Hayre, L.S.; Lauterbach, K.; and Mohebbi, C. Prepayment models and methodologies. In Frank J. Fabozzi (ed.), *Advances and Innovations in the Bond and Mortgage Markets*. Chicago: Probus Publishing, 1989.

19. Hilton, R.W. The determinants of information value: synthesizing some general results. *Management Science*, 27, 1 (1981), 57–64.

20. Ijiri, Y., and Itami, H. Quadratic cost–volume relationship and the timing of demand information. *The Accounting Review*, 48, 3 (October 1973), 724–737.

21. Kang, P., and Zenios, S.A. Complete prepayment models for mortgage-backed securities. *Management Science*, 38, 11 (November 1992), 1665–1685.

22. Law, A. M., and Kelton, W.D. *Simulation Modeling and Analysis*. New York: McGraw-Hill, 1982.

23. Leinweber, D. Intelligent trading systems. In W.H. Wagner (ed.), *The Complete Guide to Securities Transactions: Enhancing Investment Performance and Controlling Costs*. New York: John Wiley, 1989.

24. Lippmann, R.P. An introduction to computing with neural nets. *IEEE Journal of Acoustics, Speech, and Signal Processing* (April 1987), 4–22.

25. Martin, D. Early warning of bank failure: a logit regression approach. *Journal of Banking and Finance*, 1, 3 (1977), 249–276.

26. Meyer, P.A., and Pifer, H.W. Prediction of bank failures. *Journal of Finance*, 25, 4 (September 1970), 853–868.

27. Navratil, F.J. The estimation of mortgage prepayment rates. *Journal of Financial Research*, 8, 2 (Summer 1985), 107–117.

28. Neter, J.; Wasserman, W.; and Kutner, M.H. *Applied Linear Statistical Models*. Homewood, IL: Richard D. Irwin, 1985.

29. Pinkus, S.M.; Hunter, S.M.; and Roll, R. An introduction to the market and mortgage analysis. In F.J. Fabozzi and T.D. Garlicki (eds.), *Advances in Bond Analysis & Portfolio Strategies*. Chicago: Probus Publishing, 1987.

30. Richard, S.F., and Roll, R. Prepayments on fixed-rate mortgage-backed securities. *Journal of Portfolio Management* (Spring 1989), 73–82.

31. Salchenberger, L.M.; Cinar, E.M.; and Lash, N.A. Neural network: a new tool for predicting thrift failures. *Decision Sciences*, 23, 4 (July–August 1992), 899–916.

32. Salomon Brothers, Inc. *Mortgage Security Prepayment Rate Profile*, D. Walsh and M. Beveridge (eds.). Various issues beginning in April 1987, ending in February 1990.

33. Schmerken, I. Software to ride out interest-rate turbulence. *Wall Street Computer Review* (March 1990).

34. Schmerken, I. Wall Street struggles to balance global risk. *Wall Street Computer Review* (August 1990).

35. Shale, T. The great risk management systems failure. *Euromoney* (February 1989).

36. Tam, K.Y., and Kiang, M.Y. Managerial applications of neural networks: the case of bank failure predictions. *Management Science*, 38, 7 (July 1992), 926–947.

37. Waldman, M., and Gordon, M. Determining the yield of a mortgage security. In F.J. Fabozzi (ed.), *The Handbook of Mortgage-Backed Securities*. Chicago: Probus Publishing, 1985.

38. Wasserman, P.D. *Neural Computing*. New York: Van Nostrand Reinhold, 1989.

39. White, H. Consequences and detection of misspecified non-linear regression models. *Journal of the American Statistical Association*, 76, 374 (June 1981), 419–433.

40. White, H. Economic prediction using neural networks: the case of IBM daily stock returns. *Proceedings of Second IEEE International Conference on Neural Nets*, 1988, San Diego, II–451:II–458.