Causal Parameters and Policy Analysis In Economics: A Twentieth Century Retrospective

James J. Heckman
Department of Economics, University of Chicago
and The American Bar Foundation

June, 1999
Revised July, 1999

This research was supported by NSF 97-09-873 to the University of Chicago. I have benefited from comments received in discussions with Jennifer Boobar, Ray Fair, Ed Glaeser, Claudia Goldin, Lars Hansen, Larry Katz, Steve Levitt, Costas Meghir, Judea Pearl, José Scheinkman, Christopher Sims and Edward Vytlacil. For historical background on the Cowles Commission, I have relied on Christ (1952), Epstein (1987) and Morgan (1990).

Abstract

The major contributions of twentieth century econometrics to knowledge were the definition of causal parameters within well defined economic models in which agents are constrained by resources and markets and causes are interrelated, the analysis of what is required to recover causal parameters from data (the identification problem), and clarification of the role of causal parameters in policy evaluation and in forecasting the effects of policies never previously experienced. This paper summarizes the development of these ideas by the Cowles Commission, the response to their work by structural econometricians and VAR econometricians, and the response to structural and VAR econometrics by calibrators, advocates of natural and social experiments, and by nonparametric econometricians and statisticians.

JEL CODE C:10

1. Introduction

This paper considers the definition and identification of causal parameters in economics and their role in econometric policy analysis. It assesses different research programs designed to recover causal parameters from data.

At the beginning of this century, economic theory was mainly intuitive and empirical support for it was largely anecdotal. At the end of the century, economics has a rich array of formal models and a high-quality data base. Empirical regularities motivate theory in many areas of economics and data are routinely used to test theory. Many economic theories have been developed as measurement frameworks to suggest what data should be collected and how they should be interpreted.

Econometric theory was developed to analyze and interpret economic data. Most econometric theory adapts methods originally developed in statistics. The major exception to this rule is the econometric analysis of the identification problem and the companion analyses of structural equations, causality, and economic policy evaluation. Although an economist did not invent the phrase "correlation does not imply causation,"[1] economists clarified the meaning of causation within well-specified models, the requirements for a causal interpretation of an empirical relationship, and the reasons why a causal framework is necessary for evaluating economic policies.[2]

The fundamental work was done by economists associated with the Cowles Commission.[3] The lasting legacy of this research program includes the concepts of exogenous (external) and endogenous (internal) variables, and the notions of "policy invariant parameters" and "structural parameters" which have entered everyday parlance inside and

[1] The phrase is generally attributed to Karl Pearson.

[2] For example, the artificial intelligence community has just begun to appreciate the contributions of econometrics to the definition and identification of causal relationships. See the papers in Glymour and Cooper (1999) and the paper by Pearl (1998).

[3] The Cowles Commission was founded by Alfred Cowles to promote the synthesis of mathematics and economics. Cowles and the Cowles Commission played a leading role in creating the Econometric Society. It was originally based in Colorado Springs and had a loose organizational arrangement with Colorado College. It was relocated to the University of Chicago from 1939 to 1955. See Christ (1952, reprinted 1995), Epstein (1987) and Morgan (1990) for valuable histories of econometrics and the role of the Cowles Commission in defining modern econometrics.

outside of economics.

Just as the ancient Hebrews were "the people of the book," economists are "the people of the model." Formal economic models are logically consistent systems within which hypothetical "thought experiments" can be conducted to examine the effects of changes in parameters and constraints on outcomes. Within a model, the effects on outcomes of variation in constraints facing agents in a market setting are well defined. Comparative statics exercises formalize Marshall's notion of a *ceteris paribus* change which is what economists mean by a causal effect. In his own words,

> *"It is sometimes said that the laws of economics are 'hypothetical'. Of course, like every other science, it undertakes to study the effects which will be produced by certain causes, not absolutely, but subject to the condition that other things are equal and that the causes are able to work out their effects undisturbed. Almost every scientific doctrine, when carefully and formally stated, will be found to contain some proviso to the effect that other things are equal; the action of the causes in question is supposed to be isolated; certain effects are attributed to them, but only on the hypothesis that no cause is permitted to enter except those distinctly allowed for "(A. Marshall, 1920, p. 36).*

The "other things are equal" or *ceteris paribus* clause is a cornerstone of economic analysis.[4]

Defining causality within a model is relatively straightforward when the causes can be independently varied.[5] Defining causality when the causes are interrelated is less straightforward and is a major achievement of econometrics. Recovering causal parameters from data is not straightforward. An important contribution of econometric thought was the formalization of the notion developed in philosophy that many different theoretical models and hence many different causal interpretations may be consistent with the same data. In economics, this is called the problem of identification. The econometric analysis of the identification problem clarifies the limits of purely empirical knowledge. It makes precise the idea that correlation is not causation by using fully specified economic models as devices for measuring and interpreting causal parameters. It presents conditions under which the hypothetical variations mentioned in the quotation from Marshall, or the structural

[4] Marshall himself does not use the phrase *"ceteris paribus"* in his book.

[5] Marini and Singer (1988) present a valuable summary of the rancorous and confusing debates about the nature of causal laws developed in model-free fields.

parameters of well-specified economic models, can in principle be identified from data. Different *a priori* assumptions can identify the same causal parameter or identify different causal parameters. The key insight in the literature of twentieth century econometrics was the discovery of the conditional nature of empirical knowledge. The justification for interpreting an empirical association causally hinges on the assumptions required to identify the causal parameters from the data.

This paper proceeds in the following way. (1) The concept of a causal parameter within a well posed economic model is defined in an economic setting that respects the constraints imposed by preferences, endowments, and social interactions through markets. By a well posed economic model, I mean a model that specifies all of the input processes, observed and unobserved by the analyst, and their relationship. My definition of causal parameters formalizes the quotation from Marshall. This formalization is a more appropriate framework for economic causal analysis than other frameworks developed in statistics that do not recognize constraints, preferences, and social interactions (*i.e.* are not based on formal behavioral models). The concept of identification of a causal parameter is discussed using the market demand-supply example that motivated thinking about the identification problem through the first half of the twentieth century. This example emphasizes the consequences of interdependence among economic agents, but has some special features that are not essential for understanding the fundamental nature of the identification problem. A more general statement of the identification problem is given than appears in the published literature. The role of causal parameters in policy analysis is clarified.

(2) The paper then assesses the response in the larger economics community to the Cowles Commission research program. The Cowles group developed the linear equation simultaneous equations model (SEM) that is still presented in most econometrics textbooks. It extensively analyzed one form of the identification problem that most economists still think of as *the* identification problem. It focused attention on estimation of Keynesian macro models and on the parameters of market-level supply and demand curves. By the mid-1960s, the Cowles research program was widely perceived to be an intellectual success but an empirical failure.

This led to two radically different responses. The first was the VAR or "innovation accounting" program most often associated with the work of Sims (1972, 1980, 1986) that objected to the "incredible" nature of the identifying assumptions used in the Cowles Commission models and advocated application of more loosely specified economic models based on developments in the multivariate time series literature. This research program systematically incorporated time series methods into macroeconometrics and produced more accurate descriptions of macro data than did its Cowles predecessors. Its use of economic theory was less explicit but it drew on the dynamic economic models developed in the 70s and 80s to motivate its statistical decompositions.

At about the same time, and more explicitly motivated by the development of a macroeconomics based on dynamic general equilibrium theory under uncertainty, structural equations methods based on explicit parameterization of preferences and technology replaced the Cowles paradigm for market aggregates and for Keynesian general equilibrium systems. The notion of a structural or causal parameter survived but it was defined more precisely in terms of preference and technology parameters and new methods for recovering them were proposed. Nonlinear dynamic econometric models were developed to incorporate the insights of newly developed economic theory into frameworks for economic measurement and to incorporate rational expectations into the formulation and estimation of models. This approach emphasizes the clarity with which identifying assumptions are postulated and advocates an approach to estimation that tests and rejects well-posed models. It is ambitious to attempt to identify and estimate economically interpretable "policy invariant" structural parameters that can be used to ascertain the impacts of a variety of policies.

The empirical track record of the structural approach is, at best, mixed. Economic data, both micro and macro, have not yielded many stable structural parameters. Parameter estimates from the structural research program are widely held not to be credible. The empirical research program of estimating policy invariant structural parameters in the presence of policy shifts remains to be implemented. The perceived empirical failures of well-posed structural models have often led to calls for abandonment of the structural approach in many applied fields, and not to the development of better structural models in those fields.

Part of the continuing popularity of the VAR program is that it sticks more closely to the data and in that sense is more empirically successful than structuralist approaches. At the same time, its critics argue that it is difficult to interpret the estimates obtained from application of this program within the context of well-specified economic models and that the Cowles vision of using economics to evaluate economic policy and interpret phenomena has been lost. In addition, the data summaries reported by VAR econometricians are often transparent and the choice of an appropriate data summary requires knowledge of multivariate time series methods. Hence, the time series data summaries often have a black-box quality about them and judgements about fit are often mysterious to outsiders.

The tension between the goal of producing accurate descriptions of the data and the goal of producing counterfactual causal analyses for interpretation and policy prediction is a lasting legacy of the research of the Cowles Commission, and a major theme of this essay. It might be said that the theoretical reach of the Cowles analysts exceeded their empirical grasp. They developed a vision of empirical economics that has been hard to realize in practice.

Three very different responses to the perceived lack of empirical success of the structural research program and the lack of economic interpretability and apparent arbitrariness in the choice of VAR models emerged in the 1980s. All stress the need for greater transparency in generating estimates, although there is disagreement over what transparency means. At the risk of gross over-simplification, these responses can be classified in the following way. The first response is the calibration movement, which responds to the perceived inability of formal structural econometric methods to recover the parameters of economic models from time-series data and the perceived overemphasis on statistics to the exclusion of economics in the application of VAR models. This economic-theory-driven movement stresses the role of simple general equilibrium models with parameters determined by introspection, simple dynamic time-series averages, or by appeal to micro estimates. Calibrators emphasize the fragility of macro data and willingly embrace the conditional nature of causal knowledge. They explicitly reject "fit" as a primary goal of empirical economic models and emphasize interpretability over fit.

The calibrators have been accused of being too casual in their use of evidence. Sample averages from trended time series are used to determine parameters and when tested, the time series fits of the calibrated models are often poor. The microestimates that are sometimes used in this literature are often taken out of the contexts that justify them.

The second response is the nonparametric research program in econometrics and the earlier "sensitivity analysis" research in statistics that views the functional forms and distributional assumptions maintained in conventional structural (and nonstructural) approaches as a major source of their lack of credibility and seeks to identify the parameters of economic models nonparametrically or to examine the sensitivity of estimates to different identifying assumptions. The nonparametric identification analyses conducted within this research program clarify the role of functional forms and distributional assumptions in identifying causal parameters. Using hypothetical infinite samples, it separates out what can in principle be identified without functional form and distributional assumptions from what cannot. Many question the practical empirical relevance of nonparametric theory in the limited sample sizes available to most economists. Others question the novelty of the approach. Some form of bounding or sensitivity analysis has always been practiced by most careful empirical economists. Sensitivity analysis is a cornerstone of calibration econometrics.

A third, more empirical, approach to causal analysis has also emerged under the general rubric of the "natural experiment" movement. This popular movement searches for credible sources of identifying information for causal parameters, using ideal random experiments as a benchmark. It rejects the use of structural econometric models because, according to its adherents, such models do not produce credible estimates and impose arbitrary structure onto the data. In addition, the severe computational costs of estimating most structural models make the simpler estimation methods advocated by this group more appealing because findings can be easily replicated. The economic theory used to interpret data is typically kept at an intuitive level.

In many respects, this group has much in common with advocates of the VAR approach.

Both approaches are strongly empirically grounded. However, natural experimenters prefer simpler data summaries than those produced from modern time-series models. One goal, shared in common with the nonparametric econometricians and the statisticians who advocate sensitivity analysis, is to carefully locate what is "in the data" before any elaborate models are built or econometric identification assumptions are invoked.

In this literature, the "causal parameters" are often defined relative to an instrumental variable defined by some "natural experiment" or, in the best case scenario, by a social experiment. The distinction between variables that determine causes and variables that enter causal relationships is sometimes blurred. Accordingly, in this literature the definition of a causal parameter is not always clearly stated, and formal statements of identifying conditions in terms of well-specified economic models are rarely presented. Moreover, the absence of explicit structural frameworks makes it difficult to cumulate knowledge across studies conducted within this framework. Many studies produced by this research program have a "stand alone" feature and neither inform nor are influenced by the general body of empirical knowledge in economics. This literature emphasizes the role of causal models for interpreting data and analyzing existing policies, not for making the counterfactual policy predictions that motivated the research program of the Cowles Commission. That goal is viewed as impossible.

In order to make this paper accessible to a general audience, I discuss only the simplest models and deliberately avoid elaborate formal arguments. This strategy risks the danger of gross over-simplification of some very subtle points. It is hoped that the points made using simple models capture the essential features of the important contribution of econometrics to the understanding of causality, identification and policy analysis.

2. Causal Parameters, Identification, and Econometric Policy Evaluation

A major contribution of twentieth century econometrics was the recognition that causality and causal parameters are most fruitfully defined within formal economic models and that comparative statics variations within these models formalize the intuition in Marshall's quotation and most clearly define causal parameters. A second major contribution was the formalization of the insight developed in philosophy that many models are consistent with the same data and that restrictions must be the placed on models to use the data to recover causal parameters. A third major contribution was the clarification of the role of causal models in policy evaluation.

2.1 Causal Parameters

Within the context of an economic model, the concept of a causal parameter is well defined. For example, in a model of production of output Y based on inputs X that can be independently varied, we write the function $F : R^{N} \to R^{1}$ as

(1) $$Y = F(X_1, ..., X_N)$$

where $X = (X_1,...,X_N)$ is a vector of inputs defined over domain D ($X \in D$). They play the roles of the causes, *i.e.* factors that produce Y.[6] These causes are the primitives of the

[6] Philosophers would no doubt claim that I am begging the question of defining a causal parameter by assuming the existence of economic models like (1). My point is that given such models, discussions of causality become trivial. The whole goal of economic theory is to produce models like (1) and I take these as primitives. The multiplicity of possible models for the same phenomenon is the reason why a multiplicity of possible causal relationships may exist for the same phenomenon.

A more abstract approach to the definition of a causal relation that does not require specification of a function F or a well specified economic model builds on the work of Simon (1952) and Sims (1977) and specifies properties of the input space (X) and the output space (Y) and their relationship. The crucial idea is that inputs can be manipulated in ways that do not affect the structure of the causal relation but that affect the realized outputs.

Thus consider an abstract space S of possible features of models, both inputs and outputs. Consider two sets of restrictions: $A \subset S$ restricts inputs and $B \subset S$ restricts outputs. Suppose that S is mapped into two spaces: $P_X : A \to X$; $P_Y : B \to Y$. Then (A,B) defines a *causal ordering from* X to Y if A restricts X (if at all) but not Y and B restricts Y (if at all) without further restricting X. More formally (A, B), restrictions on S, determine a causal ordering from X to Y iff $P_Y(A) = Y$ and $P_X(A \cap B) = P_X(A)$. Geweke (1984) and Sims (1977) provide examples. The leading example is $S = \{(x, y) \in R^2\}$, $X = a$ (corresponds to A), $y + bx = c$ (corresponds to B). (A,B) is a causal ordering from x to Y because A determines x without affecting y. B along with A, determines y without further restricting x. There may be many pairs of restrictions on S that produce the same causal ordering. A version of this example with uncorrelated error terms across the two equations produces the causal chain model.

The Simon-Sims definition of a causal order is for a given pair of restrictions (A, B). The notion of causality is intimately involved with the idea of a stable relationship *i.e.* that if A is changed, the outcome will still be $A \cap B$ with B (the input-output relation) unchanged. Otherwise, when A is changed a different causal ordering may result. To guarantee that this does not occur we require the following condition: For any $A \subset S$ which constrains only X (*i.e.* $P_X^{-1}(P_X(A)) = A$), (A, B) determines a causal ordering from X to Y. (This is sometimes called "$B \subseteq S$" accepts X as input"). Thus a full specification of a causal model entails a description of admissible input processes and the notion that B is unchanged when A is manipulated (and hence the X is changed). For the model to be "correct," the set $B \subseteq S$ must be such that if B accepts X as an input, and when any set $C \subseteq X$ is implemented (A is manipulated), then relevant economic theory. Assuming that each input can be freely varied, so there are no functional restrictions connecting the components of X, the change in Y produced from the variation in X_j holding all other inputs constant is the causal effect of X_j. If F is differentiable in X_j, the marginal causal effect of X_j is

$$(2) \qquad \frac{\partial Y}{\partial X_j} = F_j(X_{1,...,}X_j, ..., X_N)\,|_{X=x.}$$

If F is not differentiable, finite changes replace derivatives. Nothing in this definition requires that any or all of the X_j be observed. Moreover, the X_j may be stochastic. Agents may make decisions about subsets of the X based only on expectations about the remaining X. In this case, realized X components enter (1) and we define the causal parameter in an *ex post* sense.[7] A variety of causal mechanisms can be contemplated even in this simple

---

$P_Y(P_X^{-1}(C) \cap B)$ is "true" *i.e.* in some sense depicts reality. This more general notion does not require that functions connecting causes to effects be specified.

[7] From Billingsley (1986), we know that if Y is a random variable, and X is a vector of random variables, then Y is measureable with respect to X if and only if $Y = F(X)$. Thus if we claim that an outcome is "explained" by X in this sense, then a causal relationship like (1) is automatically produced. Saying that Y is measurable with respect to X is not enough to define a causal function, however. If $Y = X + Z$, then $X = Y - Z$. Y is measurable with respect to X and Z; X is measureable with respect to Y and Z. Economic theory produces causal functions in which the inputs (or externally specified X variables) affect outputs. Different conceptual experiments define different causal relations. Thus consider a microeconomic demand curve where Y is the quantity of a good demanded and X is a vector of price, income and preference parameters. In the conceptual experiment where the agent is a price taker, and preferences and incomes are externally specified, $Y = g(X)$ is the Marshallian demand curve, and the X are the causal variables. In a different conceptual experiment, the roles of these variables may be reversed. Thus in a choice experiment examining "willingness to accept" functions, quantity Y may be specified externally and the minimum price the consumer would be willing accept to give up a unit of Y (a component of X) is the outcome of interest. Variations in Y *cause* a component of X to vary, say X_1, the reservation price. Depending on the exact question, the answer to the second problem may, or may not, be derived by inverting the $g(X)$ function specified in the first problem interchanging the roles of Y and the first component of X,X_1. Thus if income effects are small, $g(X)$ is the utility constant demand function. Varying quantities to produce associated marginal willingness to accept values would entail inverting g to obtain $X_1 = \varphi(Y, \tilde{X})$ where $\tilde{X} = (X_2, ..., X_J)$, assuming that a local implicit function theorem is satisfied (so in particular $\frac{\partial \varphi}{\partial Y} = \left(\frac{\partial g}{\partial X_1}\right)^{-1}$ where $\frac{\partial g}{\partial X_1} \neq 0$). However, if income effects are nonzero, the causal function required to answer the willingness to pay question cannot be obtained simply by inverting g. One would have to derive the Hicksian demand from the Marshallian demand and derive φ from the Hicksian demand.

In a production function example, $Y = F(X)$. If inputs X are externally specified, F is a causal function. To determine the amount of X_1 required to produce at output Y holding $(X_2, ..., X_J)$ at pre-specified values, one would invert F to obtain $X_1 = M(Y, X_2, ..., X_J)$, assuming $\frac{\partial F}{\partial X_1} \neq 0$. M is a causal function associated setting, because variations in the prices of inputs and outputs can cause X to vary. All of the parametric variations entertained in the microeconomic theory of the firm are possible sources of causal variation.

The assumption that the components of X can be varied independently is strong but essential to the definition of a causal parameter. The admissible variation may be local or global.[8] Restrictions on the admissible variation of the variables affect the interpretation of causal effects. For example, in a Leontief, or fixed-coefficient production model it is necessary to vary all inputs to get an effect from any. Thus an increase in X_j is necessary to increase Y but is not sufficient.[9] More generally, social and economic constraints operating on a firm may restrict the range of admissible variations so that a *ceteris paribus* change in one coordinate of X is not possible. Entire classes of variations for different restrictions on domain D can be defined but in general these are distinct from the *ceteris paribus* variations used to define a causal law.[10] The domain D is sometimes just one point as a consequence of the properties of a model, as I demonstrate below.

Model (1) with no restrictions among the X defines a model of potential outcomes. This can be linked to models of causal effects based on potential outcomes presented in the "treatment effect" literature by choosing the X values to correspond to different treatments.[11] When (1) is separable in X, we can write it as

---

with the conceptual thought experiment that $Y, X_2, ..., X_J$ are externally specified while X_1 is determined.

The crucial idea is that causal functions are derived from a conceptual experiment where externally specified causes are varied. There are as many causal functions as there are conceptual experiments.

[8] A formal definition of global variation independence is that the domain of D is the Cartesian product $\underline{X}_1 \times \underline{X}_2 \times \underline{X}_3, ..., \times \underline{X}_N$ where $\underline{X}_i$ is the domain of X_i and there is no restriction across the values of the X_i. When the X cross terms satisfy this restriction they are termed "variation free". The local version imposes this requirement only in neighborhoods.

[9] This corresponds to the concept of the "conjuncts of causality". See Marini and Singer (1988).

[10] One can define many different restricted "effects" depending on the restrictions imposed on D.

[11] The most direct way is to define X_1 as a treatment indicator and to define $Y_{x_1} = F_{x_1}(X_2, ..., X_N)$ as the potential outcome for treatment $X_1 = x_1$. Thus the models of potential outcomes of Neyman (1935), Fisher (1935), Cox (1959) and Rubin (1978) are versions of the econometric causal model. Galles and Pearl (1998) establish the formal equivalence of these two frameworks. Pearl (1998) presents a synthesis of these two approaches using directed acyclic graph theory. Thus the contrast sometimes made between "structural" and "causal" models formulated at the individual level is a false one. See Heckman (2000) for further discussion. Imbens and Angrist (1994) present a precise formulation of the Rubin model. The

$$Y = \sum_{j=1}^{N} \varphi_j(X_j)$$

and the causal effect of X_j can be defined independently of the level of the other values of X. Such separability is especially convenient if some of the X_j are not observed, because it avoids the need to define causal parameters in terms of unobserved levels of factors. For this reason, separable econometric models are widely used, and were the exclusive focus of the Cowles Commission analysts.

A major advance in thinking about causal parameters came when early econometric analysts recognized the possibility that Y and some or all of the components of X could be jointly determined or interrelated. This imposed severe restrictions on the causal parameters that can be defined in such models because it restricts the possibilities of variation in the causes. Controlled variation is the key idea in defining a causal parameter. The paradigm for this analysis was a model of market demand and supply:

(3) $\quad Q^D = Q^D(P^D, Z^D, U^D) \quad$ Demand

(4) $\quad Q^S = Q^S(P^S, Z^S, U^S) \quad$ Supply

where Q^D and Q^S are vectors of goods demanded and supplied at prices P^D and P^S respectively. (Throughout much of this paper, little is lost expositionally in thinking of the Q and P as scalars.) Z^D, Z^S, U^D and U^S are shifters of market demand and supply equations (*i.e.* determinants of demand and supply). They are determined outside of the markets where the P and Q are determined and are called external variables.[12] The P and Q are called internal variables. They may include distributions of the characteristics of consumers and producers. The U are causes not observed by the analyst; the Z are observed. In this section of the paper, there is no distinction between Z and U. This distinction is traditional, and useful in later sections so I make it here.

statistical models ignore the constraints across potential outcomes induced by social interactions and by resource constraints *i.e.* the potential restrictions on D. Heckman (2000) discusses the relationships among population treatment effect parameters, structural equations models and causal models.

[12] The term external variable appears to originate in Wright (1934). Frisch (1933) wrote about autonomous relationships. Given the numerous conflicting definitions of "exogenous" and "endogenous" variables documented by Leamer (1985), the "internal-external" distinction is a useful one for focusing on what is determined in a model and what is specified outside of it.

In Marshall's model of industry equilibrium, (3) is the demand for a good by a representative consumer while (4) is the supply function of the representative price-taking firm that maximizes profit given production technology (1) and factor prices. Assume that Q^{D} and Q^{S} are single-valued functions. If an equilibrium exists, $Q = Q^{\mathsf{D}} = Q^{\mathsf{S}}$ and $P = P^{\mathsf{D}} = P^{\mathsf{S}}$. If (P, Q) is uniquely determined as a function of the Z and U, the model is said to be "complete".[13]

The meaning of a causal parameter in (3) and (4) is the same as in the analysis of equation (1). If prices are fixed outside of the market, say by a government pricing program, we can *hypothetically* vary P^{D} and P^{S} to obtain causal effects for (3) and (4) as partial derivatives or as finite differences of prices holding other factors constant.[14] As in the analysis of the production function, the definition of a causal parameter does not require any statement about what is actually observed or what can be identified from data. As before, the definition of a causal parameter only requires a hypothetical model and the assumption that prices can be varied within the rules specified by the model. A statistical justification of (3) and (4) interprets (3) as the conditional expectation of Q^{D} given $P^{\mathsf{D}}, Z^{\mathsf{D}}$ and U^{D}, and interprets (4) as the conditional expectation of Q^{S} given P^{S}, Z^{S}, U^{S}.[15] Since we condition on all causes, these conditional expectations are just deterministic functional relationships. The effect of P^{D} on Q^{D} holding Z^{D} and U^{D} constant is different from the effect of P^{D} on Q^{D} not holding U^{D} constant, that is $E(Q^{\mathsf{D}} \mid P^{\mathsf{D}}, Z^{\mathsf{D}}, U^{\mathsf{D}}) \neq E(Q^{\mathsf{D}} \mid P^{\mathsf{D}}, Z^{\mathsf{D}})$. In the early investigations of causal models, and most models in current use, linear equation versions of (3) and (4) were used, so causal parameters could be defined independently of the levels of the causal variables.

As a matter of model specification, we might require that candidate causal functions obey certain restrictions. We might require that (3) and (4) have at least one solution $P = P^{\mathsf{D}} = P^{\mathsf{S}}$ and $Q = Q^{\mathsf{D}} = Q^{\mathsf{S}}$, so there is at least one market equilibrium. Other restrictions

[13] Koopmans and Hood (1953).

[14] Both (3) and (4) have well defined interpretations for their inverse functions. Thus in (3), P^{D} is the competitive price that would emerge if quantity Q^{D} were dumped on the market. In (4) P^{S} is the minimum price that competitive firms would accept to produce an externally specified Q^{S}.

[15] The justification for this is given in footnote 7.

like positivity of (the diagonals of) $\dfrac{\partial Q^{\mathsf{S}}}{\partial P^{\mathsf{S}}}$ (supply increasing in price) or negativity of (the diagonals of) $\dfrac{\partial Q^{\mathsf{D}}}{\partial P^{\mathsf{D}}}$ (downward sloping demand) might be imposed.

In the analysis of equations (3) and (4), one special case plays a central role. It is the model that equates demand and supply. In the important special case when prices and quantities are assumed to obey an equilibrium relationship, there is no meaning attached to a "causal" effect of a price change because the model restricts the domain (D) of P and Q to a single value if equilibrium is unique. Price and quantity are internal (or *endogenous)* variables jointly determined by the Z^{D}, Z^{S}, U^{S} and U^{D}. External (or *exogenous)* variables $(Z^{\mathsf{D}}, Z^{\mathsf{S}}, U^{\mathsf{D}}, U^{\mathsf{S}})$ determine (P, Q) but are not determined by them.

Holding everything else fixed in equilibrium (all other determinants of demand and supply), the prices and quantities are fixed. Thus, in equilibrium, the price of good j cannot be changed unless the exogenous or forcing variables, $Z^{\mathsf{D}}, Z^{\mathsf{S}}, U^{\mathsf{S}}, U^{\mathsf{D}}$, are changed. More formally, under completeness, we can obtain the reduced forms:

5(a) $\quad P = P(Z^{\mathsf{D}}, Z^{\mathsf{S}}, U^{\mathsf{D}}, U^{\mathsf{S}})$

5(b) $\quad Q = Q(Z^{\mathsf{D}}, Z^{\mathsf{S}}, U^{\mathsf{D}}, U^{\mathsf{S}})$.

The concept of an externally-specified variable is a model-specific notion. It entails specification of 5(a) and 5(b) as causal relationships in the sense of (1) to replace (3) and (4) when $Q^{\mathsf{D}} = Q^{\mathsf{S}}$ and $P^{\mathsf{D}} = P^{\mathsf{S}}$. In a fully nonparametric setting, this requires that the variables on the right-hand sides have no functional restrictions connecting them.[16] It also entails the notion that within the model, Z^{D} and Z^{S} can be independently varied for each given value of U^{D} and U^{S} (*i.e.* it is possible to vary Z^{D} and Z^{S} within the model holding U^{D} and U^{S} fixed).[17]

Assuming that some components of Z^{D} do not appear in Z^{S}, that some components of Z^{S} do not appear in Z^{D}, and that those components have a nonzero impact on price, one

[16] If functional forms (*e.g.* linearity) are maintained, some forms of dependence can be tolerated (*e.g.* nonlinear relationships among the variables in a linear model).

[17] Formally, the support of $(Z^{\mathsf{D}}, Z^{\mathsf{S}})$ is assumed to be the same for all values of $(U^{\mathsf{D}}, U^{\mathsf{S}})$. In this section, the Z^{D} and Z^{S} enter symmetrically with U^{D} and U^{S} so we should also require that the support of $(U^{\mathsf{D}}, U^{\mathsf{S}})$ is assumed to be the same for all values of $(Z^{\mathsf{D}}, Z^{\mathsf{S}})$ or, more generally, we might require that all variables be variation free in the sense of footnote 8.

can use the variation in the excluded variables to vary the P^{D} or P^{S} in equations (3) and (4) while holding the other arguments of those equations fixed. With this variation, one can define the causal parameters of the effect of P^{D} on Q^{D} and the effect of P^{S} on Q^{S}. Assuming differentiable functions, and letting $Z_{\mathsf{e}}^{\mathsf{S}}$ be a variable excluded from Z^{D}, and for notational simplicity assuming only a single market,

$$\frac{\partial Q^{\mathsf{D}}}{\partial P^{\mathsf{D}}} = \left(\frac{\partial Q}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}\right) \Big/ \left(\frac{\partial P}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}\right)$$

where the right-hand side expressions come from 5(a) and 5(b).[18] Defining $Z_{\mathsf{e}}^{\mathsf{D}}$ comparably,

$$\frac{\partial Q^{\mathsf{S}}}{\partial P^{\mathsf{S}}} = \left(\frac{\partial Q}{\partial Z_{\mathsf{e}}^{\mathsf{D}}}\right) \Big/ \left(\frac{\partial P}{\partial Z_{\mathsf{e}}^{\mathsf{D}}}\right).$$

Under these conditions, we can recover the price derivatives of (3) and (4) even though an equilibrium restriction connects $P^{\mathsf{D}} = P^{\mathsf{S}}$. The crucial notion in defining the causal parameter for price variation, when the market outcomes are characterized by an equilibrium relationship, is variation in external variables that affect causes (the P^{D} and P^{S}, respectively, in these examples) but that do not affect causal relationships (*i.e.* that are excluded from the relationship in question).[19] If an external variable is excluded from the causal relationship so it does not directly affect the causal relationship, the causal law is said to be invariant with respect to variations in that external variable. If the variable in question is a policy variable, the causal relationship is said to be "policy invariant."

[18] Proof: Differentiate (3) with respect to $Z_{\mathsf{e}}^{\mathsf{S}}$ to obtain

$$\frac{\partial Q^{\mathsf{D}}}{\partial Z_{\mathsf{e}}^{\mathsf{S}}} = \frac{\partial Q^{\mathsf{D}}}{\partial P^{\mathsf{D}}}\frac{\partial P^{\mathsf{D}}}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}.$$

Using equilibrium values ($P^{\mathsf{D}} = P^{\mathsf{S}} = P$) substitute from 5(a) to obtain $\frac{\partial P^{\mathsf{D}}}{\partial Z_{\mathsf{e}}^{\mathsf{S}}} = \frac{\partial P}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}$ and from 5(b) to obtain $\frac{\partial Q^{\mathsf{D}}}{\partial Z_{\mathsf{e}}^{\mathsf{S}}} = \frac{\partial Q}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}$. Assuming $\frac{\partial P}{\partial Z_{\mathsf{e}}^{\mathsf{S}}} \neq 0$, we obtain $\frac{\partial Q^{\mathsf{D}}}{\partial P^{\mathsf{D}}} = \left(\frac{\partial Q}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}\right) \Big/ \left(\frac{\partial P}{\partial Z_{\mathsf{e}}^{\mathsf{S}}}\right)$. If there are several $Z_{\mathsf{e}}^{\mathsf{S}}$ variables that satisfy the stated conditions, each defines the same causal parameter.

[19] The definition of a causal parameter crucially depends on independent variation. In the equilibrium setting under consideration, without an exclusion restriction, the equilibrium quantities cannot be independently varied. Thus no independent variation is possible. However, if we consider a disequilibrium setting, where prices (or quantities) are set externally, say through government policy or a social experiment, then the causal parameter can be defined, as before.

Variations in the included Z variables have direct effects (holding all other variables in (3) or (4) constant) and indirect effects (through their effects on the endogenous variables via 5(a) and 5(b)). The direct effects of Z can be computed by compensating for changes in the P induced by the changes in the included components of Z by varying the excluded components of Z. These direct and indirect effects play a crucial role in path analysis developed by Wright (1934) and widely used in sociology. (See Blalock, 1964).[20] The direct causal effects are called structural. Both direct and indirect effects are causal, and are defined by well-specified variations.[21]

As a consequence of the potentially interdependent nature of some causes, a new terminology was created. Structural causal effects are defined as the direct effects of the variables in the behavioral equations. Thus the partial derivatives of (3) and (4) are termed structural effects. When these equations are linear, the coefficients on the causal variables are called structural parameters and they fully characterize the structural effects. In more general nonlinear models, the derivatives of the structural (or behavioral) equation no longer fully characterize the structural relationship. The parameters required to fully characterize the structural relationship are termed structural parameters. A major difference between the Cowles group, which worked exclusively with linear equations, and later analysts working with explicitly parameterized economic models, is in the definition of a structural parameter and the separation between the concept of a structural effect from the concept of a structural parameter.[22]

Both structural equations and reduced form equations can be used to generate causal parameters. They differ in what is held constant in the sense of Marshall. Reduced form

[20] Path analysis estimates the direct effect of structural variables and the direct effects of external variables as well as their indirect effects operating through structural variables. The "total effect" of an external variable is the sum of the direct effect and the indirect effects operating through all of the endogenous variables in a relationship.

[21] In this section, there is no distinction between the Z and the U, and the variations discussed in the text can be defined using any excluded or included variables, observed or unobserved. I use the observed-unobserved notation only because it is more familiar and will play a role in the next subsection.

[22] The term "deep structural parameter" was introduced in the 1970s to distinguish between the derivatives of a behavioral relationship used to define causal effects and the parameters that generate the behavioral relationship.

relationships can be defined without the exclusion restrictions required to define structural relationships.

Functional relationships among variables that are invariant to variations in external variables are central to the definition and identification of causal laws in the case when some variables of a system of equations are interdependent. The notion of invariant relationships motivated the Cowles Commission definition of a structural equation. It also motivated the econometric estimation method of instrumental variables using empirical counterparts to the hypothetical relationships.

These notions all have counterparts in dynamic settings, where the variables are time-dated. Time-series notions of causality as developed by Granger (1969) and Sims (1972), are conceptually distinct and sometimes at odds with the notion of causality based on controlled variation that is presented in this paper and at the heart of the quotation from Marshall presented in the introduction. The time-series literature on causality uses time dating of variables (temporal precedence relationships) to determine empirical causes and does not define or establish *ceteris paribus* relationships. Thus letting t denote time, past Y_{t} is said to cause future X_{t} if past Y_{t} helps predict future X_{t} given past X_{t} using some statistical goodness-of-fit criterion. Such causality can arise if future X_{t} determines past Y_{t} as often arises in dynamic economic models.[23] The "causality" determined from such testing procedures does not correspond to causality as defined in this paper, and in this instance is in direct conflict with it.

The limited role of the time-series tests for causality within articulated causal dynamic models is discussed by Hansen and Sargent (1980). The dynamic structural models derived from economic theory of the sort analyzed by Hansen and Sargent provide the framework for defining causality as used in this paper and for conducting counterfactual policy analysis.

[23] In a perfect foresight model like that of Auerbach and Kotlikoff (1987), future prices determine current investment. Time-series causality tests would reveal that investment "causes" future prices which is precisely the wrong conclusion for the concept of causality used in this paper. Hamilton (1994, pp. 306-309) presents an example in which Granger causality is in opposition to the causal interpretation in the sense of this paper and another example in which Granger causality is in accord with the definition of causality used in this paper.

I do not exposit these models only because of my self-imposed limitation on the technical level of this paper.

2.2 Identification: Determining Causal Models From Data

The formalization of models, the definition of causal and structural laws, and the notion of structural laws that are invariant with respect to variation in excluded external variables were important contributions of economics. Even more important was the clarification of the limits of empirical knowledge.[24] An identification problem arises because many alternative structural models are consistent with the same data, unless restrictions are imposed. Empirical knowledge about structural models is contingent on these restrictions.

The first studies of this problem were in the context of the supply-demand model of equations (3) and (4), assuming equilibrium ($P^{\mathsf{S}} = P^{\mathsf{D}}$ and $Q^{\mathsf{S}} = Q^{\mathsf{D}}$). This case is still featured in econometrics textbooks. The identification problem is particularly stark in this setting if there are no Z^{D} or Z^{S} variables, and if the U^{D} and U^{S} are set to zero, so there is no problem of the unobservables U^{D} or U^{S} being correlated with P or Q.[25] In this case, two equations, (3) and (4), relating Q to P coexist. (The demand curve and the supply curve respectively). They contain the same variables. Empirically, there is no way to determine either relationship from the joint distribution of Q and P unless extra information (restrictions on models) is available.[26]

Although the identification problem was first systematically explored in this context, it is a much more general problem and it is useful to consider it more generally.[27] In its essence, it considers what particular models within a broader class of models are consistent with a given set of data or facts. More specifically, consider a model space M which is the class of all models that are considered as worthy of consideration. Elements $m \in M$ are possible theoretical models. There are two attributes of a model, corresponding to what one can observe about the model in a given set of data and what one would like to know

[24]Other fields independently developed their own analyses of the identification problem in more specialized settings. (Koopmans and Reirsol, 1950).

[25]Identification problems can arise even if there are no error terms in the model.

[26]Morgan (1990) discusses early attempts to solve this problem using *ad hoc* statistical conventions.

[27]This framework is based on my interpretation of Barros (1988).

about the model. Define functions $g : M \longrightarrow T$ and $h : M \longrightarrow S$ where T and S are spaces chosen so that $g(M) = T$ and $h(M) = S$. S is the source or data space and T is the target space. For any specific model $m \in M$, $h(m) = s \in S$ represents those characteristics of the model that can be observed in the available data. The map $g(m) = t \in T$ applied to the model gives the characteristics of a model that we would like to identify. Some parameters may be of interest while others are not, and models may only be partially identified. Only if one is interested in determining the entire model would $T = M$ and then g would be an identity map.

Many models m may be consistent with the same source space, so h is not one to one. In this abstract setting, the identification problem is to determine whether elements in T can be uniquely determined from elements in S. Elements in T and S are related by the correspondence $f \equiv g \circ h^{-1}$. The identification problem arises because for some $s \in S$, $f(s)$ may have more than one element in T. In that case more than one interpretation of the same evidence is available. If we limit attention to T via g, rather than M, f is more likely to map points of S into points of M. The goal of identification analysis is to find restrictions, R, to modify f to f^{R} so that $f^{\mathsf{R}}(s)$ has at most one element in T, *i.e.* so that only one story is possible given the data.

In the usual form of the identification problem, restrictions are imposed on the admissible models in M so $R \subseteq M$. For each $s \in S$ define $f^{\mathsf{R}}(s) = g(h^{-1}(s) \cap R)$. If R is chosen so that for all $s \in S$, $f^{\mathsf{R}}(s)$ has at most one element in T, R forms a set of identifying restrictions. Thus $R \subseteq M$ identifies g from h when for any $s \in S, f^{\mathsf{R}}(s) = g(h^{-1}(s) \cap R)$ contains at most one element in T.[28] If R is too restrictive, for some values of s, $f^{\mathsf{R}}(s) = \varnothing$, the empty set, so R may be incompatible with some s, *i.e.* some features of the model augmented by R are inconsistent with some data. In this case, R forms a set of over-identifying restrictions. Otherwise, if $f^{\mathsf{R}}(s)$ contains exactly one element in T for each $s \in S$, R forms a set of just-identifying restrictions.[29]

[28]In principle, restrictions can also be placed on T but these restrictions are often less easy to interpret. In the context of a demand function, T could include both the price and income elasticities, and we might restrict the income elasticity to be known.

[29]Note that $f^{\mathsf{R}}(s) = \varnothing$ iff $h^{-1}(s) \cap R = \varnothing$ since we assume $g(m) \neq \varnothing$ for all $m \in M$. Thus a set of

For a given set of data, it will usually be necessary to restrict the model via R to produce unique identification. Alternatively, enlarging the data (expanding the source space S) may also produce identification. Enlarging S may entail larger samples of the same distributions of characteristics or expanding the data to include more variables.[30] In the classical linear-in-parameters, simultaneous-equations model, exclusion restrictions are used to define R. The source space is the joint distribution of (P, Q) given $(Z^{\mathsf{D}}, Z^{\mathsf{S}})$. Under normality, all of the information in the source space is contained in the means and covariances of the data.

In the supply-demand example, the model space M consists of all admissible models for supply and demand, and the target space T could be the price elasticities of supply and demand. Even when $U^{\mathsf{D}} = U^{\mathsf{S}} = 0$, the model of equations (3) and (4) is not identified if $Z^{\mathsf{D}} = Z^{\mathsf{S}}$. As first noted by Tinbergen (1930, reprinted 1995), if Z^{D} and Z^{S} each contain variables not in the other, but that determine P and Q *via* 5(a) and 5(b), the model is identified. In this deterministic setting, the variation in P induced by the external variables in Z that are not in the equation being analyzed substitutes for variation in P that would occur if market equilibrium conditions did not govern the data. If there are unobservables in the model, their effects on Z must be controlled to use this variation. Conventional statistical assumptions made to effect such control are that $(U^{\mathsf{D}}, U^{\mathsf{S}})$ are statistically independent or mean independent of $(Z^{\mathsf{D}}, Z^{\mathsf{S}})$.[31]

One sample counterpart to this identification argument is the method of instrumental variables.[32] Variation in the excluded Z induces variation in the P. This variation is essential whether or not there are error terms in the model. Many different estimation methods can be used to induce the required variation in the data. An instrument - defined as a source of variation - need not necessarily be used as an instrumental variable to empirically

identifying restrictions forms a set of just-identifying restrictions iff $h(R) = S$.

[30]Most modern analyses of identification assume that sample sizes are infinite so that enlarging the sample size is not informative. However, in any applied problem this distinction is not helpful. Having a small sample (*e.g.* fewer observations than regressors) can produce an identification problem.

[31]Mean independence is the condition $E(U^{\mathsf{D}} \mid Z^{\mathsf{D}}, Z^{\mathsf{S}}) = 0$ and $E(U^{\mathsf{S}} \mid Z^{\mathsf{D}}, Z^{\mathsf{S}}) = 0$.

[32]The method was independently developed by Philip Wright (1928) and Olav Reirsol (1945). Epstein (1987) speculates that Sewall Wright, the famous geneticist, inventor of path analysis and son of Philip Wright, was the likely inventor of this method. See also Morgan (1990) and the intellectual history reported in Goldberger (1972).

identify causal parameters. Social experiments that do not alter the structural relationship being studied can induce the required variation.[33] The method of control functions that explicitly models the dependence of the error term on the right-hand side variables can do so as well.[34] A much richer class of restrictions R can also produce identification including restrictions on covariances, restrictions on coefficients and restrictions on functional forms. (Fisher, 1966). In dynamic models, cross-equation restrictions produced from theory and restrictions on the time-series processes of the unobservables provide additional sources of information. (Hansen and Sargent, 1980). Restrictions on the covariances of the innovations of vector-autoregressive models play a crucial identifying role in VAR frameworks. (Sims, 1980).

A major lesson of the econometric literature on identification that is still not well understood by many empirical economists is that just-identifying assumptions are untestable. Different restrictions R_i, $i = 1, ..., I$ that secure just-identification produce I different stories about the parameters in the target space T for the same element s in data source space S. Different admissible models m produced by different R_i that are just-identifying are compatible with the same elements of the source space S, and all empirical knowledge is in S. Accordingly, no empirical test of just-identifying restrictions is possible. They must be justified by an appeal to intuition or prior information and the justification is not statistical in nature. The extra restrictions on the source space are testable. Tests of identification reported in the empirical literature are always tests of over-identifying assumptions and they are based on maintained identifying assumptions although they are frequently and

[33] Heckman, LaLonde and Smith (1999) show that randomization is an instrumental variable.

[34] Instead of purging the endogeneous variables of the errors, this method models the errors in terms of the right-hand side endogeneous variables and all of the regressors. In a two-equation, simultaneous-equations system: (*) $Y_1 = \alpha Y_2 + \beta Z_1 + U_1$ and $Y_2 = \gamma Y_1 + \delta Z_2 + U_2$ where $E(U_1, U_2 \mid Z_1, Z_2) = 0$, the method of control functions forms $E(U_1 \mid Y_2, Z_1, Z_2, \varphi)$, where φ is some unknown parameter vector and enters it as a conditioning function in (*): $Y_1 = \alpha Y_2 + \beta Z_1 + E(U_1 \mid Y_2, Z_1, Z_2, \varphi) + (U_1 - E(U_1 \mid Y_2, Z_1, Z_2, \varphi))$. With sufficient variation in $E(U_1 \mid Y_2, Z_1, Z_2, \varphi)$, which in the absence of parametric restrictions clearly requires variation in Z_2, when Z_1, is held fixed, it is possible to identify α, β and φ. See Heckman and Robb (1986). Nonparametric versions of the method exist. The method underlies an entire class of nonparametric selection bias estimators. See Heckman, LaLonde and Smith (1999) and Heckman (2000).

confusingly referred to as tests of identifying conditions.[35]

Models that are not fully identified may contain subsets of parameters that can be identified. As first noted by Wright (1934) and Marschak and Andrews (1944), models that are not identified may still contain information on ranges of structural parameter values. This insight is a focus of recent research which I discuss below.

2.3 Econometric Policy Evaluation: Marschak and the Lucas Critique

In the context of structural economic models, identification necessarily precedes testing of specific causal hypotheses. Thus if an effect cannot, in principle, be identified it is not possible to test whether the effect is present in the data.[36] Structural parameters have a clear economic interpretation. Estimates of them can be used to test theories if identifying assumptions are maintained, to interpret empirical relationships, to perform welfare analysis (*e.g.* compute consumer surplus) and to improve the efficiency of estimates and forecasts if models are overidentified. They can also be used as frameworks for empirical analyses that facilitate the accumulation of evidence across different studies. These benefits are to be set against the costs of making identifying assumptions.

One benefit that motivated much of the early econometric literature, and one that still motivates much econometric research, is the ability to use structural models to predict the consequences and evaluate the effects of alternative economic policies. Formal econometrics was perfected by economists greatly influenced by the Great Depression. Early pioneers like Tinbergen and Frisch were policy activists who were optimistic about the ability of enlightened social planners to control the business cycle through the introduction of new policies based on empirically justified econometric models.[37]

The clearest statement of the benefit of structural econometric models for social planning is by Marschak (1953). Later work by Lucas (1976) builds on Marschak's analysis by

[35] Tests for endogeneity in simultaneous-equations models are always predicated on the existence of at least one just-identifying exclusion restriction that is used as an instrument.

[36] If the causal effect is identified within a range of values that excludes the no-effect value, exact identification of the causal effect is not required to test the null hypothesis of no-effect. One simply asks if the identified range includes the value implied by null hypothesis.

[37] For an illuminating discussion of the work of Frisch and his committment to social planning, see O. Bjerkholt (1998), J. Chipman (1998), and the other papers in S. Strom (1998).

updating it to incorporate explicit models of intertemporal decision making under uncertainty.

The goals of econometric policy evaluation are to consider the impact of policy interventions on the economy, to compute their consequences for economic welfare and to forecast the effect of new policies never previously experienced. If a policy has previously been in place, and it is possible to adjust outcome variables for changes in external circumstances unrelated to the policy that also affect outcomes, econometric policy evaluation is just a table look-up exercise. Structural parameters are not required to forecast the effects of the policy on outcomes although causal effects may still be desired for interpretive purposes. It is enough to know the reduced forms 5(a) and 5(b). What happened before will happen again. If the same basic policy is proposed but levels of policy parameters are different from what has been experienced in the past, some interpolation or extrapolation of past relationships is required. A disciplined way to do this is to impose functional forms on estimating equations.

The theory of econometric policy evaluation typically proceeds by assuming that policy parameters are external variables.[38] Under this assumption, one can use reduced forms like 5(a) and 5(b) to forecast the effects of policies on outcomes, using historical relationships to predict the future.

Marschak's case for knowing structural parameters was for their use in predicting the effects of a new policy, never previously experienced. Such a problem is intrinsically difficult and of the same character as the problem of forecasting the demand for a new good, never previously consumed. Without the ability to project past experience into knowledge of the future, the problem is intractable. The essence of this problem is succinctly summarized in a quotation from Knight:

> *"The existence of a problem in knowledge depends on the future being different from the past, while the possibility of a solution of a problem of knowledge depends on the future being like the past" (Knight, 1921, p. 313).*

[38]Marschak noted, but did not develop, the notion that policies themselves might be internal or endogenous.

Marschak analyzed a class of policies that can be evaluated using structural equations. The prototype was the analysis of the effect of a commodity tax never previously experienced on equilibrium prices and quantities. Since there is no previous experience with the policy, the table look-up method is not available to solve this problem.

Marschak operationalizes Knight's quotation by noting that in general a commodity tax changes the price of the good. For a proportional tax, τ, the after-tax price paid by a consumer is $P^* = P(1 + \tau)$ where P is the price received by producers. With structural equations in hand, one can in principle accurately forecast the effect of a new tax on market aggregates using structural equations (3) and (4) modified to incorporate the tax wedge. Simply insert the tax wedge and solve the structural equations for the new equilibrium prices and quantities.

Under the conditions sketched in Section 2.2, it is possible to use external variation in excluded variables in the pre-policy period to identify these equations from equilibrium observations. A potentially serious problem is that the functions (3) and (4) may be nonlinear, and the range of historical price variation may not be sufficient to trace out those functions over the range of values that are relevant in the new policy regime. Marschak avoids this problem by assuming linear structural equations. In his case, a finite set of data points determine these equations over their entire domain of definition. A fully nonparametric approach would require sufficient variation in P in the pre-tax sample used to fit the structural equations to encompass the relevant support of P in the period of the policy.

By estimating the structural parameters in the pre-policy period and by linking the policy variation to price variation, Marschak solves the problem of forecasting the effect of a new tax policy on market aggregates. It is important to note that both of his steps are necessary. Knowledge of structural parameters is not enough. It is also necessary to somehow link the variation that will be induced by the new policy to some variation within the model in the pre-policy sample period so "the future will be like the past." In Marschak's example, tax and price variation are on the same footing.

If this link cannot be made, the problem of forecasting the impact of a new policy cannot be solved using an econometric model. Consider a policy that injects random taxation into the deterministic environment considered by Marschak. If the uncertainty induced by the policy is of a fundamentally different character than anything previously experienced, an empirically based solution to the policy forecasting problem is intractable.[39]

In his seminal article, Marschak made another important point. Different criteria for evaluating a policy require different parameters. (Different policies require knowledge of different target spaces T). Given an assumption that describes a proposed policy in terms of variation experienced in a pre-policy period, the answers to different economic decision problems require different structural parameters or combinations of structural parameters. It may not be necessary to determine the full structure to answer a particular policy evaluation question. Thus if we seek to determine the effect of an externally determined price change on consumer welfare, it is only necessary to identify the demand curve, a weaker requirement than full system identification. Marschak presents an extensive discussion of decision problems requiring different amounts of structural information including some where only ratios of some of the structural parameters are required.

A quarter of a century later, Lucas (1976) returned to the policy evaluation problem addressed by Marschak. Among other contributions, Lucas updated Marschak's analysis to incorporate explicit models of decision making under uncertainty and to incorporate endogenous expectations. Lucas criticized the *practice* of econometric policy evaluation in the late 1960s and early 1970s because it was careless in modelling the expectations of agents and did not account for the effects of policy changes on the expectations of the agents about future shocks. He noted that different stochastic processes for external forcing variables produced by different economic policies would in general induce different behavioral responses and that an adequate model of policy forecasting should account for this.

[39] The parallel problem of estimating the demand for a new good entails the same difficulties. Lancaster (1971) solves the problem by assuming that new goods are rebundlings of the same characteristics as in the old goods and proposes estimation of the demand for characteristics and distributions of population preference parameters to generate forecasts. Quandt (1970) and Domencich and McFadden (1975) use similar schemes in the analysis of discrete choice to forecast the demand for new goods. In Domencich and McFadden, unobserved attributes of new goods are assumed to be independent realizations of unobserved attributes for old goods.

Econometric models that ignore expectations are misspecified or "incomplete" in the precise sense of the Cowles economists. A missing internal or endogenous variable would make an estimated misspecified "structure" appear to be empirically unstable when the distribution of the forcing variables was changed, say by changes induced by new economic policies.

Lucas claimed that this type of model misspecification accounted for the parameter drift observed in many empirical macro models although the empirical evidence for his claim is controversial.[40] The source of the parameter instability is model misspecification due to omitted endogenous expectations variables and any omitted external forcing variables associated with the equation determining expectations.[41] Thus estimates of misspecified structural equations will appear to be non-invariant in response to changes in the distributions of the forcing variables. Such changes may come from changes in policies but this is only one possible source of change in the forcing variables.[42]

To recast the Lucas critique in the language of a classical linear simultaneous equations model, write

(6) $\Gamma Y + BX = U$

where Y is a vector of endogenous variables and X is a vector of exogenous or external forcing variables, and U is a vector of unobserved forcing variables. If Γ has an inverse, the system is said to be complete and a reduced form Y as a function of X is well defined. As noted by Hansen and Sargent (1980), modern dynamic theory in general requires nonlinear models with cross-equation restrictions and lagged (and sometimes future) values of Y, X and U, so (6) is too simple to capture the class of dynamic structural models discussed by Lucas.

[40] See the evidence in Fair (1994). The evidence that policy shifts affect the parameters of fitted macro models is at best mixed. However, the supply shocks of the 1970s definitely affected the stability of fitted macrorelationships.

[41] As demonstrated *e.g.* by White (1987), estimated misspecified models will appear to exhibit parameter instability if the distributions of the external forcing variables change over the sample period.

[42] Lucas compares two steady states - one before the policy change and one after - and does not analyze the short-run responses to the regime shift that are required for explaining the time-series data accompanying a regime shift.

However, it is useful to make Lucas' point in this framework if only to make an analogy. To do so, partition $Y = (Y_1, Y_2)$ into two components corresponding to the included and omitted endogenous variables. Think of Y_2 as the expectations variables determined by the model under rational expectations. Then partition $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$, $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$, $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, and $U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$ conformably. Assuming Γ_{11} has an inverse

$$Y_1 = -\Gamma_{11}^{-1}\Gamma_{12}Y_2 - \Gamma_{11}^{-1}B_1X_1 + \Gamma_{11}^{-1}U_1.$$ [43]

By omitting Y_2 and any determinants of Y_2, the misspecified model for Y_1 will exhibit within-sample instability if the distribution of the forcing variables changes in different sample periods. These changes could arise because of policy interventions or because of external shocks to the economy unrelated to policy interventions. If changes in the forcing variables are modeled along with the unobserved expectations of the model, there should be no drift in estimated parameters.

Marschak established that structural models are only a necessary ingredient for evaluating a new policy. Even if correctly specified structural equations are estimated, econometric policy evaluation will be ineffective in evaluating new policies that involve variation in variables that previously did not vary and that cannot be linked to variables in the model that have varied. The problems of forecasting the effects of a new policy are profound and many would argue they are impossibly hard. Inadvertently, the Cowles group fashioned a powerful argument against the possibilities of empirically based social planning - a major goal of the early econometricians.

3. The Cowles Research Program as A Guide to Empirical Research and Responses To Its Perceived Limitations

By the mid-1950s, the Cowles research program defined mainstream theoretical econometrics. It was widely recognized that it needed to be augmented to include a richer array of

[43] Wallis (1980) presents a systematic analysis of the formulation and estimation of the rational expectations models within a linear in-parameters Cowles-like model augmented to account for serial correlation in the equation errors. He does not, however, develop the consequences of model misspecification that are discussed in this paper.

dynamic models and to account for serial correlation in aggregate time series. Granger and Newbold (1977) and Zellner and Palm (1974), among others, made important extensions of the Cowles Commission framework to time-series settings integrating the work of Frisch (1933) and Slutsky (1937) on propagation mechanisms and the dynamic stochastic theory of Mann and Wald (1943) and Phillips (1966) into the Cowles Commission framework.

The sources of identification for the macro models were controversial. In an influential paper, T.C. Liu (1960) argued that the exclusion restrictions used in the Cowles models were artificial, that economic theory would require the inclusion of many more variables than those found in the existing models of the day, and that structural models were fundamentally underidentified. He carefully scrutinized the *ad hoc* nature of identification in the influential Klein-Goldberger model (1955). In his view, the quest for structural estimation was quixotic, and hence the goal of estimating the impact of a new policy was an impossible dream. He advocated the use of reduced-form equations for making economic forecasts. As subsequent generations of macro models based on the Cowles program became larger, the criticism of them became more intense. Their computational complexity and fragile identification attracted unfavorable discussion that elaborated on the negative commentary of Liu. (See the discussion in Brunner, 1972).

At the same time, as documented by Epstein (1987) and Morgan (1990), the Cowles models were perceived to be empirical failures. Estimated parameter instability and the practice of refitting empirical macro models over short time periods indicated that the official rhetoric of econometrics was not followed in practice. Naive forecasting models often did better than fully articulated structural models estimated under the Cowles paradigm. This evidence motivated the "Lucas critique" as one possible explanation for the observed parameter instability.

Although the Cowles Commission was successful in pressing the potential importance of endogeneity bias (the feedback between the U and the Y in equation (6)), the empirical importance of this problem compared to the other problems in estimating macro models was never clearly established. Indeed Basmann (1974), one inventor of the method of two-stage least squares, claimed that endogeneity bias was of second-order importance compared to the basic measurement error in the macrodata. In his invited lecture at the 1970 World Congress of the Econometric Society, he plotted measurement error boxes around the least squares and endogeneity - corrected estimates of the consumption function from an influential article by Haavelmo (1947). The slight change in the OLS fitted line that arose from correcting for simultaneity bias was dwarfed by the intrinsic variation in the data due to measurement error.[44] Simultaneity bias was a second-order problem in Haavelmo's data, although it was the focal point of Cowles econometrics.[45]

Two radically different responses emerged to the perceived failure of the Cowles research program. One response was the development of more explicitly parameterized structural economic models. The other response, developed by Granger (1969) and Sims (1972, 1977, 1980, 1986), was based on systematic application of vector autoregression time-series methods that had been fully developed after the basic Cowles program had been completed but which originated in the work of Mann and Wald (1943). The vector autoregression approach (VAR) econometrically implemented Frisch's vision (1933) of dynamic propagation mechanisms for business cycle research.

A rough characterization of the two responses is that structuralists adhered to and refined the theory often at the expense of obtaining good fits to the data. VAR econometricians stuck to the data as summarized by vector autoregressions and used economic theory as an informal guide for interpreting statistical decompositions. Covariance restrictions on time-series processes replaced the *a priori* restrictions on behavioral equations invoked by the Cowles Commission. Innovation accounting replaced causal analysis.

Structuralists adopted a deductive, theory-oriented approach, that estimated structural models with testable implications that were frequently rejected in the data. The rejections

[44] Appreciation of the importance of measurement error in economic data goes back to the work of Morgenstern (1950). The importance and consequences of measurement error in microeconomic data is a running theme of the work of Zvi Griliches. For a recent comprehensive analysis of measurement error in survey data see Bound, Brown and Mathiowetz (2000).

[45] A quotation from Klein (1960), as presented in Epstein (1987, p. 119), reinforces this point:
"The building of institutional reality into *a priori* formulations of economic relationships and the refinement of basic data collection have contributed much more to the improvement of empirical econometric results than have more elaborate methods of statistical inference".

were to be used as a basis for improvements in the models following the methodology of Popper (1959), discussed below in Section 4. Interpretability, counterfactual policy evaluation and welfare analyses were featured in this approach. Advocates of VAR approaches started with models that fit the data and imposed minimal ground-up restrictions on time series processes that were only loosely motivated by economic theory. Both groups retained the original Cowles emphasis of using state-of-the-art formal statistical methods to describe the data.

The VAR approach starts with a vector autoregression for a k-dimensional vector Z_t :

(7) $Z_t = C_1 Z_{t-1} + C_2 Z_{t-2} + ... + C_q Z_{t-q} + \varepsilon_t$

where $E(\varepsilon_t \varepsilon_t') = V$, $E(\varepsilon_t) = 0$ and ε_t is uncorrelated with all variables dated $t-1$ and earlier. Under standard conditions, $C_1, ..., C_q$ and V can be estimated by least squares. Estimates of (7) summarize the time-series data by matrices of regression parameters and the variance V. This framework extends the Cowles model of equation (6) to a time-series setting. The "structural" version of (7) is

(8) $A_0 Z_t = A_1 Z_{t-1} + ... + A_q Z_{t-q} + U_t$

where A_0 is assumed to be non-singular and the A_i are $k \times k$ matrices of constants and $E(U_t U_t') = D$ and $A_0 \varepsilon_t = U_t$. The original Cowles model assumed that $A_i = 0$, $i > 1$,

$$Z_t = [Y_t, X_t] \text{ and } A_0 = \begin{pmatrix} \Gamma & B \\ 0 & I \end{pmatrix}.$$

The "0" in the lower left block made X_t exogenous or external - it determined Y_t but was not determined by it.

Equations (7) and (8) extend the Cowles framework to account for serial correlation and general dynamic responses across equations. Equation (7) and (8) are connected by the relationships $C_i = A_0^{-1} A_i$ and $V = A_0^{-1} D (A_0^{-1})'$. Restrictions on the A_i and D serve to identify the structural model (8). The link to an explicit dynamic economic theory is at best weak. (Hansen and Sargent, 1991). Fully specified dynamic economic models can rarely be cast in the form of equations (7) and (8).

Innovation accounting-estimating the effects of innovations in U_t on the time paths of the Z_t- is prominently featured in this literature. Such accounting takes the place of simple comparative statics exercises in the original Cowles models and traces out the dynamic response of exogenous shocks on the outcome variables using (8). In order to recover the U_t from the estimated ε_t to perform such accounting exercises, it is necessary to impose some structure on the model (8). While Sims (1980) and others criticize the practice of "incredible" identification as practiced by Cowles econometricians, to outsiders the substitute sources of identification advocated in this literature look no more credible and often appear to be of the same character.

Many different identifying conventions only loosely motived by economic theory have been assumed. The leading approach is to adopt a causal chain approach and assume that D is diagonal and A_0 is triangular. As noted by many analysts, such restrictions are intrinsically arbitrary. Other conventions, as summarized in Christiano, Eichenbaum and Evans (1999) include imposing other restrictions on the A_i and the D to accord with features of the model viewed to be intuitively satisfactory such as long-run neutrality of certain variables.[46] (See also Shapiro and Watson, 1988.) Alternative identifying assumptions produce different estimates of the importance of the innovations. The goals of estimating the impacts of new policies never previously experienced or of assessing the welfare consequences of existing policies are abandoned.[47] Vector autoregression data summaries are often not transparent to outsiders. An air of mystery and controversy surrounds the generation of the "facts" as summarized by (7). The appeal to modern time-series methods appears to substitute the black box of Cowles identification with the black box of time-series identification.

A second response to the perceived empirical failure of the Cowles research program was to develop more explicit structural models and to exploit new sources of micro data. About the same time that disillusionment was setting in with the Cowles methods, a vast new array of micro data in the form of panels and cross-sections became available to economists. Orcutt (1952) had forcefully argued that time-series variation was too limited to empirically recover the parameters of structural models and that use of micro data cross-sections and

[46] Neutrality restricts certain sums of coefficients to be zero.

[47] Sims (1986) proposes a form of time series interpolation/extrapolation to assess the impacts of new policies, but does not discuss the construction of welfare measures.

panels would greatly aid in this regard. New sources of microdata became available, in part through the pioneering efforts of the Institute for Social Research at the University of Michigan, and in part because government agencies disseminated microfiles from Census and Current Population survey data.

Microeconometrics began to flourish as a separate field. Regression analysis as synthesized by Goldberger (1964) was the tool of choice for analyzing this data. A wealth of empirical regularities was presented and new econometric methods were developed to solve problems of censoring, self-selection and limited dependent variables that arose in the analysis of micro data, and to develop new methods to exploit cross-section and panel data.[48]

With the advent of computers, the pace of empirical work increased dramatically in macro-and microeconomics. This forever changed the scale of the empirical enterprise in economics and created the phenomenon of a software-driven applied econometrics. Long tables of regression coefficients with different conditioning variables became commonplace in research papers and interpretation of estimates was difficult.[49] "Holding variables constant in the sense of linear regression" became the operational empirical counterpart of Marshall's *ceteris paribus* clause. Endogeneity bias was always a concern, and many of the careful empirical studies of the day accounted for such bias, although the precision of the estimates was usually greatly reduced when corrections for endogeneity were made, especially in the micro studies where instruments were usually weak.

The flood of often uninterpretable estimates that arose from this research activity produced a new demand for structural models to organize the data and facilitate comparisons of estimates across studies. Jorgenson's (1963) powerfully simple model of the demand for capital showed the value of structural econometric models for interpreting data and focusing empirical research on essential economic ideas. In one variable — the user cost of capital — Jorgenson was able to summarize the essential features of an important theory

[48] Amemiya (1985) and Hsiao (1986) provide comprehensive discussions of these new methods.

[49] Milton Friedman in a private conversation recalled the days at the NBER in the 1930s when extensive formal discussions were required to justify the substantial cost of adding an additional regressor to an equation.

of investment. The simplicity and the elegance of his empirical analysis contrasted sharply with the long and uninterpretable tables of regression coefficients reported in studies of investment prior to Jorgenson's. Although serious empirical objections were raised against his theory, its interpretability was never questioned.

Jacob Mincer's study of female labor supply (1962) was equally influential in demonstrating the power of a simple structural model to organize evidence and predict phenomena. By introducing wage and income effects into the empirical analysis of labor supply, he was able to reconcile empirical anomalies between time-series and cross-section estimates of female labor supply, and was able to explain the time series of female labor supply using a parsimonious, economically interpretable framework.

The interpretability of structural estimates and the value of structural models in constructing counterfactuals spurred the structural estimation movement in the 1970s and early 1980s.[50] The "Lucas Critique" suggested that properly specified structural models would avoid the problem of parameter drift in estimated macro models. This instability was especially pronounced in the estimation of the "Phillip's Curve" in the late 60s and early 70s. The quest for policy invariant structural equations in macroeconomics stimulated a flood of theoretical papers and a trickle of empirical ones.[51] A parallel movement emerged in microeconomics where research on labor supply, selection bias, the returns to schooling, the causal effect of unions on wages, and the determinants of unemployment flourished.

The theoretical distinctions introduced in the post-Cowles Commission structural estimation literatures are of fundamental lasting importance. They demonstrate what can be extracted from the data provided sufficient prior knowledge is assumed. These models extend the Cowles paradigm by specifying preferences, technology, and expectations and by addressing the identification of structural parameters in a variety of empirical settings.

At the same time, the empirical track record from the structural research program is not impressive. Computational limitations have hindered progress. Typically only the

[50] This dating is only rough. In industrial organization, structural modelling began to be practiced in the late 1980s and is an active frontier area of research in that field.

[51] See the collection edited by Lucas and Sargent (1981).

simplest of structural models can be estimated and these are often rejected in the data, a point vigorously emphasized by advocates of the VAR program. Few structural models have been estimated that systematically account for the vast array of economic policies that confront the agents being analyzed. In practice, analysts seeking policy-invariant structural parameters to conduct policy evaluation typically have ignored the policies in place when estimating the structural parameters. Moreover, the functional forms that facilitate structural estimation are often inconsistent with the data.

Euler equation estimation methods developed in the 1980s and 1990s became popular because they avoid the computational cost of full structural estimation and enable analysts to estimate one structural parameter under weaker assumptions than are required to estimate full structural equation systems. In this sense, the Euler equation methods are more robust than full system structural estimates.[52] By estimating a few parameters of empirical models that are not capable of generating forecasts of the levels of the variables, it is possible to avoid embarrassing confrontations with the data. At the same time, pursuit of this research program was tantamount to abandonment of the Cowles program for macro policy evaluation.[53]

In many quarters of economics, the evidence from structural econometric models is held to be unreliable. An influential book by a leading empirical labor economist, H. Gregg Lewis (1986), is typical of the response to structural econometrics by many empirical economists. Reviewing estimates of the structural (causal) effect of unionism on wages that correct for self-selection into union status — the average effect of unionism on those who were unionized holding their observed and unobserved personal characteristics fixed — Lewis found a wide range of estimates from different econometric methods, and different conditioning variables. This variability led him to dismiss structural equation methods as unreliable, and to advocate a return to more familiar least squares methods that apparently generate more robust estimates. Similar findings were reported for the estimates of structural labor

[52] However, the range of estimates produced in the Euler equation literature itself vary widely. See *e.g.* Browning and Lusardi (1996).

[53] Summers (1991) presents a vigorous critique of the Euler equation research program and questions its contributions to economic knowledge.

supply parameters. (Killingsworth, 1983, Table 4.3)

In fairness to the structural approach, some of the variability reported in the empirical literature arises from the imposition of false over-identifying restrictions about distributions and unobservables that could have been tested and relaxed, but typically were not. Part of the variability in the estimates emphasized by Lewis (1986, Table 4.2) arises from the variation in the conditioning variables and choices of combinations of functional forms and identifying assumptions. Different conditioning sets and different estimators define different structural parameters.[54] This is an intrinsic feature of structural models, not of econometric methods. Lewis was frustrated by the failure of a purely empirical approach to solve a causal problem, and the difficult task of working backward from the assumptions embedded in any particular empirical study to explain why its results are different from any other.

Lewis' evidence on the sensitivity of estimates of causal parameters to alternative identifying assumptions does not demonstrate the impossibility of estimating a causal effect of unions on wages. Assuming that the functional form of population wage equations is known, the only safe conclusions that can be drawn from his study are (a) selection bias is an empirically important problem in estimating the causal effect of unions on wages, and (b) different identifying assumptions produce different estimates of the causal parameters.[55] Only if selection bias is *not* a problem will different methods for solving the selection problem estimate the same causal effect. The only way to "solve" the problem of parameter variability reported in the empirical literature in labor economics is to develop different economic models to evaluate the plausibility of the different assumptions used to generate the estimates. No purely empirical method is available. Agreeing to report only least squares estimates establishes a convention that is easy to follow but that evades the stated causal question.

Lewis' reaction to the variability of structural estimates under different identifying as-

[54] This problem is discussed in Heckman (1990) and Heckman, LaLonde, and Smith (1999).

[55] These are my conclusions and not his. Note that the assumption of correct specification of population wage functions was not questioned by Lewis. In labor economics, it is presumed that the "Mincer equation" is the correct functional form for an earnings equation. See Mincer (1974).

sumptions is typical of the reaction to structural models by many economists. In application, structural econometricians often impose onto the data many assumptions not intrinsic to the economics of the problem for the sake of computational convenience. In many applications of the method, and of VAR methods, there is an appeal to formal statistical methods that has "black box" feature and the numbers produced using them are often not perceived to be transparent or easily replicable. The quest for transparency underlies all of the recent research programs in econometrics, although there is no agreement over what constitutes transparency.

Methods developed in nonparametric econometrics and sensitivity analysis in statistics in principle reduce some of this arbitrariness. Nonparametric identification analyses reveal whether causal distinctions are made solely on the basis of distributional assumptions or on the basis of functional form assumptions. They open up the black box of parametric econometrics to establish the sources of identification of economic models.[56]

Those promoting calibration as a method for determining the parameters of structural models emphasize the value of securing transparent estimates for structural equations and further emphasize the conditional — model dependent — nature of empirical knowledge. Like Lewis, they reject the black box features of structural estimation. Unlike Lewis, they seek to resolve empirical ambiguities by developing the economic theory. Unlike the VAR econometricians, the calibrators focus on a few main correlations and means and build explicit models to account for them. Unlike the structural econometricians, and the VAR econometricians, formal statistical models are not used by this group and the goodness-of-fit of models is deemphasized as a model evaluation criterion.

A third response to the structural equation program is that of the more empirically oriented natural experiment movement, which is loosely allied with, and takes its inspiration from the social experiment movement. Like the calibrators, practitioners of this approach seek transparent estimates obtained from simple econometric methods because

[56] For example, Flinn and Heckman (1982) establish the nonparametric non-identifiability of the stationary job search model of unemployment relative to data on accepted wages and unemployment durations and demonstrate the extreme sensitivity of structural estimates derived from this model to alternative distributional assumptions.

transparency and simplicity promote replicability and understanding. Like the VAR econometricians, this group attempts to ground its activities in the data, although it does not use time series methods, and tends to deemphasize formal discussions of econometric methods. A hallmark of this approach is the quest for plausible sources of external variation to solve identification problems, with the ideal being a randomization that does not alter the causal relationship being studied. Like the nonparametric econometricians, advocates of natural experiments distinguish what is in the data from what has to be added to it to obtain estimates of causal parameters. The emphasis is on recovering causal parameters and not on evaluating the effects of new policies or the welfare consequences of policies already in place. I discuss each of these research programs in turn.

3.1. Understanding the Limitations of the Data: Bounding and Sensitivity Analysis

In response to the variation of the estimates produced from alternative specifications of parametric econometric models applied to the same data, econometricians have increasingly emphasized separating aspects of an estimation that require out-of-sample extrapolation or imposition of difficult-to-justify functional forms, from aspects of an estimation that are based on sample information and are nonparametrically identified.

Exploring the limits of identification, this line of work also considers the restrictions placed on ranges of parameter values when models are not formally identified. The paradigm for this line of work goes back to the research of Marschak and Andrews (1944) who present ranges of estimates for the structural parameters of firm production functions that are consistent with biased least squares estimates.[57] Marschak and Andrews work within the context of an under-identified parametric model. More recent work emphasizes more nonparametric approaches to perform this type of bounding and sensitivity analysis.

Recent prototypes for this separation of conjectural from factual information are studies by Smith and Welch (1986), Glynn, Laird and Rubin (1986), Holland (1986) and Rosenbaum (1995) who analyze the standard selection problem that is at the heart of Lewis' problem of recovering the union-nonunion wage differential. Smith and Welch consider

[57] Wright (1934) presents ranges of estimates for under-identified path coefficient models.

identification of means. Glynn, Laird and Rubin, Holland, and Rosenbaum consider identification of entire distributions.

To illustrate these ideas in the simplest possible setting, let $f(W \mid D = 1)$ be the density of outcomes (*e.g.* wages) for persons who work ($D = 1$ corresponds to work). Suppose that we know $\Pr(D = 1 \mid Z)$ where Z is a vector of determinants of work. Hence we know $\Pr(D = 0 \mid Z)$. Missing is $f(W \mid D = 0)$ *e.g.* wages of non-workers.[58] In order to estimate $E(W \mid Z)$, Smith and Welch (1986) use the law of iterated expectations to obtain

$$E(W \mid Z) = E(W \mid D = 1, Z) \Pr(D = 1 \mid Z) + E(W \mid D = 0, Z) \Pr(D = 0 \mid Z).$$

To estimate the left-hand side of this expression, it is necessary to obtain information on the missing component $E(W \mid D = 0, Z)$. Smith and Welch propose and implement bounds on $E(W \mid D = 0, Z)$ *e.g.*

$$W_{\mathsf{L}} \leq E(W \mid D = 0, Z) \leq W^{\mathsf{U}},$$

where W^{U} is an upper bound and W_{L} is a lower bound.[59] Using this information, they construct the bounds

$$E(W \mid D = 1, Z) \Pr(D = 1 \mid Z) + W_{\mathsf{L}} \Pr(D = 0 \mid Z) \leq E(W \mid Z)$$
$$\leq E(W \mid D = 1, Z) \Pr(D = 1 \mid Z) + W^{\mathsf{U}} \Pr(D = 0 \mid Z).$$

By doing a sensitivity analysis, they produce a range of values for $E(W \mid Z)$ that are explicitly dependent on the range of values assumed for $E(W \mid D = 0, Z)$.[60]

Glynn, Laird, and Rubin (1986) present a sensitivity analysis for distributions using Bayesian methods under a variety of different assumptions. Letting F denote a distribution, by the law of iterated expectations,

$$F(W \mid Z) = F(W \mid D = 1, Z) \Pr(D = 1 \mid Z) + F(W \mid D = 0, Z) \Pr(D = 0 \mid Z).$$

[58] In Lewis' unionism problem, there are two sets of missing data: the wages that nonunion workers would earn if they were union workers and the wages that union workers would earn if they were nonunion workers. I analyze the selection problem for one case only to simplify the exposition.

[59] In their problem there are plausible ranges of wages which dropouts can earn.

[60] Later work by Manski (1995) and Robins (1989) develops this type of analysis more systematically. Balke and Pearl (1997) present an elegant approach to the problem of incorporating multiple sources of prior information using linear programming methods.

Sensitivity analysis using alternative values of $F(W \mid D = 0, Z)$ is performed to determine a range of values of $F(W \mid Z)$. Holland (1986) and Rosenbaum (in a series of papers starting in the early 1980s and summarized in 1995) consider more classical sensitivity analyses that vary the ranges of parameters of models.

The objective of the Smith-Welch bounds and the Bayesian and classical sensitivity analyses is to clearly separate what is known from what is conjectured about the data, and to explore the sensitivity of reported estimates to the assumptions used to secure them. Work on nonparametric identification by Heckman (1990), Heckman and Honoré (1990) and Heckman and Smith (1998) examines nonparametric assumptions required to identify selection models from infinite samples. This work establishes "what is in the data" and what has to be imposed on it to make different causal distinctions.

Some of the analysis presented in the recent literature is nonparametric although in practice, high-dimensional nonparametric estimation is not feasible. However, feasible parametric versions of these methods run the risk of imposing the false parametric structure used in estimating the structural models that is so deeply criticized by advocates of this approach. These methods also depend critically on the choice of conditioning variables Z. In principle, all possible choices of the conditioning variables should be explored, especially in computing bounds for all possible models, although in practice this is never done. If it were, the range of estimates produced by the bounds would be substantial.[61]

A fully nonparametric approach in deriving estimates or bounds for causal parameters

[61] The term "nonparametric" is actually something of a misnomer. In finite samples, all nonparametric methods are parametric. The choices of the parameterization are dictated by properties possessed by these functions as samples become infinite, and not necessarily by their economic interpretability in small samples of the kind that are available — the criterion adopted in the older parameteric structural literature. The semiparametric econometric literature relaxes some of the parametric assumptions of parametric models, while retaining the remaining parametric structure. Semiparametric estimation methods are more suited for samples of the size available to economists and for that reason are more widely used. In many empirical applications, simple parametric methods applied to analyze the data are often more convincing, and accurate, than nonparametric or semiparametric methods based on arbitrary bandwidths chosen by mechanical mean square error criteria. Ironically, the variety of nonparametric approximation schemes produces a variety of small sample approximations all of which converge to the same functions but which often exhibit very different properties in the small samples to which they are applied. See the discussion in Heckman, LaLonde and Smith (1999).

is unlikely to produce useful results, although semiparametric models to compute bounds and estimate causal models are beginning to be used on a wide scale. Explorations of bounds with simple functional forms of the sort advocated and implemented by Marschak and Andrews (1944) are likely to be more informative.[62] Unless some assumptions about functional forms are maintained, as in the parametric or semiparametric literatures, economic data are unlikely to be very informative about causal parameters, especially when there are many possible conditioning variables.

3.2. The Calibration Movement

The development of dynamic economic theory and computable general equilibrium theory in the mid 1960s posed a serious challenge to structural econometrics. Unlike the simple static theory of demand, or the simple Keynesian models that defined the structural econometrics of the 1940s and 1950s, recursive dynamic theory does not typically produce simple functional forms for estimating equations. Vector autoregression models like (7) and (8) rarely capture the dynamic economic theory in an explicit way. (Hansen and Sargent, 1980, 1991). Problems of estimating the parameters of large-scale static general equilibrium models are equally formidable.

Many of these new dynamic models produce a rich dynamics with qualitative properties that depend critically on values assumed by parameters. Thus, while dynamic theory needs good empirical input to determine which qualitative properties are empirically relevant, at the same time there are computational problems that make estimation of the required parameters a formidable task. This computational challenge is so great that a new field of computational economics, concerned solely with simulation of these models, has opened up. (See, *e.g.* Amman, Kendrick and Rust, 1996).

In an attempt to anchor the theory in data, and to use the theory to produce counterfactual policy simulations, calibration has come into use on a wide scale. In static general equilibrium models, the practice is to pick simple functional forms (typically Cobb-Douglas) and fit one cross-section exactly using share data. (See Dawkins, Srinivasan and Whalley,

[62] Bounds for the regression coefficients in errors in variables models are developed by Klepper and Leamer (1984).

1999). In the real business cycle models, parameters are picked from time-series averages to match the parameters of simple models that produce growth steady states (see, *e.g.* Cooley and Prescott, 1995). In other branches of this literature, calibrators pick parameters from micro studies. This practice has been criticized because the source micro models are often based on assumptions about the economic environments that are incompatible with the assumptions of the calibrated macro model. (Hansen and Heckman, 1996).

This research program emphasizes interpretability of the estimates in terms of economic models and subjects the calibrated models to rigorous internal consistency checks. Certain features of data (elements of the source space S) become the focus of attention while others are ignored. Overall fit of the model is deemphasized and formal testing programs are explicitly rejected. There is no attempt to account for the time-series correlations in the fashion of the VAR econometricians. Current approaches to producing the estimates for this class of causal models are casual, although in its defense, the econometrics is transparent and often avoids the black-box features of standard structural econometrics and the problems of determining the appropriate representation of economic time series that plague the VAR approach. Some have argued that the calibrators are transparently wrong. When real business cycle models fit by calibration have been subject to rigorous empirical testing of the sort used in VAR econometrics, they have performed poorly in terms of goodness-of-fit criteria. (Watson, 1993).

Given the weak empirical foundations for these models, it is not surprising that the policy counterfactuals based on them are controversial and few outside the subfield take the estimates of the welfare consequences of policies produced by this line of research very seriously. At the same time, the models are intellectually interesting frameworks and demonstrate what is logically possible. Anchoring them in data, however loosely, gives them some plausibility. Routinely performed sensitivity analyses reveal which parameters are crucial and which are not important (Auerbach and Kotlikoff, 1987). These findings serve to direct the attention of empirical analysts toward estimating economically important parameters.

At the time of this writing, it is unclear whether the calibration movement is a transitional stage that will be replaced by a more formal structural econometric research program, or a permanent fixture of the economics landscape. In the life cycle of any class of general equilibrium models, it is likely that calibration will be the first stage of empirical implementation and that formal structural estimation will follow for the subclass of calibrated models that attract the most professional attention.

3.3 The Natural Experiment Movement

A third response to the perceived failure of structural econometrics is the natural experiment movement. Somewhat controversially, I include in this group advocates of social experiments. This group, like the nonparametric econometricians, the statisticians who advocate sensitivity analysis, and the calibrators, seeks transparency in its use of empirical evidence. The movement is organized around the principle of finding good instruments (in the sense of section 2.2) as inputs into a standard instrumental variable estimator of causal parameters that is simple to estimate and easy to replicate. Randomization is often held to be the ideal instrument.[63]

Unlike the calibrators, members of this movement are much less concerned about estimating structural parameters derived from economic theory. The emphasis is on identifying causal parameters, intuitively defined. By not tying the empirical work to any specific economic model, the evidence produced from this approach appears to be relevant to a wider class of models although the link to any specific model is weaker. Simultaneity and self-selection are recognized as potentially important problems, but simple solutions to them are sought using transparent, replicable, instrumental variables methods or difference-in-differences methods.[64] Given the emphasis on intuitively defined causal parameters as opposed to structural parameters, this group eschews formal presentations of economic theory to motivate empirical models, as is favored by the calibrators, or explicit statements

[63]Social experiments are sometimes alleged to be the "gold standard" for causal inference. For one discussion of the problems arising in implementing social experiments and for the interpretation of social experiments as instruments, see Heckman, LaLonde and Smith (1999).

[64]Differences-in-differences are a form of instrumental variables method. See Blundell, Duncan and Meghir (1998) and Heckman, LaLonde and Smith (1999). Both papers question the validity of the method for evaluating many social programs.

of identifying assumptions which characterize the analyses of nonparametric econometricians.[65]

Applications of this approach often run the risk of producing estimates of causal parameters that are difficult to interpret. Like the evidence produced in VAR accounting exercises, the evidence produced by this school is difficult to relate to the body of evidence about the basic behavioral elasticities of economics. The lack of a theoretical framework makes it difficult to cumulate findings across studies, or to compare the findings of one study with another. Many applications of this approach produce estimates very similar to biostatistical "treatment effects" without any clear economic interpretation.[66] The less explicit discussion about economic models and sources of identification that characterizes much of the research conducted in this style sometimes creates the impression that the reported empirical evidence is more robust than the empirical evidence produced from research programs that adopt a more explicitly qualified approach in interpreting evidence.

This group, like the VAR econometricians, stresses empirical credibility, intuitive plausibility, and replicability. Analysts working within this paradigm have produced an important body of factual knowledge. Like the nonparametric econometricians, this group emphasizes the limits of empirical knowledge. Counterfactual questions about the effects of new policies, never previously implemented, are viewed as empirically impossible to answer. Explanation is emphasized over prediction. When prediction is performed, it is done using interpolation or extrapolation of existing estimates rather than the formal methods advocated in the econometric policy evaluation literature.

4. The Adequacy of the Cowles Research Program as a Guide to Learning from Data

At the time the Cowles analysts were developing their body of theory, classical statistical

[65] However, the choice of instruments is highly controversial. In an important paper, Bound and Jaeger (1999) question the validity of exclusion restrictions used to justify the instruments employed in the recent literature on estimating the return to schooling in the assumed presence of ability bias.

[66] However, some structural approaches are subject to the same criticism. Thus in the structural union wage literature, the "treatment" of unionism is often estimated rather than the direct economic effects of unionism on technology, labor supply and factor substitution.

inference as embodied in the work of Neyman and Pearson (1933) had a virtual intellectual monopoly in statistics. That approach emphasized the formal testing of models arrived at through *a priori* means. Haavelmo (1944) synthesized the Cowles approach with the Neyman-Pearson approach in his Nobel-Prize-winning foundational essay.

Few empiricists now embrace the Cowles research program advanced by Haavelmo that remains the credo of most structural econometricians and is implicitly advocated in most econometrics textbooks. The Haavelmo program is a vision of induction on causal parameters produced from well-defined structural economic models derived from explicitly articulated axioms. This approach to empirical analysis and model testing seems foreign to many empirically oriented economists who favor more interaction between models and data than is envisioned in the Haavelmo paradigm. This difference in approach to empirical analysis is a major dividing line between structural econometricians and most other empirical economists.

Classical statistics separates the act of constructing a model from the act of verifying it. Two fundamental assumptions underlie this approach. The first assumption is that a set of consequential facts (the source space S) is fixed in advance of conducting an empirical analysis or in advance of writing down formal economic models to describe the available data. The second assumption is that the set of models that live in model space M is fixed in advance of looking at the data. The task of constructing theoretical models is assumed to be divorced from the task of using data to test them.

Separation between model formation and data analysis does not describe much empirical research activity in economics or in any other scientific field. Yet it is a cornerstone of classical statistics and the Popper (1959) falsificationist program which is the official paradigm of econometrics and classical statistics. The Haavelmo program offers a very rigid vision of empirical analysis. It ignores the interactive nature of the empirical learning process that moves between theory and data, a process that is central to the act of creating empirical knowledge. It is not informative about what steps to take in response to the rejection of a model, nor about the implications of any response for the interpretation of test statistics used in subsequent tests.[67]

The set of available models used to analyze a given body of data is never fixed in advance of looking at it. Inspection of the data usually suggests new models and the new models may in turn suggest collecting new data — or more careful examination of old data including addition of "extra" variables to the source space S — in the light of new theoretical insights. The Haavelmo program does not describe this empirical learning process. Nor does a Bayesian version of it. Neither captures the act of discovery of new models not previously contemplated that are suggested by analysis of the data or the insight that a model may produce about new sources of data to test it.[68] While there are serious problems in using the data that suggest a theory to test that theory, even more important problems arise from refusing to learn from the data in revising economic models.

Sample reuse vitiates standard testing procedures, and reported t values do not convey the information traditionally assigned to them. Bayesian methods are more flexible in this regard but like the classical methods they do not allow for surprise (i.e. new discoveries previously thought to be impossible or not thought about prior to seeing the data).[69] Only tests on fresh data with different distributions of forcing variables provide convincing verification of a model.[70]

In many areas of economics there are few precisely formulated models that are known

[67] Friedman (1953, p. 12, footnote 11) was an early critic of the Haavelmo program. He claimed that the choice of a class of models often forced the conclusions of an empirical study. Morgan (1990) presents an illuminating discussion of the failure of the Cowles group to produce a model for learning from the data.

[68] The program of model testing and "encompassing" advocated by Hendry and Richard (1982) appeals to the Haavelmo program and suffers from the same criticism. It proposes *a priori* specification of a large class of models and use of classical significance tests to test down from a general model. This influential research program does not account for non-nested hypotheses and does not account for learning about new models not previously contemplated (and placed in the *a priori* encompassing specification). Tests conducted within the Hendry and Richard program are sensitive to the order in which they are performed.

[69] However, Bayesian methods applied to events with positive prior probabilities allow for surprise in the sense that posteriors can differ greatly from priors on specific outcomes. In this sense Bayesians have a formal method for measuring surprise.

[70] Replications of the same model on the same type of data (*i.e.* data with the same distributions of variables) are unlikely to be informative. One can replicate the same misspecified results in each instance. Only if analysts use models on data sets with different distributions of external forcing variables will differences in models due to misspecification be detected.

in advance of looking at the data. Usually a lot of vague *a priori* notions seem equally plausible as theoretical points of departure. In such settings, identification analyses are necessarily informal because the theory is not tightly specified and the VAR and natural experiment research programs may be good starting points for initial data explorations.

The econometric paradigm of the identification problem and the notion of a precise theory constructed in advance of the evidence do not apply to areas of social science where knowledge is not settled. It may describe a mature science built on numerous previous empirical studies conducted under other research paradigms where empirical regularities about a phenomenon have accumulated. In this regard the vision of empirical research in economics conducted within the paradigm of the Haavelmo program is a very limited one. Econometricians operating within the Haavelmo paradigm too easily forget that *a priori* theories specified in advance of looking at the data are often just condensations of accumulated empirical knowledge acquired using crude empirical methods.

Leamer's anthropology of econometric practice (1978) and the recent study of the use of "tacit knowledge" in econometric practice by Magnus and Morgan (1999), demonstrate that in practice the Haavelmo program of specifying models in advance of the data is rarely used although test statistics are reported as if it had been used. While interest in causal models and causal questions motivates empirical studies, the rules for induction to generate empirical causal models that were prescribed by Haavelmo and the Cowles group are typically ignored.[71]

No successful mechanical algorithm for discovering causal or structural models has yet been produced, and it is unlikely that one will ever be found.[72] At the same time, it is unlikely that the quest for a mechanical algorithm for determining causality from data will ever be abandoned.[73] The tension between the use of tacit knowledge and formal

[71]One contributing factor was the emergence of rival schools of statistical inference and the breakdown in the consensus about the appropriate way to do empirical research.

[72]This debate has a counterpart in the debate in the artificial intelligence community over whether machines can think. (See Dreyfus and Dreyfus, 1986).

[73]See Glymour and Cooper (1999) for a variety of algorithms for mechanically producing causal parameters from data.

algorithmic methods is likely to be a permanent feature of empirical research in economics. It arises because in most empirical studies there is always more knowledge about a problem being studied than appears in the sampling distributions of the measured variables being analyzed or in well-specified Bayesian priors. The best empirical work in economics uses economic theory as a framework for integrating all of the available evidence, tacit and algorithmic, to tell a convincing story.

As documented by Leamer (1978), creative empirical work is often presented as if it has been conducted within the Haavelmo paradigm, while in fact, it is not. The format of the Haavelmo program is merely used as a reporting style to accord with the official rhetoric of econometrics.[74]

5. Conclusions And a Vision of the Future

In the smoke of battle over the "correct" way to do empirical research in economics, it is easy to lose sight of the important and enduring contributions that twentieth century econometrics has made to knowledge. Its fundamental contributions include establishing formally that causality is a property of a model, that many models may explain the same data and that assumptions must be made to identify causal or structural models. It extended nineteenth century notions of causality by recognizing the possibility of interrelationships among causes. Econometrics clarified the conditional nature of causal knowledge and the impossibility of a purely empirical approach to analyzing causal questions. The information required to evaluate economic policies and to forecast the effects of new policies was carefully delineated.

Economists respond differently to these universally agreed-upon principles. Some embrace a style that emphasizes the conditional nature of causal knowledge while others embrace a style that deemphasizes it. All agree that identification of structural or causal relationships is a difficult task. The apparent empirical failure of many structural research efforts, developed in the 1970s and 1980s, to produce credible estimates of economic parameters and causal relationships motivates current research programs. A desire for clearer, more transparent, sources of identification of causal effects is a common goal.

[74] See McCloskey (1985) on the important role of rhetoric in economic discourse.

Disagreements arise over the interpretation of what constitutes transparency, reporting styles, the identifying assumptions that different groups are willing to make, and the emphasis placed on linking empirical work to explicit economic models. Some of these differences arise from differences in the intended audiences for the empirical work based on these paradigms. Evidence that is convincing in public policy forums may not be convincing to professional economists trained in the use of modern statistical methods. In many public policy settings, precise statements of identifying conditions and qualified presentations of evidence are unwelcome.

The different approaches to empirical research in economics have much to learn from each other. Structural methods will be more widely accepted if sensitivity analyses are conducted, the consequences of functional form and distributional assumptions are investigated, and nonparametric or semiparametric methods are used. With the continuing decline in computer costs, such sensitivity studies will become feasible. The natural experiment movement will gain a wider following if it becomes more integrated with economics. It will produce a cumulative body of knowledge if economic theory is used to guide the choice of estimating equations and to report estimates.[75] Similarly, social experiments that use experimental variation to identify economic parameters rather than program-specific treatment effects are more likely to produce cumulative knowledge.[76]

The calibration movement will produce more credible empirical estimates if it draws on the estimates produced from a more data-sensitive structural econometrics movement and a more economics-sensitive natural experiment movement, if it uses microestimates

[75] The analyses of Card (1999) and Heckman and Vytlacil (1998) demonstrate how simple economic models of the returns to schooling clarify the benefits and limitations of instrumental variables methods. In particular, Heckman and Vytlacil show that Card's implicit assumption that tuition costs together with foregone earnings are not a cost of schooling is critical to the successful application of the standard form of the method of instrumental variables to estimating the rate of return to schooling. Bound and Jaeger (1999) raise serious questions about the exclusion restrictions endorsed by Card. Blundell, Duncan and Meghir (1998) demonstrate the value of economic models in interpreting policy parameters estimated by difference-in-differences methods. Rosenzweig and Wolpin (1980) demonstrate the power of natural experiment twins data to test an otherwise unidentified economic model of fertility.

[76] The original goal of the social experimentation movement was to recover structural parameters to perform econometric policy evaluation. See Orcutt and Orcutt (1968) and Cain and Watts (1973). See the discussion in Heckman, LaLonde and Smith (1999).

obtained from empirical models consistent with the macro-general equilibrium frameworks and if it subjects calibrated models to the rigorous time-series tests advocated by VAR econometricians. The bounding and sensitivity analysis movement is likely to be more influential if it relies on explicit economic models and uses economically interpretable models to conduct semiparametric bounding and sensitivity analyses.[77]

It is far from obvious that these divergent empirical practices will ever converge to a common practice for extracting causal parameters and conducting policy analysis. The Cowles Commission presented a bold program of causal analysis and policy evaluation that has proved difficult to operationalize in practice. Later developments in dynamic general equilibrium theory and game theory have not made structural estimation any easier. Responding to this difficulty is a major source of anxiety and tension in empirical economics. Those most strongly motivated by its theoretical vision are less easily discouraged by the empirical failings of the Cowles program. Those most strongly motivated to describe the economic world, rather than to explain it, or predict the effects of new policies, are acutely aware of the empirical limitations of highly structured models.

Some of the disagreement that arises in interpreting a given body of data is intrinsic to the field of economics because of the conditional nature of causal knowledge. The information in any body of data is usually too weak to eliminate competing causal explanations of the same phenomenon. There is no mechanical algorithm for producing a set of "assumption free" facts or causal estimates based on those facts. In reaching this understanding, economists part company with statisticians who, as a group, still fail to understand this important lesson of twentieth century econometrics, and advocate purely empirical approaches for determining causal relationships.[78] The only certain routes for eliminating some of the disagreement among economists who maintain different identify-

[77]The analysis of Marschak and Andrews (1944) is a good parametric paradigm. A more recent analysis is that of Viverberg (1993).

[78]There is considerable irony in the observation that some econometricians in the recent "treatment effect" literature have turned for guidance to statistics to obtain frameworks for interpreting causal laws while statisticians and experts in artificial intelligence such as many writing in Glymour and Cooper (1999) have turned to Cowles econometrics for clear definitions of causality within a model.

ing assumptions are through collecting better data and stating differences in identifying assumptions more clearly.[79] Economic theory as a framework for interpretation and synthesis is an inseparable part of good empirical research in economics. These are major lessons of twentieth century econometrics.

[79] Throughout this essay, I have taken it as self-evident that the introduction of new data sources has greatly enriched economic knowledge. To cite only a few major contributors, the research of Kuznets on national income (1937), and economic growth (1973), Summers and Heston (1991) on international comparisons and Griliches (1995) on research and development have had a major influence on our profession.

6. References

Amemiya, Takeshi, *Advanced Econometrics*, (Cambridge, MA: Harvard University Press, 1985).

Amman, Hans, David Kendrick and John Rust, *Handbook of Computational Economics: Volume 1*, (Amsterdam: North Holland, 1996).

Auerbach, Allen and Lawrence Kotlikoff, *Dynamic Fiscal Policy*, (Cambridge: Cambridge University Press, 1987).

Balke, Alexander and Judea Pearl, "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association,* XCII, (1997), 1171-1176.

Basmann, Robert, "Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal," in Michael Intriligator and David Kendrick, eds., *Frontiers of Quantitative Economics*, Chapter 4, (New York: North Holland, 1974).

Barros, Ricardo Paes, "Minimal Just-Identifying Assumptions in Nonparametric Selection Models," unpublished manuscript, (Rio de Janeiro: IPEA, 1988).

Billingsley, Patrick, *Probability and Measure, Second Edition,* (New York: John Wiley, 1986).

Bjerkholt, Olav, "Ragnar Frisch and the Foundation of the Econometric Society and *Econometrica*," 26-57 in Steinar Strom, ed., *Econometrics and Economic Theory in the 20th Century*, (Cambridge: Cambridge University Press, 1998).

Blalock, Hubert, *Causal Inferences in Nonexperimental Research*, (Chapel Hill: University of North Carolina Press, 1964).

Blundell, Richard, Alan Duncan and Costas Meghir, "Estimating Labor Supply Response Using Tax Reforms," *Econometrica*, LXVI, (1998), 827-862.

Bound, John and David Jaeger, "Do Compulsory Schooling Attendance Laws Alone Explain the Association Between Quarter of Birth and Earnings, forthcoming in S. Polachek, ed., *Research in Labor Economics*, (1999).

Bound, John, Charles Brown and Nancy Mathiowetz, "Measurement Error in Survey Data," in James Heckman and Edward Leamer, eds., *Handbook of Econometrics: Volume 5*, (Amsterdam: North Holland, 2000).

Browning, Martin and Anamaria Lusardi, "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, XXXIV, (1996), 1797-1855.

Brunner, Karl, *Problems and Issues in Current Econometric Practice,* (Columbus, OH: College of Administrative Science Monograph AA-6, 1972).

Cain, Glenn and Harold Watts, "Summary and Overview," in Glenn Cain and Harold Watts, eds., *Income Maintenance and Labor Supply*, (Chicago: Markham, 1973).

Card, David, "The Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics: Volume 3*, (Amsterdam: North Holland, 1999).

Chipman, John, "The Contributions of Ragnar Frisch to Economics and Econometrics," in Steinar Strom, ed., *Econometrics and Economic Theory in the 20th Century*, (Cambridge: Cambridge University Press, 1998).

Christ, Carl, "History of the Cowles Commission 1932-1952," *Economic Theory and Measurement*, (Chicago: Cowles Commission For Research in Economics, 1952), reprinted in *Econometrics, Macroeconomics and Economic Policy: Selected Papers of Carl Christ*, (Oxford: Elgar, 1995).

Christiano, Larry, Martin Eichenbaum and George Evans, "Monetary Policy Shocks: What Have We Learned and to What End?" in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, (Amsterdam: North Holland, 1999).

Cooley, Thomas and Edward Prescott, "Economic Growth and Business Cycles," in Thomas Cooley, ed., *Frontiers of Business Cycle Research,* (Princeton: Princeton University Press, 1995).

Cox, David Roxbee, *The Design of Experiments,* (New York: John Wiley, 1959).

Dawkins, Christina, T.N. Srinivasan and John Whalley, "Calibration" in James Heckman and Edward Leamer, eds., *Handbook of Econometrics: Volume 5*, (Amsterdam: North Holland, 1999).

Domencich, Thomas and Daniel McFadden, *Urban Travel Demand*, (Amsterdam: North Holland, 1975).

Dreyfus, Hubert and Stuart Dreyfus, *The Power of Human Intuition and Expertise in the Era of the Computer*, (New York: Free Press, 1986).

Epstein, Roy, *A History of Econometrics,* (Amsterdam: North Holland, 1987).

Fair, Ray, *Testing Macroeconometric Models,* (Cambridge: Harvard University Press, 1994).

Fisher, Ronald Aylmar, *Design of Experiments,* (New York: Hafner, 1935).

Fisher, Franklin, *The Identification Problem in Economics,* (New York: McGraw Hill, 1966).

Flinn, Christopher and James Heckman, "New Method for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, XVIII, (1982), 115-168.

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

Frisch, Ragnar, "Propagation Problems and Impulse Problems in Dynamic Economics," in *Essays in Honor of Gustav Cassell,* (London: Allen and Urwin, 1933).

Galles, David and Judea Pearl, "An Axiomatic Characterization of Counterfactuals," *Foundations of Science*, III, (1998), 151-182.

Geweke, John, "Inference and Causality in Economic Time Series Models," in Zvi Griliches and Michael Intriligator, eds., *Handbook of Econometrics, Volume II*, (Amsterdam: North Holland, 1984).

Glynn, Robert, Nan Laird and Donald Rubin, "Selection Modeling vs. Mixture Modeling with Nonignorable Response," in Howard Wainer, ed., *Drawing Inferences from Self-Selected Samples,* (New York: Springer-Verlag, 1986).

Glymour, Clark and Gregory Cooper, *Computation, Causation and Discovery,* (Cambridge, MA: MIT Press, 1999).

Goldberger, Arthur, *Econometric Theory,* (New York: John Wiley, 1964).

_____, "Structural Equation Methods in the Social Sciences," *Econometrica*, XL, (1972), 979-1001.

Granger, Clive, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, XXXVII, (1969), 424-438.

Granger, Clive and Paul Newbold, *Forecasting Economic Time Series,* (New York: Academic Press, 1977).

Griliches, Zvi, "R&D and Productivity: Econometric Results and Measurement Issues," in Paul Stoneman, ed., *Handbook of the Economics of Innovation and Technological Change*, (Oxford: Basil Blackwell, 1995), 52-89.

Haavelmo, Trygve, "Methods of Measuring the Marginal Propensity to Consume," *Journal of the American Statistical Association*, XLII, (1947), 105-122.

______, "The Probability Approach in Econometrics," *Econometrica*, XII, Supplement (1944).

Hamilton, James, *Time Series Analysis,* (Princeton: Princeton University Press, 1994).

Hansen, Lars and James Heckman, "The Empirical Foundations of Calibration," *Journal of Economic Perspectives*, X, (1996), 87-104.

Hansen, Lars and Thomas Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, II, (1980), 7-46.

______, "Two Difficulties in Interpreting Vector Autoregressions," in *Rational Expectations Econometrics*, (Boulder, Colorado: Westview Press, 1991).

Heckman, James, "Varieties of Selection Bias," *American Economic Review*, LXXX, (1990), 313-318.

______, "Econometric Evaluation of Social Programs," in James Heckman and Edward Leamer, eds., *Handbook of Econometrics*, *Vol. V*, (Amsterdam: North Holland, 2000).

Heckman, James and Bo Honoré, "The Empirical Content of the Roy Model," *Econometrica*, LVIII, (1990), 1121-1149.

Heckman, James and Richard Robb, "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in Howard Wainer, ed., *Drawing Inference From Self-Selected Samples*, (New York: Springer-Verlag, 1986), 63-107.

Heckman, James and Jeffrey Smith, "Evaluating the Welfare State," in Steinar Strom, ed., *Econometric and Economic Theory in the 20th Century*, (Cambridge: Cambridge University Press, 1998).

Heckman, James and Edward Vytlacil, "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling when the Return is Correlated with Schooling," *Journal of Human Resources*, XXXIII, (1998), 974-987.

Heckman, James, Robert LaLonde and Jeffrey Smith, "The Economics and Econometrics of Active Labor Market Programs," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics: Volume 3*, Chapter 31, (Amsterdam: North Holland, 1999).

Hendry, David and Jean-Francis Richard, "On the Formulation of Empirical Models in Dynamic Econometrics," *Annals of Applied Econometrics*, XX, (1982), 3-34.

Holland, Paul, "A Comment on Remarks by Rubin and Hartigan," in Howard Wainer, ed., *Drawing Inferences From Self-Selected Samples*, (New York: Springer-Verlag, 1986), 149-152.

Hsiao, Cheng, *Analysis of Panel Data*, (Cambridge: Cambridge University Press, 1986).

Imbens, Guido and Joshua Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, LXII, (1994), 467-476.

Jorgenson, Dale, "Capital Theory and Investment Behavior," *American Economic Review*, LIII, (1963), 247-259.

Klein, Lawrence, "Single Equation vs. Equation System Methods of Estimation in Econometrics," *Econometrica*, XXVIII, (1960), 866-871.

Klein, Lawrence and Arthur Goldberger, *An Econometric Model of the United States, 1929-1952,* (Amsterdam: North Holland, 1955).

Klepper, Steven and Edward Leamer, "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, LII, (1984), 163-183.

Killingsworth, Mark, *Labor Supply*, (Cambridge: Cambridge University Press, 1983).

Knight, Frank, *Risk, Uncertainty and Profit,* (New York: Houghton Mifflin Co, 1921).

Koopmans, Tjalling and William Hood, "The Estimation of Simultaneous Linear Economic Relationships," in William Hood and Tjalling Koopmans, eds., *Studies in Econometric Method*, (New York: John Wiley, 1953).

Koopmans, Tjalling and Olav Reirsol, "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, XVIII, (1950), 165-181.

Kuznets, Simon, *National Income and Capital Formation,* (New York: National Bureau of Economic Research, 1937).

_____, "Modern Economic Growth: Findings and Reflections," *American Economic Review*, LXIII, (1973), 247-258.

Lancaster, Kelvin, *Consumer Demand: A New Approach,* (New York: Columbia University, 1971).

Leamer, Edward, *Specification Searches*, (New York: John Wiley, 1978).

_____, "Vector Autoregression for Causal Inference?" *Carnegie-Rochester Conference Series on Public Policy,* XXII, (1985), 255-303.

Lewis, H. Gregg, *Unionism and Relative Wages Effects: A Survey,* (Chicago: University of Chicago Press, 1986).

Liu, Ta-Chung, "Under Identification, Structural Estimation and Forecasting," *Econometrica*, XXVIII, (1960), 855-865.

Lucas, Robert, "Econometric Policy Evaluation: A Critique," *The Phillips Curve and Labor Markets,* I, (1976), Carnegie Rochester Conference Series on Public Policy, 19-46.

Lucas, Robert and Thomas Sargent, *Essays on Rational Expectations and Econometric Practice,* (Minneapolis: University of Minnesota Press, 1981).

Magnus, Jan and Mary Morgan, *Methodology and Tacit Knowledge,* (New York: John Wiley, 1999).

Mann, Henry and Abraham Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, XI, (1943), 173-220.

Manski, Charles, *The Identification Problem in the Social Sciences,* (Cambridge: Harvard University Press, 1995).

Marini, Margaret and Burton Singer, "Causality in the Social Sciences," *Sociological Methodology*, XVIII, (1988), (Washington, DC: American Sociological Association), 347-410.

Marschak, Jacob, "Economic Measurements for Policy and Prediction," in William Hood and Tjalling Koopmans, eds., *Studies in Econometric Method*, (New York: John Wiley, 1953), 1-26.

Marschak, Jacob and William Andrews, "Random Simultaneous Equations and the Theory of Production," *Econometrica*, XII, (1944), 143-205.

Marshall, Alfred, *Principles of Economics, Ninth Edition*, (London: McMillan, 1920), 36.

McCloskey, Donald, *The Rhetoric of Economics*, (Madison: University of Wisconsin Press, 1985).

Mincer, Jacob, "Labor Force Participation of Married Women," in H. Gregg Lewis, ed., *Aspects of Labor Economics,* (Princeton: Princeton University Press, 1962).

_____, *Schooling, Experience and Earnings,* (New York: Columbia University Press, 1974).

Morgan, Mary, *The History of Econometric Ideas,* (Cambridge: Cambridge University Press, 1990).

Morganstern, Oskar, *On the Accuracy of Economic Observations*, (Princeton: Princeton University Press, 1950).

Neyman, Jerzy, "Statistical Problems in Agricultural Experiments," Supplement to the *Journal of the Royal Statistical Society*, II, (1935), 107-180.

Neyman, Jerzy and Egon Pearson, "On the Testing of Statistical Hypotheses in Relation to Probability A Priori," *Proceedings of the Cambridge Philosophical Society*, XXIX, (1933), 492-510.

Orcutt, Alice and Guy Orcutt, "Experiments for Income Maintenance Policies," *American Economic Review*, LVIII, (1968), 754-772.

Orcutt, Guy, "Towards Partial Redirection of Econometrics," *Review of Economics and Statistics*, XXXIV, (1952), 195-200.

Pearl, Judea, "Graphs, Causality and Structural Equations Models," *Sociological Methods and Research*, XXVII, (1998), 226-249.

Phillips, Allan W., "Estimation of Systems of Difference Equations with Moving Average Disturbances," in Alan Ray Bergstrom, ed., *Stability and Inflation*, (New York; Wiley, 1966), 181-199.

Popper, Karl, *The Logic of Scientific Discovery*, (London: Hutchinson, 1959).

Quandt, Richard, *The Demand For Travel,* (Lexington, MA: D.C. Heath, 1970).

Reirsol, Olaf, "Confluence Analysis by Means of Instrumental Sets of Variables," *Arkiv for Mathematik Astronomi och Fysik*, XXXII(A), (1945), 1-119.

Robins, Jamie, "The Analysis of Randomized and Non-randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in Lee Sechrest, Howard Freeman, and Albert Mulley, eds., *Health Service Research Methodology: a Focus on AIDS*, (Washington, DC: U.S. Public Health Service, 1989), 113-150.

Rosenbaum, Paul, *Observational Studies*, Chapter 4, (New York: Springer-Verlag, 1995).

Rosenzweig, Mark and Kenneth Wolpin, "Testing The Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, XLVIII, (1980), 227-240.

Rubin, Donald, "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, VI, (1978), 34-58.

Shapiro, Matthew and Mark Watson, "Sources of Business Cycle Fluctuations" in Stanley Fischer, ed., *NBER Macroeconomics Annual,* (Cambridge, MA: MIT Press, 1988).

Simon, Herbert, A., "On The Definition of the Causal Relation," *Journal of Philosophy*, 49, (1952), 517-527.

Sims, Christopher, "Money, Income and Causality," *American Economic Review*, LXII, (1972), 540-552.

____, "Exogeneity and Causal Ordering in Macroeconomic Models" in C. A. Sims, ed., *New Methods In Business Cycle Research: Proceedings From A Conference.* (Minneapolis: Federal Reserve Bank of Minneapolis, 1977).

_____ , "Macroeconomics and Reality," *Econometrica,* XLVIII, (1980), 1-48.

_____ , "Are Forecasting Models Usable for Policy Analysis?" *Quarterly Review of the Federal Reserve of Minneapolis*, Winter, (1986), 2-16.

Slutsky, Eugene, "The Summation of Random Causes as the Source of Cyclic Processes," *Econometrica,* V, (1937), 105-146.

Smith, James and Finis Welch, "Closing the Gap: Forty Years of Economic Progress for Blacks," (Santa Monica, CA: Rand Corporation, 1986).

Strom, Steinar, *Econometrics and Economic Theory in the 20th Century*, (Cambridge: Cambridge University Press, 1998).

Summers, Lawrence, "The Scientific Illusion in Empirical Macroeconomics," *Scandanavian Journal of Economics*, XCII, (1991), 129-148.

Summers, Robert and Alan Heston, "The Penn World Table (Mark V): An Expanded Set of International Comparisons," *Quarterly Journal of Economics*, CVI, (1991), 327-368.

Tinbergen, Jan, "Bestimmung und Deutung Von Angesbotscurven," *Zeitshrift Für Nationalökonomies, I*, (1930), 669-679, reprinted and translated in David Hendry and Mary Morgan, eds., *The Foundations of Econometric Analysis*, (Cambridge: Cambridge University Press, 1995).

Viverberg, William, "Measuring the Unidentified Parameter of the Roy Model of Selectivity," *Journal of Econometrics*, LVII, (1993), 69-90.

Wallis, Kenneth, "Econometric Implications of the Rational Expectations Hypothesis," *Econometrica*, XLVIII, (1980), 49-74.

Watson, Mark, "Measures of Fit for Calibrated Models," *Journal of Political Economy*, CVI, (1993), 1011-1041.

White, Hal, *Estimation, Inference and Specification Analysis for Econometricians,* (New York: Academic Press, 1987).

Wright, Philip Green, *The Tariff on Animal and Vegetable Oils,* (New York: MacMillan, 1928).

Wright, Sewall, "The Method of Path Coefficients," *Annals of Mathmatical Statistics*, V, (1934), 161-215.

Zellner, Arnold and Franz Palm, "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics*, II, (1974), 17-54.