CAUSAL PARAMETERS AND POLICY ANALYSIS IN ECONOMICS: A TWENTIETH CENTURY RETROSPECTIVE*

JAMES J. HECKMAN

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.

I. 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 database. 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 clari-

---

* This research was supported by NSF 97-09-873 to the University of Chicago. I have benefited from comments received in discussions with Jeff Biddle, Jennifer Boobar, Ray Fair, Edward Glaeser, Claudia Goldin, Lars Hansen, Lawrence Katz, Steven 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].

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

The Quarterly Journal of Economics, February 2000

fied 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 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 [Marshall, 1961, p. 36].

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

Defining causality within a model is relatively straightfor-

---

[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 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.

ward when the causes can be independently varied.[4] 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 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 to outputs. 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

---

[4] 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.

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 [1972,1977,1980,1986] 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 seventies and eighties 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 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 in its 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 abandoned by adherents of this research program. In addition, the data summaries reported by VAR econometricians are often not transparent, and the choice of an appropriate data summary requires knowledge of multivariate time series methods. Hence, the time series data summaries produced by this approach often have a black-box quality about them, and judgments 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 oversimplification, 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 oversimplification 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.

II. 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 then 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.

II.1. Causal Parameters

Within the context of a well-specified 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 $F: R^N \rightarrow R^1$ as

$$Y = F(X_1, \ldots, X_N), \tag{1}$$

where $X = (X_1, \ldots, X_N)$ 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.[5] These causes are the primitives of the relevant

---

[5] 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: $P_X: A \rightarrow X$; $P_Y: B \rightarrow Y$. 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), 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

$$\frac{\partial Y}{\partial X_j} = F_j(X_1, \ldots, X_j, \ldots, X_N)|_{X=x}. \tag{2}$$

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.[6] A variety

---

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). This definition can be modified to apply only to certain subsets, and not all A. 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 $P_Y(P_X^{-1}(C) \cap B)$ is "true," i.e., in some sense depicts reality. This more general definition does not require that functions connecting causes to effects be specified.

[6] 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 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

of causal mechanisms can be contemplated even in this simple 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.[7] 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.[8] 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.[9] The domain D is sometimes just one point as a consequence of the properties of a model, as I demonstrate below.

---

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 $\tilde{X} = (X_2, \ldots, X_J)$, assuming that a local implicit function theorem is satisfied (so in particular $\partial\varphi/\partial Y = (\partial g/\partial X_1)^{-1}$, where $\partial g/\partial X_1 \neq 0$). 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 $(X_2, \ldots, X_J)$ at prespecified values, one would invert F to obtain $X_1 = M(Y, X_2, \ldots, X_J)$, assuming that $\partial F/\partial X_1 \neq 0$. M is a causal function associated with the conceptual thought experiment that $Y, X_2, \ldots, X_J$ 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.

[7] A formal definition of global variation independence is that the domain of D is the Cartesian product $X_1 \times X_2 \times X_3, \ldots, \times X_N$, where X_i is the domain of X_i and there is no restriction across the values of the X_i. When the X terms satisfy this restriction, they are termed "variation free." The local version imposes this requirement only in neighborhoods.

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

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

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.[10] When (1) is separable in X, we can write it as

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

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. The paradigm for this analysis was a model of market demand and supply:

$$Q^D = Q^D(P^D, Z^D, U^D) \qquad \text{Demand} \tag{3}$$

$$Q^S = Q^S(P^S, Z^S, U^S) \qquad \text{Supply,} \tag{4}$$

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

---

[10] The most direct way is to define X_1 as a treatment indicator and to define $Y_{x_1} = F_{x_1}(X_2, \ldots, X_N)$ 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. Imbens and Angrist [1994] present a precise formulation of the Rubin model. The statistical models ignore the constraints across potential outcomes induced by social interactions and by resource constraints, i.e., the potential restrictions on D. Heckman and Vytlacil [2000] discuss the relationships among population treatment effect parameters, structural equations models, and causal models.

determined and are called external variables.[11] 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.

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, $Q = Q^D = Q^S$, and $P = P^D = P^S$. If (P,Q) is uniquely determined as a function of the Z and U, the model is said to be "complete" [Koopmans and Hood 1953].

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.[12] 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^D, Z^D, and U^D, and interprets (4) as the conditional expectation of Q^S given P^S, Z^S, and U^S.[13] 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, $E(Q^D|P^D,Z^D,U^D) \neq E(Q^D|P^D,Z^D)$. In the early investigations of causal models, and most models in current use, linear equation versions of (3) and (4) were used, so

---

[11] 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.

[12] 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.

[13] The justification for this is given in footnote 6.

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 $P = P^D = P^S$ and $Q = Q^D = Q^S$, so there is at least one market equilibrium. Other restrictions like positivity of (the diagonals of) $\partial Q^S/\partial P^S$ (supply increasing in price) or negativity of (the diagonals of) $\partial Q^D/\partial P^D$ (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^D,Z^S,U^D,U^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^D, Z^S,U^S, U^D, are changed. More formally, under completeness, we can obtain the reduced forms:

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

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

The concept of an externally specified variable is a model-specific notion. It entails specification of (5a) and (5b) as causal relationships in the sense of (1) to replace (3) and (4) when $Q^D = Q^S$ and $P^D = P^S$. In a fully nonparametric setting, this requires that the variables on the right-hand sides have no functional restrictions connecting them.[14] 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).[15]

---

[14] 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).

[15] Formally, the support of (Z^D,Z^S) is assumed to be the same for all values of (U^D,U^S). In this section the Z^D and Z^S enter symmetrically with U^D and U^S so

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 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_e^S be a variable excluded from Z^D and for notational simplicity assuming only a single market,

$$\frac{\partial Q^D}{\partial P^D} = \left(\frac{\partial Q}{\partial Z_e^S}\right) \Big/ \left(\frac{\partial P}{\partial Z_e^S}\right),$$

where the right-hand side expressions come from (5a) and (5b).[16] Defining Z_e^D comparably,

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

Under these conditions, we can recover the price derivatives of (3) and (4) even though an equilibrium restriction connects $P^D = P^S$. 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

---

we should also require that the support of (U^D, U^S) is assumed to be the same for all values of (Z^D, Z^S) or, more generally, we might require that all variables be variation free in the sense of footnote 7.

[16] Proof: Differentiate (3) with respect to Z_e^S to obtain

$$\frac{\partial Q^D}{\partial Z_e^S} = \frac{\partial Q^D}{\partial P^D}\frac{\partial P^D}{\partial Z_e^S}.$$

Using equilibrium values ($P^D = P^S = P$), substitute from (5a) to obtain $\partial P^D/\partial Z_e^S = \partial P/\partial Z_e^S$ and from (5b) to obtain $\partial Q^D/\partial Z_e^S = \partial Q/\partial Z_e^S$. Assuming that $\partial P/\partial Z_e^S \neq 0$, we obtain

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

If there are several Z_e^S variables that satisfy the stated conditions, each defines the same causal parameter.

the relationship in question).[17] 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."

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 internal variables via (5a) and (5b)). 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.[18] These direct and indirect effects play a crucial role in path analysis developed by Wright [1934] and widely used in sociology (see Blalock [1964]).[19] The direct causal effects are called structural. Both direct and indirect effects are causal, and are defined by well-specified variations.

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 (or causal) 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 a 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

---

[17] 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.

[18] The required compensation for the excluded variables may not be achievable and depends on the curvature of the functions, the magnitude of the change in the included Z, and the support of the excluded Z.

[19] 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.

definition of a structural parameter and the separation between the concept of a structural (or causal) effect from the concept of a structural parameter.[20]

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

Functional relationships among variables that are invariant to variations in excluded 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. 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.[21]

The limited role of the time-series tests for causality within articulated causal dynamic models is discussed by Hansen and

---

[20] 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.

[21] 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.

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

II.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.[22] 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 ($P^S = P^D$ and $Q^S = Q^D$). 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.[23] 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.[24]

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.[25] 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

---

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

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

[24] Morgan [1990] discusses early attempts to solve this problem using ad hoc statistical conventions.

[25] This framework is based on my interpretation of Barros [1988].