UC Berkeley

UC Berkeley Previously Published Works

Title

Information recovery and causality: A tribute to George Judge

Permalink

<https://escholarship.org/uc/item/5zz731q1>

Authors

Rausser, G
Bessler, DA

Publication Date

2016-10-05

DOI

10.1146/annurev-resource-121615-032137

Peer reviewed

Information Recovery and Causality: A Tribute to George Judge

by Gordon Rausser and David A. Bessler

Abstract

In Professor Judge's pursuit of information recovery and isolating causality in noisy effects observational data, there is a critical distinction between deductive versus inductive empirical analysis. For the former, we bring together a synthesis of the literature that has emerged since Koopmans' measurement with theory philosophy, and for the latter, we present a host of methodologies which attempt to isolate the causal mechanisms that exist in patterns revealed in noisy measurement data. The deductive focus is limited by available theoretical constructs while the inductive focus is fraught with "data mining" complications.

JEL Codes: C51, C52, C53, C55

Keywords: Econometrics, Causality, Forecasting, Big data, Econometric paradigms, Maintained hypotheses, Data mining

1. Introduction

Academic researchers focusing on the four major themes of *The Annual Review of Resource Economics*, when faced largely with non-experimental data, have constantly consulted the remarkable work of Professor George Judge and his many joint authors over the years. There have been many advancements orchestrated by the intellectual leadership of Professor Judge. These consultations have run the gamut, including, inter alia, estimation and inference from simultaneous equations models; data mining and the need for pre-test estimators with expanded confidence intervals; inequality restrictions that may result from economic theory; the use of prior information from outside of our sample data; the need to impose stochastic restrictions; the interpretation and estimation of transition probabilities from one state or category to another; the specification and estimation of Markoff processes; a complete assessment of the distinction between traditional and Stein-rule estimators; articulation of the sampling properties, both asymptotic and finite for a number of estimators; improved methodology for prediction in the presence of multicollinearity and heteroscedasticity obstacles; generalized moment-based estimation and inference; and extending the combination of prior information and sample data in the context of Bayesian analysis.

In the decade following the completion of his Ph.D. dissertation, operations research provided a common methodology for evaluating spatial and temporal prices and allocations. Here, George's work with colleagues reached the pinnacle of spatial equilibrium models with the publication of his extraordinary, co-authored paper, "Equilibrium among Spatially Separated Markets: A Reformulation." This work and his two books published on the topic integrated more

than two decades of work by agricultural economists. Following his retirement from teaching, George moved in still new directions, collaborating not only with fellow economists, but physicists, information theorists, and a number of statisticians. His recent work emphasizes the connection between physical thermodynamic systems and economic systems, where the amount of information revealed from a particular system is referred to as entropy.

Even though Professor Judge has retired from teaching, he has never retired from research. He is actively engaged in researching the ill-posed, inverse information recovery problem with information theoretic methods. This problem pervades all of our empirical analyses, given the unknown mathematical form of the distribution of the underlying, indirect, noisy effects measured by the sample data. As George notes in his article published in this volume, any attempts to measure causal influence require the solution of a stochastic inverse problem. This is because the data comes to us in the “effects domain” while our interest lies in the “causal domain.” An ill-posed problem is simply defined as one where the number of measurements, or data points, is smaller than the number of unknown parameters to be estimated. This new focus of George and his colleagues has resulted in numerous refereed journal articles and a major book, *An Information Theoretic Approach to Econometrics*, co-authored with Ron Mittelhammer.

George Judge, for much of his career, followed a clear and rigorous path for implementing a research agenda laid-out by the Cowles Commission at Chicago and Yale during the 1940s and 1950s. This work is documented in well over 100 refereed journal articles, 16 books – and we’re sure that others are on their way – and his editorial guidance, primarily

through *The Journal of the American Statistical Association* and *The Journal of Econometrics*. He has collaborated with a large number of colleagues and provided mentorship for numerous graduate students at Connecticut, Oklahoma State, Illinois, and Berkeley. He has forged a legacy that few can ever match. Yet he pushes on, not content with the wealth of estimators and models that have been advanced over the last seven decades. Now at 90, he has cultivated a new path, as noted above, using an information theoretic entropy basis for information recovery and inference for noisy measurements contained in observed sample data. He now explicitly asks us to recognize the stochastic nature of the inverse information recovery problem that, basically in applications, is ill-posed. In essence, he asks us to understand the uncertainty in our uncertainty, taking us back to I.J. Good's (1965, 1983) probabilities of our probabilities.

In our profession's pursuit of this path, this article argues that there is an important empirical distinction between deductive versus inductive applications, all the while searching for the underlying causal mechanisms faced by the inverse information recovery problem. We bring together a synthesis of literature related to Professor Judge's article in this volume, one focus which accepts the measurement with theory philosophy of Koopman's (1947) and a second focus which attempts to ferret out the causal mechanisms through patterns in the noisy measurement data by induction. The former focus is limited by available theoretical constructs while the latter focus is fraught with "data mining" complications. Ultimately, the empirical identification of causality from either or both of these two focuses should be measured in terms of the out-of-sample, predictive performance.

2. Deductive Methodology

While the process of formulating hypotheses has been formalized in science and economics over the centuries, there persists substantial disagreement among philosophers and economists on matters of “validity,” “truth,” “reality,” and thus model validation. The 18th century philosopher David Hume (1748) gave structure to the scientific process in *An Enquiry Concerning Human Understanding*, which argued the necessity of reasoning, rather than merely observation, and advanced the fundamental notion of induction. An intellectual leader in this field of inquiry, Karl Popper (2005) wrote almost 200 years later that “[a] scientist, whether theorist or experimenter, puts forward statements, or systems of statements, and tests them step by step. In the field of the empirical sciences, more particularly, he constructs hypotheses, or systems of theories, and tests them against experience by observation and experiment” (p. 3). It may be tempting for economists to comb through the available data, looking for information that appears to be supportive of their position. This approach, however, should be viewed with skepticism, since it is not consistent with sound scientific practice, which requires the investigator first to specify a theory with testable predictions and then to test those predictions with empirical observation. Popper writes that “a hypothesis can only be empirically *tested* — and only *after* it has been advanced.” (p. 9)¹ This sentiment has been echoed elsewhere in the literature (Christ and Rausser, 1973), “We are never justified in inferring theories from empirical observations” (p. 273). Patterns are readily observed in the real world, and paying casual attention to them, rather than following a formal process of logical theory, model specification and estimation, may lead to spurious inferences (Shapiro, 1973, p. 252).

1 (emphasis in the original).

There are a number of useful concepts introduced by other philosophers (cf. [3, 7, 8, 11, 12, 13, 19]) which substantiate some and refute other elements of Popper's views. These concepts assist us in understanding the process of model specification, estimation, selection, and evaluation. At this stage, we restrict our remarks to non-experimental or observation data in the context of deductive, Cowles Commission econometric methodology. The methodology turns on paradigms and economic theory; model specification, estimation, and evaluation.

3. Paradigms and Economic Theory

The application of Thomas Kuhn's (1970) notion of a paradigm, which was originally formulated for natural sciences, has become fashionable in economics over the past several decades.² In the present context, it is useful to note that a paradigm typically defines a large set of possible hypotheses and makes no claims for the validity of any particular member of that set. A paradigm in economics, as in the natural sciences, "defines the type of relationships to be investigated and the methods and abstractions which are regarded as legitimate within a particular problem area" (Christ and Rausser, 1973, p. 274). Accepted economic paradigms include the assumption that agents on average act rationally and that the laws of supply and demand apply. Paradigms, unlike the hypotheses to which they give rise, cannot be validated by experimental or statistical methods.³ For example, the maximum principle or rationality is not a hypothesis but a paradigm; and although a specific hypothesis embodying some version of the

² A paradigm, in the natural sciences as well as in economics, defines the type of relationships to be investigated and the methods and abstractions which are regarded as legitimate within a particular problem area.

³ Some of the members may be mutually exclusive alternatives.

maximum principle can, in effect, be tested, the maximum principle paradigm cannot.⁴ This does not suggest that the maximum principle ought to be abandoned. On the contrary, since abstractions from details are essential, any usable model must be a misspecification of the system to which it refers. We have no option but to construct models which fall short of a complete specification of the system under examination. In this sense all model representations, econometric or otherwise, are partially reduced forms (due to omission of variables, distortion of relationships, aggregation, etc.) even though we frequently refer to them as structural models. It is always possible to imagine a more fundamental explanation of the phenomenon under examination involving more equations and thus endogenous variables. Hence it appears reasonable to suggest that (a) economic models cannot be judged solely by the resemblance between their specification and the systems which they are designed to represent, and (b) the choice of different model specifications of the same system by different economists implies no presumption that one of them must be in error. For these reasons, it is safer to investigate the “sufficiency” of models rather than their “realism,” i.e., to inquire whether the constructed model is, for the purposes designed, adequately sufficient. As noted in a recent “Entropic Dynamics” review[CITATION Cat15 \l 1033]⁵,

“The laws of physics provide a framework for processing information about nature. From this perspective physical models are mere tools that are partly discovered and partly designed with our own very human purposes in mind. This

4 For example, the conditions under which the firm profit-maximization formulation of the maximum principle is unique, viz., the conditions of long-run static equilibrium and perfect knowledge, are never likely to be met.

5 The authors thank George Judge for recommending this article to them.

approach is decidedly pragmatic: when tools happen to be successful we do not say that they are *true*; we say that they are *useful*.”

The requisite conditions for any particular theoretical formulation are never likely to be met and thus they should be no more than a subsidiary concern. The important issue is how extensively such conditions can be violated without seriously impugning the results obtained. Few would attempt to argue that Marshall’s static equilibrium formulation, or for that matter Jorgenson’s investment theory,⁶ is internally consistent; however, some would contend that such theories are reasonably accurate and useful depending, of course, upon the purposes of the analysis.

The impossibility of falsifying any empirical theory on the basis of observation is far more reasonable if couched specifically in terms of paradigms and not testable hypotheses (whether of the nested or non-nested type). Various paradigms of microeconomic theory (e.g., perfect competition, imperfect competition, or the original behavioral formulation of Cyert and March (1963)) or of macroeconomic theory (e.g., neoclassical, neoKeynesian) are not directly confronted with empirical evidence. Although some would argue otherwise, a disregard for empirical evidence is generally appropriate for conflicting paradigms; in a conflict of paradigms the disagreement does not center on specific hypotheses but on the types of hypotheses and the type of data that might be presented for empirical investigation. Furthermore, it is difficult if

⁶ It is interesting to note that Nerlove (1972, p.223) has characterized Jorgenson’s formulation “... as a perfect paradigm of the ad hoc superposition of a lag structure on what is an essentially static theory and the resulting difficulties in the interpretation of empirical lag structures.”

not impossible to isolate criteria for judging among paradigms that are not products of the paradigms themselves.⁷ Throughout much of economics there is a dominance of market paradigms which have conditioned most economists to view the firm as a basic element rather than as an economic system worthy of investigation.⁸

To be sure, since economics must typically operate in an “open system,” where noncontrollable factors can always interfere with the operation of economic forces, economists can never be certain that the requisite conditions necessary to examine their hypotheses are actually obtained. Moreover, as economic systems may not be explicable as a simple, or even a weighted, sum of separate effects, it is the contrast, not the comparison, between economics and the experimental sciences which is illuminated. Kuhn’s analysis, which is directed chiefly to the physical sciences, requires some modification when applied to the special case of economics and other social sciences. For these sciences it is not surprising that a paradigm, once established, is especially difficult to overthrow. As its continued usefulness depends not only on the existence of unsolved problems but also on the prospects of their eventual solution by its application, there is no unequivocal standard by which it, the paradigm, can be refuted. As emphasized by Kuhn, the conflict among paradigms depends not only on their relative success, however measured, in explaining certain phenomena, but also on judgements about which phenomena ought to be explained.

7 The apparently objective notion of rationality is in fact very heavily paradigm-dependent.

8 As Kuhn (1970) suggests, if criteria employed in an existing paradigm are used in judging between it and a new paradigm, the latter will typically perform rather badly.

From the standpoint of the model builder, the relationship among a paradigm (P), maintained hypotheses (Ω), and testable hypotheses (ω) might be simply characterized as $\omega \in \Omega \subseteq P$. That is, the paradigm helps to define a set of possible maintained hypotheses from which those most consistent with economic logic will be selected.⁹ Not all hypotheses are, in fact, testable, given the limitations of real-world observations data; from the selected maintained hypotheses, it is therefore necessary to isolate a still smaller set which represents the testable hypotheses.¹⁰ In this setting the model builder in a sense becomes a prisoner not only of the maintained hypotheses but also of the paradigm within which he operates. Nevertheless, to deductively glean some information from a particular body of data, a paradigm is needed as are maintained hypotheses involving in most cases arbitrary judgements with respect to forms of functional relationships, nature of lag distributions, and the like. An important part of the art of empirical econometrics is knowing how much to include in the maintained hypothesis or model specification; the use of intuition, logic, and the “pragmatic test of validity” are often necessary to restrict the range of possible specifications.¹¹ Some elements of the maintained hypothesis may, of course, be inconsistent with the parent population generating the observed sample observations. To ascertain potential inconsistencies, we have no recourse but to relax and test certain presumed crucial maintained hypotheses. This is perhaps the only objective grounds for preferring one set of maintained hypotheses over some other set, assuming both are equally

⁹ Hayashi (2000): “Sometimes the maintained hypothesis is somewhat loosely referred to as ‘the model’” (p. 33).

¹⁰ Christ and Rausser (1973): “the selected maintained hypotheses ... isolate a still smaller set which represents the testable hypotheses” (p. 275).

¹¹ Clearly, if too much is assumed, there may be little or nothing left to test, while if too little is assumed, it may not be possible to reach any conclusion or else the analysis may become hopeless complex.

reasonable from a theoretical, intuitive, and/or technical standpoint.¹² All of this typically takes place within the context of a particular paradigm. Except perhaps for the case in which an alternative or new paradigm is, as Kuhn (1970, 13, p. 95) notes, “simply a higher level theory than those known before, one that linked together a whole group of lower level theories without substantially changing any,” the possibility of relaxing maintained elements of one paradigm and moving into another is remote. Braithwaite (1953) has provided an exposition of hierarchy among hypotheses. In a probabilistic sense, there is of course a type I error associated with rejecting a true hypotheses; nevertheless, for many testable hypotheses which come to mind, the associated type I errors are so small that they can be rejected with little loss, regardless of model purpose.

Antitrust class proceedings illustrate a case in which an accepted paradigm gives way to maintained hypotheses of liability and common impact under which economic experts can test for the presence of classwide damages. Economic experts engaged in a class-certification analysis generally apply the accepted paradigm of neoclassical economics. Under this paradigm, most agents are rational and act in their self-interest in order to maximize their utility.¹³ In this setting, an economic expert may either test or assume the liability hypothesis that the

12 The chief difficulty is that there are in principle an infinite number of hypotheses capable of explaining a given finite body of data.

13 Neoclassical economics assumes that man, *homo economicus* if you will, has consistent preferences (if he prefers apples to oranges and oranges to nuts, then he prefers apples to nuts) and makes choices that maximize his utility at all times. An alternative paradigm is that of “behavioral economics.” Under this paradigm, agents “have only limited or ‘bounded’ rationality and therefore sometimes make choices that satisfy their preferences but do not maximize their utility. This idea flows from the observation that humans do not possess the cognitive capacity required to process all the information necessary to maximize utility at all times; instead, they use heuristics or ‘shortcuts’ to make decisions that sometimes fail to jibe with the predictions of neoclassical economics.” However, although its influence in legal scholarship is growing, “behavioral economics has not yet affected judicial decisions in the United States in any substantive area of law,” Ginsburg and Moore (2010), at 89.

defendants acted as alleged and in a fashion that would be expected to cause injury to the putative class. This then establishes a maintained hypothesis (a hypothesis that is plausibly determined or assumed to be true) of liability, which can serve as the foundation for the expert to assess the hypothesis that such behavior had a common impact on the putative class. If the results of the inquiry support the common impact hypothesis, then the expert can approach the damage analysis and quantify classwide damages armed with the demonstrated maintained hypothesis of common impact. This type of analytical progression is generally widely accepted in economics and most other fields of scientific inquiry.

Returning to the philosophical foundations, Popper's views on inductive logic (infinite regress or a priorism), deductive logic, testability criteria (simplicity), demarcation criteria (between science and metaphysical speculation), and falsification (his major emphasis), though influential, have not won universal acceptance. In a probabilistic context the "problem of induction" loses much of its relevance. I. J. Good has gone so far as to argue that Popper's view destroys not only scientific induction but science itself. The formal basis for this statement and some of the major arguments against Popper's philosophy of induction may be found in Good (1968).

In Jeffreys' (1971, 1973) view deductive logic is simply encompassed within inductive logic. Hence, Jeffreys' rules of inductive inference contain many elements of Popper's deductive approach, with the latter offering special limiting cases of statement forms obtained by the use of inductive logic. The implications of these rules for Bayesian analysis have been illustrated in many places. However, what is not generally recognized is that there are several formal

analogies between Popper's degree of corroboration and Bayes' theorem on probability (Good, 1968). There are costs as well as benefits associated with the Jeffreys' approach. The costs are concerned with *inter alia*, the difficulty of characterizing appropriate initial or prior probability distributions, the vague process of generalization, metaphysical problems, and the potential effects on the accumulation of empirical knowledge. Nevertheless, within the confines of a particular paradigm, the process of revising probabilities associated with propositions as new information evolves is the essence of learning from experience. This process is formalized and made operational by Bayes' theorem.

Ultimately what matters is the purpose for which a model has been constructed. Of course, there is a nonmutually exclusive nature of models constructed for explanation, prediction, and decision purposes. Models which are to be utilized for decision purposes obviously require, either explicitly or implicitly, a criterion function for the policy makers (which represents an ordering for the states of the economy under examination) while models constructed strictly for prediction purposes do not require such information. Worland (1972) has argued that it is possible to refute economic theory or paradigms only in a decision context, i.e., a theoretical anomaly in the Kuhn sense is not possible in a science such as economics; only a policy anomaly leading to the rejection of one paradigm and the replacement by another is possible.

In addition to the obvious temporal asymmetry between explanation and prediction (Rescher, 1968), complex systems models often require several distinct theories in order to predict the behavior of related empirical phenomena. In the latter instance, as noted in Bear

and Orr (1967), the distinction between the objectives of explanation and prediction can be most meaningful. Popper objects to treating theory as nothing but an instrument for generating prediction statements. He regards the explanation of observable phenomena as crucial in the development of higher-level hypotheses.

In the final analysis, Kuhn does not emphasize falsification as a crucial element in scientific progress and finds no support for Popper's idealization of science—falsifying instances seldom lead to the revocation of theory. A paradigm is adhered to until a superior alternative appears to supplant it. In a conflict among paradigms, however, the “validity of assumptions” is a legitimate criterion, and thus Kuhn's view is not consistent with Friedman's position.¹⁴ Popper places his entire emphasis on falsification (at least in principle) and fails to recognize the costs involved in renouncing a predictively useful theory for which there is no adequate replacement.

When applying scientific principles to economic analysis, it is important to keep in mind “the nature of economic data.” As emphasized by econometrician Jeffrey Wooldridge (2002, p. 3), “rarely can we run a controlled experiment” to uncover a causal relationship between one economic variable and another. Instead, economic data that are generated and recorded as part of real-world interactions, rather than laboratory experiments, consist of two basic components: a systematic signal component that represents a causal relationship between measureable economic variables (for instance, the effect of increased income levels on price, through increased demand) and a random noise component.

¹⁴ “Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots as if he knew the complicated mathematical formulas that would give the optimum directions of travel” Friedman, Milton, “The Methodology of Positive Economics” in *Essays in Positive Economics* page 21.

The random noise component can be due to measurement error or it can be due to idiosyncratic variations that are not readily explained by any economic model, yet are not relevant to the purpose of the model, be it explanation, prediction or decision making. It is accepted in the profession that “[n]ot all possible variables that might influence the dependent variable [say, price] can be included ... some cannot be measured, and others may make little difference” (Rubinfeld, 2011, p. 314). As a result, “[n]o model could hope to encompass the myriad essentially random aspects of economic life” (Greene, 2012, p. 6). The fact that there is random variation present in the detailed non-experimental data available means that economists should exercise caution in the identification of causal relationships (when their inferences would be potentially affected by random variation) and the development of hypotheses based on such relationships.

4. The Road to Induction and Data Mining

The blueprint that was laid out by the Cowles Commission and adopted, almost without question, across the entirety of the economics profession since the 1940's, commencing with theory, then proceeding to estimation and testing. The “blueprint” Cowles provided was fundamentally different from that carried-out by Arthur Burns and Wesley Mitchell (1946) at the National Bureau of Economic Research (NBER). These researchers began with aggregate data and, without the (explicit) foreplay of economic theory, looked for patterns or regularities in the data. Koopmans (1947, pp. 161-72) first characterized such work as Kepler-style research. He advocated the work of the Cowles Commission, which he characterized as Newton-style research, as the latter began with a prior theory and then proceeded to estimation and testing.

This “Newton-style” research was consistent with the logic of science as suggested by Popper (1959).

A critical member of (and Koopmans’ colleague at) the Cowles Commission who advanced the basic probability approach to econometrics was Trygve Haavelmo (1944). He characterized economic data as experimental or passive observations. This single characteristic (how we obtain our data) provided Haavelmo quite different paths for our approach to economic science.¹⁵ While much of Haavelmo’s 1944 treatise is certainly supportive of the Koopmans’ chain: theory → data → estimation and testing, his treatment of passive observations places theory last in a causal chain, following the data, in particular:

... in the second case (passive observations) we can only try to adjust our theories to the reality that appears before us. And what is the meaning of a design of experiments in this case? It is this: We try to choose a theory and a design of experiments to go with it, in such a way the resulting data would be those which we get by passive observations of reality. And to the extent that we succeed in doing so we become master of reality – by passive agreement (Haavelmo, 1944, pp. 14-15).

Generally, through the decades of the 1940s, 1950s, and 1960s, econometricians did not follow Haavelmo’s “passive agreement with the observations” recommendation; indeed neither did Haavelmo (see Girshick and Haavelmo, 1947).

The 1970’s showed us the consequences of ignoring Haavelmo’s (1944) advice. Work by Nelson (1972) and Cooper and Nelson (1975) shows clearly that predictions of important macro

15 This clear distinction in how are data are collected (experiment versus passive observations) in Haavelmo’s work appears to take him (us) back to the work of his major professor at Oslo, Ragnar Frisch (1932): “economic theory is about to enter into a phase ... in which theory draws its fundamental conceptions from actual observation technique,” (p. 99).

variables in the economy from simple regularities (ARIMA representations) of non-structural patterns performed well in comparison to predictions from the (more structural) theoretically-based models, following usual standard Cowles Commission practices. Moreover, Rausser and Just (1980) and Just and Rausser (1981) showed that predictions of commodity prices from forward futures market prices outperformed predictions from structural econometric models.

In agriculture, similarly uncomfortable results were found. Rausser writes of the 1970's:

To the U.S. government officials who were struggling to control inflation... the tremendous increase in food prices was indeed a bitter disappointment. At this juncture, it became crystal clear that the constructed models of the USDA were no longer viable. The forecasts generated by these models appear to be outliers in comparison to the actual behavior of the system. (Rausser, 1982, p. 2).

The work of Nelson (1972) and Cooper and Nelson (1975) compared large-scale macro models against small univariate autoregressive integrated moving average (ARIMA) models as suggested by Box and Jenkins (1976). While such forecasting comparisons offer economists pause, their univariate forms left little room in our efforts to inform policy choices. It wasn't until the late 1970's that a clear multivariate representation was offered in terms of the vector autoregression (VAR) on observational data. At that time (approximately 1980) econometricians began to separate themselves along the lines hinted at in Haavelmo's 1944 offering. The field of experimental economics followed the path laid out earlier by Smith (1962). Here the more traditional approach to hypothesis testing (described above), beginning with theory then to estimation and testing continued to hold sway. Work with observational data, which focused on regularities, eschewed *a priori* claims of exogeneity, letting the data have a role in ultimate specification. Sims (1980) marked the beginning of modern attempts to meet

Haavelmo's suggestion for modeling passive observations (observational data). Here all variables in a small vector of time sequenced data are estimated as a vector autoregression. Unique is Sims' reluctance to apply a priori identifying restrictions, not wishing to exclude some variables in the representation of others. All variables were endogenous. Further, the endogenous paths of influence are not restricted a priori.¹⁶ The VAR model, of course, results in an abundance of parameters (Sims calls them profligately parameterized) which limited the numbers of variables that could be effectively considered in early VAR studies. There was effort among practitioners to submit this new representation to out-of-sample forecasting tests, both by Sims and his students and many others (Doan, Litterman and Sims, 1983; Kling and Bessler, 1985, 1986; and Allen, 1994).

This early VAR piece by Sims might be regarded as a second major step in data mining in economics; keeping Burns and Mitchell firmly in their place as the first. The difference, of course, is that Sims approached the question of economic dynamics with much statistical theory and the benefit of the thirty year prelude that was the Cowles Commission.

4.1 Data Mining and Machine Learning

Data Mining and Machine Learning are modern terms that summarize the Kepler stage of modeling, as described by Koopmans (1947). In contrast to the early NBER work of Burns and Mitchell or the early vector autoregression work of Sims and others, modern Data Mining and Machine Learning look for pattern recognition over what may be large data sets. Tuffery's (2011) definition makes this abundantly clear

Data mining is the set of methods and techniques for exploring and analyzing data sets (which oftentimes are large), in an automatic or semiautomatic way, in order to find among these data certain unknown or hidden rules, associations, or

¹⁶ Much similar to the suggestion of Liu (1960) some twenty years prior.

tendencies; special systems output the essential of the useful information while reducing the quantity of the data. (p. 4)

Maimon and Rokach (2005,) concur with their more parsimonious definition, “Data mining is a term coined to describe the process of shifting through large databases for interesting patterns and relationships,” (p. 2).

The goal of the mining task is to extract useful knowledge from these large data bases, which can show its mettle in a particular decision problem. Often the output results in a set of forecasts in near real time. Again, Tuffery writes, “... the main data mining and data analysis methods can be divided into two large families: descriptive methods and predictive methods,” (p. 167). Descriptive methods summarize the (large) data set as groups (with no pre-assigned endogenous variable). Predictive methods look to predict endogenous variables, based on these groupings. Maimon and Rokach (2005) discuss similar functions in terms of unsupervised learning (grouping techniques such as clustering and factor modeling) and supervised learning, modeling the relationship between input (independent) variables and output (dependent) variables. What evidence do we use to decide the “dependent” / “independent” characteristic? While some authors do not utilize Machine Learning to inquire about causality among the set of grouped variables, others do. Modern versions of data mining look for patterns in the data that are predictive (hold up beyond the sample data for which they were first uncovered).

As suggested previously, Sims (1980) recognized the wasteful nature of the VAR representation. Recent work has addressed this problem through dimension reduction (data grouping) method: principle components, cluster, and/or factor analysis. The factor augmented

VAR models of Stock and Watson (2002) and Bernanke et al (2005) study over one hundred variables, the dimension of which is reduced to less than ten factors, augmented with a particular set of policy instruments (federal funds rate, for example). This allows the researcher to study a large number of series, a feat not possible in the original presentation by Sims (1980).

17

To do unrestricted forecasting within Tuffery's "predictive family" of methods of data mining, there is no need to identify *the underlying causal structure*, as the chain rule of forecasting provides minimum mean squared error forecasts of the (presumed) variable of interest. This is precisely the path followed by Stock and Watson (2002). If one is interested in considering the conditional path on $Y_{1,t+k}$ (label this $Y^f_{1,t+k}$) given the firm (state or country) places restriction on $Y_{2,t+k}$, $k=1,2,3, \dots$, (label this set of restrictions on $Y_{2,t+k}$ as $Y^r_{2,t+k}$) she must identify the structure on contemporaneous innovations (or contemporaneous variables in a VAR). It is less than clear how one ought to assign such causal structure. Of course, early Cowles Commission researchers did this via *a priori* theory. What the early VAR-workers needed was an explicit treatment of "cause". In the following subsection, we discuss "cause" in terms of its general characteristics and specific (machine learning) algorithms which attempt to capture it in the data.

4.2 Causality and Common Impact

17 An additional form of dimension reduction that is commonly followed by econometricians, but not data miners, is reduced rank regression; see for example, Anderson, T. W. (1951). Anderson, of course, was a highly productive member of the Cowles Commission group. In modern day parlance we refer to this area of study in times series econometrics as study of cointegration, see Johansen (1988).

Attempting to capture causal relationships is at the heart of economics. It is universally recognized by economists and statisticians that correlation does not imply causality; yet in certain economic circles, this critical distinction is often overlooked. The most obvious example of this shortcoming in antitrust class proceedings is the extensive use of univariate analysis (where pricing is investigated along only one dimension), rather than multivariate analysis (which attempts to control for each of the observable characteristics that would be expected to materially influence price). Instead of investigating only how prices change as one price determining variable changes (univariate analysis), multivariate analysis seeks to explain how price changes as a multitude of determining variables change. So, instead of looking solely at price points graphed over time, a multivariate analysis can investigate how various economic and transaction characteristics (product strength, distance shipped, fuel costs, and so on) cause prices to change — both over time and across different transactions. A frequent method to implement multivariate analysis is to use a multiple regression model, which can “back out” the influence of a multitude of economic and transaction characteristics on price (Greene, 2012). Such multivariate approaches can be used to demonstrate that factors that cause prices to vary can be controlled for, as they might otherwise confound isolation of any wrongful price manipulation. The “backing out,” however, brings forth a problem with observational data, as there may be relationships among the “right hand side variables” or other missing variables that “cause” both the right hand side variables and the dependent variable. In other words, economic variables are correlated because they are simultaneously determined by some other factor. As a result, we should “use econometric methods to effectively hold other factors fixed” (Wooldridge, 2002, p. 3). As Wooldridge emphasizes: “[t]he notion of *ceteris paribus* — i.e.,

holding all other (relevant) factors fixed— is at the crux of establishing a *causal relationship*. Simply finding that two variables are correlated is rarely enough to conclude that a change in one variable causes a change in another” (*ibid.*). The *ceteris paribus* clause is maintained by design in experimental studies (by random assignment of values to the right hand side variables).

However, even the utility of multivariate regression analysis can be limited because in some instances not all of the factors that influence the variable of interest can be observed in the available data. The approach adopted by econometricians to “identify” a causal relationship in the presence of unexplained variation in the variable is referred to in the profession as an “identification strategy.” Ideally, an identification strategy involves a “natural experiment”¹⁸ in which a “treated” group is compared to an otherwise identical or near identical “untreated” group, in order to identify the effect of the treatment.

An alternative to *a priori* identification or to assuming “treated and non-treated groups” as otherwise identical (the notion of identical is of course problematic with observational data in many cases) is given in modern work from computer science. Here one begins with a set of theoretically (*a priori* plausibly) related variables and allows patterns in the data to inform the underlying structure, allowing the investigator to be in compliance with Haavelmo’s admonition to summarize the patterns in the data. Algorithms of inductive causation (Bayes’ Nets¹⁹) are

18 The term “natural experiment” is used because, of course, econometricians cannot conduct these experiments in the laboratory; rather we exploit instances where they have occurred naturally in the real world. See, Coleman and Langenfeld (2008)

19 Bayes’ Nets (or Bayesian Networks) are directed acyclic graphs that maintain the probability updating rule of Thomas Bayes (see Pearl, 2009, p. 14). In particular, these networks are acyclic (recursive) flows of information among a set of causally sufficient variables.

recipes or procedures (developed in the machine learning literature) which can be followed to yield a causal graph over large numbers of variables, if such a causal structure exists in the data. Work in this area is ongoing at several places around the world, for example: The Computer Science Laboratory at UCLA under the direction of Judea Pearl (2000, 2009) and the Department of Philosophy at Carnegie Mellon University under the guidance of Clark Glymour (Spirtes, 2000). Pearl's work described in his book *Causality*, results in what he labels as the IC (Inductive Causation) algorithm; Spirtes et al offer PC algorithm.

Both IC and PC algorithms give outputs as "Bayes Nets" connecting graphical representations to the underlying probability representation on the variables studied. One begins with a problem for which she has a universe of measures (observations on) variables that prior theory, problem definition, or common sense²⁰ suggest as related. This limiting or circumscription (McCarthy, 1980) of variables is defined as the set over which the net is to be constructed. Both algorithms rest on a set of three assumptions governing the underlying process studied: Causal Sufficiency, Causal Markov Condition and Faithfulness. Causal Sufficiency assumes there is no omitted variable, say D, which causes two or more of the variables studied (A,B,C). It is relaxed in more advanced learning algorithms (FCI for example)²¹. The Causal Markov Condition assumes that the underlying joint probability distribution on the variables under study can be factored as the product of the conditional distributions on each variable. So, for example, if there is a causal

20 There appear to be no clear guidelines indicating what bounds are to be placed on the circumscribed set. Indeed one might adopt Feyerabend's (1975) admonition to accept even folklore or other "non-scientific" priors on what is admissible to the circumscribed set (not that we are suggesting such here). Below we discuss "causal sufficiency" which does put a (possibly unknowable) restriction on the set studied.

21 FCI: Fast Causal Inference Algorithm (Spirtes, Glymour and Scheines, 2000) does not require causal sufficiency. See Malinvaud (1980, p. 50) who recognizes that the researcher does not know all the variables that interact with the measured variables in her circumscribed set (that is, she may look for nets over sets which include latent variables).

fork (to be read variable B causes both A and C) $A \square B \square C$, the corresponding joint distribution on variables A, B, and C is given as $P(A,B,C) = P(B)P(A|B)P(C|B)$. And finally PC and IC algorithms rest on an assumption of Faithfulness: if there is zero correlation between two variables that value (0) arises because there is no relation between the two variables and not because of cancellations of deeper “structural” parameters.²²

Primary in the operation of these algorithms on Gaussian data is that the correlation structure on causal forks ($A \square B \square C$) and causal chains ($A \square B \square C$ or $A \square B, \square C$) is the same. The three structure are said to be observationally equivalent, so one is not able (short of having either good time ordering knowledge or other a prior information) to assign causal direction by looking at the data. In essence, (using rho to represent correlation), $\rho(A,B) \neq 0$, $\rho(B,C) \neq 0$, $\rho(A,C) \neq 0$ and $\rho(A,C|B) = 0$, in all three representations, given above. The correlation structure on causal inverted forks, however, is different ($A \square B \square C$). On inverted forks one has $\rho(A,B) \neq 0$, $\rho(B,C) \neq 0$ and $\rho(A,C|B) \neq 0$. Here variable B is known as a collider. Conditioning on a collider opens up communication between previously independent variables. PC (Spirtes et al, 2000) and IC (Pearl, 2000 and 2009) algorithms are machine learning programs that sort through data sets, investigating causal colliders amongst variables to begin to build up a causal pattern. In addition, latent variable versions (FCI: Fast Causal Inference Algorithm) of this basic “search for colliders” set-up are provided by both schools (UCLA and CMU). Early work by Bessler and Akleman (1998) took Swanson and Granger’s (1997) suggestion to study the innovations from a VAR using these algorithms. The result is a “structural VAR” where the analyst relies on the data

²² Hoover (2005) suggests that rational expectations may call into question “faithfulness.” This suggests a plausible motivation for experimental knowledge from Haavelmo’s first class of experiments.

to not only determine lag structures but contemporaneous relations, as well. As a result, this places us in closer agreement with Haavelmo's 1944 recommendation.

A simple example, which may well cohere with many people's prior beliefs, is given in Roh and Bessler's study of 1947 – 1974 yearly data from Peltzman's (1975) study of Traffic Fatalities in the US. Here the authors begin with 1947 – 1974 data on Total US Traffic Fatalities, Number of Young Drivers, Cost of Safety Devices, Speed Limit, Alcohol Consumption, Mileage Driven in the US, and US Disposable Income. We focus here on the three variable sub-model on Traffic Fatalities, Alcohol Consumption and Speed. Both PC and IC algorithms begin with a complete undirected graph (every variable is connected to every other variable with no arrow heads):

Each edge in this complete graph is challenged by a test of unconditional correlation equal to zero (and conditional correlations in the general case). If the correlation is deemed not to be significantly different from zero, the edge, the connecting line, is removed. In our case (using Fisher's Z) the edges between Traffic Fatalities and Alcohol Consumption and Speed and Alcohol Consumption are not removed, as their corresponding correlations are significantly different from zero (at very low p-values); whereas the edge between Speed and Alcohol

Consumption is not deemed significantly different from zero at modest significance levels (correlation is -.05, p-value of .84). Since this last correlation between Alcohol Consumption and Speed is (deemed to be) zero, we know that the causal model connecting these three variables is not a chain: Alcohol Consumption $\rightarrow$ Traffic fatalities $\rightarrow$ Speed (or the same three with arrows reversed). Further, the model is not a causal fork (Alcohol Consumption $\rightarrow$ Traffic Fatalities $\rightarrow$ Speed). In Gaussian data chains and forks are observationally equivalent. In both cases here (Forks and Chains), information on Alcohol Consumption would pass through to Speed, so we would see a non-zero correlation between Alcohol Consumption and Speed (which we do not). The model holding these three variables together appears to be an inverted fork (Alcohol Consumption $\rightarrow$ Traffic Fatalities $\rightarrow$ Speed).

Other Machine Learning algorithms for capturing causal structure building off of model scoring on observational equivalent structures, using greedy search (GES algorithm) and search on non-Gaussian data (LiNGAM), offer analysts more opportunities to explore possible causal structures with observational data. PC, IC and scoring (GES-type search) algorithms often generate the same causal structure. But agreement is not always the case. When disagreement is found on the causal structure emanating from distinct machine learning algorithms, study or (short of a random assignment experiment) presentation of results under all structures obtained (if this is a small number) is a natural next step.

4.3 Forecasting

As indicated above forecasting is a component of Data Mining, as it was in much of the time series econometrics work of the 1980's. Both time series econometrics and data mining

are subject to the criticism that the data are intimately involved in model specification. Theory is not used, as in the earlier Cowles Commission work, to suggest, a priori, an explicit model. As with the Machine Learning algorithms and with Data Mining in general, theory is used rather casually to suggest a causally sufficient set of variables to investigate. Given a tendency of models to over-fit the data, a recommended step before accepting the model as 'sufficient' or 'useful' is to explore model performance with a fresh data set, what some have called out-of-sample (or post-fit-sample) data.

Here several questions need to be answered: 1. the type of forecast (point or probability forecast?) and 2. the horizon (one-step or multiple-steps ahead?). Point forecasts have been the most widely studied. Metrics of performance, while generally matched with the user's utility function (and the underlying model parameter estimation method applied (Zellner, 1986)), most often apply quadratic loss. Root Mean Squared is the most popular; its corresponding test of significant differences is the test recommended by Diebold and Mariano (1995). Other forms of testing significance difference and/or information blockage are forecast encompassing (Harvey, Leybourne and Newbold, 1998) and d-separation (Bessler and Wang, 2012)²³. Forecast horizon studied is usually one-step ahead. Of course the problem under study may suggest other horizons, especially in agriculture where seasonality is not a trivial issue.

23 Forecast encompassing refers to the idea that under some circumstances a model's (person's) forecasts contain no useful information, relative to forecasts from another model (person). Forecasts from the latter are said to encompass forecasts from the former. This idea is an improvement over tests showing that the preferred model has significantly lower mean squared errors (MSE) (Diebold and Mariano 1995), as even if one model has significantly lower MSE, the model with the higher MSE may yet contribute useful information (if its forecasts are not encompassed) in a composite (combination forecast). Bessler and Wang (2012) apply these ideas to directional separation and screening off of forecasts in graphical representations of US Federal Reserve (Greenbook) and Survey of Professional Forecasters forecasts of US inflation and GDP and US agricultural prices.

The researcher also needs to decide on point or probability forecasts. As suggested above most forecast evaluations are point forecasts (usually mean squared error or mean squared percentage error). In some settings such a forecast may be not sufficient. In financial modeling, we sometimes want a description of the left tail of the forecasted distribution (value at risk) or in an economic development context, the density below some subsistence level. Here work of Dawid (1984), the earlier work of psychologists (Lichtenstein, Fischhoff and Phillips, 1982) and statisticians (Savage, 1971) is helpful. Calibration metrics (based on the probability integral transform) and scoring rules (based on matched utility of users) can focus attention on the issued probabilities and test or score the competitors relative to what actually occurs.

Another relevant issue is what alternative might be employed. The answer depends on the *milieu* of the problem under study and purpose of the constructed model. The choice of an alternative to compare forecasts from the data mining model will depend on what might be labeled the “next best alternative”. Whether point forecasts or probability forecasts, one has to be sufficiently familiar with the environment in which the model is to be used to provide an alternative. In studying probability forecasts of the Bank of England, Casillas-Olvera and Bessler (2006) had near contemporaneous set of forecasts from a set of professional forecasters. If a contemporaneous set of forecasts are unavailable, analysts are oftentimes left to use univariate autoregressive representation as the alternative.

Any model selected is only tentatively accepted if it outperforms alternatives in out-of-sample forecasting exercises (in probability forecasts it is well-calibrated and performs well in

scoring (Good, 1952 and Winkler and Murphy 1967)).²⁴ Here no claim is made to “truth” or “correct specification”; rather the most that can be claimed is that the current best performing model awaits the challenge of the next model specification. Thus the machine learning and data mining extensions of the time series revolution are best viewed as instrumentalist attempts to understand the world with observational data. The “Gold Standard” with respect to such understanding is (ideally) the random assignment experiment; having no such opportunities in much of our work, we take a more humble stance on what can ultimately be known.

5. Conclusion

Not surprisingly, dollar amounts matter in economics. This fact makes the distinction between *economically* significant effects and *statistically* significant effects meaningful. Since econometric analysis is conducted by selecting a sample of data from a population, and statistical precision increases with sample size, the availability of very large proprietary transaction-level datasets means that it is easy to find a *statistically* significant effect. The large amount of data will allow any small price difference between groups of proposed class members to be detected.²⁵ As a result, after finding a statistically significant effect, the economist should ascertain whether or not that effect is also *economically significant*. To be significant, the effect is usually sizeable in dollar terms. Economically small effects are of no

24 Good 1952 proposes motivating an individual's probability forecast according to the log rule: $S = \log(1-p')$, where S is the individual's score, and p' is the stated (revealed) probability on the event that actually occurs, S represents the amount of “information lost through not knowing with certainty what will happen” (Good 1983, page 11). That is, he/she is motivated to set his/her revealed probability p' (probability revealed) equal to his/her ‘true’ subjective probability, p . This same motivational rule has been proposed (along with other “proper scoring rules”) as an evaluation metric to assess after the fact “goodness” in probability assessment (Winkler and Murphy 1968).

25 With a very large amount of data, some academics even question the usual threshold for assessing statistical significance. Judge et al, 1985).

importance and should be ignored by economists, even if they are statistically significant (McCloskey and Ziliak, 2008).

With respect to model specification, selection and estimation: (1) Particularly if a model is constructed for decision purposes, the method of estimation should be a function of model purpose. (2) Recognizing that sample data are typically employed to assist in model specification, loss due to incorrect decisions, and thus the mean squared error or risk measures for gauging estimator performance, Stein-like estimators might best be utilized. Along these same lines, as suggested in Leamer (1974), more econometric research effort might “best” be devoted to developing statistical procedures for drawing inferences based on false models.

More specifically with respect to model evaluation: (1) Should various nonparametric measures be categorized at least by the general purposes of explanation, prediction, and decision? Recognizing that evaluation is properly treated as a purpose-dependent process, such a categorization would appear most useful. (2) Assuming a constructed model is to be utilized for decision purposes and since it is very unlikely that a unique or single valued criterion function can be obtained, what should be the role of policy maker(s) or model user(s) in the evaluation process? In this regard it seems that it is only possible to capture an objective function set (or welfare correspondence) which, of course, complicates but nevertheless provides a formal initial basis for the policy analysis.

Acknowledgements

We would like to thank George Judge, Sofia Berto Villas-Boas and David Zilberman for their valuable comments on a previous draft of this paper.

Literature Cited

- Allen, P. Geoffrey. (1994). Economic forecasting in Agriculture. *International Journal of Forecasting* 10(1), 81-135.
- Anderson, T. W. (1951), "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Mathematical Statistics*, 22, 327- 351.
- Bear, Donald VT, and Daniel Orr. (1967). Logic and expediency in economic theorizing. *The Journal of Political Economy*. 188-196.
- Bernanke, B.S., J. Boivin, P. Elias. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120, 387-422.
- Bessler, D.A. and Akleman, D. G. (1998.) Farm prices, retail prices, and directed graphs: Results for pork and beef. *American Journal of Agricultural Economics*, 80(5), 1145-1150.
- Bessler, D.A. and Z. Wang. (2012). D-separation, forecasting, and economic science: A conjecture. *Theory and Decision*, 295-314.
- Box, G.E.P. and G. Jenkins. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day. (Earlier versions and sub-sections of this book came out as early as the mid-1960's as technical reports as cited in their 1976 edition).
- Braithwaite, Richard Bevan. (1953). *Scientific Explanation*. Cambridge: Cambridge University Press.
- Burns, A. F. and W. C. Mitchell. (1946). Plan of treating secular, seasonal and random movements. In A.F Burns and W.C. Mitchell (Eds.), *Measuring Business Cycles* (pp. 23-36). New York: National Bureau of Economic Research, Studies in Business Cycles.

- Casillas-Olvera, G. and D. A. Bessler. (2006). Probability forecasting and central bank accountability. *Journal of Policy Modeling*, 28, 223-234.
- Caticha, A. (2015). Entropic dynamics. *Entropy* 17(9), 6110-6128.
- Christ, Carl F. and Gordon C. Rausser. (1973). The validity and verification of complex system models: Discussion on Shapiro and Fromm. *American Journal of Agricultural Economics*, 55(2), 271-79.
- Coleman, M., and J. Langenfeld. (2008). Natural experiments, issues in competition law and policy. *ABA Section of Antitrust Law*, 1, 743-772.
- Cooper, J. P. and C. R. Nelson. (1975). The ex ante prediction performance of the st. louis and frb-mit-penn econometric models and some results on composite predictors. *Jo Money Credit and Banking* 44, pages 1-32.
- Cyert, Richard M., and James G. March. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs: Prentice-Hall.
- Dawid, P. (1984). Statistical theory: a prequential approach. *Journal of the Royal Statistical Society* 147:278-92.
- Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.
- Doan, T., Robert B. Litterman, and Christopher A. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1), pp 1-100

Feyerabend, Paul. (1975). *Against Method: Outline of an Anarchist Theory of Knowledge*, New Left Books.

Friedman, Milton. (1953). *Essays in Positive Economics*, Chicago The University of Chicago Press.

Frisch, R. (1932). New orientation of economic theory: Economics as an experimental science., *Nordic Statistical Journal*, 4, pp 97-111.

Ginsburg, D., and D. Moore. (2010). The future of behavioral economics in antitrust jurisprudence. *Competition Policy International*, 6(1) (Spring), 89-102.

Girschick, M.A. and T. Haavelmo. (1947). Statistical analysis of the demand for food: Examples of simultaneous equations estimation. *Econometrica*, 15, pp. 79-110.

Good, I.J. (1952). Rational Decisions. *Journal of the Royal Statistical Society*, series B, 41:107-114.

Good, I. J. (1960). Review of KR Popper: The logic of scientific discovery. *Mathematical Reviews*, 21, pp. 1171-1173.

Good, I. J. (1965). *The Estimation of Probabilities*. Cambridge: MIT Press.

Good, I. J. (1968). Corroboration, explanation, evolving probability, simplicity and a sharpened razor. *British Journal for the Philosophy of Science*, pp. 123-143.

Good, I. J. (1983). *Good Thinking*. University of Minnesota Press.

Greene, W. (2012). *Econometric Analysis* (7th ed.). Prentice Hall.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, 12, pp. 1- 118.

Harvey, D., S. J. Leybourne and P. Newbold. (1998). Tests for forecast encompassing *Journal of Business and Economic Statistics*, 16, pp. 254-259.

Hayashi, F. (2000). *Econometrics*, Princeton University Press.

Hoover, K. D. (2005). Automatic inference of the contemporaneous causal order of a system of equations. *Econometric Theory*, 21, pp. 69-77.

Hume, D. (1748). *An enquiry concerning human understanding*. London.

Jeffreys, Harold. (1971). *The Theory of Probability* (3rd ed.). Oxford: Clarendon Press.

Jeffreys, Harold. (1973). *Scientific Inference*. Cambridge: Cambridge University Press

Johansen, S. (1988). Statistical analysis of cointegrating vectors, *Journal of economic Dynamics and Control*, 12:231-54.

Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lutkepohl and Tsoung-Chao Lee. (1985). *The Theory And Practice Of Econometrics*, (2nd ed.). p. 131.

Just, Richard E., and Gordon C. Rausser (1981). Commodity price forecasting with large-scale econometric models and the futures market. *American Journal of Agricultural Economics* 63(2): 197-215.

Kling, John L. and David A. Bessler. (1985). A comparison of multivariate forecasting procedures for economic time series. *International Journal of Forecasting*, 1(1), pp. 1-80.

Koopmans, T. J. (1947). Measurement without theory. *Review of Economics and Statistics*, 29 (1947).

- Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions*, (2nd ed.). Chicago: The University of Chicago Press.
- Leamer, Edward E. (1974). False models and post-data model construction. *Journal of the American Statistical Association*, 69(345), pp. 122-131.
- Lichtenstein, S., B. Fischhoff and L.D. Phillips. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic and A. Tversiy (Eds) *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Liu, T.C. (1960). Underidentification, structural estimation and forecasting. *Econometrica*, 28, pp. 855-65.
- Maimon, O. and L. Rokach. (2005). *The Data Mining and Knowledge Discovery Handbook*. New York: Springer.
- Malinvaud, E. (1980). *Statistical Methods of Econometrics*. Amsterdam: North Holland.
- McCarthy, J. (1980). Circumscription - A form of non-monotonic reasoning. *Stanford Artificial Intelligence Laboratory Memo: AIM - 334*, (February 1980).
- McCloskey, Deirdre N. and Stephen T. Ziliack (2008). Signifying nothing: Reply to Hoover and Siegler. *Journal of Economic Methodology*, 15, pp. 39-55.
- Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review*, 62, pp. 902-17.
- Nerlove, Marc. (1972). Lags in economic behavior. *Econometrica: Journal of the Econometric Society*, pp. 221-251.

- Pearl, J. (2000 and 2009). *Causality: Models, Reasoning and Inference*, (1st and 2nd eds.) New York: Cambridge University Press.
- Peltzman, S. (1975). The effect of automobile safety regulation, *Journal of Political Economy* 83:677- 725.
- Popper, K. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.
- Popper, K. (2005/1935). The logic of scientific discovery. In *Routledge Classics*. Taylor & Francis e-Library
- Rausser, C. G. (1982). *New Directions in Econometric Modeling and Forecasting in U.S. Agriculture*, New York: North Holland.
- Rausser, Gordon and R. E. Just. (1980). Futures versus commercial econometric models. In *Futures Trading Seminar*, edited by Anne Peck. Chicago: Board of Trade of the City of Chicago.
- Rescher, Nicholas. (1968). *Topics in Philosophical Logic*.
- Roh, J.W., D.A. Bessler and Roy F. Gilbert. 1999. "Traffic Fatalities, Peltzman's Model and Directed Graphs," *Accident Analysis and Prevention* 31:55-61.
- Rubinfeld, D. (2011). Reference guide on multiple regression. In *Reference Manual on Scientific Evidence*, (3rd ed.). Federal Judicial Center.
- Savage, L. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66, pp. 783-801.
- Shapiro, H. (1973). Is verification possible? The evaluation of large econometric models. *American Journal of Agricultural Economics*, 55(2) (May), pp. 250-258.
- Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48, pp. 1-48.

- Smith, V.L. (1962). An experimental study of competitive market behavior, *Journal of Political Economy* 70: 111-137.
- Solow, R. M. (1971). Discussion. *American Economic Review*, 71, (May), pp. 63-65.
- Spirtes, P., C. Glymour, and R. Scheines. (2000). *Causation Prediction and Search*. MIT Press.
- Stock, J. H., and M. W. Watson. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, pp. 1167-1179.
- Swanson, N. R., and C. W. J. Granger. (1997). Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association*, pp. 357-367.
- Tuffery, S. (2011). *Data Mining and Statistics for Decision Making*. Chichester: John Wiley and Sons Ltd.
- Winkler, R. L. and A. H. Murphy (1968). 'Good' Probability Assessors, *Journal of Applied Meteorology* 7:751-758.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Worland, Stephen T. (1972). Radical political economy as a 'scientific revolution'. *Southern Economic Journal*, pp. 274-284.
- Zellner, A. (1986). A tale of forecasting 1001 series: The Bayesian knight strikes again. *International Journal of Forecasting*, 92, pp. 491-494.