Principles of Forecasting: A Handbook for Researchers and Practitioners,
J. Scott Armstrong (ed.): Norwell, MA: Kluwer Academic Publishers, 2001.

The Forecasting Dictionary

Updated: October 23, 2000

J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia PA 19104

"But 'glory' doesn't mean "a nice knock-down argument," Alice objected.
"When I use a word," Humpty Dumpty said, in a rather scornful tone, "it means just what I choose it to mean—neither more nor less."
"The question is," said Alice, "whether you can make words mean so many different things."
"The question is," said Humpty Dumpty, "which is to be master—that's all."

Through the Looking Glass
Lewis Carroll

This dictionary defines terms as they are commonly used in forecasting. The aims, not always met, are to:

?? provide an accurate and understandable definition of each term,

?? describe the history of the term,

?? demonstrate how the term is used in forecasting,

?? point out how the term is sometimes misused, and

?? provide research findings on the value of the term in forecasting.

Acknowledgments

Geoff Allen and Robert Fildes inspired me to develop a comprehensive forecasting dictionary, and they provided much advice along the way. *The Oxford English Dictionary* was helpful in developing this dictionary, partly for definitions but, more important, for ideas as to what a dictionary can be. Most of the authors of *Principles of Forecasting* provided definitions. Definitions were also borrowed from the glossaries in Armstrong (1985) and Makridakis, Wheelwright and Hyndman (1998). Stephen A. DeLurgio added terms and revised definitions throughout. Eric Bradlow reviewed all statistical terms, and Fred Collopy reviewed the complete dictionary. Sandy D. Balkin, Robert G. Brown, Christopher Chatfield, Philip A. Klein, and Anne B. Koehler also made good suggestions, and many others provided excellent help. The Forecasting Dictionary was posted in full text on the Forecasting Principles website in October 1999 and e-mail lists were notified in an effort to obtain further peer review; many suggestions were received as a result. Mary Haight, Ling Qiu, and Mariam Rafi provided editorial assistance.

Abbreviations and Acronyms

Following are commonly used symbols. I give preference to Latin letters rather than Greek.

Symbol	**Description**	**Symbol**	**Description**
A	Actual value of a forecasted event	MdRAE	Median Relative Absolute Error
? ,???	alpha, beta, and gamma: smoothing factors in exponential smoothing for average, trend, and seasonality, respectively, they represent the weights placed on the latest value	MSE	Mean Square Error
APE	Absolute Percentage Error	*n*	sample size (number of observations, that is the number of decision units or number of years in a time series)
ARMA	AutoRegressive Moving Average	OLS	Ordinary Least Squares
ARIMA	AutoRegressive Integrated Moving Average	PI	Prediction Interval
b	measure of the impact of variable *x* on the dependent variable *Y* in regression analysis	*p*	probability
e	error	*r*	correlation coefficient
F	Forecast value	R^2	coefficient of determination
G	Growth or trend (it can be negative)	RAE	Relative Absolute Error
GMRAE	Geometric Mean of the Relative Absolute Error	RMSE	Root Mean Square Error
h	forecast horizon	*S*	Seasonal factor
j	period of the year	*t*	time; also a measure of statistical significance
MAD	Mean Absolute Deviation	*v*	number of variables
MAE	Mean Absolute Error	*w*	weighting factor
MAPE	Mean Absolute Percentage Error	*X*	explanatory or causal variable
$\overline{\text{MAPE}}$	Adjusted Mean Absolute Percentage Error; in which the denominator is the average of the forecasted and actual values. Also called the Symmetric MAPE.	*Y*	dependent variable (variable to be forecasted)

Terms

Underlined terms are defined elsewhere in the dictionary.

Terms are linked to relevant pages in *Principles of Forecasting* using PoF xxx.

Acceleration. A change in the trend, also including a negative change (deceleration). Although there have been attempts to develop quantitative models of acceleration for forecasting in the social and management sciences, these have not been successful. Of course, if one has good knowledge about its cause and its timing, acceleration can be a critical part of a forecast. Consider this when skydiving and you need to predict when to open a parachute. PoF xxx

Accuracy. See forecast accuracy.

ACF. See autocorrelation function.

Actuarial prediction. A prediction based on empirical relationships among variables. See econometric model.

Adaptive Conjoint Analysis (ACA). A method conceived by Rich Johnson (of Sawtooth Software, Inc.) in which self-explicated data are combined with paired-comparison preferences to estimate respondents' utility functions. ACA is a computer-interactive method in which the self-explicated data collected from a respondent influence the characteristics of the paired objects shown to the respondent. PoFxxx

Adaptive parameters. A procedure that reestimates the parameters of a model when new observations become available.

Adaptive response rate. A rule that instructs the forecasting model (such as exponential smoothing) to adapt more quickly when it senses that a change in pattern has occurred. In many time-series forecasting methods, a trade-off can be made between smoothing randomness and reacting quickly to changes in the pattern. Judging from 12 empirical studies (Armstrong 1985, p. 171), this strategy has not been shown to contribute much to accuracy, perhaps because it does not use domain knowledge. PoFxxx

Adaptive smoothing. A form of exponential smoothing in which the smoothing constants are automatically adjusted as a function of forecast errors. (See adaptive response rate.) PoFxxx

Additive model. A model in which terms are added. See also multiplicative model.

Adjusted Mean Absolute Percentage Error ($\overline{\text{MAPE}}$). The absolute error is divided by the average of the forecast and actual values. This has also been referred to as the Unbiased Absolute Percentage Error (UAPE) and as the symmetric MAPE (sMAPE).

Adjusted R^2. (See also $\bar{R}^2$.) R^2 adjusted for loss in the degrees of freedom. R^2 is penalized by adjusting for the number of parameters in the model compared to the number of observations. At least three methods have been proposed for calculating adjusted R^2: Wherry's formula $[1-(1-R^2)\cdot(n-1)/(n-v)]$, McNemar's formula $[1-(1-R^2)\cdot(n-1)/(n-v-1)]$, and Lord's formula $[1-(1-R^2)(n+v-1)/(n-v-1)]$. Uhl and Eisenberg (1970) concluded that Lord's formula is most effective of these for estimating shrinkage. The adjusted R^2 is always preferred to R^2 when calibration data are being examined because of the need to protect against spurious relationships. According to Uhl and Eisenberg, some analysts recommend that the adjustment include all variables considered in the analysis. Thus, if an analyst used ten explanatory variables but kept only three, R^2 should be adjusted for ten variables. This might encourage analysts to do a priori analysis. PoFxxx

Adjustment. A change made to a forecast after it has been produced. Adjustments are usually based on judgment, but they can also be mechanical revisions (such as to adjust the level at the origin by half of the most recent forecast error).

AIC (Akaike Information Criterion). A goodness of fit measure that penalizes model complexity (based on the number of parameters). The method with the lowest AIC is thought to represent the best balance of accuracy and complexity. Also see BIC, the Bayesian Information Criterion, which imposes a stronger penalty for complexity.

AID (Automatic Interaction Detector). A procedure that makes successive two-way splits in the data to find homogeneous segments that differ from one another. Also called tree analysis. Predictions can be made by forecasting the size and typical behavior for each segment. As its name implies, this procedure is useful for analyzing situations in which interactions are important. On the negative side, it requires much data so that each segment (cell size) is large enough (certainly greater than ten, judging from Einhorn's [1972] results). The evidence for its utility in forecasting is favorable but limited. Armstrong and Andress (1970) analyzed data from 2,717 gas stations using AID and regression. To keep knowledge constant, exploratory procedures (e.g., stepwise regression) were used. Predictions were then made for 3,000 stations in a holdout sample. The MAPE was much lower for AID than for regression (41% vs. 58%). Also, Stuckert (1958) found trees to be more accurate than regression in forecasting the academic success of about one thousand entering college freshmen. See also segmentation. PoFxxx

Akaike Information Criterion. See AIC.

Algorithm. A systematic set of rules for solving a particular problem. A program, function, or formula for analyzing data. Algorithms are often used when applying quantitative forecasting methods.

Amalgamated forecast. A seldom-used term that means combined forecast. See combining forecasts.

Analogous time series. Time-series data that are expected to be related and are conceptually similar. Such series are expected to be affected by similar factors. For example, an analyst could group series with similar causal forces. Although such series are typically correlated, correlation is not sufficient for series to be analogous. Statistical procedures (such as factor analysis) for grouping analogous series have not led to gains in forecast accuracy. See Duncan, Gorr and Szczyula (2001). PoFxxx

Analogy. A resemblance between situations as assessed by domain experts. A forecaster can think of how similar situations turned out when making a forecast for a given situation (see also analogous time series). PoFxxx

Analytic process. A series of steps for processing information according to rules. An analytic process is explicit, sequential, and replicable.

Anchoring. The tendency of judges' estimates (or forecasts) to be influenced when they start with a "convenient" estimate in making their forecasts. This initial estimate (or anchor) can be based on tradition, previous history, or available data. In one study that demonstrates anchoring, Tversky and Kahneman (1974) asked subjects to predict the percentage of nations in the United Nations that were African. They selected an initial value by spinning a wheel of fortune in the subject's presence. The subject was asked to revise this number upward or downward to obtain an answer. The information-free initial value had a strong influence on the estimate. Those starting with 10% made predictions averaging 25%. In contrast, those starting with 65% made predictions averaging 45%. PoFxxx

Anticipations. See expectations.

A posteriori analysis. Analysis of the performance of a model that uses actual data from the forecast horizon. Such an analysis can help to determine sources of forecast errors and to assess whether the effects of explanatory variables were correctly forecasted. PoFxxx

A priori analysis. A researcher's analysis of a situation before receiving any data from the forecast horizon. A priori analysis might rely on domain knowledge for a specific situation obtained by interviewing experts or information from previously published studies. In marketing, for example, analysts can use meta-analyses to find estimates of price elasticity (for example, see Tellis 1988) or advertising elasticity (Sethuraman and Tellis 1991). To obtain information about prior research, one can search the *Social Science Citation Index (SSCI)* or *A Bibliography of Business and Economic Forecasting* (Fildes, Dews and Howell 1981). The latter contains references to more than 4,000 studies taken from 40 journals published from 1971 to 1978. A revised edition was published in 1984 by the

Manchester Business School, Manchester, England. It can guide you to older sources that are difficult to locate using electronic searches. Armstrong (1985) describes the use of a priori analysis for econometric models. PoFxxx

AR model. See AutoRegressive model.

ARCH model. (Autoregressive conditionally heteroscedastic model.) A model that relates the current error variance to previous values of the variable of interest through an autoregressive relationship. ARCH is a time-series model in which the variance of the error term may change. Various formulations exist, of which the most popular is GARCH.

ARIMA. (AutoRegressive Integrated Moving Average model.) A broad class of time-series models that, when stationarity has been achieved by differencing, follows an ARMA model. See stationary series.

ARMA model. (AutoRegressive Moving Average.) A type of time-series forecasting model that can be autoregressive (AR), moving average (MA), or a combination of the two (ARMA). In an ARMA model, the series to be forecast is expressed as a function of previous values of the series (autoregressive terms) and previous error terms (the moving average terms). PoFxxx

Assessment center tests. A battery of tests to predict how well an individual will perform in an organization. Such tests are useful when one lacks evidence on how a candidate has performed on similar tasks. The procedure is analogous to combining forecasts. Hinrichs (1978) conducted a long-term follow-up of the predictive validity of assessment centers. PoFxxx

Asymmetric errors. Errors that are not distributed symmetrically about the mean. This is common when trends are expressed in units (not percentages) and when there are large changes in the variable of interest. The forecaster might formulate the model with original data for a variety of reasons such as the presence of large measurement errors. As a result, forecast errors would tend to be skewed, such that they would be larger for cases when the actual (for the dependent variable) exceeded the forecasts. To deal with this, transform the forecasted and actual values to logs and use the resulting errors to construct prediction intervals (which are more likely to be symmetric), and then report the prediction intervals in original units (which will be asymmetric). However, this will not solve the asymmetry problem for contrary series. For details, see Armstrong and Collopy (2000). PoFxxx

Asymptotically unbiased estimator. An estimator whose bias approaches zero as the sample size increases. See biased estimator.

Attraction market-share model. A model that determines market share for a brand by dividing a measure of the focal brand's marketing attractiveness by the sum of the attractiveness scores for all brands assumed to be in the competitive set. It is sometimes referred to as the US/(US + THEM) formulation. PoFxxx

Attributional bias. A bias that arises when making predictions about the behavior of a person (or organization) based upon the person's (or organization's) traits, even when the situation is the primary cause of behavior. (See Plous, 1993, Chapter 16.)

Autocorrelation. The correlation between values in a time series at time t and time t-k for a fixed lag k. Frequently, autocorrelation refers to correlations among adjacent time periods (lag 1 autocorrelation). There may be an autocorrelation for a time lag of one period, another autocorrelation for a time lag of two, and so on. The residuals serve as surrogate values for the error terms. There are several tests for autocorrelated errors. The Box-Pierce test and the Ljung-Box test check whether a sequence of autocorrelations is significantly different from a sequence of zeros; the Durbin-Watson statistic checks for first-order autocorrelations. PoFxxx

Autocorrelation function (ACF). The series of autocorrelations for a time series at lags 1, 2, ... A plot of the ACF against the lag is known as the correlogram. ACF can be used for several purposes, such as to identify the presence and length of seasonality in a given time series, to identify time-series models for specific situations, and to determine whether the data are stationary. See stationary series.

Automatic forecasting program. A program that, without user instructions, selects a forecasting method for each time series under study. Also see batch forecasting. The method-selection rules differ across programs but are frequently based on comparisons of the fitting or forecasting accuracy of a number of specified methods. Tashman and Leach (1991) evaluate these procedures. PoFxxx

Automatic Interaction Detector. See AID. PoFxxx

AutoRegressive (AR) model. A form of regression analysis in which the dependent variable is related to past values of itself at varying time lags. An autoregressive model would express the forecast as a function of previous values of that time series data (e.g., $Y_t = a + bY_{t-I} + e_t$, where a and b are parameters and e_t is an error term). PoFxxx

AutoRegressive Conditionally Heterosedastic model. See ARCH.

Availability heuristic. A rule of thumb whereby people assess the probability of an event by the ease with which they can bring occurrences to mind. For example, which is more likely – to be killed by a falling airplane part or by a shark? Shark attacks receive more publicity, so most people think they are more likely. In fact, the chance of getting killed by falling airplane parts is 30 times higher. Plous (1993, Chapter 11) discusses the availability heuristic. This heuristic can produce poor judgmental forecasts. It can be useful, however, in developing plausible scenarios. PoFxxx

Backcasting. Predicting what occurred in a time period prior to the period used in the analysis. Sometimes called postdiction, that is, predicting backward in time. It can be used to test predictive validity. Also, backcasting can be used to establish starting values for extrapolation by applying the forecasting method to the series starting from the latest period of the calibration data and going to the beginning of these data. See Armstrong (2001d) and PoFxxx

Backward shift operator. A notation aid where the letter B denotes a backward shift of one period. Thus, B operating on X_t (noted as BX_t) yields, by definition, X_{t-1}. Similarly BB or B^2 is the same as shifting back by two periods. A first difference (X_t-X_{t-1}) for a time series can be denoted $(1 – B) X_t$. A second-order difference is denoted $(1 – B)^2 X_t$. See differencing.

Baffelgab. Professional jargon that confuses more than it clarifies. Writing that sounds impressive while saying nothing. The term bafflegab was coined in 1952 by Milton A. Smith, assistant general counsel for the American Chamber of Commerce. He won a prize for the word and its definition: “multiloquence characterized by a consummate interfusion of circumlocution and other familiar manifestations of abstruse expatiation commonly utilized for promulgations implementing procrustean determinations by governmental bodies.” Consultants and academics also use bafflegab. Armstrong (1980a) showed that academics regard journals that are difficult to read as more prestigious than those that are easy to read. The paper also provided evidence that academics rated authors as more competent when their papers were rewritten to make them harder to understand. Researchers in forecasting are not immune to this affliction. PoFxxx

Base period. See calibration data.

Base rate. The typical or average behavior for a population. For example, to predict the expected box-office revenues for a movie, use those for a typical movie. PoFxxx

Basic research. Research for which the researcher has no idea of its potential use and is not motivated by any specific application. This is sometimes called pure research. One assumption is that eventually someone will find out how to use the research. Another assumption is that if enough researchers do enough research, eventually someone will discover something that is useful. PoFxxx

Basic trend. The long-term change in a time series. The basic trend can be measured by a regression analysis against time. Also called secular trend. PoFxxx

Batch forecasting. Forecasting in which a prespecified set of instructions is used in forecasting individual time series that are part of a larger group of time series. The forecasting method may be predesignated by the user or may rely on automatic forecasting. If the group has a hierarchical structure (see product hierarchy), the batch-processing program may allow reconciliation of item and group-level forecasts. For details and relevant software programs, see Tashman and Hoover (2001). PoFxxx

Bayesian analysis. A procedure whereby new information is used to update previous information. PoFxxx

Bayesian Information Criterion. See BIC.

Bayesian methods. A recursive estimation procedure based on Bayes' theorem that revises the parameters of a model as new data become available.

Bayesian pooling. A method that improves estimation efficiency or speed of adapting time-varying parameters models by using data from analogous time series. PoFxxx

Bayesian Vector AutoRegressive (BVAR) model. A multivariate model whose parameters are based on observations over time and a cross-section of observational units that uses a set of lagged variables and Bayesian methods.

Benchmark forecasts. Forecasts used as a basis for comparison. Benchmarks are most useful if based on the specific situation, such as forecasts produced by the current method. For general purposes, Mentzer and Cox (1984) examined forecasts errors for various levels in the product hierarchy and for different horizons as shown here:

Typical Errors For Sales Forecasts (Entries are MAPEs)

	Forecast Horizon		
Level	Under 3 Months	3 Months to 2 Years	Over 2 Years
Industry	8	11	15
Corporate	7	11	18
Product group	10	15	20
Product Line	11	16	20
Product	16	21	26

Source: Mentzer and Cox's (1984) survey results from 160 corporations are crude because most firms do not keep systematic records. Further, the study was ambiguous in its definitions of the time interval. "Under 3 months" probably refers to "monthly" in most cases, but the length of time is not apparent for "Over 2 years."

BFE (Bold Freehand Extrapolation). The process of extending an historical time series by judgment. See judgmental extrapolation.

Bias. A systematic error; that is, deviations from the true value that tend to be in one direction. Bias can occur in any type of forecasting method, but it is especially common in judgmental forecasting. Researchers have identified many biases in judgmental forecasting. Bias is sometimes a major source of error. For example, Tull (1967) and Tyebjee (1987) reported a strong optimistic bias for new product forecasting. Some procedures have been found to reduce biases (Fischhoff and MacGregor 1982). Perhaps the most important way to control for biases is to use structured judgment.

Biased estimator. An estimate in which the statistic differs from the population parameter. See asymptotically unbiased estimator.

BIC (Bayesian Information Criterion). Also called the Schwarz criterion. Like the AIC, the BIC is a criterion used to select the order of time-series models. Proposed by Schwarz (1978), it sometimes leads to less complex models than the AIC. Several studies have found the BIC to be a better model selection criterion than the AIC.

BJ methods. See Box-Jenkins methods.

Bold Freehand Extrapolation. See BFE.

Bootstrapping. In forecasting, bootstrapping typically refers to judgmental bootstrapping. Bootstrapping is also a term used by statisticians to describe estimation methods that reuse a sample of data. It calls for taking random samples from the data with replacement, such that the resampled data have similar properties to the original sample. Applying these ideas to time-series data is difficult because of the natural ordering of the data. Statistical bootstrapping methods are computationally intensive and are used when theoretical results are not available. To date, statistical bootstrapping has been of little use to forecasters, although it might help in assessing prediction intervals for cross-sectional data. PoFxxx

Bottom-up. A procedure whereby the lowest-level disaggregate forecasts in a hierarchy are added to produce a higher-level forecast of the aggregate. (See also segmentation.) PoFxxx

Bounded values. Values that are limited. For example, many series can include only non-negative values. Some have lower and upper limits. (Percentages are limited between zero and one hundred.) When the values are bounded between zero and one, consider using a transformation such as the logit. If a transformation is not used, ensure that the forecasts do not go beyond the limits. PoFxxx

Box-Jenkins (BJ) methods. The application of autoregressive-integrated-moving average (ARIMA) models to time-series forecasting problems. Originally developed in the 1930s, the approach was not widely known until Box and Jenkins (1970) published a detailed description. It is the most widely cited method in extrapolation, and it has been used by many firms. Mentzer (1995) found that analysts in 38% of the 205 firms surveyed were familiar with BJ, it was used in about one-quarter of these firms, and about 44% of those familiar with it were satisfied. This satisfaction level can be compared with 72% satisfaction with exponential smoothing in the same survey. Contrary to early expectations, empirical studies have shown that it has not improved forecast accuracy of extrapolation methods. PoFxxx

Box-Pierce test. A test for autocorrelated errors. The Box-Pierce Q statistic is computed as the weighted sum of squares of a sequence of autocorrelations. If the errors of the model are white noise, then the Box-Pierce statistic is distributed approximately as a chi-square distribution with $h - v$ degrees of freedom, where h is the number of lags used in the statistic and v is the number of fitted parameters other than a constant term. It is sometimes known as a portmanteau test. Another portmanteau test is the Ljung-Box test, which is a version of the Box-Pierce test.

Brainstorming. A structured procedure for helping a group to generate ideas. The basic rules are to suspend evaluation and to keep the session short (say ten minutes). To use brainstorming effectively, one should first gain the group's agreement to use brainstorming. Then, select a facilitator who

- encourages quantity of ideas,
- encourages wild or potentially unpopular ideas,
- reminds the group not to evaluate (either favorably or unfavorably),
- does not introduce his or her own ideas, and
- records all ideas.

When people follow the above procedures carefully, brainstorming greatly increases the number of creative ideas they suggest in comparison with traditional group meetings. This is because it removes some (but not all) of the negative effects of the group process. Brainwriting (individual idea generation) is even more effective than brainstorming, assuming that people will work by themselves. One way to do this is to call a meeting and then allocate, say, ten minutes for brainwriting. Brainwriting is particularly effective because everyone can generate ideas (i.e., no facilitator is needed). The sources of the ideas are not identified. Brainstorming or brainwriting can be used with econometric models to create a list of explanatory variables and to find alternative ways of measuring variables. It can also be used to create a list of possible decisions or outcomes that might occur in the future, which could be useful for role-playing and expert opinions. Brainstorming is often confused with "talking a lot," which is one of the deplorable traits of unstructured or leaderless group meetings.

Brier score. A measure of the accuracy of a set of probability assessments. Proposed by Brier (1950), it is the average deviation between predicted probabilities for a set of events and their outcomes, so a lower score represents higher accuracy. In practice, the Brier score is often calculated according to Murphy's (1972) partition into three additive components. Murphy's partition is applied to a set of probability assessments for independent-event forecasts when a single probability is assigned to each event:

$$B = c(1 - c) + \frac{1}{N}\sum_{t=1}^{T} n_t(p_t - c_t)^2 - \frac{1}{N}\sum_{t=1}^{T} n_t(c_t - c)^2,$$

where c is the overall proportion correct, c_t is the proportion correct in category t, p_t is the probability assessed for category t, n_t is the number of assessments in category t, and N is the total number of assessments. The first term reflects the base rate of the phenomenon for which probabilities are assessed (e.g., overall proportion of correct forecasts), the second is a measure of the calibration of the probability assessments, and the third is a measure of the resolution. Lichtenstein, Fischhoff and Phillips (1982) provide a more complete discussion of the Brier score for the evaluation of probability assessments.

Brunswick lens model. (See lens model.)

Business cycle. Periods of economic expansion followed by periods of economic contraction. Economic cycles tend to vary in length and magnitude and are thought of as a separate component of the basic pattern contained in a time series. Despite their popularity, the use of business cycles has not been shown to lead to more accurate forecasting. PoFxxx

BVAR model. See Bayesian Vector AutoRegression model.

Calibrate. (1) To estimate relationships (and constant terms) for use in a forecasting model. (See also fit.) Some software programs erroneously use the term *forecast* to mean calibrate. (2) To assess the extent to which estimated probabilities agree with actual probabilities. In that case, calibration curves plot the predicted probability on the *x*-axis and the actual probability on the *y*-axis. A probability assessor is perfectly calibrated when the events or forecasts assigned a probability of X occur X percent of the time for all categories of probabilities assessed.

Calibration data. The data used in developing a forecasting model. (See also fit.) PoFxxx

Canonical correlations. A regression model that uses more than one dependent variable and more than one explanatory variable. The canonical weights provide an index for the dependent variables but without a theory. Despite a number of attempts, it seems to have no value for forecasting (e.g., Fralicx and Raju, 1982, tried but failed).

Case-based reasoning. Reasoning based on memories of past experiences. Making inferences about new situations by looking at what happened in similar cases in the past. (See analogy.)

Causal chain. A sequence of linked effects; for example, A causes B which then causes C. The potential for error grows at each stage, thus reducing predictive ability. However, causal chains lead judgmental forecasters to think the outcomes are more likely because each step seems plausible. Causal chains are useful in developing scenarios that seem plausible. PoFxxx

Causal force. The net directional effect domain experts expect for a time series over the forecast horizon. Armstrong and Collopy (1993) classified them as growth, decay, opposing, regressing, or supporting forces. The typical assumption behind extrapolation is supporting, but such series are rare. Armstrong, Adya and Collopy (2001) discuss evidence related to the use of causal forces. PoFxxx

Causal model. A model in which the variable of interest (the dependent variable) is related to various explanatory variables (or causal variables) based on a specified theory.

Causal relationship. A relationship whereby one variable, *X*, produces a change in another variable, *Y*, when changes in *X* are either necessary or sufficient to bring about a change in *Y*, and when the change in *X* occurs before the change in *Y*. Einhorn and Hogarth (1982) discuss causal thinking in forecasting.

Causal variable. A variable, *X*, that produces changes in another variable, *Y*, when changes in *X* affect the probability of *Y* occurring, *and* a theory offers an explanation for why this relationship might hold.

Census Program X-12. A computer program developed by the U.S. Bureau of the Census. (See X-12 ARIMA decomposition.) The program is available at no charge; details can be found at hops.wharton.upenn.edu/forecast

Census II. A refinement of the classical method that decomposes time series into seasonal, trend, cycle, and random components that can be analyzed separately. The Census II method X-11 decomposition, has been superseded by the X-12-ARIMA decomposition method. The programs contain excellent procedures for seasonal adjustments of historical data. However, the developers did not seem to be concerned about how these factors should be used in forecasting.

Central limit theorem. The sampling distribution of the mean of *n* independent sample values will approach the normal distribution as the sample size increases regardless of the shape of the population distribution. This applies when the sample size is large enough for the situation. Some people suggest 30 as adequate for a typical situation.

Chow test. A test that evaluates whether a subsample of data, excluded from the model when it was estimated, can be regarded as indistinguishable from the data used for estimation. That is, it measures whether two samples of data are drawn from the same population. If so, the coefficient estimates in each sample are considered to be identical. For details, see an econometric textbook. An alternative viewpoint, which some favor, would be to use a priori analysis to decide whether to combine estimates from different sets of data.

Classical decomposition method. A division of a time series into seasonal, trend, and error components. These components can then be analyzed individually. See also Census II. PoFxxx

Classification method. (See segmentation.)

Clinical judgment. (See expert opinions.)

Coefficient. An estimate of a relationship in an econometric model.

Coefficient of determination. See R^2.

Coefficient of inequality. See Theil's *U*.

Coefficient of variation. The standard deviation divided by the mean. It is a measure of relative variation and is sometimes used to make comparisons across variables expressed in different units. It is useful in the analysis of relationships in econometric or judgmental bootstrapping models. Without variation in the data, one may falsely conclude that a variable in a regression analysis is unimportant for forecasting. Check the coefficients of variation to see whether the dependent and explanatory variables have fluctuated substantially. If they have not, seek other ways of estimating the relationships. For example, one might use other time-series, cross-sectional, longitudinal or simulated data. Alternatively, one could use a priori estimates as relationships, basing these on prior research or on domain knowledge.

Cognitive dissonance. An uncomfortable feeling that arises when an individual has conflicting attitudes about an event or object. The person can allay this feeling by rejecting dissonant information. For example, a forecast with dire consequences might cause dissonance, so the person might decide to ignore the forecast. Another dissonance-reduction strategy is to fire the forecaster.

Cognitive feedback. A form of feedback that includes information about the types of errors in previous forecasts and reasons for these errors. PoFxxx

Coherence. The condition when judgmental inputs to a decision-making or forecasting process are internally consistent with one another. For example, to be coherent, the probabilities for a set of mutually exclusive and exhaustive events should sum to unity.

Cohort model. A model that uses data grouped into segments (e.g., age 6 to 8, or first year at college, or start-up companies) whose behavior is tracked over time. Predictions are made for the cohorts as they age. Cohort models are commonly used in demographic forecasting. For example, an analyst could forecast the number of students entering high school in six years by determining the number of students currently in the third-grade cohort in that region (assuming no deaths or net migration). PoFxxx

Cointegration. The co-movement of two or more non-stationary variables over time. If two variables are cointegrated, regression of one variable on the other results in a set of residuals that is stationary. Existence of this long-run equilibrium relationship allows one to impose parameter restrictions on a Vector AutoRegressive model (VAR). The restricted VAR can be expressed in various ways, one of which is the error correction model. With more than two non-stationary variables, it is possible to have more than one long-run equilibrium relationship among them.

Combining forecasts. The process of using different forecasts to produce another forecast. Typically, the term refers to cases where the combining is based on an explicit, systematic, and replicable scheme, such as the use of equal weights. If subjective procedures are used for averaging, they should be fully disclosed and replicable. Combining forecasts should not be confused with combining forecasting *methods*. Combining is inexpensive and almost always improves forecast accuracy in comparison with the typical component. It also helps to protect against large errors. See Armstrong (2001e) and PoFxxx

Commensurate measure. An explicit measure that is common to all elements in a category. If the category is a set of candidates for a job and the task is to select the best candidate, a commensurate measure would be one that all candidates have in common, such as their grade-point average in college. When trying to predict which candidate will be most successful, selectors tend to put too much weight on commensurate measures, even if the measures are irrelevant, thus reducing forecast accuracy (Slovic and McPhillamy 1974). PoFxxx

Comparison group. A benchmark group used for comparison to a treatment group when predicting the effects of a treatment. See control group.

Compensatory model. A model that combines variables (cues) to form a prediction. It is compensatory because high values for some cues can compensate for low values in other cues. Adding and averaging are compensatory models.

Composite forecast. A combined forecast. (See combining forecasts.)

Composite index. A group of indicators that are combined to permit analysts to monitor economic activity. In business-cycle analysis, composite indexes of leading, coincident, and lagging indicators have similar timing and are designed to predict turning points in business cycles. See cyclical data.

Conditional forecast. A forecast that incorporates knowledge (or assumptions) about the values of the explanatory variables over the forecast horizon. Also called an ex post forecast.

Confidence interval. An expression of uncertainty. The likelihood that the true value will be contained with a given interval. The 95% confidence level is conventional but arbitrary; ideally, one would choose a limit that balances costs and benefits, but that is seldom easy to do. In forecasting, the term *confidence interval* refers to the uncertainty associated with the estimate of the parameter of a model, while the term *prediction interval* refers to the uncertainty of a forecast. Confidence intervals play a role in judgmental bootstrapping and econometric models by allowing one to assess the uncertainty for an estimated relationship (such as price elasticity). This, in turn, might indicate the need for more information or for the development of contingency plans.

Conjoint analysis. A methodology that quantifies how respondents trade off conflicting object characteristics against each other in a compensatory model. For example, alternative products could be presented to subjects with the features varied by experimental design. Subjects would be asked to state their preferences (through ratings, rankings, intentions, or choices). The importance of each feature is assessed by statistical analysis. Software packages are available to aid the process. See Wittink and Bergestuen (2001) and PoFxxx

Conjunction fallacy. The notion that the co-occurrence of two events is more likely than the occurrence of either event alone. When people are asked to predict the outcomes of events, the added detail, especially when representative of the situation, leads them to increase their estimate of the likelihood of their joint occurrence. For example, in one study, people thought that President Reagan was more likely to provide more federal support for unwed mothers *and* cut federal support to local governments than he was to simply provide more federal support for unwed mothers (Tversky and Kahneman 1983). See representativeness.

Conjunctive model. A nonlinear model that combines variables (cues) to ensure that scores on all variables must be high before the forecast generated by the model will be high.

Consensus. Agreement of opinions; the collective unanimous opinion of a number of persons. A feeling that the group's conclusion represents a fair summary of the conclusions reached by the individual members.

Consensus seeking. A structured process for achieving consensus. Consensus seeking can be useful in deciding how to use a forecast. It can help groups to process information and to resolve conflicts. In practice, complete unanimity is rare. However, each individual should be able to accept the group's conclusion. Consensus seeking requires the use of a facilitator who helps the group to follow these guidelines:

- Avoid arguing for your own viewpoint. Present your position logically, then listen to the other members.
- Do not assume that someone must win when the discussion reaches a stalemate. Instead, restate the problem or generate new alternatives.
- Do not change your mind simply to avoid conflict. Be suspicious when agreement seems to come too quickly. Explore the reasons, and be sure that everyone accepts the solution.
- Avoid conflict-reducing techniques, such as majority vote, averages, coin flips, and bargaining. When a dissenting member finally agrees, do not think the group must give way to their views on some later point.
- Differences of opinion are natural and expected. Seek them out and involve everyone in a discussion of them. A wide range of opinions increases the chance that the group will find a better solution.

Alternatively, consensus has been used to assess the level of agreement among a set of forecasts. Higher consensus often implies higher accuracy, especially when the forecasts are made independently. Ashton (1985) examined two different forecast situations: forecasts of annual advertising sales for *Time* magazine by 13 Time, Inc. executives given forecast horizons for one, two, and three quarters, and covering 14 years; and forecasts by 25 auditors of 40 firms' problems, such as bankruptcy. Using two criteria, correlations and mean absolute deviation, she compared the actual degree of agreement (between forecasts by different judges) against the accuracy of these judges. She also compared each judge's degree of agreement with all other judges and related this to that judge's accuracy. Agreement among judges did imply greater accuracy and this relationship was strong and statistically significant. This adds evidence for using consensus as a proxy for confidence. PoFxxx

Conservatism. The assumption that things will proceed much as they have in the past. Originally a political term that involved resistance to change. Conservatism is useful when forecasts involve high uncertainty. Given uncertainty, judgmental forecasters should be conservative and they typically are. Some quantitative procedures, such as regression analysis, provide conservative estimates. PoFxxx

Consistent trends. A condition that occurs when the basic trend and the recent trend extrapolations are in the same direction. The basic trend is long term, such as that obtained by a regression against time. The recent trend is short term, such as that obtained with an exponential smoothing model with a heavy weight on the most recent data. Extrapolations of trends are expected to be more accurate when trends are consistent, as discussed under inconsistent trends. PoFxxx

Construct validity (or **conceptual validity** or **convergent validity).** Evidence that an operational measure represents the concept. Typically assessed by examining the correspondence among different operational measures of a concept. PoFxxx

Consumer heterogeneity. Differences among people, either in terms of observable characteristics, such as demographics or behavior, or in terms of unobservable characteristics, such as preferences or purchase intentions. In some forecasting settings, it may be helpful to capture these types of differences as well as the factors that affect the future behavior of individuals.

Contextual information. Information about explanatory variables that could affect a time-series forecast. The contextual information that the forecaster has is called domain knowledge. PoFxxx

Contrary series. A series in which the historical trend extrapolation is opposite in direction to prespecified expectations of domain experts. For example, domain experts might think that the causal forces should drive the series up, but the historical trend is headed down. Contrary series can lead to large errors. Evidence to date suggests that statistical trend estimates should be ignored for contrary series (Armstrong and Collopy 1993). In addition, contrary series are expected to have asymmetric errors, even when expressed in logs (Armstrong and Collopy 2000). See Armstrong, Adya and Collopy (2001). PoFxxx

Contrast group. See comparison group.

Control group. A group of randomly assigned people (or organizations) that did not receive a treatment. If random assignment of treatments is not possible, look for a comparison group.

Convenience sample. A sample selected because of its low cost or because of time pressures. Convenience samples are useful for pretesting intentions surveys or expert opinion studies. However, it is important to use probability samples, not convenience samples, in conducting intentions studies.

Correlation (*r*). A standardized measure of the linear association between two variables. Its values range from –1, indicating a strong negative relationship, through zero, which shows no relationship, to +1, indicating a strong positive association. The correlation coefficient is the covariance between a pair of standardized variables. Curtis and Alf (1969) and Ozer (1985) argue that r is a better measure of predictive ability than R^2 (but neither is very useful for time-series data). A strong correlation does not imply a causal relationship.

Correlation matrix. A set of correlation coefficients presented in the form of a matrix. Most computer programs that perform multiple regression analysis show the correlation coefficients for each pair of variables. They can be useful for assessing multicollinearity.

Correlogram. Graphical representation of the autocorrelation function.

Covariance. A measure of the variation between variables, say X and Y. The range of covariance values is unrestricted. However, if the X and Y variables are first standardized, then covariance is the same as correlation and the range of covariance (correlation) values is from –1 to +1.

Criterion variable. See dependent variable.

Cross-correlation. A standardized measure of association between values in one time series and those of another time series. This statistic has the characteristics of a regular correlation coefficient.

Cross-sectional data. Data on a number of different units (e.g., people, countries, firms) for a single time period. Cross-sectional data can be used to estimate relationships for a forecasting model. For example, using cross-sectional data from different countries, one could assess how prices affect liquor sales. PoFxxx

Cross-validation. A test of validity that consists of splitting the data using probability sampling, estimating the model using one subsample, and testing it on the remaining subsample. More elaborate approaches such as double cross-validation and the jackknife are discussed in Armstrong (2001d).

Croston's method. See intermittent series.

Cue. A variable. In judgmental forecasting, a cue refers to a variable perceived by an expert.

Cumulative error. The total of all forecast errors (both positive and negative) over the forecast horizon. For example, for forecasts for the next five years, the analyst would sum the errors (with signs) for the five forecasts. This will approach zero if the forecast is unbiased.

Cumulative forecasting. The total value of a variable over several horizon periods. For example, one might forecast total sales over the next year, rather than forecast sales for each of the 12 months.

Current status. The level at the origin of the forecast horizon.

Curve fitting. To fit historical time-series data to a functional form such as a straight line or a polynomial.

Cusum. Cumulative sum of forecast errors. The cusum is used in tracking signals.

Cyclical data. Time-series data that tend to go through recurring increases and decreases. See also business cycle. This term is generally not used for seasonal variations within a year. Although it is difficult to forecast cycles, knowledge that a time series is subject to cycles may be useful for selecting a forecasting method and for assessing uncertainty. (See also long waves.) See Armstrong (2001c), Armstrong, Adya and Collopy (2001), and PoFxxx

Cyclical index. A number, usually standardized to have a mean of 100, that can help to identify repetitive patterns. It is typically applied to annual time-series data, but can also be used for shorter periods, such as hours within a day. (See also seasonal index)

Damp. To reduce the size of an effect, as in "to damp the trend" (as contrasted to dampening, which would imply some type of moisturizing and thus be senseless, or worse, for forecasters). Damped estimates are useful in the presence of uncertainty. Thus, in making extrapolations over long horizons, one should damp. Seasonal factors can also be damped if there is uncertainty. In addition, the effects in an econometric model can be damped in light of uncertainty about the forecasts of the explanatory variables. See mitigation and Armstrong (2001c). PoFxxx

Damped trend. See damp.

Data Generating Process (DGP). A model of the system under investigation that is assumed to represent the system and to be responsible for the observed values of the dependent variable. It is important to remember that the model is based on assumptions; for real-world data in the social sciences, one can only guess at the DGP.

Decay forces. Forces that tend to drive a series down. For example, the costs for such technical products as computers might fluctuate over time, but as long as the underlying forces are downward, they are classified as decay. See Armstrong, Adya and Collopy (2001). PoFxxx

Deceleration. A decrease in the trend. See acceleration.

Decomposition. The process of breaking a problem into subproblems, solving them, and then combining the solutions to get an overall solution. MacGregor (2001) provides evidence on the value of this procedure for judgmental forecasting. Typically, decomposition refers to multiplicative breakdowns, but sometimes it applies to additive breakdowns. Additive breakdowns, however, are usually called disaggregate forecasting or segmentation. Time series are often decomposed by level, trend, cycle, seasonality, and error. PoFxxx

Degrees of freedom. The number of observations minus the number of parameters in a regression analysis. It is sensible to include all variables considered for use in the model, not just those in the final version. The larger the number of coefficients estimated, the larger the number of constraints imposed in the sample and the smaller the number of observations left to provide precise estimates of the regression coefficients. A greater number of degrees of freedom is often thought to provide more reliable estimates, but the relationship to reliability is weak. (See adjusted R^2.)

Delphi technique. A method for obtaining independent forecasts from an expert panel over two or more rounds, with summaries of the anonymous forecasts (and perhaps reasons for them) provided after each round. Delphi has been widely used in business. By applying well-researched principles, Delphi provides more accurate forecasts than unstructured groups (Rowe and Wright 1999). The process can be adapted for use in face-to-face group meetings, and is then called mini-Delphi or Estimate-Talk-Estimate (ETE). Rowe and Wright (2001) provide principles for the use of the Delphi technique. PoFxxx

Demand. The need for a particular product or component. Demand can come from a number of sources (e.g., customer order or producer's good). Demand can be forecast for each level in a supply chain. At the finished-goods level, demand data often differ from sales data because demand does not necessarily result in sales (e.g., if there is no stock there may be unfulfilled demand).

Dependent variable. The variable that is to be forecast; that is, the variable of interest to the researcher. In regression analysis, it is the variable on the left side of the equation.

Deseasonalized data. See seasonal adjustment.

Detrend. To remove an upward or downward trend from a time series. Frequently, this is done by regressing a series against time, then using the trend coefficient to remove the trend from the observations. Detrending data can reveal patterns in the data. Detrending should be done prior to making seasonal adjustments. PoFxxx

Devil's advocate. A procedure whereby one person in a group is assigned the task of trying to find everything that might be wrong in a forecast (or a plan), while the rest of the group defends it. This should be done as a structured approach, perhaps with this role rotating among group members. (Someone adopting this role without permission from the group can become unpopular.) Use the devil's advocate procedure only for *short* time periods, say 20 minutes or less if done in a meeting. Cosier's (1978) experiment showed that groups that used the devil's advocate procedure obtained more accurate predictions than those who solely argued in favor of a forecast. One would also expect the devil's advocate procedure to improve the calibration of prediction intervals. According to Cosier (1978) and Schwenk and Cosier (1980), the "attack" is best presented in written form and in an objective manner; the use of strong emotionally laden criticism should be avoided. This research is consistent with findings that peer review leads to improvements in research papers. PoFxxx

DGP. See Data Generating Process.

Diagnostic checking. A step in time-series model building where the estimated errors of a model are examined for independence, zero mean, constant variance, and other assumptions.

Dickey-Fuller test. A test to determine whether a time series is stationary or, specifically, whether the null hypothesis of a unit root can be rejected. A time series can be nonstationary because of a deterministic trend (a stationary trend or TS series) or a stochastic trend (a difference stationary or DS series) or both. Unit root tests are intended to detect stochastic trend, although they are not powerful at doing so, and they can give misleading inferences if a deterministic trend is present but is not allowed for. The augmented Dickey-Fuller test, which adds lagged dependent variables to the test equation, is often used. Adding the lagged variables (usually at the rate corresponding to $n/3$, where n is the sample size) removes distortions to the level of statistical significance but lowers the power of the test to detect a unit root when one is present. There is a difference between forecasting with trend-stationary (TS) and difference-stationary (DS) models (though probably little difference in point forecasts and intervals for short horizons, h = 1 or 2). The point forecasts of a TS series change by a constant amount (other things being equal) as the forecast horizon is incremented. Their prediction intervals are almost constant. The point forecasts of a DS series are constant as the horizon is increased (like naive no-change forecasts), other things being equal, while the prediction intervals widen rapidly. There is a vast literature on unit roots. The expression "unit root test$" ($ indicates a wildcard) generated 281 hits in the *Econolit* database of *OVID* (as of mid-December, 1999), although when it was combined with "forecast$," the number fell to 12. Despite this literature, we can say little about the usefulness of a unit-root test, such as the Dickey-Fuller test, as part of a testing strategy to improve forecasting accuracy. Meese and Geweke (1984) examined 150 quarterly and monthly macroeconomic series and found that forecasts from detrended data (i.e., assuming TS) were more accurate than forecasts from differenced data. Campbell and Perron (1991) conducted a Monte Carlo simulation with an ARMA (1,1) Data Generating Process and samples of 100. When there was an autoregressive unit root or near unit root (.95 or higher), an autoregressive model in differences forecasted better at h = 1 and h = 20 horizons. When there was an autoregressive unit root and the moving average parameter was 0.9 or less, the model in differences was also better. Otherwise the AR model in levels with a trend variable was better. Since most economic series appear to contain a unit root, the Campbell and Perron study seems to call for using a DS model, exactly the opposite of the strategy indicated by Meese and Geweke. But what if the parameter values are unknown? Campbell and Perron also considered a mixed strategy: Use a levels model if the augmented Dickey-Fuller test and the Phillips-Perron test for a unit root were both rejected at the five percent level of significance; otherwise use a model in differences. Such a strategy gave almost as good results as using the better model given knowledge of the parameter values. This slender evidence provides some support for using a unit-root test to select a forecasting model. Maddala and Kim (1998) provide a helpful summary. PoFxxx

Differencing. A time series of successive differences ($X_t - X_{t-1}$). When a time series is non-stationary, it can often be made into a stationary series by taking first differences of the series. If first differences do not convert the series to stationary form, then one can create first differences of first differences. This is called second-order differencing. A distinction is made between a second-order difference and a second difference ($X_t - X_{t-2}$). See backward shift operator.

Diffusion. The spreading of an idea or an innovation through a population. Typically, an innovation such as television is initially used by a small number of people. The number of new users per year increases rapidly, then, after stabilizing, decreases as unsatisfied demand for the innovation dies away. Meade and Islam (2001) examine the use of diffusion models for time-series extrapolation. Rogers (1995), based on an extensive review of the literature, updated his conclusions that the speed of diffusion depends on: (1) the relative advantage of the product over existing products, (2) compatibility with existing solutions, (3) divisibility (the user can try part of the idea), (4) communicability, (5) complexity, (6) product risks (will it actually provide the benefits?), and (7) psychological risks (e.g., will people laugh at me if I adopt this new product or idea?).

Diffusion index. The percentage of components in a selected collection of time-series indicators that are increasing. Given one hundred components of the same size, the index would be 40 percent when 40 were expanding, and zero when none were increasing.

Disaggregation. See segmentation.

Disconfirming evidence. Evidence that refutes one's beliefs or forecasts. Substantial evidence shows that people do not use disconfirming evidence effectively, especially if received on a case-by-case basis. Tetlock (1999), in a long-term study of political, economic, and military forecasts, shows how people use a variety of belief-system defenses, which makes learning from history a slow process. PoFxxx

Discontinuity. A large shift in a time series that is expected to persist. The effect is usually a change in level but can also be a change in trend. Trend discontinuities are difficult to estimate, so it might be best to assume that the change occurred only in the level, although this is speculative. Discontinuities play havoc with quantitative approaches to extrapolation (Armstrong and Collopy 1992). PoFxxx

Discrete event. A one-time event that causes outliers or changes in time-series patterns. Examples of such events are a factory closing, a hurricane, or a change in the products offered.

Discriminant analysis. A variation of regression analysis used to predict group membership. The dependent variable is based on categorical data. The simplest variation is a dependent variable with two categories (e.g., "accepted bribe" vs. "did not accept bribe," "bid accepted" vs. "bid rejected," or "survived medical operation" vs. "died"). PoFxxx

Disjunctive model. A nonlinear judgment model that combines variables (cues) to ensure, say, that at least one cue must take on a high value before the forecast generated by the model will be high.

Domain expert. A person who knows a lot about the situation being forecast, such as an expert in automotive marketing, restaurant management, or the weather in a given region.

Domain knowledge. Expert's knowledge about a situation, such as knowledge about a brand and its market. This knowledge is a subset of the contextual information for a situation. PoFxxx

Double cross-validation. A procedure used to test predictive validity, typically with longitudinal or cross-sectional data. The data to be analyzed are split into two roughly equal subsets. A model is estimated on one subset and its ability to forecast is tested on the other half. The model is then estimated for the other subset, which is then used to forecast for the first subset. This procedure requires a large sample size. (Also see jackknife.)

Double moving average. A moving average of a series of data that already represents a moving average. It provides additional smoothing (the removal of more randomness than an equal-length single moving average).

Dummy variable. An explanatory variable that assumes only two values, 0 or 1. In a regression analysis, the coefficient of a dummy variable shows the average effect on the level of the dependent variable when the dummy variable assumes the value of 1. For example, a dummy variable might represent the presence or absence of capital punishment in a geographical region, and its regression coefficient could show the effect of capital punishment on the level of violent crime. More than two categories can be handled by using additional dummy variables; for example, to represent three political affiliations (e.g., Republican, Democrat, or Other) in a model to predict election outcomes, one could use two dummy variables ("Republican or not?" and "Democrat or not?"). One needs v-1 dummy variables to represent v variables. PoFxxx

Durbin-Watson statistic. A measure that tests for autocorrelation between error terms at time t and those at t + 1. Values of this statistic range from 0 to 4. If no autocorrelation is present, the expected value is 2. Small values (less than 2, approaching 0) indicate positive autocorrelation; larger values (greater than 2, approaching 4) indicate negative autocorrelation. Is autocorrelation important to forecasting? It can tell you when to be suspicious of tests of statistical significance, and this is important when dealing with small samples. However, it is difficult to find empirical evidence showing that knowledge of the Durbin-Watson statistic leads to accurate forecasts or to well-calibrated prediction intervals. Forecasters are fond of reporting the D-W statistic, perhaps because it is provided by the software package. Do not use it for cross-sectional data as they have no natural order.

Dynamic regression model. A regression model that includes lagged values of the explanatory variable(s) or of the dependent variable or both. The relationship between the forecast variable and the explanatory variable is modeled using a transfer function. A dynamic regression model can predict what will happen if the explanatory variable changes.

Eclectic research. A set of research studies having the same objective but using procedures that differ substantially from one another. This has also been called the multi-trait multi-method approach, convergent validation, and methodological triangulation. By varying the approach, one hopes to identify and compensate for mistakes and biases. Eclectic research can be used to estimate parameters for econometric models and to assess their construct validity. Armstrong (1985, pp. 205-214) provides examples and evidence on its value. PoFxxx

Econometric method. Originally, the application of mathematics to economic data. More specifically, the statement of theory followed by the use of objective measurement methods, usually regression analysis. The econometric method might be viewed as the thinking-man's regression analysis. It consists of one or more regression equations. The method can be used in economics, in other social sciences (where some people refer to these as "linear models"), and in the physical sciences. It can be applied to time series, longitudinal, or cross-sectional data. For a detailed description of econometric methods, see Allen and Fildes (2001). PoFxxx

Econometric model. One or more regression equations used to capture the relationship between the dependent variable and explanatory variables. The analyst should use a priori analysis to specify a model (or a set of feasible models) and then calibrate the model parameters by minimizing the sum of the squared errors in the calibration data. The parameters can also be estimated by minimizing the least absolute values.

Economic indicator. A time series that has a reasonably stable statistical relationship to the whole economy or to time series of particular interest. Coincident indicators are often used to identify turning points in aggregate economic activity and leading indicators to forecast such turning points.

Efficient. The characteristic of a forecast or estimate that cannot be improved by further analysis of the calibration data.

Elasticity. A measure of the relationship between two variables. Elasticity expresses the percentage change in the variable of interest that is caused by a 1% change in another variable. For example, an income elasticity of +1.3 for unit automobile sales means that a 1% increase in income will lead to an increase of 1.3% in the unit sales of automobiles. It is typically easier to think about elasticities than about marginal propensities (which show the unit change in the dependent variable Y when X is changed by one unit).

Encompassing model. A model whose forecast errors explain the errors produced by a second model.

Endogenous variable. A variable whose value is determined within the system. For example, in an econometric model, the market price of a product may be determined within the model, thus making it an endogenous variable. (See also exogenous variable.)

Ensemble. The average of a set of forecasts. This term is used in weather forecasting. See combining forecasts.

Environment. Conditions surrounding the situation. The environment includes information about the ranges and distributions of cues, the correlations among them, and the relations between the cues and the event being judged. In judgmental forecasting, the environment includes constraints on information available to the judge and on actions the judge may take, as well as time pressures, requirements for documentation, and anything else that might affect cognitive processes. Alternatively, environment refers to the general situation when using an econometric model.

Equilibrium correction model. See error correction model.

Error term. The difference between the actual values and the forecasted values. The error term is a random variable at time t whose probability distribution is assumed to have a mean of zero and is usually assumed to have a constant variance at all time periods and a normal distribution.

Error correction model. A model that explains changes in the dependent variable in terms of changes in the explanatory variables as well as deviations from the long-run relationship between the dependent variable and its determinants. Do error correction models lead to more accurate forecasts? The jury is still out. PoFxx.

Error cost function. The economic loss related to the size of errors. It is difficult to generalize about this. The suggested procedure is to leave this aspect of the problem to the planners and decision makers.

Error distribution. The theoretical probability distribution of forecast errors. It is often assumed to be normal. In the social sciences, this assumption is generally reasonable for short-interval time-series data (say, monthly or less), but not for annual data.

Error ratio. The error of a selected forecasting method divided by that for a benchmark forecast. The term is commonly used in judgmental forecasting. It is also used in quantitative forecasting. See Theil's U and Relative Absolute Error.

Estimate-Talk-Estimate (E-T-E). A structured procedure calling for independent and anonymous judgments, followed by a group discussion, and another round of individual judgments. It is also called mini-Delphi. See Delphi technique.

Estimation sample. See calibration data.

Estimation. Finding appropriate values for the parameters of an equation based on a criterion. The most commonly used criterion is minimizing the Mean Squared Error. Sometimes an iterative procedure is needed to determine parameter values that minimize this criterion for the calibration data.

E-T-E. See Estimate-Talk-Estimate.

Event modeling. A feature of some exponential smoothing programs that allows the user to specify the time of one or more special events, such as irregular promotions and natural disasters, in the calibration data. For each type of special event, the effect is estimated and the data adjusted so that the events do not distort the trend and seasonal patterns of the time series. Some programs use a procedure called intervention analysis to model events.

Ex ante forecast. A forecast that uses only information that would have been available at the forecast origin; it does not use actual values of variables from later periods. This term, often used interchangeably with unconditional forecast, is what we normally think of as a forecast. It can refer to holdout data (assuming the values to be unknown) or to a situation in which the event has not yet occurred (pure ex ante). See Armstrong 2001d.

Exogenous variable. A variable whose value is determined outside of the model. For example, in an econometric model, the gross national product might be an exogenous variable.

Expectations surveys. Surveys of how people or organizations expect that they will behave in given situations. See also intentions surveys. PoFxxx

Experimental data. Data from situations in which a researcher has systematically changed certain variables. These data could come from laboratory experiments, in which the researcher controls most of the relevant environment, or field experiments, in which the researcher controls only part of the relevant environment. (See quasi-experimental data.)

Experiments. Changes in key variables that are introduced in a systematic way to allow for an examination of the effects that one variable has on another. For example, a firm could charge different prices in different geographical regions to assess price elasticity. In a sense, it involves doing something wrong (not charging the apparently best price) to learn. In addition to helping analysts develop forecasting models, experiments are useful in persuading decision makers to accept new forecasting methods. Whereas people are often willing to reject a new idea, they are less likely to reject a request to do an experiment. Armstrong (1982b) conducted an experiment in which subjects were asked to describe how they would gain acceptance of a model to predict the outcome of medical treatment for patients. Only one of the 16 subjects said that he would try an experiment. Armstrong then presented the situation as a role-playing case to 15 groups of health-care executives; only one group proposed an experiment, and this group was successful at implementing change while all other groups failed. Finally, Armstrong gave 14 groups instructions on how to propose experiments in this situation; of these, 12 were successful at gaining acceptance in role-playing exercises. PoFxxx

Expertise. Knowledge or skill in a particular task. In forecasting, this might be assessed by the extent to which experts' forecasts are more accurate than those by nonexperts. See also seer-sucker theory.

Expert opinions. Predictions of how others will behave in a particular situation, made by persons with knowledge the situation. Rowe and Wright (2001) discuss principles for the use of expert opinions. Most important forecasts rely on unaided expert opinions. Research has led to many principles to improve forecasting with expert opinions. For example, forecasters should obtain independent forecasts from 5 to 20 experts (based on research findings by Ashton 1986; Hogarth 1978; and Libby and Blashfield 1978). PoFxxx

Expert system. A model designed to represent procedures that experts use in making decisions or forecasts. Often, these procedures are supplemented by other information, such as estimates from econometric models. The term has also been applied to procedures for selecting forecasting methods. Armstrong, Adya and Collopy (2001) discuss principles for developing expert systems for forecasting. PoFxxx

Explanation effect. The increase in the perceived likelihood of an event's occurrence that results from explaining why the event might occur. This effect is relevant to conjoint analysis and to expert opinions (Arkes 2001). On the positive side, it can cause decision makers to pay attention to a possible outcome; as a result, it can contribute to scenarios. PoFxxx

Explanatory variable. A variable included in an econometric model to explain fluctuations in the dependent variable. (See also causal variable.)

Exploratory research. Research carried out without hypotheses. The data are allowed to speak for themselves. Exploratory research can be a worthless or even dangerous practice for forecasters. On the other hand, it might provide ideas that can subsequently be tested. It is most useful in the early stages of a project when one knows little about the problem.

Exponential smoothing. An extrapolation procedure used for forecasting. It is a weighted moving average in which the weights are decreased exponentially as data becomes older. For most situations (but not all), it is more accurate than moving averages (Armstrong 2001c). In the past, exponential smoothing was less expensive than a moving average because it used only a few values to summarize the prior data (whereas an *n*-period moving average had to retain all *n* values). The low cost of computer storage has reduced this advantage. When seasonal factors are difficult to measure, moving averages might be preferred to exponential smoothing. For example, a 12-month moving average might be useful in situations with much seasonal variation and less than four years of data. A comprehensive treatment of exponential smoothing is provided in Gardner (1985). See also Holt-Winters exponential smoothing method and state-space model. PoFxxx

Ex post forecast. A forecast that uses information from the situation being forecast. The actual values of the causal variables are used, not the forecasted values; however, the parameters are not updated. This term is used interchangeably with conditional forecast. It can help in assessing predictions of the effects of change in explanatory variables.

Extrapolation. A forecast based only on earlier values of a time series or on observations taken from a similar set of cross-sectional data. Principles for extrapolation are described in Armstrong (2001c). PoFxxx

Face validity. Expert opinion that a procedure represents what it purports to represent. To obtain a judgment on face validity, ask a few experts what they expect. For example, you might ask them to specify variables and relationships for an econometric model. Agreement among experts is evidence of face validity.

Facilitator. A group member whose only role is to help the group to function more effectively by following a structured procedure. One of the dominant conclusions about judgmental forecasting is that structure contributes to forecast accuracy.

Factor analysis. A statistical procedure for obtaining indices from variables by combining those that have high correlations with one another. Factor analysis has been used to develop predictive indices, but this has not been successful; Armstrong (1985, p. 223) reports on eight studies, all failures in this regard.

Feature identification. The identification of the conditions (features) of a set of data. Features can help select an extrapolation method, as described in Armstrong, Adya and Collopy (2001).

Features. Operational measures of the characteristics of time-series or cross-sectional data. Examples include basic trend, coefficient of variation, and discontinuity. PoFxxx

Feedback. Information that experts receive about the accuracy of their forecasts and the reasons for the errors. Accurate, well-summarized feedback is probably the primary basis experts have for improving their judgmental forecasts. The manner in which feedback is provided is critical because people tend to see what they want to see or what they expect. When feedback is well-summarized, frequent, and when it contains explanations for the events, judgmental forecasters can become well-calibrated. Weather forecasters receive this kind of feedback, and they are almost perfectly calibrated: it rains on 80% of the days on which they predict an 80% chance of rain (Murphy and Winkler 1984). Well-structured feedback is especially important when it involves disconfirming evidence. PoFxxx

File. A collection of data.

Filter. A process developed in engineering for eliminating random variations (high or low frequencies) in an attempt to ensure that only the true pattern remains. For example, a filter might adjust outliers to be within two or three sigmas (standard deviations) of forecasted or fitted values.

First differences. See differencing.

Fisher exact test. A nonparametric test used to assess relationships among variables in a 2 ? 2 table when samples are small. Siegel and Castellan (1988) provide details on calculating this and other nonparametric statistics.

Fit. The degree to which a model explains (statistically speaking) variations in the calibration data. Fit is likely to be misleading as a criterion for selecting and developing forecasting models, because it typically has only a weak relationship to ex ante forecast accuracy (Armstrong 2001d). Fit tends to favor complex models, and these models often do not hold up in forecasting, especially when using time-series data. Nevertheless, Pant and Starbuck (1990) found a modest relationship between fit (when using MAPE) and short-term forecasts for 13 extrapolation methods. It is more relevant when working with cross-sectional data. PoFxxx

Focus group. A group convened to generate ideas, where a facilitator uses nondirective interviewing to stimulate discussion. Fern (1982) found that such groups are most useful when, in the real situation, people's responses depend to some extent on their peers' beliefs. This could include responses to visible products, such as clothing or automobiles. Focus groups might be used to generate ideas about variables for judgmental bootstrapping or conjoint analysis when the forecasting problem involves visible products. In general, however, there are better (and less expensive) ways to obtain information, such as personal interviews. Focus groups should not be used to make forecasts. (Alas, in the real world, they are used to make poor but convincing forecasts.) PoFxxx

Forecast. A prediction or estimate of an actual value in a future time period (for time series) or for another situation (for cross-sectional data). Forecast, prediction, and prognosis are typically used interchangeably.

Forecast accuracy. The optimist's term for forecast errors.

Forecast competition. A competition in which forecasters are provided with the same calibration data, and they independently make forecasts for a set of holdout data. Ideally, prior to the competition, competitors should state hypotheses on the conditions under which their methods will be most accurate. Then they submit forecasts to an administrator who calculates the forecast errors. There have been a number of competitions for extrapolation methods (for example, see the M-Competition).

Forecast criteria. Factors used to evaluate and compare different forecasting techniques. Forecast accuracy is generally considered the most important criterion, but Yokum and Armstrong (1995) showed that others, such as ease of interpretation and cost savings, may be as important when the forecasting situation or the forecaster's role is considered.

Forecast error. The difference between the forecasted value (F) and the actual value (A). By convention, the error is generally reported as F minus A. Forecast errors serve three important functions: (1) *The development of prediction intervals*. Ideally, the errors should be obtained from a test that closely resembles the actual forecasting situation. (2) *The selection (or weighting) of forecasting methods*. Thus, one can analyze a large set of forecasts and then select based on which method produced the more accurate forecasts. In such evaluations, the error term should be immune to the way the series is scaled (e.g., multiplying one of the series by 1,000 should not affect the accuracy rankings of various forecasting methods). Generally, the error measure should also be adjusted for the degree of difficulty in forecasting. Finally, the measure should not be overly influenced by outliers. The Mean Squared Error, which has been popular for years, should not be used for forecast comparisons because it is not independent of scale and it is unreliable compared to alternative measures. More appropriate measures include the APE (and the MdMAPE when summarizing across series) and the Relative Absolute Erros (and the MdRAE when summarizing across series). (3) *Refining forecasting models*, where the error measures should be sensitive to changes in the models being tested. Here, medians are less useful; the APE can be summarized by its mean (MAPE) and the RAE by its geometric mean (GmRAE). Armstrong and Collopy (1992a) provide empirical evidence to support these guidelines, and the measures are discussed in Armstrong (2001d).

Forecast horizon. The number of periods from the forecast origin to the end of the time period being forecast.

Forecast interval. See prediction interval.

Forecast validity. See predictive validity.

Forecast variable. The variable of interest. A variable that is predicted by some other variable or variables; it is also called the dependent variable or response variable.

Forecasting. Estimating in unknown situations. *Predicting* is a more general term and connotes estimating for any time series, cross-sectional, or longitudinal data. *Forecasting* is commonly used when discussing time series.

Forecasting competition. See forecast competition.

Forecasting engine. The module of a forecasting system containing the procedures for the estimation and validation of forecasting models.

Forecasting model. A model developed to produce forecasts. It should be distinguished from a measurement model. A forecasting model may draw upon a variety of measurement models for estimates of key parameters. A forecaster might rely on different models for different parts of the forecasting problem, for example, using one model to estimate the level in a time-series forecast and another to forecast change.

Forecasting support system. A set of procedures (typically computer based) that supports forecasting. It allows the analyst to easily access, organize, and analyze a variety of information. It might also enable the analyst to incorporate judgment and monitor forecast accuracy.

Framing. The way a question is asked. Framing can have an important effect upon subjects' responses, so it is important to ensure that questions are worded properly. The first influential treatment of this issue was by Payne (1951). Much useful work followed, summarized by Sudman and Bradburn (1982). Knowledge of this work is important in conducting intentions studies, eliciting expert opinions, and using methods that incorporate judgmental inputs. Consider the effect of the wording in the following example provided by Norman R. F. Maier: "A man bought a horse for $60 and sold it for $70. Then he bought it back again for $80 and sold it for $90. How much money did he make in the horse trading business?" Almost half of the respondents answered incorrectly. Now consider this question: "A man bought a horse for $60 and sold it for $70. Then he bought a pig for $80 and sold it for $90. How much money does he make in the animal trading business?" Almost all respondents get the correct answer to this version of the question ($20). Tversky and Kahneman (1981) demonstrated biases in peoples' responses to the way that questions are framed. For example, they asked subjects to consider a hypothetical situation in which a new disease is threatening to kill 600 people. In Program A, 200 people will be saved, while in Program B, there is a one-third chance of saving all 600 people, but a two-thirds chance of saving none of them. In this case, most respondents chose Program A (which is positively framed in terms of saving lives). However, when the question was reframed with Program A leading to 400 deaths, and Program B as having a one-third chance that nobody would die and a two-thirds chance that that all would die, then the majority of respondents chose Program B (this alternative is negatively framed in terms of losing lives). This negative way of framing the question caused people to respond differently, even though the two problems are identical. This example implies that framing could play a role in writing scenarios. The discovery of biases due to framing seems to outpace research on how to avoid them. Unfortunately, telling people about bias usually does little to prevent its occurrence. Beach, Barnes and Christensen-Szalanski (1986) concluded that observed biases may arise partly because subjects answer questions other than those the experimenter intended. Sudman and Bradburn (1982) provide a number of solutions. Two procedures are especially useful: (1) pretest questions to ensure they are understood, and (2) ask questions in alternative ways and compare the responses. Plous (1993, chapter 6) provides additional suggestions on framing questions.

F-test. A test for statistical significance that relies on a comparison of the ratio of two mean square errors. For example, one can use the ratio of "mean square due to the regression" to "mean square due to error" to test the overall statistical significance of a regression model. $F = t^2$ (see t-test).

Function. A formal statement of the relationship between variables. Quantitative forecasting methods rely on functional relationships between the item to be forecast and previous values of that item, previous error values, or explanatory variables.

Functional form. A mathematical statement of the relationship between an explanatory variable (or time) and the dependent variable.

Gambler's fallacy. The notion that an unusual run of events, say a coin coming up heads five times in a row, indicates a likelihood of a change on the next event to conform with the expected average (e.g., that tails is more likely than heads on the next toss). The reason, gamblers say, is the law of averages. They are wrong. The gambler's fallacy was examined by Jarvik (1951).

Game theory. A formal analysis of the relationships between competing parties who are subject to certain rules. The Prisoner's Dilemma is one of the more popular games that had been studied. Game theory seems to provide insight into complex situations involving conflict and cooperation. Brandenburger and Nalebuff (1996) describe such situations. Although game theory has been the subject of enormous research, no evidence exists that it is helpful in forecasting. To be useful, the rules of the game must match the real world, and this is typically difficult to do. In contrast, role playing provides a way to represent the actual situation, and it has been shown to produce accurate predictions in such cases (Armstrong 2001a). PoFxxx

GARCH. A Generalized AutoRegressive Conditionally Heteroscedastic model contains an equation for changing variance. GARCH models are primarily used in the assessment of uncertainty. A GARCH equation of order (p, q) assumes that the local variance of the error terms at time t is linearly dependent on the squares of the last p values of the error terms and the last p values of the local variances. When q is zero, the model reduces to an ARCH model.

Generalized least squares (GLS). A method for estimating a forecasting model's parameters that drops the assumption of independence of errors and uses an estimate of the errors' interrelationships. In the Ordinary-Least-Squares (OLS) estimation of a forecasting model, it is assumed that errors are independent of each other and do not suffer from heteroscedasticity. Whether GLS is useful to forecasters has not been established. OLS generally provides sufficient accuracy.

Genetic algorithm. A class of computational heuristics that simulate evolutionary processes using insights from population dynamics to perform well on an objective function. Some analysts speculate that competition among forecasting rules will help to develop a useful forecasting model, but it is difficult to find empirical support for that viewpoint.

Global assessment. An overall estimate (in contrast to an explicit estimate of parts of a problem). An expert forecast made without an explicit analysis. (See also intuition.)

Goodness of fit. A measure of how well a model explains historical variations in calibration data. PoFxxx

Growth cycle. See deviation cycle.

Growth forces. Forces that tend to drive a series up. For example, actively marketing a product and participating in a developing market are growth forces. Growth forces could be found for products such as computers since the 1960s. PoFxxx

Heteroscedasticity. Nonconstant variances in a series (e.g., differing variability in the error terms over the range of data). Often found when small values of the error terms correspond to small values of the original time series and large error terms correspond to large values. This makes it difficult to obtain good estimates of parameters in econometric models. It also creates problems for tests of statistical significance. Log-log models generally help to reduce heteroscedasticity in economic data.

Heuristic. From the Greek word, meaning *to discover* or *find.* Heuristics are trial-and-error procedures for solving problems. They are simple mental operations that conserve effort. Heuristics can be used in representing expert systems.

Hierarchical model. A model made up of submodels of a system. For example, a hierarchical model of a market like automobiles could contain models of various submarkets, like types of automobiles, then brands.

Hierarchy of effects. A series of psychological processes through which a person becomes aware of a new product or service and ultimately chooses to adopt or reject it. Hierarchy of effects models can be used to forecast behavioral changes, such as programs to reduce smoking. These processes consist of sequential stages, including awareness, knowledge, liking, preference, and choice. Forecasting models can be developed for each of these stages by including policy variables critical to that stage (e.g., promotions for awareness, informational advertising for knowledge, and comparative advertising for liking).

Hindsight bias. A tendency to exaggerate in hindsight how accurately one predicted or would have been able to predict by foresight. Sometimes referred to as the “I knew it all along” effect. Forecasters usually "remember" that the forecasts were more accurate. Because of hindsight bias, experts may be overconfident about later forecasts. To reduce hindsight bias, ask forecasters to explicitly consider how past events might have turned out differently. Much research on hindsight bias was apparently stimulated by Fischhoff (1975), which was cited by about 400 academic studies as of the end of 1999. A meta-analysis was published by Cristensen-Szalanski (1991). For a discussion of principles relating hindsight bias to forecasting, see Fischoff (2001) and PoFxxx

Hit rate. The percentage of forecasts of events that are correct. For example, in conjoint analysis, the hit rate is the proportion of correct choices among alternative objects in a holdout task.

Holdout data. Data withheld from a series that are not used in estimating parameters. These holdout data can then be used to compare alternative models. See post-sample evaluation and ex ante forecast. For a discussion of the types of holdout data, see Armstrong (2001d).

Holdout task. In conjoint analysis, respondents use holdout data to make choices from sets of alternative objects described on the same attributes (Wittink and Bergesteum 2001). Ideally, holdout choice sets have characteristics that resemble actual choices respondents will face in the future.

Holt's exponential smoothing method. An extension of single exponential smoothing that allows for trends in the data. It uses two smoothing parameters, one for the level and one for the trend. (See discussion in Armstrong 2001c.)

Holt-Winters' exponential smoothing method. An extension of Holt's exponential smoothing method that includes seasonality (Winters 1960). This form of exponential smoothing can be used for less-than-annual periods (e.g., for monthly series). It uses smoothing parameters to estimate the level, trend, and seasonality. An alternative approach is to deseasonalize the data (e.g., via Census Program X-12), and then use exponential smoothing. There is little evidence on which seasonality procedure is most accurate. See state-space model.

Homoscedasticity. Variability of error that is fairly constant over the range of the data.

Horizon. See forecast horizon.

Identification. A step in building a time-series model for ARMA and ARIMA in which one uses summary statistics, such as autocorrelation functions or partial autocorrelation functions, to select appropriate models for the data. The term is also used for econometric models.

Illusion of control. An erroneous belief that one can control events. People who have no control over events often think they can control them. As Mark Twain said in describing a fight. “Thrusting my nose firmly between his teeth, I threw him heavily to the ground on top of me.” Even gamblers have an illusion of control (Langer and Roth 1975).

Inconsistent trends. A condition for time series when the basic (long-term) trend and the recent (short-term) trend are forecasted to be in opposite directions. When it occurs, trend extrapolation is risky. One strategy is to blend the two trends as one moves from the short to the long term. A more conservative strategy is to forecast no trend. For evidence on how inconsistent trends affect forecast errors, see Armstrong, Adya and Collopy (2001). See also consistent trends. PoFxxx

Independent variable. A variable on the right-hand side of a regression. It can be used as a predictor. It includes time, prior values of the dependent variable, and causal variables**.** See explanatory variable.

Index numbers. Numbers that summarize the level of economic activity. For example, the Federal Reserve Board Index of Industrial Production summarizes a number of variables that indicate the overall level of industrial production activity. Index numbers can control for scale in forecasting.

Index of Predictive Efficiency (IPE). IPE = (E1-E2)/ E1, where E1 is the error for the benchmark forecast, which might be based, say, on the method currently used. The measure was proposed by the sociologists, Ohlin and Duncan (1949), for cross-sectional data. The comparison to a benchmark is also used in Theil’s U and in the Relative Absolute Error.

Inductive technique. A technique that searches through data to infer statistical patterns and relationships. For example, judgmental bootstrapping induces rules based on forecasts by an expert.

Initializing. The process of selecting or estimating starting values when analyzing calibration data.

Innovation. In general, something new. Forecasters use the term to refer to the disturbance term in a regression or to an event that causes change in a time series. (Also see diffusion.)

Input-output analysis. An examination of the flow of goods among industries in an economy or among branches of an organization. An input-output matrix is used to show interindustry or interdepartmental flows of goods or services in the economy, or in a company and its markets. The matrix can be used to forecast the effects of a change in one industry on other industries (e.g., the effects of a change in oil prices on demand for cars, then steel sales, then iron ore, and then limestone.) Although input-output analysis led to one Nobel prize (Wassily Leontief’s in 1964), its predictive validity has not been well-tested. However, Bezdek (1974), in his review of 16 input-output forecasts in seven countries made between 1951 and 1972, concluded that input-output forecasts were more accurate than those from alternative techniques.

Instabilities. Changes resulting from unidentified causes in the pattern of a time series, such as a discontinuity or a change in the level, trend, or seasonal pattern.

Integrated. A characteristic of time-series models (the I in ARIMA models) in which one or more of the differences of the time-series data are included in the model. The term *integrated* is used because the original series may be recreated from a differenced series by summation.

Intentions survey. A survey of how people say they will act in a given situation. See also expectations surveys and Juster scale. Especially useful for new products, but also used to supplement behavioral data (such as sales) as shown in Armstrong, Morwitz and Kumar (2000). See Morwitz (2001). PoFxxx

Interaction. A relationship between a predictor variable (X_1) and the dependent variable (Y) that depends upon the level of another predictor variable (X_2). (There may be main effects as well.) To address problems containing interaction, consider a program such as AID. It is difficult to find evidence that interaction terms in regression analysis contribute to forecast accuracy.

Intercept. The constant term in regression analysis. The regression's intersection with the *Y*-axis. If the explanatory variable *X* is 0, then the value of the forecast variable, *Y*, will be the intercept value. The intercept has no meaning in the traditional log-log model; it is simply a scaling factor.

Interdependence. A characteristic of two or more variables that are mutually dependent. Thus, a change in the value of one of the variables would correlate with a change in the value of the other variable. However, correlation does not imply interdependence.

Intermittent demand. See intermittent series.

Intermittent series. A term used to denote a time series of non-negative integer values where some values are zero. For example, shipments to a store may be zero in some periods because a store's inventory is too large. In this case, the demand is not zero, but it would appear to be so from the data. Croston's method (Croston 1972) was proposed for this situation. It contains an error that was corrected by Rao (1973). Willemain et al. (1994) provide evidence favorable to Croston's method. Other procedures such as aggregating over time can also be used to solve the problem. See Armstrong (2001c). PoFxxx

Interpolation. The process of using some observations to estimate missing values in a series.

Interrater reliability. The amount of agreement between two or more raters who follow the same procedure. This is important for judgmental forecasting or for assessing conditions in a forecasting problem or when using judgmental inputs for an econometric model.

Interrupted series. See intermittent series.

Interval scale. A measurement scale where the intervals are meaningful, but the zero point of the scale is not meaningful (e.g., the Fahrenheit scale for temperature).

Intervention analysis. A procedure to assess the effects on the forecast variable of large changes such as a new advertising campaign, strike, or reduced tax. Intervention models can use dummy variables to represent interventions.

Intuition. A person's immediate apprehension of an object without the use of any reasoning process. An unstructured judgmental impression. Intuitions may be influenced by subconscious cues. When one has much experience and there are many familiar cues, intuition can lead to accurate forecasts. However, it is difficult to find published studies in which intuition is superior to structured judgment.

Ipsative scores. An individual's rating of the relative importance of an item compared with other items. Ipsative scores do not allow for comparisons among people; e.g., Lloyd likes football better than basketball, while Bonnie likes basketball better than football. Does Bonnie like basketball better than Lloyd likes basketball? You do not have enough information to answer that question. Hence, when using intentions or preferences to forecast, ipsative scores can be misleading and difficult to interpret. Guard against this problem by finding other ways for framing questions.

Irregular demand. See intermittent series.

Jackknife. A procedure for testing predictive validity with cross-sectional data or longitudinal data. Use N-1 observations to calibrate the forecasting model, then make a forecast for the remaining observation. Replace that observation and draw a new observation. Repeat the process until predictions have been made for all observations. Thus, with a sample of 57 observations, you can make an out-of-sample forecast for each of the 57 observations. This procedure is also called N-way cross validation.

Judgmental adjustment. A subjective change that a forecaster makes to a forecast produced by a model. Making such changes is controversial. In psychology, extensive research on cross-sectional data led to the conclusion that one should not subjectively adjust forecasts from a quantitative model. Meehl (1954) summarized a long stream of research on personnel selection and concluded that employers should not meet job candidates because that would lead them to improperly adjust a model's prediction as to their success. In contrast, studies on economic time series show that judgmental adjustments sometimes help, although mechanical adjustments seem to do as well. Armstrong (1985, pp. 235-238) summarizes seven studies on this issue. The key is to identify the conditions under which to make adjustments. Adjustments seem to improve accuracy when the expert has knowledge about the level. Judgmental adjustments are common. According to Sanders and Mandrodt's (1990) survey of forecasters at 96 US corporations, about 45% of the respondents claimed that they always made judgmental adjustments to statistical forecasts, while only 9% said that they never did. The main reasons the respondents gave for revising quantitative forecasts were to incorporate "knowledge of the environment" (39%), "product knowledge" (30%), and "past experience" (26%). While these reasons seem sensible, such adjustments are often made by biased experts. In a survey of members of the International Institute of Forecasters, 269 respondents were asked whether they agreed with the following statement: "Too often, company forecasts are modified because of political considerations." On a scale from 1 = "disagree strongly" to 7 = "agree strongly," the mean response was 5.4. (Details on the survey are provided in Yokum and Armstrong 1995.) In Fildes and Hastings' (1994) survey of 45 managers in a large conglomerate, 64% of them responded "forecasts are frequently politically motivated." For a discussion on principles for making subjective adjustments of extrapolations, see Sanders and Ritzman (2001). PoFxxx

Judgmental bootstrapping. An inductive method of assessing how a person makes a judgmental decision or forecast. The model is inferred statistically by regressing the factors used by an expert against the expert's forecasts. The procedure can also be used for forecasts by a group. See Armstrong (2001b). PoFxxx

Judgmental extrapolation. A subjective extension of time-series data. A time series extended by freehand, also known as bold free hand extrapolation (BFE). This can be done by domain experts, who can use their knowledge as well as the historical data. Most research to date, however, has been done with subjects having no domain knowledge. Interestingly, naive extrapolations have often proven to be as accurate as quantitative extrapolations, perhaps because subjects see patterns that are missed by the quantitative methods. This finding is difficult to believe. In fact, the first paper reporting this finding was soundly rejected by the referees and was published only because the editor, Spyros Makridakis, overrode the referees. The paper (Lawrence, Edmundson and O'Conner 1985) went on to become one of the more highly cited papers in the *IJF* and it stimulated much useful research on the topic. Judgmental extrapolations can sometimes be misleading. In a series of studies, Wagenaar (1978) showed that people can misperceive exponential growth. For a simple example, ask people to watch as you fold a piece of paper a few times. Then ask them to guess how thick it will be if you fold it another 40 times. They will usually reply that it will be a few inches, some say a few feet, and occasionally someone will say a few miles. But if they calculated it, they would find that it would extend past the moon. Despite the above findings, when the forecaster has substantial domain knowledge, judgmental extrapolation may be advantageous, especially when large changes are involved. For a discussion of principles related to judgmental extrapolation, see Webby, O'Connor and Lawrence (2001).

Judgmental forecasting. A subjective integration of information to produce a forecast. Such methods can vary from unstructured to highly structured.

Judgmental revision. See judgmental adjustment.

Jury of executive opinion. Expert opinions produced by executives in the organization.

Juster scale. An 11-point scale for use in expectations surveys and intentions surveys. The scale was proposed by Juster (1964, 1966), who compared an 11-point scale with a 3-point scale (definite, probable, maybe) in measuring intentions to purchase automobiles. Data were obtained from 800 randomly selected respondents, the long scale being administered to them a few days after the short scale. Subsequent purchasing behavior of these respondents indicated that the longer probability scale was able to explain about twice as much of the variance among the subsequent behavior of the judges as was the shorter scale. In addition, the mean value of the probability distribution for the 800 respondents on the 11-point scale provided a better estimate of the purchase rate for this group than the short scale. Day et al. (1991) concluded that Juster's 11-point purchase probability scale provides substantially better predictions of purchase behavior than intention scales. They based their conclusion on the evidence from their two

New Zealand studies and prior research by Juster (1966), Byrnes (1964), Stapel (1968), and Gabor and Granger (1972). PoFxxx

Kalman filter. An estimation method (for fitting the calibration data) based on feedback of forecast errors that allows model parameters to vary over time. (See state space model.)

Kendall rank correlation. A nonparametric measure of the association between two sets of rankings. It is an alternative to the Spearman rank correlation. Siegel and Castellan (1988) describe this measure and its power. This statistic is useful for comparing methods when the number of forecasts is small, the distribution of the errors is unknown, or outliers exist, such as with financial data. (See statistical significance.)

Lag. A difference in time between an observation and a previous observation. Thus, Y_{t-k} lags Y_t by k periods. See also lead.

Lagged values. See lag.

Lagging index. A lagging index is a summary measure of aggregate economic activity. The last measured indication of a business cycle turning point is sometimes an indication of the next business cycle turn. Some people speculate that the lagging index, when inverted, might anticipate the next business cycle turn.

Lead. A difference in time between an observation and a future observation. Thus, Y_{t+k} leads Y_t by k periods. See also lag.

Lead time. The time between two related events. For example, in inventory and order entry systems, the lead time is the interval between the time an order is placed and the time it is delivered (also called delivery time).

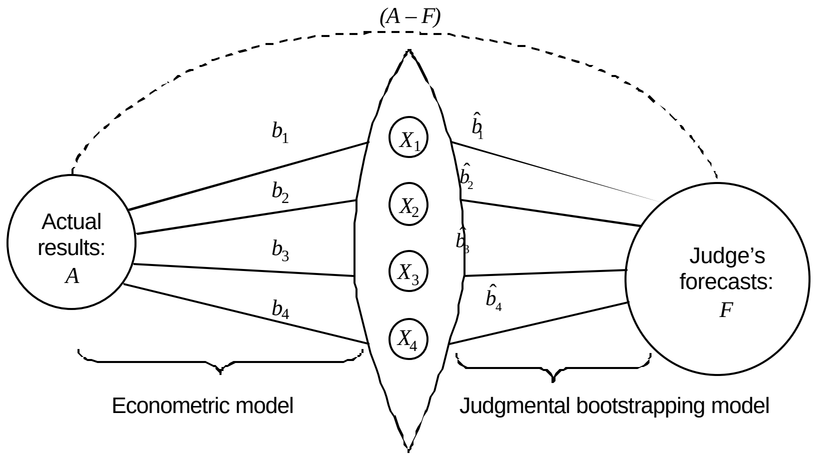
Leading indicator. An economic indicator whose peaks and troughs in the business cycle are thought to lead subsequent turning points in the general economy or some other economic series. But do they really? Here is what William J. Bennett, former U.S. Secretary for Education, said about the U.S. Census Bureau's Index of Leading Economic Indicators in the *Wall Street Journal* on 15 March 1993: "These 11 measurements, taken together, represent the best means we now have of . . . predicting future economic trends." This appears to be a common viewpoint on leading economic indicators. Research on leading economic indicators began in the late 1930s. In 1950, an index of eight leading indicators was developed using data from as far back as 1870. Use of the method spread to at least 22 countries by the end of the century. By the time the U.S. Commerce Department turned the indicators over to the Conference Board in the early 1990s, there had been seven revisions to improve the data. There has long been criticism of leading indicators. Koopmans (1947), in his review of Burns and Mitchell's early work, decried the lack of theory. Few validation studies have been conducted. Auerbach (1982), in a small-scale test involving three-month-ahead ex-ante forecasts of unemployment, found that the use of leading indicators reduced the RMSE slightly in tests covering about 24 years. Diebold and Rudebusch (1991) examined whether the addition of information from the Composite Leading Index (CLI) can improve upon extrapolations of industrial production. They first based the extrapolations on regressions against prior observations of industrial production and developed four models. Using monthly data from 1950 through 1988, they then prepared ex ante forecasts for one, four, eight, and twelve periods ahead using successive updating. The extrapolations yielded a total of 231 forecasts for each model for each forecast horizon. The results confirmed prior research showing that ex post forecasts are improved by use of the CLI. However, inclusion of CLI information reduced ex ante forecast accuracy, especially for short-term forecasts (one to four months ahead). Their findings are weak as they come from a single series. In general then, while leading indicators are useful for showing where things are now, we have only weak evidence to support their use as a forecasting tool. For more on leading indicators, see Lahiri and Moore (1991). PoFxxx

Least absolute values. Regression models are usually estimated using Ordinary Least Squares (OLS). An alternative method is to minimize the sum of absolute errors between the actual observation and its "predicted" (fitted) value for calibration data, a procedure known as least absolute value estimation (LAV). According to Dielman (1986), the LAV method as a criterion for best fit was introduced in 1757. About half a century later, in 1805, least squares was developed. Using Monte Carlo simulation studies, Dielman concluded that, in cases in which outliers are expected, LAV provides better forecasts than does least squares and is nearly as accurate as least squares for data that have normally distributed errors.

Least squares estimation. The standard approach for estimating parameters in a regression analysis, based on minimizing the sum of the squared deviations between the actual and fitted values of the criterion (dependent) variable in the calibration data. (See Ordinary Least Squares.)

Lens model. A conceptual model, proposed by Brunswick (1955), that shows how an expert receives feedback in a situation. The model is related to judgmental bootstrapping and econometric methods, as shown here.

The Brunswick Lens Model of Feedback



The X's are causal variables. The solid lines represent relationships. The b's represent estimated relationships according to the actual data, while the $\hat{b}$'s represent relationships as seen by the judge. The dashed line represents feedback on the accuracy of the judge's predictions. The judgmental bootstrapping model can provide feedback to the judge on how she is making forecasts. The econometric model provides information on the actual relationships. Actual outcomes and a record of forecasts are needed to assess accuracy. Given that the econometric model provides better estimates of relationships, one would expect that such feedback would be the most effective way to improve the accuracy of an expert's forecasts. Newton (1965), in a study involving the prediction of grade-point averages for 53 students, found that feedback from the econometric model was more effective in improving accuracy than was feedback about accuracy or information from the bootstrapping model. For a further discussion on the use of the lens model in forecasting, see Stewart (2001).

Level. The value of a time series at the origin of the forecast horizon (i.e., at time t_0). The current situation.

Lewin's change process. Efforts to implement change should address three phases: Unfreezing, change, and refreezing. In discussing this process, Lewin (1952) used an analogy to ice; it is difficult to change the shape of ice unless you first unfreeze it, then change it and refreeze it. Similarly, when trying to introduce a new forecasting procedure, first ask the clients what they are willing to change (unfreezing). To change, propose experiments. Refreezing involves rewarding new behavior (e.g., showing that the new forecasting procedure continues to be useful). For the change to succeed, the clients should have control over the three stages (for example, they would define how to determine whether the new forecasting method was successful). A number of studies show that change efforts in organizations are more successful when they address the three phases explicitly (e.g., see review of studies provided in Armstrong 1982b). This process can also be used when seeking changes as a result of a forecast. PoFxxx

Linear model. A term used (especially by psychologists) to denote a regression model. The linear model is typically based on causal relationships that are linear in the parameters. In other words, the variables might be transformed in various ways, but these transformed variables are related to each other in a linear fashion, such as $Y = a + b_1 x_1 + b_2 x_2$. See econometric model.

Ljung-Box test. A version of the Box-Pierce test for autocorrelated errors.

Local trend. See recent trend.

Logarithmic transformation. By taking logs of the dependent and explanatory variables, one might be able to remove heteroscedasticity and to model exponential growth in a series. In such a model, the coefficients represent elasticities that are constant over the forecast range; this is a standard assumption in economics.

Logistic. A special case of diffusion in which the probability of a population member adopting an innovation is proportional to the number of current adopters within the population. It is a mathematical representation of "keeping up with the Joneses." If the number of adopters is Y_t and a is the saturation level, then the equation

$$Y_t = \frac{a}{1 + ce^{-bt}}$$

describes the growth of the number of adopters of the innovation over time (b and c are constants controlling the rate of growth). For a discussion of the logistic and related diffusion curves for forecasting, see Meade and Islam (2001).

Logit. A transformation used when the values for the dependent variable are bounded by zero and one, but are *not equal to zero or one.* (The log of zero is minus infinity and it cannot be computed.) Thus, it is appropriate for series based on percentages, such as market-share predictions. Transform the dependent variable as follows:

$$\text{logit } (Y) = \log\left(\frac{p}{1 - p}\right)$$

Log-log model. A model that takes the logs (to the base e or base 10) of the Y and X variables. (See logarithmic transformation.) Econometric models are often specified as log-log under the assumption that elasticities are constant. This is done to better represent behavioral relationships, to make it easier to interpret the results, to permit a priori analysis, and to better represent the relationships.

Longitudinal data. Data that represent a collection of values recorded between at least two times for a number of decision units. (See panel data.) For example, one might examine data on 30 countries in 1950 and on the same countries in 2001 in order to determine whether changes in economic well-being are related to reported happiness levels.

Long range. The period of time over which large changes are expected. Long range for the bread industry might be 20 years, while long range for the internet industry might be one year.

Long-run effect. The full effect that a change in a causal variable has on the dependent variable. In a regression model, where $Y = a + bX$, a shift in X has an instantaneous effect (of b) on Y. In dynamic regression, there are lags in either X or Y in the model. A shift in X also has a long-run effect, which may either amplify or damp the short-run effect. When using causal variables in a forecasting model, one is typically concerned with long-run effects. Thus, it is inadvisable to formulate a model on first differences.

Long-run relationship. An effect of a predictor (X) on the dependent variable (Y) that is expected to hold over a long forecast horizon. (See long-run effect.)

Long waves. Very long-term business cycles. A Russian economist, Nikolai D. Kondratieff, introduced the term in a series of papers in the 1920s arguing that "on the basis of the available data, the existence of long waves of cyclical character is very probable." Kondratieff (1935) presented no theory as to why cycles of 40 to 60 years should be characteristic of capitalist countries, but he did associate various "empirical characteristics" with phases of his long waves, which he professed to find in France, England, the United States, Germany, and the "whole world." According to his predictions, a long decline would have begun in the 1970s and continue until the first decade of the 21st century. People actually paid attention to such strange ideas.

Loss function. An expression that represents the relationship between the size of the forecast error and the economic loss incurred because of that error. PoFxxx

MAD (Mean Absolute Deviation). An estimate of variation. It is an alternative to the standard deviation of the error. The ratio of standard deviation to MAD is 1.25 for normal distributions, and it ranges from 1.0 to 1.5 in practice. See Mean Absolute Error.

Market potential. The maximum total sales that might be obtained for a given product. (Also see saturation level.)

Markov chains. A method of analyzing the pattern of decision-making units in moving from one behavior state to another. Construct a transition matrix to show the proportion of times that the behavior in one trial will change (move to another state) in the next trial. If the transition process remains stable and if the sample of actors is representative of the entire population, the matrix can be used to forecast changes. However, there is a problem. Forecasts are most useful when changes occur. But given the assumption of stability, Markov chains are risky for predicting behavior when organizations make efforts to change behavior and thus to change the transition matrix. Markov chains have been recommended for predictions in marketing when people are assumed to go through various states in using a product (e.g., trial, repeat purchase, and adoption) and for cases in which consumers purchase different brands. Early published applications of Markov chains covered problems such as predicting changes in the occupational status of workers, identifying bank loans that will go into default, and forecasting sales in the home-heating market. Despite many research publications on Markov chains, I have been unable to find accounts of research that supports their predictive validity. Armstrong and Farley (1969) compared Markov chains with simple extrapolations in forecasting store visits and Markov chains produced no gains in accuracy. PoFxxx

Martingale. A sequence of random variables for which the expected value of the series in the next time period is equal to the actual value in the current time period. A martingale allows for non-constant variance; a random walk does not.

Maximum likelihood estimation. A method of estimating the parameters in an equation by maximizing the likelihood of the model given the data. For regression analysis with normally distributed errors, maximum likelihood estimation is equivalent to Ordinary Least Squares estimation.

M-Competition. The term used for the series of three comparative studies of extrapolation methods organized by Spyros Makridakis, starting with the 1,001 time-series competition in Makridakis et al. (1982) and including Makridakis et al. (1993) and Makridakis and Hibon (2000). In each study, a number of different experts prepared extrapolations for holdout data. The accuracies of the various methods were then compared by the study's lead author. Raw data and information about these competitions can be found at hops.wharton.upenn.edu/forecast. PoFxxx

Mean Absolute Deviation. See MAD and mean absolute error.

Mean Absolute Error (MAE). The average error when ignoring signs. This can be useful in assessing the cost of errors, such as for inventory control (also called MAD).

Mean Absolute Percentage Error (MAPE). The average of the sum of all the percentage errors for a data set, taken without regard to sign. (That is, the absolute values of the percentage errors are summed and the average is computed.)

Mean Percentage Error (MPE). The average of all of the percentage errors for a given data set. The signs are retained, so it serves as a measure of bias in a forecasting method.

Mean Squared Error (MSE). The sum of the squared forecast errors for each of the observations divided by the number of observations. It is an alternative to the mean absolute deviation, except that more weight is placed on larger errors. (See also Root Mean Square Error.) While MSE is popular among statisticians, it is unreliable and difficult to interpret. Armstrong and Fildes (1995) found no empirical support for the use of the MSE or RMSE in forecasting. Fortunately, better measures are available as discussed in Armstrong (2001d).

Measurement error. Failures, mistakes, or shortcomings in the way a concept is measured.

Measurement model. A model used to obtain estimates of parameters from data. For example, an estimate of price elasticity for a product from household survey data. The measurement model is not the same as the forecasting model.

Median. The value of the middle item in a series of items arranged in order of magnitude. For an even number of items, it is the average of the two in the middle. Medians are often useful in forecasting when the historical data or the errors contain outliers.

Meta-analysis. A systematic and quantitative study of studies. In meta-analysis, an "observation" is a finding from a study. Meta-analysis was applied in 1904 by Karl Pearson, who combined data from British military tests and concluded that the then-current practice of vaccination against intestinal fever was ineffective (Mann 1994). Although meta-analysis had also been used for decades in personnel psychology, Glass (1976) introduced the term. In meta-analysis, one uses documented procedures to (1) search for studies, (2) screen for relevant studies, (3) code results (a survey of the authors of the studies can be used to help ensure that their findings have been properly coded), and (4) provide a quantitative summary of the findings. The primary advantages of meta-analysis are that it helps to obtain all relevant studies and that it uses information in an objective and efficient manner. Cooper and Rosenthal (1980) found that meta-analysis was more effective than traditional (unstructured) literature reviews. Meta-analyses are useful in making generalizations, such as which forecasting method is best in a given situation. Meta-analyses are also useful when estimating relationships for an econometric model (see a priori analysis). When aggregating results across studies with small sample sizes, it may be useful to follow the procedures for assessing statistical significance described by Rosenthal (1978). Since 1980, meta-analysis has been popular in many fields.

Mini-Delphi. See Estimate-Talk-Estimate.

Misspecification test. A test that indicates whether the data supporting the building of the model violate assumptions. When an econometric model is estimated, for example, it is generally assumed that the error term is independent of other errors (lack of autocorrelation) and of the explanatory variables, and that its distribution has a constant variance (homoscedasticity).

Mitigation. The reduction of the effects of a factor on a forecast. It is useful to mitigate the forecast of changes when one faces uncertainty in the forecast. In econometric models, this can be done by reducing the magnitude of a relationship or by reducing the amount of change that is forecast in the explanatory variable. It is difficult to find studies on mitigation. However, in Armstrong (1985, pp. 238-242), mitigation produced large and statistically significant error reductions for predictions of camera sales in 17 countries over a six-year horizon. The concept has been valuable in extrapolation, where it is called damping. This term is similar to the term shrinking, and it avoids confusion with the term shrinkage.

Model. A representation of the real world. In forecasting, a model is a formal statement about variables and relationships among variables.

Monte Carlo simulation. A procedure for simulating real-world events. First, the problem is decomposed; then a distribution (rather than a point estimate) is obtained for each of the decomposed parts. A trial is created by drawing randomly from each of the distributions. The procedure is repeated for many trials to build up a distribution of outcomes. Monte Carlo simulation can be used to estimate prediction intervals.

Months for Cyclical Dominance (MCD). The number of months, on average, before the cyclical change dominates the irregular movement in a time series. The MCD is designed to offset the volatility in a time series so that cyclical phases can be seen (Shiskin 1957).

Moving average. An average of the values in the last *n* time periods. As each new observation is added, the oldest one is dropped. A smoothed estimate of the level can be used to forecast future levels. Trends can be estimated by averaging changes in the most recent *n'* periods (*n'* and *n* generally differ). This trend can then be incorporated in the forecast. The value of *n* reflects responsiveness versus stability in the same way that the choice of smoothing constant does in exponential smoothing. For periods of less than a year, if the data are subject to seasonal variations, *n* should be large enough to contain full cycles of seasonal factors. Thus, for monthly data, one could use 12, 24, or 36 months, and so on. Differential weights can be applied, as is done by exponential smoothing. PoFxxx

Moving origin. See successive updating.

Multicollinearity. A measure of the degree of correlation among explanatory variables in a regression analysis. This commonly occurs for nonexperimental data. Parameter estimates will lack reliability if there is a high degree of covariation between explanatory variables, and in an extreme case, it will be impossible to obtain estimates for the parameters. Multicollinearity is especially troublesome when there are few observations and small variations in the variables. PoFxxx

Multiple correlation coefficient. Often designated as *R*, this coefficient represents a standardized (unit free) relationship between $\overset{?}{Y}$ and Y ($\overset{?}{Y}$ is the result when Y is regressed against explanatory variables $X_1, X_2, \ldots X_k$). It is customary to deal with this coefficient in squared form (i.e., R^2). See R^2 and adjusted R^2.

Multiple hypotheses. The strategy whereby a study compares two or more reasonable hypotheses or methods. Although it goes back to a paper published by T. C. Chamberlin in 1890 (reprinted in Chamberlain 1965), it is used occastionally in the social sciences. Results are often not meaningful in absolute terms, so the value of an approach (or theory) should be judged relative to current practice or to the next best method (or theory). PoFxxx

Multiple regression. An extension of simple regression analysis that allows for more than one explanatory variable to be included in predicting the value of a forecast variable. For forecasting purposes, multiple regression analysis is often used to develop a causal or explanatory model. (See econometric method.)

Multiplicative model. A model in which some terms are multiplied together. An alternative is an additive model.

Multi-state Kalman Filter. A univariate time-series model designed to react quickly to pattern changes. It combines models using Bayesian estimation.

Multivariate ARMA model. ARMA models that forecast several mutually dependent time series. Each series is forecast using a function of its own past, the past of each of the other series, and past errors. See dynamic regression model.

Naive model. A model that assumes things will behave as they have in the past. In time series, the naive model extends the latest observation (see random walk model). For cross-sectional data, the base rate can serve as a naive model.

Neftci probability approach. A technique for forecasting business-cycle turning points developed by Neftci (1982). It signals cyclical turning points by calculating the likelihood that the economic environment has changed. A turning-point probability signal occurs when the estimated probability reaches some preset level of statistical confidence (say 90% or 95%). The likelihoods are based on (1) the probability that the latest observation comes from a recession (or a recovery) sample, (2) the chance of recession (or recovery) given the length of the current cyclical phase in comparison to the historical average, and (3) the comparison of 1 and 2 with the previous month's probability estimate.

Neural networks. Information paradigms inspired by the way the human brain processes information. They can approximate almost any function on a closed and bounded range and are thus known as universal function approximators. Neural networks are black-box forecasting techniques, and practitioners must rely on ad hoc methods in selecting models. As a result, it is difficult to understand relationships among the variables in the model. Franses and Van Dijk (2000) describe how to compute elasticities from neural nets. See Remus and O'Connor (2001). PoFxxx

NGT. See Nominal Group Technique.

Noise. The random, irregular, or unexplained component in a measurement process. Noise can be found in cross-sectional data as well as in time-series data.

Nominal dollars. Current values of dollars. To properly examine relationships for time-series data, dollar values should be expressed in real (constant) dollars; that is, they should be adjusted for inflation. A complicating factor for adjusting is that the U.S. government has overstated inflation by about one percent per year.

Nominal Group Technique (NGT). A group of people who do not communicate with one another as they make decisions or forecasts. Such groups are used in the Delphi technique, as described by Rowe and Wright (2001).

Nominal scale. Measurement that classifies objects (e.g., yes or no; red, white, or blue; guilty or innocent).

Noncompensatory model. A model that employs a nonlinear relationship combining cues to make a forecast. It is noncompensatory because low (high) values for some cues cannot be offset in their contribution by high (low) values in other cues. Conjunctive and disjunctive models are two noncompensatory models.

Nondirective interviewing. A style of interviewing in which the interviewer asks only general questions and encourages the interviewee to discuss what he considers important. The interviewer probes for additional details and does not introduce ideas or evaluate what is said. This approach is useful in determining what factors enter into a person's decision making. Thus, it could help in identifying variables for judgmental bootstrapping, conjoint analysis, or econometric models. It can also be useful in developing a structured questionnaire, such as might be used for intentions surveys. Here are some guidelines for the interview.

> Start by explaining what you would like to learn – e.g., "what factors cause changes in the sales of your primary product?" If a general opener does not draw a response, try something more specific – e.g., "perhaps you could describe how product x did last year?"
>
> During the interview:
>
> - Do not evaluate what the interviewee says. If he feels that he is being judged, he is likely to reveal less.
> - Let the interviewee know that you're interested in what he says and that you understand. To find out more about a particular subject that is mentioned by the interviewee, ask for elaboration – e.g., "that's interesting, tell me more." Or you may use a reflection of the interviewee's comments – "You seem to think that . . ." often picking up the last few words used by the interviewee.
> - Do not interrupt. Let the interviewee carry the conversation once he gets going.
> - Do not bring in your own ideas during the interview.
> - Do not worry about pauses in the conversation. People may get uncomfortable during pauses, but do not be in a hurry to talk if it is likely that the interviewee is thinking.

Nonexperimental data. Data obtained with no systematic manipulation of key variables. Regression analysis is particularly useful in handling such data as it assesses the partial effects of each variable by statistically controlling for other variables in the equation. If the variables do not vary or the explanatory variables are highly correlated with one another, nonexperimental data cannot be used to estimate relationships.

Nonlinear estimation. Estimation procedures that are not linear in the parameters. Nonlinear techniques exist for minimizing the sum of squared residuals. Nonlinear estimation is an iterative procedure, and there is no guarantee that the final solution is the best for the calibration data. What does this have to do with forecasting in the social sciences? Little research exists to suggest that nonlinear estimation will contribute to forecast accuracy, while Occam's razor suggests that it is a poor strategy.

Nonlinearity. A characteristic exhibited by data that shows substantial inflection points or large changes in trends.

Nonparametric test. A test of statistical significance that makes few assumptions about the distribution of the data. A nonparametric test is useful for comparing data when some observations (or some forecast errors) are outliers and when the error distributions depart substantially from normal distributions.

Nonresponse bias. A systematic error introduced into survey research, for example, in intentions surveys, because some people in the sample do not respond to the survey (or to items in a questionnaire). Because those interested in the topic are more likely to respond, it is risky to assume that nonresponders would be similar to responders in reporting about their intentions. To avoid this bias, obtain high response rates. By following the advice in Dillman (2000), one should be able to achieve well over a 50% response rate for mail surveys, and often as much as 80%. To estimate nonresponse bias, try to get responses from a subsample of nonrespondents. Armstrong and Overton (1977) provide evidence showing than an extrapolation of trends across waves in responses to key questions, such as "How likely are you to purchase . . .?" will help to correct for nonresponse error.

Nonstationarity. See stationary series.

Nowcasting. Applying a forecasting procedure to obtain an estimate of the current situation or level at the origin. Nowcasting is especially important when data are subject to much error and when short-term forecasts are needed. It is also useful when a model may provide a poor estimate of the level; for example, regression analysis often provides poor estimates of the level at t_0 for time-series data. Combined estimates can improve the estimate of the current level. These can draw upon extrapolation, judgment, and econometric models. Such a procedure can help to reduce forecast error, as shown in Armstrong (1970). PoFxxx

Null hypothesis. A proposition that is assumed to be true. One examines outcomes (e.g., from an experiment) to see if they are consistent with the null hypothesis. Unfortunately, the null hypothesis is often selected for its convenience rather than for its truth. The rejection of an unreasonable null hypothesis (or nil hypothesis) does not advance knowledge. For example, testing against the null hypothesis that income unrelated to the sales of automobiles would be foolish at best and might even be misleading (see statistical significance). Unfortunately, null hypotheses are frequently misused in science (Hubbard and Armstrong 1992).

Number-of-attribute-levels effect. An artificial result in decompositional conjoint analysis that results from increasing the number of (intermediate) levels for an attribute in a conjoint study while holding other attribute levels constant; this increases the estimated impact of the attribute on preferences. See Wittink and Bergestuen (2001).

N-way cross validation. See jackknife.

Observation. A measurement of a characteristic for a given unit (e.g., person, country, firm) for a given period of time.

Occam's Razor. The rule that one should not introduce complexities unless absolutely necessary. "It is vain to do more what can be done with less," according to William of Occam (or Ockham) of England in the early 1300s. Occam's razor applies to theories about phenomena and methods.

OLS. See Ordinary Least Squares.

Omitted variable. An explanatory variable that should be part of a model but has been excluded. Its exclusion can lead to biased and inefficient estimates of the remaining parameters in the model. Omitting it causes no problem in the estimation of the included variables if it is constant for the calibration data, or if its variations are uncorrelated with the included variables. Its exclusion can lead to inaccurate forecasts if it changes over the forecast horizon.

Operational measure. A description of the steps involved in assigning numbers to a variable. It should be specific enough so others can carry out the same procedure. Ideally, operational procedures are representative of the concept that is being measured. Even seemingly simple concepts might be difficult to operationalize, such as estimating the price of computers year by year.

Opposing forces. Forces that are expected to move against the direction of the historical trend. An example is inventory levels relative to sales: When inventories get too large, holding costs lead managers to reduce their levels, thus opposing the trend. When inventories are too small, service suffers, prompting decisions to hold larger inventories, again, opposing the trend. See Armstrong, Adya and Collopy (2001). PoFxxx

Optimism. A state of mind that causes a respondent to forecast that favorable events are more likely to occur than is justified by the facts. Also known as wishful thinking. This has long been recognized. For example, Hayes (1936) surveyed people two weeks before the 1932 U.S. presidential election. Of male factory workers who intended to vote for Hoover, 84% predicted he would win. Of those who intended to vote for Roosevelt, only 6% thought Hoover would win. Many of us are susceptible to this bias. We think we are more likely to experience positive than negative events (Plous 1993, pp. 134-135). Warnings about the optimism bias (e.g., "People tend to be too optimistic when making such estimates") help only to a minor extent. Analogies may help to avoid optimism. PoFxxx

Ordinal scale. A method of measuring data that allows only for ranking. The intervals between observations are not meaningful.

Ordinary Least Squares (OLS). The standard approach to regression analysis wherein the goal is to minimize the sum of squares of the deviations between actual and predicted values in the calibration data. Because of its statistical properties, it has become the predominant method for regression analysis. However, it has not been shown to produce more accurate forecasts than least absolute values.

Origin. The beginning of the forecast horizon. (Also, see level.)

Outcome feedback. Information about an outcome corresponding to a forecast. For example, how often does it rain when the weather forecaster says the likelihood is 60%? (See also lens model.)

Outlier. An observation that differs substantially from the expected value given a model of the situation. An outlier can be identified judgmentally or by a statistically significant deviation.

Out-of-sample forecast. See holdout data.

Overconfidence. A state of mind that causes a forecaster to think that the probability that a forecast is correct is greater than the actual probability. This leads prediction intervals to be too narrow. Experts are overconfident because of various biases, such as an unwarranted feeling of control or a desire to see things turn out well. Overconfidence is widespread. For example, when subjects are asked how many times the letter F appears in: "Finished files are the result of years of scientific study combined with the experience of years," about half answer incorrectly. Most are sure that their answer is correct for this problem, and those who are more confident are no more accurate than those who are less confident. (The correct answer is six.) See Arkes (2001). PoFxxx

Panel. A group of experts (or decision making units) whose opinions are sought periodically. Ideally, the composition of the panel remains constant over time. In practice, it is not easy to ensure this, so rules must be set up in advance on replacing panel members. Alternatively, one can start with a large panel and then analyze only responses from those who remain for all periods. Panels are used in the Delphi technique, and they can also be used in intentions surveys and for retail sales forecasts, where periodic reports are obtained for sales at representative stores.

Panel data. Data on the same cross section measured on at least two time periods (see longitudinal data).

Parameter. The "true" value of some unknown population value (such as a relationship). Parameters can be estimated from samples of data and from a priori analysis.

Parameter stability. The conditions in which the parameters of a model estimated separately for two sets of data show no substantial or statistically significant differences. This provides some assurance that relationships are stable, but it does not ensure that they will be stable over the forecast horizon. (See Chow test.)

Parsimony. The use of as few parameters as possible in fitting a model to calibration data. (See Occam's Razor.)

Partial correlation. A measure of the association between a dependent variable and one of the explanatory variables when the effects of the other explanatory variables are held statistically constant. Multiple regression provides partial correlations, which are useful in developing econometric models.

Pattern regime. A time interval over which the parameters of a time-series model are relatively constant.

Phase Average Trend (PAT). A technique for trend-adjusting composite indexes used for the measurement of growth cycles. The PAT method is based on constructing a variable trend. Its basis is a 75-month moving average, which means that 37 months of trend are lost at the beginning and, at the end of the period being studied. The lost months are approximated by extrapolation. After estimating the trend, the forecaster next calculates the deviations from the trend, which produces an approximation of the growth cycle. This is used to calculate the phase averages which form the bases for approximating the curvilinear trend. Calculating the deviation of the original observations from this trend is the basis for determining the final growth cycle.

Plan. To develop a set of objectives and describe strategies for reaching these objectives. Planning should precede forecasting the outcomes of various plans. If none of the plans are forecast to produce satisfactory outcomes, new plans can be developed, followed by new forecasts. Armstrong (1983) discusses how forecasting contributes to the planning process. Extensive research has shown that formal planning is useful for decision-making groups, and that the better it is done, the more useful it is. Armstrong (1982a) provided a meta-analysis of this research, and the review was updated in Armstrong (1990). PoFxxx

Policy-capturing. An alternative term for judgmental bootstrapping. PoFxxx

Polynomial. A mathematical expression containing one or more terms, each of which consists of a coefficient and a variable(s) raised to some power. Thus $a + bx$ is a linear polynomial and $a + bx + cx^2$ is a quadratic polynomial in x. A polynomial of order m includes terms with powers of x up to x^m. Polynomials will typically provide excellent fits to the calibration data but it is difficult to find a case where polynomials have contributed to forecasting in the social or management sciences. On the contrary, they are associated with poor accuracy. Do not use polynomials unless there is a very strong a priori case.

Postdiction. See backcasting.

Postsample evaluation. The evaluation of a forecasting model using data that were collected at the end of the forecast horizon.

Practical significance. The importance of a result to decision making. "A difference that makes a difference." Statistical significance does not imply practical significance. Many people, including leading researchers, misinterpret statistical significance as implying practical significance. See McClosky and Ziliak (1996).

Preciseness. The level of detail in the presentation of a numerical forecast, usually thought of as the number of significant digits reported. The believability of forecasts can be influenced by how precisely they are reported. Teigen (1990) showed that more precise reporting can make a forecast more acceptable (unless such precision seems unreasonable). Teigen calls this the preciseness paradox. That is, under a wide variety of circumstances, the more precise the forecast, the more confident we are about the forecast. But the more precise the forecast, the less likely it is to turn out correct. When forecasters provide detail, they imply that they have much expertise about the topic. Thus, the preciseness paradox should be stronger for statements about the past than for predictions. It is. Consider one of Teigen's studies. He asked subjects how much confidence they would have in different informants if they visited Iceland and received the following answers to this question:

> "Owing to various price regulation measures, this year's inflation rate was down to 5%. Was it higher last year?"
>
> *Responses:*

– Olafur said "Yes, it was."
– Larus said "Yes, it was about 7%."
– Jon said "Yes, it was between 5 and 9%."

Which of these answers would you be most confident about? Teigen says that Olafur's statement is the most general, and Larus's the most exact. If Larus was right, so are Olafur and Jon. On the other hand, Olafur could have been right, while Larus and Jon were wrong (if inflation were to be 14%, for example). However, most of the subjects (16) were most confident in Larus and eight subjects were most confident in Jon, while only seven subjects were most confident in Olafur. When the statements about inflation were converted from the past to represent a forecast about next year's inflation, the confidence in the most precise forecast (by Larus) decreased (to 6 of 34 subjects) but did not disappear. Teigen suggests that this occurs because people do not expect forecasters to be able to provide precise forecasts of inflation. When people expect that experts can make good forecasts, added detail and preciseness are likely to lead them to have more confidence in the forecasts. To avoid misplaced confidence, forecasters should ensure that there is no false precision in their reports.

Precision. The exactness of a measure. For numerical forecasts, precision can be indicated by the number of significant digits. The preciseness of the report should match the precision of measurement.

Prediction. A statement regarding future events or events that are unknown to the forecaster. Generally used as synonymous with forecast. Often, but not always used when the task involves forecasting with cross-sectional data (e.g., personnel predictions).

Prediction interval. The bounds within which future observed values are expected to fall, given a specified level of confidence. For example, a 95% prediction interval is expected to contain the actual forecast 95% of the time. However, estimated prediction intervals are typically too narrow for quantitative and judgmental forecasting methods.

Predictive validity. The extent to which a model or method is useful in making forecasts. This is best assessed by comparing it with alternative methods. Determining predictive validity has long been recognized as one of the primary ways to test hypotheses (Friedman 1953). See Armstrong (2001d).

Probability sample. Elements selected from the population such that there is a known probability of an element being included. This helps to ensure that the sample is representative of the population. For example, you could obtain a list of the population, then select every n^{th} element from the list. A probability sample can help reduce sampling error for intentions surveys. It is irrelevant for expert opinions studies.

Process-tracing methods. Methods of studying human decision making and problem solving as they occur in natural settings. Protocol analysis of expert decision making is one such approach.

Product hierarchy. A family of related products or items organized at various levels. For example, a certain brand of toothpaste may come in several flavors, and each flavor may be packaged in several tube sizes. The forecaster's task is to project the volume of demand for each stock-keeping unit (sku) – a package of a specific flavor – as well as total demand for each flavor and overall demand for the brand. The forecaster should reconcile the forecasts made at each level of the hierarchy. See reconciling forecasts. PoFxxx

Product life cycle. Sales of a product are assumed to follow an S-shaped curve, growing slowly in the early stages, achieving rapid and sustained growth in the middle stages, slowing up in the mature stage, and then declining. This should help in the selection of an appropriate forecasting method. For example, different methods should be used in the concept phase than in the test marketing. However, no empirical evidence exists for this claim. An analyst who has never heard of the product life cycle might still use an appropriate forecasting method.

Production function. A causal model that relates the output from a production process (including both manufacturing and services) to the factors (explanatory variables) that contribute to the output. For example, such a model might link manufacturing output to labor, equipment, and training inputs.

Production system: Representations of conditional knowledge using IF-THEN statements. IF represents conditions, while THEN represents actions.

Prognosis. See forecast.

Projection. See extrapolation, prediction, or forecast.

Projective test. A test that asks a subject to respond to a vague stimulus. It is assumed that the subject will project his or her own expected behavior on the situation. Such tests can be useful in situations involving the prediction of socially undesirable behavior. For example, you could ask how someone else would act in a situations in which it was easy to steal money. The question might be framed, “predict how your best friend would react in the following situation,” or “write a story about the following situation . . .”

Protocol. A record of a person's thought process as they talk about it when performing a task (such as when making a forecast). The record can be made with audio records, video records, or paper and pencil. PoFxxx

Proxy variable. A variable that acts as a substitute for an unobserved explanatory variable in a model. Such variables are often used when it is infeasible or too expensive to use a more relevant operational measure. For example, income is often used as a measure of *ability to purchase* even though it does not fully capture that concept because ability to purchase depends also on wealth, unreported income, gifts, theft, and subsidies. In 1999, for example, it was estimated that poverty-level families in the U.S. (based on income) consume almost twice their income (because of government subsidies, gifts, theft, and unreported income).

Purchase intentions. A self-reported measure of whether a person or organization plans to purchase a product during a specific time period. For example, intentions surveys from key people in organizations can be used to forecast plant and equipment expenditures. See Morwitz (2001). PoFxxx

Purchase probabilities. Measures of the probability that people will purchase a product during a specific time period. Typically, they are based on self-reports. Purchase probabilities are more encompassing than purchase intentions because they also include expectations of unplanned purchases and recognize that your intentions may change over the forecast horizon. See Morwitz (2001).

Quasi-experimental data. Data in which changes are introduced naturally, rather than by a researcher. For example, governments in different countries have different levels of spending. This would allow for an analysis of the effect of government spending on growth. (Not surprisingly, this has been studied, and increased government spending is closely associated with reduced economic growth). Forecasters often rely on quasi-experimental data. In contrast to experimental data, there are many threats to validity, so eclectic research might be useful.

Quasi-rational judgment. Judgment based on both intuition and analytic processes. PoFxxx

R^2 (*R*-squared). The coefficient of determination. In regression analysis, the square of the correlation between Y (the forecast variable) and $\hat{Y}$ (the estimated Y value based on the set of explanatory variables) is denoted as R^2. R^2 can be interpreted as the proportion of variance in Y that can be explained by the explanatory variables. R^2 is appropriate only when examining holdout data (use adjusted R^2 for the calibration data). Some researchers believe that the dangers of R^2 outweigh its advantages. Montgomery and Morrison (1973) provide a rule of thumb for estimating the calculated R^2 when the true R^2 is zero: it is $R^2 = v/n$, where v is the number of variables and n is the number of observations. They showed how to calculate the inflation in R^2 and also presented a table showing sample sizes, number of variables, and different assumptions as to the true R^2. If you are intent on increasing R^2, see “rules for cheaters” in the practitioners’ section of the principles site, hops.wharton.upenn.edu/forecast. R^2 can be especially misleading for time-series data. Used with caution, R^2 may be useful for diagnostic purposes in some cases, most likely when dealing with cross-sectional data. Even then, however, the correlation coefficient is likely to be a better measure.

$\bar{R}^2$ (*R*-bar-squared). See adjusted R^2 and R^2.

Random errors. Errors that exhibit no systematic pattern.

Random sampling. A statistical sampling method for selecting elements from a population in such a way that every element within that population has the same probability of being selected. This is not an exact definition however, so here is a statistician's definition: In simple random sampling, we select a sample *n* out of N population units such that each subset of size n has the same probability of occurring. For example, if population size is N = 100, and sample size is *n* = 10, then each random sample of size 10 has probability 1/(100 choose 10) of occurring. Many misinterpret simple random sampling to mean that each unit has probability *n*/N of being in the sample; however, many sampling procedures that are not simple random sampling have this property.

Random walk model. A model in which the latest value in a time series is used as the forecast for all periods in the forecast horizon. Alternatively, it is a model stating that the difference between each observation and the previous observation is random. See naive model. Statisticians define the term as follows: A random walk is a time-series model in which the value of an observation in the current time period is equal to the value of the observation in the previous time period plus the value of an error term from a fixed probability distribution. It is a special case of a martingale.

Randomized Response Technique. A way of stating questions that permits either answers to the question or responses to random events, such as a coin toss. This approach is useful for forecasting socially undesirable behavior. For example, to forecast the profitability of a proposed chain of convenience stores, an analyst might need to forecast theft by store employees. Wimbush and Dalton (1997) examined ways to predict which job candidates might become thieves. Theft can be expected to vary by situation, by method of estimation, and by the definition of a theft (amount stolen and time period). Previous research led to widely varying estimates of theft, ranging from 28% to 62%. (Even at the lowest figure, this represents a major cost.) Wimbush and Dalton thought direct questions were not reasonable because people would lie. When they asked 210 employees on an anonymous questionnaire, “Are/were you involved in theft from your employer of from [dollar amount specified] in cash, supplies, or merchandise a month?” 28.2% admitted to theft. Wimbush and Dalton then used more appropriate methods. One was the Randomized Response Technique: they asked interviewees to flip a coin in a self-administered survey (only the respondent knows how the coin lands) and then answer the following question:

> “If your coin flip is a head OR if you are/were involved in the theft from your employer of [dollar amount specified] in cash, supplies, or merchandise a month, please put an “X” in the box to the right.”

When asked this way, the estimated percentage of thieves was 57.9%. Another approach was the Unmatched Count Technique: they gave 353 respondents two sets of questions; half had five items and the other half had six items to choose from, such as “I have been to Spain,” or “I currently have one or more cats.” The sixth item was the question used above in the direct questioning approach. They asked respondents to indicate whether any of the items was true. The percentage of thieves, as estimated by the Unmatched Count Technique, was 59.2%.

Ratio scale. A scale in which the measured intervals are meaningful and the zero point is known (e.g., the Kelvin scale for temperature is a ratio scale, as is the yardstick).

Realization. The sum of the stochastic pattern and random errors when a stochastic process is assumed to be the data-generating process. The distinction between process and realization is relevant when considering, for example, the difference between theoretical and sample autocorrelation and between theoretical and sample partial autocorrelation functions.

Recent trend. The short-term trend in a time series, often measured by exponential smoothing, where the smoothing factor puts much weight on the last observations. Also called local trend.

Reconciling forecasts (in a product hierarchy). Adjustments of forecasts to ensure that the whole will equal the sum of its parts. Such adjustments are needed because forecasting the whole will produce a different forecast than forecasting the parts and summing them. The forecaster can draw upon at least three approaches: (a) the bottom-up approach: summing model-based forecasts for each subgroup at the lowest level of the hierarchy to obtain forecasts for all group totals. (b) The top-down approach: allocating forecasts for each group total to subgroups. (c) The middle-out approach*:* In a hierarchy with three or more levels – for example, brand, flavor, package size – creating model-based forecasts for a middle-level (flavor in this case) and then reconciling higher levels (brand) using the bottom-up approach and the lower levels (package size) using the top-down approach. PoFxxx

Recursive model. A model in which the current value of one set of variables determine the current value of another, whereas previous (or lagged) values of the latter determine the current values of the former. A series of independent models to deal with causal chains. A simple example of a recursive model is:

$$Y_t = a + bX_1$$
$$X_t = c + dY_{t-1}$$

Regressing forces. Forces that move the series toward some mean. An example is a measure of the performance of a professional athlete, such as a batting average; his average for the first three games of the current season would tend to regress toward his historical average. For a new player, his average might regress to the average for new players. Regressing forces are discussed in Armstrong, Adya and Collopy (2001).

Regression. A tendency to return to a previous average. The term *regression* dates back to Francis Galton and his work concerning the heights of children in different generations. The heights of children of exceptionally tall (or short) parents "regress" to the mean of the population. So if you see a result that is far above the current level, such as a baseball player hitting 70 home runs in a season, you should forecast that the next result will not be so outstanding. (See also regression to the mean.)

Regression analysis. A statistical procedure for estimating how explanatory variables relate to a dependent variable. It can be used to obtain estimates from calibration data by minimizing the errors in fitting the data ($Y = a + b_1X_1 + b_2X_2 \ldots$). Typically, Ordinary Least Squares is used for estimation, but least absolute values can be used. Regression analysis is useful in that it shows relationships, and it shows the partial effect of each variable (statistically controlling for the other variables) in the model. As the errors in measurement increase, the regression model shrinks the magnitude of the relationship towards zero. See Allen and Fildes (2001). PoFxxx

Regression coefficients. The relationship of X to Y. In regression analysis, a forecast variable Y is modeled as a function of explanatory variables X_1 through X_k. The explanatory variables are multiplied by the regression coefficients. A regression coefficient represents the effect of an explanatory variable on the dependent variable.

Regression to the mean. The tendency for extreme observations measured in one time period to revert toward a mean value when measured during another time period. For a discussion, see Plous (1993, pp. 116-118). PoFxxx

Regressor. See explanatory variable and causal variable.

Reinforcing series. A time series in which the expected direction of movement, based on causal forces, corresponds with the direction of the statistical extrapolation. If they conflict, they are contrary series. Armstrong, Adya and Collopy (2001) discuss extrapolation for reinforcing series.

Relative Absolute Error (RAE). The absolute error of a proposed time-series forecasting model divided by the absolute error of the random walk (no-change) model. The RAE is similar to Theil's U2. The RAE can be averaged by taking a geometric mean (because the data are ratios) to get the Geometric Mean Relative Absolute Error (GMRAE). If outliers are expected, the GMRAE should be trimmed or the Median RAE (MdRAE) should be used. The GMRAE is used for calibrating models, and the MdRAE is used for comparing models. Armstrong (2001d) discusses the use of the RAE in comparing forecasts from different methods.

Reliability. The extent to which a replication of a measurement process will yield the same results. In forecasting, the extent to which a method will produce similar forecasting accuracy when used in similar situations. Tests of statistical significance do not provide good measures of reliability. PoFxxx

Replication. Application of given procedures to similar sets of data to determine whether they produce similar findings. Replications provide good evidence on reliability. PoFxxx

Representativeness. The subjective impression that one situation is similar to another situation. People often judge probability by the degree to which A resembles B. In other words, when aspects of a situation seem similar to those of another situation, they are more likely to predict that they will show similar responses to change. They do this even when they believe that the similar characteristics are irrelevant. Tversky and Kahneman (1982) use the following example to reveal this tendency:

> "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. Please check off the most likely alternative:
> ? Linda is a bank teller.
> ? Linda is a bank teller and is active in the feminist movement."

Nearly 90% of the subjects (n = 86) thought that Linda was more likely to be a bank teller *and* a feminist than to be a bank teller. Representativeness can create problems in using expert opinions or intentions surveys. It is useful in writing scenarios. (See conjunction fallacy.) PoFxxx

Residual. The difference between an actual observed value and its forecast value in univariate and multivariate models. One can draw a distinction between errors (based on the true model) and residuals, obtained by subtracting the fitted value from the actual value in the calibration data. Of course, who knows what the true model is? (See error term.)

Resolution. A measure of a probability assessor's ability to assign events into subcategories for which the proportion of events correct is different from the overall proportion of events correct. A higher resolution score reflects the ability of an assessor to discriminate between differing degrees of uncertainty in their predictions using the probability scale.

Response error. The error that occurs when respondents do not reveal their true opinions on a subject. They may misunderstand the question, fail to understand their true opinions, lie, or try to present themselves in a favorable light. Response error is a particular problem when respondents are unfamiliar with the situation, for example, when they are asked about intentions to purchase a new product. For ideas on how to reduce response error, see Sudman and Bradburn (1982).

Retrospective process tracing. A procedure in which one asks experts to describe, from memory, the steps they take in making decisions or forecasts. This can be used for developing expert systems, as discussed in Collopy, Adya and Armstrong (2001).

Revised forecast. See updated forecast.

Robust trend. A trend estimate based on medians or modified means instead of arithmetic means. Thus, trends are estimated for a series of time intervals, say the trend from year 1 to year 2, then from year 2 to year 3, and so on in the calibration data. The median trend is then selected from these estimates for use in the forecasting model. Use of a robust trend requires three or more trend estimates. The forecast is the current level plus the estimated trend. The robust trend protects against outliers. Thus, it can be expected to be useful for noisy data. Little validation research has been done for the robust trend. However, Fildes et al. (1998) found that the robust trend produced fairly accurate forecasts (compared to other extrapolation methods) for some monthly telecommunications data (which, at the time, were characterized by declining trends). They also used a factor to take into account the size and sign of the differences between the individual trend estimates and their median. It is not known whether this adjusting factor contributes to accuracy. Following Occam's razor, the adjustment factor should be avoided until it has been tested.

Role. See role playing and role taking.

Role playing. A technique whereby people play roles and enact a situation in a realistic manner. Role playing can be used to predict what will happen if various strategies are employed. It is especially relevant when trying to forecast decisions made by two parties who are in conflict. Armstrong (2001a) provides principles for the use of role playing in forecasting and shows that role playing is substantially more accurate than expert opinions. It is also expected to be more accurate than game theory.

Role reversal. A technique in which decision makers take the role of their opponent in a conflict situation. Because decision makers might lack awareness about their opponent's thinking, it may be useful to have the decision makers exchange roles. This might lead to better predictions about their opponent's behavior (Armstrong 2001a) and to acceptance of these predictions.

Role taking. A technique in which people are asked to think how they could behave in a certain role without acting it out. For example, they could be asked to make forecasts "assuming that you were the President of the U.S." Taking roles can affect forecasts. Cyert, March and Starbuck (1961) divided subjects into two groups of 16 each. Each group was assigned a different role. Subjects were given the role as the chief cost analyst for a manufacturing concern and were asked to produce cost forecasts on the basis of preliminary estimates provided by two assistants in whom they had equal confidence. Other subjects were given the role as the chief market analyst and were also asked to provide sales forecasts. The data were identical for both roles. Seldom did the analysts simply average the estimates from their two assistants – the expected behavior if no role had been assigned. The cost analysts forecasted on the high side, and the market analysts forecasted on the low side. Roles can also affect the acceptability of forecasts as shown by Wagenaar and Keren (1986). The subjects' acceptance of information depended upon the roles they were assigned. In this study of 388 subjects, the two roles were either individual decision maker ("a parent") or a societal decision maker ("minister of traffic"). Half of the subjects were given each role and were provided with either anecdotal or statistical evidence on the need for safety belts in the back seats of automobiles. The anecdotal evidence was a three-sentence description of a traffic accident in which a seven-year old girl died because she did not wear a seat belt. The statistical evidence was two sentences stating that 150 children die each year in motor vehicle accidents and that this could be reduced to 50 if seat belts were used in the back seats. The societal decision makers were more influenced by the statistical evidence (62% favoring the use of seat belts) than were the individual decision-makers (47% favoring the use of seat belts). PoFxxx

Rolling horizon. See successive updating.

Root Mean Squared Error (RMSE). The square root of the Mean Squared Error.

Round. One of a series of successive administrations of a given instrument to a panel. At least two rounds are used in the Delphi technique to solicit forecasts from experts. Each round leads to improved accuracy, although with diminishing marginal returns. Rounds are expected to be especially useful when the panel is small (say about five), when misinterpretations are likely, the problem is difficult, and the experts are heterogeneous. (See Rowe and Wright 2001.)

Rule-based forecasting. A type of expert system that is applied to time-series extrapolation. Rules based on forecasting expertise and domain knowledge are used to combine alternative extrapolations. Armstrong, Adya and Collopy (2001) describe principles for rule-based forecasting. PoFxxx

Runs test. A nonparametric test for time-series data that detects tendencies for the errors to be in one direction. It can indicate that the forecasting model is providing biased forecasts, which may call for changes in the forecasting procedure.

Safety stock. Additional inventory in case demand exceeds the forecast. Because there are always forecast errors, safety stocks are necessary in some parts of the supply chain.

Sales composite. Expert opinion forecasts by members of the sales force. While the sales-force members usually have good information, they are likely to be biased, perhaps because of optimism or because of payment incentives.

Sample. A limited number of observations selected from a population.

Sampling error. The error that results from using a probability sample as opposed to using the population of all observations relevant to the given problem. It is possible to quantify this error, which is often referred to as the standard error of the estimate. Nonprobability sampling (see convenience sample) introduces error because the sample is likely to be unrepresentative. Traditional measures of probability sampling error do not account for nonresponse bias and response errors; in many practical situations, these errors are often much larger than sampling errors. Consider political polling, in which the situation is well-known to the respondents. Lau (1994) examined the errors in 56 national surveys concerning the 1992 U.S. presidential election. The sample sizes varied from 575 to 2,086. Although the errors varied substantially across the surveys, they were only weakly related to sample size. Perry (1979) estimated that the total error for U.S. political election polls was 30% larger than the sampling error. Buchanan (1986) studied 155 elections from nine countries from1949 to 1985 and estimated that sampling error, given the typical sample size of 1,500, would yield a 95% prediction interval of ?2.5%. However, the actual 95% prediction interval was ?5.1%. One would expect that the size of other errors would be even larger relative to sampling errors if the analyst were forecasting for an unusual new product rather than for a political candidate.

Saturation level. The maximum number of members of a population who will eventually adopt an innovation. Some analysts attempt to measure this limit from time-series data. Good domain knowledge can help analysts to estimate the saturation level. The use of the saturation level is discussed in Meade and Islam (2001).

Scenario. A story about what happened in the future (note the past tense). According to the *Oxford English Dictionary*, "A sketch, outline, or description of an imagined situation or sequence of events; esp. a) a synopsis of the development of a hypothetical future world war, and hence an outline of any possible sequence of future events; b) an outline of an intended course of action; (to make a scenario of a story, book, or idea; to sketch out; also scenarioize, scenarize.) The over-use of this word in various loose senses has attracted frequent hostile comment." For example, *scenario* is used as a substitute for the word *alternative* in spreadsheet talk. Scenarios can distort one's perception of the likelihood of future events, and for that reason, they should not be used to make forecasts. Instead, they can be used to gain acceptance of forecasts. Scenarios can help to get people to think about the unthinkable or to consider what they would do given an unfavorable forecast. It can lead to contingency plans. One of the earliest uses of scenarios relates to the Battle of Dorking:

> In 1872, there was a German invasion of Britain. The British armies and fleet, it will be remembered, were at that time scattered across the world – putting down mutiny in India, protecting Canada from the United States, and guarding Ireland against Emperor Napoleon III. As a result, the home defenses were minimal on that morning in March when the German boats set out across the North Sea. What Royal Navy was left in British waters soon succumbed to the German mines and torpedoes – weapons that had been developed in secrecy. British land forces suffered not only from lack of numbers, but also from inadequate training and discipline, combined with an outdated philosophy of warfare. The great stand at the Battle of Dorking failed: The Germans conquered the British.

This story is, of course, false. It was written by G. T. Chesney and was published in *Blackwood's Magazine* in 1871. At that time, it was a plausible forecast. The publication of "The Battle of Dorking" created a political sensation. Prime Minister Gladstone attacked both the plausibility of the forecast and the wisdom of publishing such an alarmist view. Debate followed, and changes took place as a result. (The story has been passed along by Encel, Marstrand and Page, 1975, pp. 63-64.) Gregory and Duran (2001) discuss principles for using scenarios in forecasting.

Schwarz criterion. See BIC.

***S*-curve.** See S-shaped curve.

Seasonal adjustment. The process of removing recurrent and periodic variations over the course of a year. Also called deseasonalizing the data. Seasonal adjustments are discussed in Armstrong (2001c).

Seasonal difference. The difference calculated between seasonal factors that are separated by one year (e.g., four quarters, 12 months). Thus, if monthly data are used, a seasonal difference would be the difference for values separated by 12 months. See differencing.

Seasonal exponential smoothing. See Holt-Winters exponential smoothing method.

Seasonal factors. Numbers that indicate systematic variations within a year.

Seasonal index. See seasonal factors.

Seasonality. Systematic cycles within the year, typically caused by weather, culture, or holidays. PoFxxx

Secular trend. See basic trend.

Seer-sucker theory. "No matter how much evidence exists that seers do not exist, seers will find suckers." Proposed, along with relevant evidence, in Armstrong (1980b).

Segmentation. The division of a heterogeneous population into homogenous groups. See AID, bottom-up, and decomposition. Segmentation can produce substantial improvements in accuracy as shown in Armstrong (1985, chapter 9). Various techniques can be used to forecast the segments. For example, one useful strategy is to develop separate econometric models for each segment (Armstrong, 1985, pp. 284-287).

Self-confidence. A person's assessment of the likelihood that his predictions are correct. Useful for tasks for which the forecaster gets good feedback; otherwise, self-confidence and accuracy are not closely related for individual forecasters (Plous 1993, pp. 225-227). Self-confidence rises rapidly as groups discuss problems and as people receive more information. However, this rise is often unrelated to gains in accuracy (Oskamp, 1965). Self-confidence ratings are useful for assessing prediction intervals in situations where the forecaster gets excellent feedback. PoFxxx

Self-defeating prophecy. A forecast that proves to be false because of actions resulting from the forecast. By forecasting a potential disaster, a person or organization can take steps to ensure that it does not occur. In 1985, Ravi Batra forecast the great depression of 1990 (see Armstrong 1988). Batra might claim that thanks to his forecast, corrective actions were taken and the depression was avoided, but that would be far-fetched.

Self-fulfilling prophecy. A forecast that affects what actually happens so that it becomes true. This is related to the Pygmalion Principle or, as in *My Fair Lady*, a woman who is treated like a lady becomes a lady. Rosenthal and Jacobson (1968) present evidence on this phenomenon. In many cases, the effects can be beneficial. Sherman (1980) found that when people were asked how they would respond in a given situation, they tended to cast themselves in a responsible and favorable manner. Then, when later faced with the situation, they tended to live up to their predictions. Scenarios can be used to present forecasts so as to create self-fulfilling prophecies.

Sensitivity analysis. An analysis in which variations are introduced to the explanatory variables on the parameters in a model to examine what effect they have upon the variable of interest. This includes variations in the parameters or in values of the explanatory variables.

Serial correlation. See autocorrelation.

Setwise regression. Using sets of variables rather than one variable at a time for developing a model. (Also see stepwise regression.)

Shrinkage. The loss of predictive validity that results when moving from the calibration data to tests on holdout data. Do not confuse this with shrinking.

Shrinking. To modify an estimate by moving it towards a benchmark. For example, one could shrink parameters of a model based on an individual product toward those of a model built from a general class of products. In general, shrinking reduces the size of effects (e.g., the trend or the difference from a typical case). Shrinking is useful where uncertainty exists. (See mitigation.) PoFxxx

Sigmoid. An S-shaped curve or S-curve. Curves describing the growth of the number of adopters of an innovation. The elongated S-shape grows from near zero to approach the saturation level over time. The equation of the logistic process is an example. Meade and Islam (2001) discuss the use of such curves in forecasting.

Significant digits. Numerical digits other than zero. Unfortunately, when presenting forecasts, analysts sometimes let the computer determine the number of digits to report. The use of many digits gives a false sense of precision. In the social sciences, one is often uncertain about the second digit, yet analysts often provide four or five digits, such as a forecast of 14,332. The figure 14,000 might be regarded as "less scientific," but it is easier to read and remember. A good rule of thumb is to use three significant digits unless the measures do have greater precision and the added precision is needed by the decision maker.

Simple regression. An analytical procedure based on the assumption of a linear relationship between a single explanatory variable and a dependent variable. The relationship in a simple regression is typically estimated using the method of Ordinary Least Squares (OLS).

Simulated data. Artificial data constructed to represent a situation. Simulated data are used in conjoint analysis to represent such things as product designs. They can be used in judgmental bootstrapping to see how changes would affect an expert's forecasts. Simulated data have been used in extrapolation to see how various methods perform when the data come from a known process (for example, to examine effects of a discontinuity or high uncertainty).

Simulated test markets. An artificial laboratory setting that attempts to study aspects of behavior. Respondents are exposed to different types of marketing stimuli and are asked to evaluate the product (perhaps through a choice task or through purchase intention questions). In many cases, this exercise is followed by a related set of tasks and questions several weeks later designed to gauge consumers' reactions after trying the new product, and to obtain their longer-term reactions to the marketing activities.

Simultaneous causality. The situation in which *X* causes changes in *Y*, which in turn, causes changes in *X*. It occurs when the time periods used in the analysis (e.g., years rather than weeks) are so long that the direction of causality appears to be simultaneous. One way that has been used to model such a situation is to use simultaneous equations.

Simultaneous equations. Equations within a model in which a number of dependent variables appear as explanatory variables in more than one equation. These dependent variables are simultaneously determined by other dependent variables and explanatory variables in the system. For example, increased demand for a product could lead to increased sales, which lead to economies of scale, which lead to lower costs, which lead to lower prices, which then lead to increased sales for the product. Research on simultaneous equations was popular in the 1950s. According to Christ (1960) however, this approach had not been useful in forecasting and this conclusion has not changed since then.

Single-equation model. A model in which a single dependent variable is determined by the explanatory variables in one equation.

SKU. Stock-keeping unit.

Smoothing. Removing randomness by using some form of averaging. The term *smoothing* is used because such averages tend to reduce randomness by allowing positive and negative random effects to partially offset each other.

Smoothing constant. The weight given to the most recent observation in exponential smoothing.

Spatial diffusion. The spread of an innovation, like a new product, to new geographical areas.

Spearman rank correlation. A nonparametric measure of the association that exists between two sets of rankings. Siegel and Castellan (1988) describe this measure. (See also Kendall rank correlation.)

Special event. See discrete event.

Specification error. An error resulting from use of an inappropriate model, for example, the omission of an important variable, the inclusion of an irrelevant variable, or selection of an inappropriate functional form.

Specification tests. See misspecification tests.

Spectral analysis. The decomposition of a time-series data into a set of sine or cosine waves with differing amplitudes, frequencies, and phase angles.

Split samples. See cross-validation**.**

Spreadsheet add-ins. Utility programs that accomplish tasks not performed by the basic functions of the spreadsheet program itself. The spreadsheet add-ins provide time-series forecasting routines, such as exponential smoothing, supplement the results from the spreadsheet's multiple regression function, and provide forecasting graphics. See Tashman and Hoover (2001) for details on such programs.

Spurious relationships. Statistical relationships between variables that have no reason to be related. Such relationships are common in time series where two unrelated variables may be correlated because they are both related to another factor, such as gross national product. For example, the oft-noted strong correlation between liquor sales and teachers' salaries does not mean that an increase in teachers' salaries causes liquor sales to rise.

SSCI (Social Science Citation Index). The primary source for literature searches in the social sciences. It is useful for finding research publications on forecasting, which, in turn, are useful for a priori analysis. The *SSCI* allows for an efficient search because it does not include opinion-based articles. It also provides information on citations of papers, so if you find a paper that is useful in forecasting, you can also track down related papers.

S-shaped curve. Any one of a number of functional forms (such as the logistic curve) that starts out slowly but at an increasing rate, but then the rate slows as it approaches an asymptote (see saturation level). Such curves can be used to capture a diffusion process, as described in Meade and Islam (2001).

Standard deviation. The square root of the variance. A summary statistic, usually denoted by s, that measures variation in the sample. For data that are normal, Y ? $2s$ represents a 95% prediction interval at the origin of the forecast horizon (or, for cross-sectional data, at the mean of the calibration data).

Standard error of the estimate. A measure of the precision of an estimate for a coefficient in a regression model. It is the standard deviation for an estimate and it provides a crude measure of how reliably the relationship has been measured.

Standard error of the model. The standard deviation of the error term in the fit of a model to the calibration data. This is a poor measure for comparing the predictive validity of time-series models; use it only if no other measures can be obtained and use it with skepticism.

Standardize. To put data on a common basis by removing the effects of scale. One way to do this is to control for the variation in variables. For examp le, given a sample set of values for X, where the mean is $\bar{X}$ and the standard deviation is s, the ith value in the set, X_i, is standardized by subtracting the mean and dividing by the standard deviation.

Starting value. The initial values used to begin the estimation of exponential smoothing models in calibration data. Not to be confused with forecasts starting at the beginning of the forecast horizon, which is commonly referred to as the level at the origin. PoFxxx

State-space model. Multi-equation or matrix representation for a univariate or multivariate time series. State-space modeling is a way to handle computations for a variety of time-series models. Some forecasting methods use state-space models directly. Computations for state-space models are carried out using the Kalman filter.

Static simulation. The use of a model with actual values for the explanatory variables. In an econometric model that includes lagged values of *Y*, a static simulation uses the actual values of these lags, rather than the forecasted values. (See ex post forecast.)

Stationary series. A time series whose structure (e.g., mean, variance) does not change over time. Time-series methods often involve covariance (or weakly) stationary processes that have finite means and variances. Their means, variances, and covariances are unaffected by changes of time origin.

Statistical significance. The probability that a given result would be obtained, assuming that the null hypothesis were true. The misuses of statistical significance often outweigh its benefits (as shown for economics by McCloskey and Ziliak 1996, and for psychology by Cohen 1994 and Smith et al. 2000). Fisher (1925) claimed that significance testing is best used in conjunction with replication. However, statistical significance is useful in some aspects of forecasting, such as in determining whether to use a trend factor or whether to use seasonal factors, particularly when these involve small samples and high variation. When using statistical significance to test multiple hypotheses, such as a comparison of three or more forecasting methods, one should adjust the levels of significance (see "For Researchers" at hops.wharton.upenn.edu/forecast).

Statistical group. See Nominal Group Technique.

Stepwise regression. An automatic procedure for maximizing R^2 in multiple regression. There are several approaches to stepwise regression including forward (step-up) and backward (step-down) versions. The forward version first enters the causal variable with the highest correlation to the dependent variable, then enters the one with the highest partial correlation (given the variable already included in the model), then enters the variable with the highest partial correlation (given the two variables already included), and so on, until certain stopping rules are encountered. One common rule is to include all those and only those variables that have a *t*-statistic equal to or greater than 1. According to Haitovsky (1969), this rule maximizes the adjusted R^2. The step-down version puts all of the variables in initially, then removes the one that contributes least to R^2, next removes from the remaining variables the one that contributes least, and so on. Stepwise regression does not use much prior knowledge, other than to propose a possible set of variables and a functional form. As a result, stepwise regression should not be used for forecasting. In addition, empirical evidence does not support the use of stepwise regression for forecasting. For example, Armstrong (1985, pp. 54) developed two models to forecast camera sales per capita in each of 11 countries. Each of these models was developed using data from 19 other counties. An exploratory model used stepwise regression, drawing from a set of 15 variables, and the model with the highest R^2 was selected as the forecasting model. A theory-based model was also developed by selecting seven variables, by putting a priori constraints on the signs, and by incorporating prior estimates of magnitudes. Although the exploratory model provided the best fit to the 19-country calibration data ($\overline{R}^2$ of 99.8% vs. 99.6%), its performance in forecasting for an 11-country validation sample was inferior; its mean absolute percentage error was 52% vs. 31% for the theory-based model. The average percentage error (using the signs) of the theory-based model was also lower at 5% vs. 38%. If despite this advice, you insist on using stepwise regression and associated measures of statistical significance, use the tables provided by McIntyre et al. (1983).

Stochastic variable. A variable whose value changes. Measurements of stochastic variables reflect true changes and measurement error.

Structural break. A large change in a model that arises from a shift in the constant term or a shift in the relationship between an explanatory variable and a dependent variable.

Structural model. See causal model.

Structured judgment. An attempt to move beyond intuition in making judgmental forecasts. One approach is to formalize the way that a question is posed (e.g., decomposition, role-playing, and the Delphi technique are types of structure), the procedure for collecting responses (e.g., mail survey), and the method for summarizing the responses (e.g., averaging forecasts by of ten domain experts). PoFxxx

Successive reestimation. Reestimation of a model's coefficients each time a new observation becomes available. PoFxxx

Successive updating. Updating a model using the actual value of a new observation. Typically it refers to only an update in the level. You then obtain a sample of *h*-step-ahead forecasts based on originally estimated coefficients. Also called moving origin or rolling horizon. Armstrong (2001e) describes the use of successive updating.

Super ensemble. An average of averages. Combines a set of ensemble forecasts.

Supporting forces. Forces that reinforce the historical trend. Real-world examples of supporting forces are difficult to find because information about the trend in a series is assumed to be the dominant factor affecting behavior, and other factors are unimportant. Supporting forces might occur over specific periods for sales of fashion crazes or fad items, inflation, or for market prices such as for real estate or for internet stocks in the late 20th century. See Armstrong, Adya and Collopy (2001).

Survey error. Survey error is the total error due to sampling, nonresponse bias, and response error. Sampling error is often a small part of the total error, especially in new situations, for example, forecasting the effects of a new advertising strategy or sales of a new product. Researchers often confuse sampling error with survey error.

Surveys of consumer and business expectations. Surveys of consumers and firms as to their expectations about aspects of the economy. Such surveys have been used to forecast business conditions. George Katona at the Survey Research Center at the University of Michigan pioneered the development of surveys of consumer expectations in the 1940s. The surveys typically contain questions about what will likely happen in the next four months. Business people are asked about their expectations for salaries, profits, new orders, production, and their overall confidence levels. The IFO Institute for Economic Research in Munich (www.ifo.de) has encouraged countries to collect and analyze data on business and consumer expectations.

Suspicious pattern. A pattern in a time series that is judged by a domain expert to be behaving in an unexpected manner. The forecasts for such series should be conservative and the prediction intervals should be widened. PoFxxx

Switching model. A model composed of two (or more) submodels in which submodel A holds true in one set of circumstances, and submodel B in another (e.g., submodel A applies at time t if Y_{t-1} is greater than a specified value, submodel B if Y_{t-1} is less than that value). The purpose is to obtain accurate forecasts by using the most appropriate model for the situation.

Symmetric MAPE. See Adjusted MAPE.

Systems model. A model that tries to represent all key inputs and outputs of a situation.

Telescoping. A respondent's tendency to remember that a recent event occurred further back in time or that a distant event occurred more recently than it did. Telescoping can create problems in an intentions survey if people use their past behavior as a guide to the timing of their intentions. PoFxxx

Test market. See simulated test markets and test marketing.

Test marketing. A simulation where a product is made available to customers. For example, a simulated store that stocks the product in question. or the introduction of a product in limited (and isolated) geographical areas. While they are expensive and there are many threats to validity, the realism of test markets leads to good predictive validity.

Theil's U. Theil proposed two error measures, but at different times and under the same symbol "U," which has caused some confusion.

$$(1) \quad U = \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2\right]^{1/2}}{\left[\frac{1}{n}\sum_{i=1}^{n}A_i^2\right]^{1/2} + \left[\frac{1}{n}\sum_{i=1}^{n}P_i^2\right]^{1/2}}$$ better called U1

$$(2) \quad U = \frac{\left[\sum_{i=1}^{n}(P_i - A_i)^2\right]^{1/2}}{\left[\sum_{i=1}^{n}A_i^2\right]^{1/2}}$$ better called U2

U1 is taken from Theil (1958, pp. 31-42), where he calls U a measure of *forecast accuracy.* A_i represents the actual observations and P_i the corresponding predictions. He left it open whether A and P should be used as absolute values or as observed and predicted changes. Both possibilities have been taken up in the literature and used by different forecasters, while Theil himself applied U1 to changes. Theil (1966, chapter 2) proposed U2 as a measure of forecast quality, "where A_i and P_i stand for a pair of predicted and observed changes." Bliemel (1973) analyzed Theil's measures and concluded that U1 has serious defects and is not informative for assessing forecast accuracy regardless of being applied with absolute values of the changes. For example, when applying U1 to changes, all U1 values will be bounded by 0 (the case of perfect forecasting) and 1 (the supposedly worst case). However, the value of 1 will be obtained when a forecaster applies the simple no-change model (all P_i are zero). All other possible forecasts would lead to a U1 value lower than 1, regardless of whether the forecast method led to better or worse performance than the naive no-change model. U1 should therefore not be used and should be regarded as a historical oddity. In contrast, U2 has no serious defects. It can be interpreted as the RMSE of the proposed forecasting model divided by the RMSE of a no-change model. It has the no-change model (with U2 = 1 for no-change forecasts) as the benchmark. U2 values lower than 1.0 show an improvement over the simple no-change forecast. Some researchers have found Theil's error decomposition useful. For example, Ahlburg (1984) used it to analyze data on annual housing starts, where a mechanical adjustment provided major improvement in accuracy for the two-quarters-ahead forecast and minor improvements for eight-quarters-ahead. (See also Relative Absolute Error.) PoFxxx

Theory. A hypothesis that has received much support. In practice, theory is often used interchangeably with the word "hypothesis." Theory can be a dangerous term because it is often misused to mean "complicated and obscure arguments." Also, "theory" is often added to a paper after the study has been completed. A good theory should have predictive validity. To demonstrate how to test the predictive validity of theories, Armstrong (1991) examined theories about consumer behavior. In the *Journal of Consumer Research*, authors generally begin their papers by describing theories. Knowledge of such theories should lead one to make better forecasts. Sixteen academics in this field, presumably familiar with the theories, were asked to predict the outcomes of 20 studies with 105 hypotheses. All of these studies had been published in the *Journal of Consumer Research*, but the academics in the sample reported that they could not remember seeing them. As it turned out, their predictions were less accurate than those made by 43 high school students. Thus, contrary to the hypothesis, academic theories in consumer behavior did not have predictive validity.

Time-series data. A collection of values observed sequentially through time.

Time-series pooling method. An estimation method that pools data from analogous time series to improve the accuracy of a model for an individual time series. Pooling can be effective for estimating trends or seasonal factors for series with sparse data.

Time-varying parameter model. A specification of a forecasting model in which relationships (coefficients) change over time. It may be difficult to identify when parameters change and a time-varying parameter model might make changes in response to false signals. Some researchers advocate time-varying parameter models. Riddington (1993) systematically evaluated research on time-varying coefficients in forecasting. He "concludes conclusively that the [time-varying coefficient models] approach significantly improves forecasting performance." He reached this conclusion by summarizing results from 21 forecasting studies. However, Riddington's evidence is based only on ex post evaluations of forecast accuracy. (Ex post forecast evaluation can be useful for assessing how well models might predict the effects of changes in policy variables.) If the time-varying procedure provides substantially better parameter estimates, it might also improve ex ante forecasts. However, a common finding in this area is that refinements in the estimation of the parameters in econometric models do not contribute to ex ante accuracy. Time-varying-coefficients procedures are harder to understand, expensive, and may reduce the reliability of the model. Evidence that the parameters will change, or that they have recently changed, is unlikely to be found in the time series itself. If the structural changes are recent, then it is important to capture the changes. However, when one has only small samples (with perhaps unreliable data) and no domain knowledge data, the procedure may lead to a false identification of changes in parameters. Given the evidence to date, and modern computer capabilities, the analyst should simply rely on successive reestimation of models as more data are obtained, unless it is possible to use domain knowledge. (See adaptive parameters.)

Top-down. A procedure whereby a forecast of a disaggregate component is based on the forecast made of an aggregate variable (e.g., a forecast of menthol toothpaste based on a forecast for all toothpaste). Although this approach loses information about trends in the components (e.g., menthol flavor is becoming popular), reliability is usually better. (See also bottom-up.) PoFxxx

Tracking signal. A statistic that reveals when parameter estimates in a forecasting model are not optimal. For example, a tracking signal might be based on a graph of the ratio of the cumulative sum of the differences between the actual and forecast values to the mean absolute deviation. If the tracking signal exceeds a certain value, the series can then be flagged for examination. This concept has been used successfully in quality control. It seems sensible also for forecasting, although little research supports its use. An alternative is to use successive reestimation. PoFxxx

Trade-off analysis. An analysis based on surveys in which respondents make choices where they give up some benefits in order to receive others. See conjoint analysis. Wittink and Bergestuem (2001) discuss trade-off analysis.

Trading day. A day on which business is transacted. In many time series, the number of business days in a month (or some other specified period of time) may vary. Frequently, trading-day adjustments are needed to reflect the fact that a period (e.g., April) may not include the same number of trading days every year.

Transformation. The performance of an arithmetic operation upon a variable (e.g., taking the natural log of a variable or subtracting a constant). Data for an econometric model are often transformed by taking the logs of all variables, creating a so-called log-log model.

Treatment effect. The act of making a forecast causes a person to act differently in the future. See self-fulfilling prophesy, self-defeating prophesy, and unobtrusive measure.

Trees. A method of analyzing data by making a series of splits in the data. (See also AID.) PoFxxx

Trend analysis. Procedures for predicting trends. Trend analysis (or trend-line analysis) can be performed using different methods. For example, one can use exponential smoothing, simple regression in which time is the independent variable, robust trend, or simply the percentage change between two points in time.

True score. An accurate and valid measure of a concept. Observed test scores are rarely equal to the true scores. For example, a person's score on a test of verbal aptitude consists of her true verbal aptitude plus error.

t-test. A test of statistical significance that assumes a null hypothesis is true. See also F-test, as $F = t^2$.

Turing test. A test of face validity proposed by Turing (1950) in which an expert panel interrogates two unidentified sources – an expert system and an expert, and based on the responses, tries to determine which source is which. PoFxxx

Turning point. The point at which a time series changes direction. Determining the true turning point of a time series can be difficult. Despite their popular appeal to practitioners, turning-point measures have limited value because they do not contain information about the magnitude of changes. Furthermore, in most cases, the number of turning points is so small as to lack reliability as a measure of the comparative accuracy of forecasting methods. PoFxxx

Uncertainty. The lack of confidence associated with a forecast, which can be represented by a prediction interval. Also, the lack of confidence about a parameter estimate, which can be represented by a confidence interval. Uncertainty *cannot* be represented well by statistical significance. PoFxxx

Unconditional forecast. An estimate of what will happen in a situation when no actual data from that situation are used to produce the forecast. (See ex ante.) PoFxxx

Unit root. A measure for nonstationary time series, $Y(t)$, with a stationary transformation created by taking one (or more) first differences. If $Z(t) = Y(t) - Y(t-1)$ is a stationary series, $Y(t)$ has one unit root. (See Dickey-Fuller test.) Allen and Fildes (2001) discuss the use of unit roots in forecasting.

Unit weights. A factor of +1 or –1 used to weight predictor variables, where the signs are based on a priori information. These are often equivalent to equal weights. One may need to decide how to scale the variables. Typically, each variable's observations are transformed to standard normal deviates from the variable's mean.

Univariate time-series model. A model that uses only prior values of the series to make forecasts (see extrapolation).

Unobtrusive measure. Data obtained in situations in which the act of measurement does not affect the behavior of the object that is measured. For example, to forecast sales in shopping malls, one could secretly count the number of cars in the parking lots at various times, perhaps using photographs from high locations. The shoppers do not know they are being counted. Awareness of the measurement can change people's behavior. Fitzsimons and Morwitz (1996) present evidence on the use of intentions surveys can\affect subsequent behavior.

Updated forecast. A revision of an original forecast in light of data that became available after the original forecast was made. Updating can involve reestimation of the parameters of the model.

Updated model. A model whose level has been reestimated in light of new information. Frequent updating is important to accuracy. (See also adaptive parameters.)

Validation. In forecasting, the process of testing how accurate a model is for making forecasts. The sample data are often split into two segments, one used to estimate the parameters of the model, and the other, the holdout data, used to test the forecasts made with the model. The many variations of validation include cross-validation, n-way validation, and the jackknife. Validation can also be used to assess the usefulness of the parameters of a forecasting model. Armstrong (2001d) describes various approaches to validation.

Variance. A measure of variation equal to the mean of the squared deviations from the mean. As a result, observations with large deviations are heavily weighted.

Vector Autoregressive Model (VAR). A model in which a set of dependent variables are explained by lagged values of the same set of variables. Zellner (see Garcia-Ferrer 1998) refers to VARs as "very awful regressions," and claims that they have not been successful in forecasting. Allen and Fildes (2001) review evidence on VARs; it is weak. PoFxxx

Volatility. Large, sudden and unexplained fluctuations in time-series data.

WAG (Wild-assed guess). An intuitive forecast based on little information.

Wave. A set of responses to a mail survey request. The first wave consists of responses before they receive a second request. Similarly, the second wave consists of responses that come in after a second request, but before a third request is received. Trends across waves can be useful in analyzing nonresponse bias (Armstrong and Overton 1977). Adjustments for nonresponse bias are especially important for intentions surveys. They are generally irrelevant for expert opinions.

Weight. The importance given to a value. For example, in a four-year moving average, each year is generally given equal weight. In exponential smoothing, the weights decrease for older data. Also, weights refer to the emphasis given to components in a combined forecast. Finally, weights refer to the emphasis given to alternative parameter estimates.

Weighted Application Blank. A job application form listing various factors related to job performance. The weights on these factors can be obtained judgmentally (in a process similar to an expert system) or statistically, based on previous applicants' success (similar to econometric models) or based on the judgments of experts (in a process similar to judgmental bootstrapping). PoFxxx

Wilcoxon matched-pairs signed-ranks test. A nonparametric test used to determine whether a difference between two sets of paired data has statistical significance. This test gives more emphasis to larger differences and is almost as powerful as the *t*-test. Siegel and Castellan (1988) give details on this and other nonparametric tests that can be used to compare forecasts from two methods.

Wind-tunnel data. Data used to test alternative procedures. The M-Competition provides wind-tunnel data for extrapolation.

Winsorizing. The practice of modifying outliers in the data by making them no more extreme than the most extreme data that you believe to be relevant or accurately measured. Winsorizing data is one way to calculate a modified mean.

Winters exponential smoothing. See Holt-Winters' exponential smoothing method.

Wishful thinking. See optimism.

X-11 decomposition. A set of statistical procedures for calculating seasonal factors in time-series data. The X-11 method for time-series decomposition is part of the Census II family developed at the United States Bureau of the Census originally developed in 1960s and improved in the X-11-ARIMA method. It has now been superseded by the X-12-ARIMA method.

X-12-ARIMA decomposition. An update of the X-11 decomposition method for time-series decomposition from the Census II family. Details can be found at hops.wharton.upenn.edu/forecast.

REFERENCES

Ahlburg, D. (1984), "Forecast evaluation and improvement using Theil's decomposition," *Journal of Forecasting*, 3, 345-351.

Allen, P. G. & R. Fildes (2001), "Econometric forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Arkes, H. R. (2001), Overconfidence in judgmental forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S. (1970), "An application of econometric models to international marketing," *Journal of Marketing Research*, 7, 190-198. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (1980a), "Unintelligible management research and academic prestige," *Interfaces*, 10, No. 2, 80-86. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (1980b), "The seer-sucker theory: The value of experts in forecasting," *Technology Review*, 83 (June/July), 18-24. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (1982a), "The value of formal planning for strategic decisions: Review of empirical research," *Strategic Management Journal*, 3, 197-211. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (1982b), "Strategies for implementing change: An experiential approach," *Group and Organization Studies*, 7, 457-475. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (1983) "Strategic planning and forecasting fundamentals," in K. Albert, *The Strategic Management Handbook*. New York: McGraw Hill, pp. 2-1 to 2-32. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (1985), *Long-Range Forecasting*. New York: John Wiley. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S. (1988), "Review of *The Great Depression of 1990*, by Ravi Batra," *International Journal of Forecasting*, 4, 493-502. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S. (1990), "Review of *Corporate Strategic Planning* by Capon et al.," *Journal of Marketing*, 54, 114-119. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S. (1991), "Prediction of consumer behavior by experts and novices," *Journal of Consumer Research*, 18, 251-256. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S. (2001a), "Role-playing: A method to forecast decisions," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S. (2001b), "Judgmental bootstrapping: Inferring experts' rules for forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S. (2001c), "Extrapolation for time-series and cross-sectional data," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S. (2001d), "Evaluating forecasting methods," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S. (2001e), "Combining forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S., M. Adya & F. Collopy (2001), "Rule-based forecasting: Using judgment in time-series extrapolation," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. S. & J. G. Andress (1970), "Exploratory analysis of marketing data: Trees vs. regression," *Journal of Marketing Research*, 7, 487-492. Full text at hops.wharton.upenn.edu/forecast

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

Armstrong, J. S. & F. Collopy (1992b), "Expert opinions about extrapolation and the mystery of the overlooked discontinuities," *International Journal of Forecasting*, 8, 575-582. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. & F. Collopy (1993), “Causal forces: Structuring knowledge for time-series extrapolation,” *Journal of Forecasting*, 12, 103-115. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S. & F. Collopy (2001), “Identification of asymmetric prediction intervals through causal forces,” *Journal of Forecasting* (forthcoming).

Armstrong, J. S. & J. U. Farley (1969), “A note on the use of Markov chains in forecasting store choice,” *Management Science*, 16, B-281 to B-285.

Armstrong, J. S. & R. Fildes (1995), “On the selection of error measures for comparisons among forecasting methods,” *Journal of Forecasting*, 14, 67-71. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S., V. G. Morwitz & V. Kumar (2000), “Sales forecasts for existing consumer products and services: Do purchase intentions contribute to accuracy?” *International Journal of Forecasting*, 16, 383-397. Full text at hops.wharton.upenn.edu/forecast

Armstrong, J. S. & T. S. Overton (1977), “Estimating nonresponse bias in mail surveys,” *Journal of Marketing Research*, 14, 396-402. Full text at hops.wharton.upenn.edu/forecast

Ashton, A. H. (1985), “Does consensus imply accuracy in accounting studies or decision making?” *Accounting Review*, 60, 173-185.

Ashton, A. H. (1986), “Combining the judgments of experts: How many and which ones?” *Organizational Behavior and Human Decision Processes*, 38, 405-414.

Auerbach, A. J. (1982), “The index of leading indicators: Measurement without theory, thirty-five years later,” *Review of Economics and Statistics*, 64, 589-595.

Beach, L. R., V. E. Barnes & J. J. J. Christensen-Szalanski (1986), “Beyond heuristics and biases: A contingency model of judgmental forecasting,” *Journal of Forecasting*, 5, 143-157.

Bezdek, R. H. (1974), “Empirical tests of input-output forecasts: Review and critique,” U. S. Department of Commerce, BEA Staff paper No. 24, July 1974.

Bliemel, F.W. (1973), “Theil’s forecast accuracy coefficient: A clarification,” *Journal of Marketing Research*, 10, 444-446.

Box, G. E. P. & G. M. Jenkins (1970), *Time-Series Analysis.* San Francisco: Holden-Day. Later editions were published in 1976 and 1994, the latter with G.C. Reinsell.

Brandenburger, A. M. & B. J. Nalebuff (1996), *Co-opetition*. New York: Doubleday.

Brier, G. W. (1950), “Verification of forecasts expressed in terms of probability,” *Monthly Weather Review*, 75, 1-3.

Brunswick, E. (1955), “Representative design and probabilistic theory in functional psychology,” *Psychological Review*, 62, 193-217.

Buchanan, W. (1986), “Election predictions: An empirical assessment,” *Public Opinion Quarterly*, 50, 222-227.

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

Campbell, J. Y. & P. Perron (1991), "Pitfalls and opportunities: What macroeconomists should know about unit roots," *National Bureau of Economic Research Technical Working Paper No. 100*, also in the *NBER Macroeconomics Annual*, pp. 141-201.

Chamberlain, T. C. (1965), "The method of multiple working hypotheses," *Science*, 148, 754-759.

Christ, C. F. (1960), "Simultaneous equation estimation: Any verdict yet?" *Econometrica*, 28, 835-845.

Christensen-Szalanski, J. J. J. (1991), "The hindsight bias: A meta-analysis," *Organizational Behavior and Human Decision Processes*, 43, 147-168.

Cohen, J. (1994), "The earth is round (p < .05)," *American Psychologist*, 49, 997-1003.

Collopy, F., M. Adya & J. S. Armstrong (2001), "Expert systems for forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Cooper, H. & R. Rosenthal (1980), "Statistical versus traditional procedures for summarizing research findings," *Psychological Bulletin*, 87, 442-449.

Cosier, R. A. (1978), "The effects of three potential aids for making strategic decisions on prediction accuracy," *Organizational Behavior and Human Performance*, 22, 295-306.

Croston, J. D. (1972), "Forecasting and stock control for intermittent demand," *Operational Research Quarterly*, 23, 289-303.

Curtis, E. W. & E. F. Alf (1969), "Validity, predictive efficiency, and practical significance of selection tests," *Journal of Applied Psychology*, 53, 327-337.

Cyert, R., J. March & W. Starbuck (1961), "Two experiments on bias and conflict in organizational estimation," *Management Science*, 7, 254-264.

Day, D., B. Gan, P. Gendall & D. Esslemont (1991), "Predicting purchase behavior," *Marketing Bulletin*, 2, 18-30 (full text at marketing-bulletin.massey.ac.nz/articles/articles.asp).

Diebold, F. X. & G. D. Rudebusch (1991), "Forecasting output with the composite leading index: A real time analysis," *Journal of the American Statistical Association*, 86, 603-610.

Dielman, T. E. (1986), "A comparison of forecasts from least absolute value and least squares regression," *Journal of Forecasting*, 5, 189-195.

Dillman, D. (2000), *Mail and Internet Surveys*. New York: John Wiley.

Duncan, G., W. L. Gorr & J. Szczypula (2001), "Forecasting analogous time series," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Einhorn, H. J. (1972), "Alchemy in the behavioral sciences," *Public Opinion Quarterly*, 36, 367-378.

Einhorn, H. J. & R. M. Hogarth (1982), "Prediction diagnosis and causal thinking in forecasting," *Journal of Forecasting*, 1, 23-36.

Encel, S., P. K. Marstrand & W. Page (1975), *The Art of Anticipation*. London: Martin Robertson.

Fern, E. F. (1982), "The use of focus groups for idea generation: The effects on group size, acquaintanceship and moderator on response quantity and quality," *Journal of Marketing Research*, 19, 1-13.

Fildes, R., D. Dews & S. Howell (1981), *A Bibliography of Business and Economic Forecasting*. Westmead, England: Gower.

Fildes, R. & R. Hastings, (1994), “The organization and improvement of market forecasting,” *Journal of the Operational Research Society*, 45, 1-16.

Fildes, R., M. Hibon, S. Makridakis & N. Meade (1998), "Generalizing about univariate forecasting methods: Further empirical evidence," *International Journal of Forecasting*, 14, 339-358.

Fischhoff, B. (1975), “Hindsight ? foresight: The effect of outcome knowledge on judgment under uncertainty,” *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-297.

Fischhoff, B. (2001), “Learning from experience: Coping with hindsight bias and ambiguity,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Fischhoff, B. & D. G. MacGregor (1982), “Subjective confidence in forecasts,” *Journal of Forecasting*, 1, 155-172.

Fisher, R. A. (1925), *Statistical Methods for Research Workers*. London: Oliver & Boyd.

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

Fralicx, R. & N. S. Raju (1982), “A comparison of five methods for combining multiple criteria into a single composite,” *Educational and Psychological Measurement*, 42, 823-827.

Franses, P. H. & D. J. C. Van Dijk (2000), *Non-linear Time Series: Models in Empirical Finance*. Cambridge, U. K.: Cambridge University Press.

Friedman, M. (1953), “ The methodology of positive economics,” in *Essays in Positive Economics.* Chicago: U. of Chicago Press.

Gabor, A. & C. W. J. Granger (1972), “Ownership and acquisition of consumer durables: Report on the Nottingham consumer durables project,” *European Journal of Marketing*, 6 (4), 234-248.

García-Ferrer, A. (1998), “Professor Zellner: An interview for *the International Journal of Forecasting,” International Journal of Forecasting*, 14, 303-312.

Gardner, E. S. Jr. (1985), “Exponential smoothing: The state of the art,” *Journal of Forecasting*, 4, 1-28.

Glass, G.V. (1976), “Primary, secondary and meta-analysis of research,” *Educational Researcher*, 5, 3-8.

Gregory, L. & A. Duran (2001), “Scenarios and acceptance of forecasts,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Haitovsky, Y. (1969), “A note on the maximization of R^2,” *American Statistician*, 23, (Feb.), 20-21.

Hayes, S. P. (1936), “The predictive ability of voters,” *Journal of Social Psychology*, 7, 185-191.

Hinrichs, J. R. (1978), “An eight-year follow-up of a management assessment center,” *Journal of Applied Psychology*, 63, 596-601.

Hogarth, R. M. (1978), “A note on aggregating opinions,” *Organizational Behavior and Human Performance*, 21, 40-46.

Hubbard, R. & J. S. Armstrong (1992), “Are null results becoming an endangered species in marketing,” *Marketing Letters*, 3, 127-136. Full text at hops.wharton.upenn.edu/forecast

Jarvik, M. E. (1951), "Probability learning and negative recency effect in the serial anticipation of alternative symbols," *Journal of Experimental Psychology*, 291-297.

Juster, F. T. (1964), *Anticipations and Purchases: An Analysis of Consumer Behavior.* Princeton, N. J.: Princeton University Press.

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

Kondratieff, N. D. (1935), "The long waves in economic life," presented in an English translation in *Review of Economics and Statistics*, 17 (6), pp.105-115. Reprinted in J. J. Clark and M. Cohen, *Business Fluctuations Growth and Stabilization.* New York: Random House, 1963, p. 44.

Koopmans, T. (1947), "Measurement without theory," *Review of Economics and Statistics*, 29 (Aug.), 161-172.

Lahiri, K. & G. H. Moore (1991), *Leading Economic Indicators: New Approaches and Forecasting Records*. Cambridge, UK: Cambridge University Press.

Langer, E. J. & J. Roth (1975), "Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task," *Journal of Personality and Social Psychology*," 32, 951-955.

Lau, R. (1994), "Analysis of the accuracy of 'trial heat' polls during the 1992 presidential election," *Public Opinion Quarterly*, 58, 2-20.

Lawrence, M. J., R. H. Edmundson & M. J. O'Connor (1985), "An examination of judgmental extrapolation of time series," *International Journal of Forecasting*, 1, 25-35.

Lewin, K. (1952), "Group decision and social change," in G. E. Swanson, T. Newcomb & E. L. Hartley (eds.), *Readings in Social Psychology* (2nd ed.). New York: Holt.

Libby, R. & R. K. Blashfield (1978), "Performance of a composite as a function of the number of judges," *Organizational Behavior and Human Performance*, 21, 121-129.

Lichtenstein, S., B. Fischhoff & L. Phillips (1982), "Calibration of probabilities: The state of the art to 1980," in Kahneman, D., P. Slovic & A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

MacGregor, D. G. (2001), "Decomposition in judgmental forecasting and estimation," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Maddala, G. S. & I. M. Kim (1998), *Unit Roots, Cointegration and Structural Change*. Cambridge, UK: Cambridge University Press.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen & R. Winkler (1982), "The accuracy of extrapolation (time-series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1, 111-153.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen & R. Winkler (1984), *The Forecasting Accuracy of Major Time Series Models.* Chichester: John Wiley.

Makridakis, S., C. Chatfield, M. Hibon, M. Lawrence, T. Mills, K. Ord & L. F. Simmons (1993), "The M2-Competition: A real-time judgmentally based forecasting study," *International Journal of Forecasting*, 9, 5-22.

Makridakis, S. & M. Hibon (1979), “Accuracy in forecasting: An empirical investigation” (with discussion), *Journal of the Royal Statistical Society*, Series A, 142, 97-145.

Makridakis, S. & M. Hibon (2000), “The M3-Competition: Results, conclusions and implications,” *International Journal of Forecasting*, 16, 451-476.

Makridakis, S., S. C. Wheelwright & R. J. Hyndman (1998), *Forecasting Methods and Applications*. New York: John Wiley.

Mann, C. N. (1994), “Can meta-analysis make policy?” *Science*, 266 (November 11), 960-962.

McCloskey, D. N. & S. T. Ziliak (1996), “The standard error of regressions, “*Journal of Economic Literature*, 34, 97-114.

McIntyre, S. H., D. B. Montgomery, V. Srinivasan & B. A. Weitz (1983), “Evaluating the statistical significance of models developed by stepwise regression,” *Journal of Marketing Research*, 20, 1-11.

Meade, N. & T. Islam (2001), “Forecasting the diffusion of innovations: Implications for time-series extrapolatioin,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Meehl, P. (1954), *Clinical versus Statistical Prediction.* Minneapolis: University of Minnesota Press.

Meese, R. & J. Geweke (1984), "A comparison of autoregressive univariate forecasting procedures for macroeconomic time series," *Journal of Business and Economic Statistics*, 2, 191-200.

Mentzer, J. T. & J. E. Cox, Jr. (1984), “Familiarity, application, and performance of sales forecasting techniques,” *Journal of Forecasting*, 3, 27-36,

Mentzer, J. T. & K. B. Kahn (1995), “Forecasting technique familiarity, satisfaction, usage, and application” *Journal of Forecasting*, 14, 465-476.

Montgomery, D. & D. Morrison (1973), “A note on adjusting R^2,” *Journal of Finance*, 28, 1009-1013.

Morwitz, V. G. (2001), “Methods for forecasting from intentions data,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Murphy, A. H. (1972), “Scalar and vector partitions of the probability score (Part I), Two state situation,” *Journal of Applied Meteorology*, 11, 273-282.

Murphy, A. H. & R. L. Winkler (1984), “Probability forecasting in meteorology,” *Journal of the American Statistical Association*, 79, 489-500.

Neftci, S. N. (1982), “Optimal prediction of cyclical downturns,” *Journal of Economic Dynamics and Control*, 4, 225-241.

Newton, J. R. (1965), “Judgment and feedback in a quasi-clinical situation,” *Journal of Personality and Social Psychology*, 1, 336-342.

Ohlin, L. E. & O. D. Duncan (1949), “The efficiency of prediction in criminology,” *American Journal of Sociology*, 54, 441-452.

Oskamp, S. (1965), “Overconfidence in case study judgments,” *Journal of Consulting Psychology*, 29, 261-265.

Ozer, D. J. (1985), “Correlation and the coefficient of determination,” *Psychological Bulletin*, 97, 307-315.

Pant, P. N. & W. H. Starbuck (1990), "Innocents in the forest: Forecasting and research methods," *Journal of Management*, 16, 433-460.

Payne, S. (1951), *The Art of Asking Questions.* Princeton: Princeton University Press.

Perry, P. (1979), "Certain problems with election survey methodology," *Public Opinion Quarterly*, 43, 312-325.

Plous, S. (1993), *The Psychology of Judgment and Decision Making*. New York: McGraw Hill.

Rao, A. V. (1973), "A comment on forecasting and stock control for intermittent demands," *Operational Research Quarterly*, 24, 639-640.

Remus, W. & M. O'Connor (2001), "Neural networks for time-series forecasting," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Riddington, G. L. (1993), "Time varying coefficient models and their forecasting performance," *Omega*, 21, 573-583.

Rogers, E. M. (1995), *Diffusion of Innovations.* New York: The Free Press.

Rosenthal, R. (1978), "Combining results of independent studies," *Psychological Bulletin*, 85, 185-193.

Rosenthal, R. & L. Jacobson (1968), *Pygmalion in the Classroom*. New York: Holt, Rinehart and Winston.

Rowe, G. & G. Wright (1999), "The Delphi technique as a forecasting tool: Issues and analysis," *International Journal of Forecasting*, 15, 351-371 (includes commentary).

Rowe, G. & G. Wright (2001), "Expert opinions in forecasting: Role of the Delphi technique," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Sanders, N. R. & K. B. Mandrodt (1990), "Forecasting practices in U. S. corporations," *Interfaces*, 24 (2), 92-100.

Sanders, N. R. & L. Ritzman (2001), "Judgmental adjustments of statistical forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Schwarz, G. (1978), "Estimating the dimensions of a model," *Annals of Statistics*, 6, 461-464.

Schwenk, C. R. & R. Cosier (1980), "Effects of the expert, devil's advocate, and dialectical inquiry methods on prediction performance," *Organizational Behavior and Human Performance*, 26, 409-423.

Sethuraman, P. & G. J. Tellis (1991), "An analysis of the tradeoff between advertising and price discounting," *Journal of Marketing Research*, 28, 160-174.

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

Shiskin, J. (1957), "Electronic computers and business indicators," *Journal of Business*, 30 (4), 219-267.

Siegel, S. & N. J. Castellan, Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. New York: McGraw-Hill.

Slovic, P. & D. J. McPhillamy (1974), "Dimensional commensurability and cue utilization in comparative judgment," *Organizational Behavior and Human Performance*, 11, 172-194.

Smith, L. D., L. A. Best, V. A. Cylke & D. A. Stubbs (2000), "Psychology without *p* values," *American Psychologist*, 55, 260-263.

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

Stewart, T. R. (2001), "Improving reliability of judgmental forecasts," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Stuckert, R. P. (1958), "A configurational approach to prediction," *Sociometry*, 21, 225-237.

Sudman, S. & N. N. Bradburn (1982), *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey Bass.

Tashman, L. J. & J. Hoover (2001), "Diffusion of forecasting principles: An assessment of forecasting software programs," in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Tashman, L. J. & M. L. Leach (1991), "Automatic forecasting software: A survey and evaluation," *International Journal of Forecasting*, 7, 209-230.

Teigen, K. H. (1990), "To be convincing or to be right: A question of preciseness," in K. J. Gilhooly, M. T. G. Keane, R. H. Logie & G. Erdös (eds.), *Lines of Thinking*. Chichester: John Wiley, pp. 299-313.

Tellis, G. J. (1988), "The price elasticity of selective demand: A meta-analysis of econometric models of sales," *Journal of Marketing Research*, 25, 331-341.

Tetlock, P.E. (1999), "Theory-driven reasoning about plausible pasts and probably futures in world politics: Are we prisoners of our preconceptions?" *American Journal of Political Science*, 43, 335-366.

Theil, H. (1958), *Economic Forecasts and Policy*. Amsterdam: North Holland.

Thiel, H. (1966), *Applied Economic Forecastin*g. Chicago: Rand McNally.

Tull, D. S. (1967), "The relationship of actual and predicted sales and profits in new-product introductions," *Journal of Business*, 40, 233-250.

Turing, A. M. (1950), "Computing machinery and intelligence," *Mind*, 59, 443-460.

Tversky, A. & D. Kahneman (1974), "Judgment under uncertainty," *Science*, 185, 1122-1131.

Tversky, A. & D. Kahneman (1981), "The framing of decisions and the psychology of choice," *Science*, 211, 453-458.

Tversky, A. & D. Kahneman (1982), "Judgments of and by representativeness," in D. Kahneman, P. Slovic & A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press.

Tversky, A. & D. Kahneman (1983), "Extensional vs. intuitive reasoning: the conjunctive fallacy in probability judgment," *Psychological Review*, 90, 293-315.

Tyebjee, T. T. (1987), "Behavioral biases in new product forecasting," *International Journal of Forecasting*, 3, 393-404.

Uhl, N. & T. Eisenberg (1970), "Predicting shrinkage in the multiple correlation coefficient," *Educational and Psychological Measurement*, 30, 487-489.

Wagenaar, W. A. (1978), “Intuitive predictions of growth,” in D. F. Burkhardt & W. H. Ittelson (eds.), *Environmental Assessment of Socioeconomic Systems.* New York: Plenum.

Wagenaar, W. A. & G. B. Keren (1986), "The seat belt paradox: Effect of accepted roles on information seeking," *Organizational Behavior and Human Decision Processes*, 38, 1-6.

Webby, R., M. O’Connor & M. Lawrence (2001), “Judgmental time-series forecasting using domain knowledge,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Willemain, T. R., C. N. Smart, J. H. Schockor & P. A. DeSautels (1994), “Forecasting intermittent demand in manufacturing: A comparative evaluation of Croston’s method,” *International Journal of Forecasting*, 10, 529-538.

Wimbush, J. C. & D. R. Dalton (1997), “Base rate for employee theft: Convergence of multiple methods,” *Journal of Applied Psychology*, 82, 756-763.

Winters, P. R. (1960), “Forecasting sales by exponentially weighted moving averages,” *Management Science*, 6, 324-342.

Wittink, D. R. & T. Bergestuen (2001), “Forecasting with conjoint analysis,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Press.

Yokum, T. & J. S. Armstrong (1995), “Beyond accuracy: Comparison of criteria used to select forecasting methods,” *International Journal of Forecasting*, 11, 591-597.