Why Functional Form Matters: Revealing the Structure in Structural Models in Econometrics

Damien Fennell†‡

This paper argues that econometricians' explicit adoption of identification conditions in structural equation modelling commits them to read the functional form of their equations in a strong, nonmathematical way. This content, which is implicitly attributed to the functional form of structural equations, is part of what makes equation structural. Unfortunately, econometricians are not explicit about the role functional form plays in signifying structural content. In order to remedy this, the second part of this paper presents an interpretation of the functional form based on Herbert Simon's definition of causal order. This begins to set out just what the functional form of structural equations represents.

1. Introduction. Econometricians use structural models to make inferences about the underlying 'structure' of the economy. Their aim is to identify functional relationships between variables of interest that are robust to certain changes and interventions. The ambition is that these models, once suitably confirmed, will allow policy makers to intervene and effectively control part of the economy.

Intuitively these structural models are hypotheses about the causes in the economy. And yet, until relatively recently, econometricians have shied away from causal language when discussing structural equation models.[1]

†To contact the author, please write to: Centre for Philosophy of Natural and Social Science, London School of Economics, London WC2A 2AE, United Kingdom; e-mail: d.j.fennell@lse.ac.uk.

‡I would like to thank Nancy Cartwright and attendants at UCSD Graduate Seminar 2006 for helpful comments. I also want to thank the AHRC for supporting the research for this paper.

1. Hoover (2004) has an interesting historical account of the absence of causal language in economic journals during the twentieth century, where he postulates that this avoidance of causal language is in part due to the influence of Humean skepticism about causes.

Philosophy of Science, 74 (December 2007) pp. 1033–1045. 0031-8248/2007/7405-0040$10.00
Copyright 2007 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

Unfortunately, this has obscured the meaning of structural models in econometrics.[2] It is unfortunate because it leaves ambiguity in the interpretation and use of models that are widely used for economic policy making. For better and more accountable policy making, ideally one should be clear about what economic structural models mean and how their use is (to be) justified.

Analysing structural models in econometrics should not simply be of interest to econometricians and those influenced by their policies. It is also interesting for philosophers of science, not only because it make explicit what some scientists do (here econometricians), but because it brings out an alternative approach to modelling causal relations—one which permits the representation of equilibrium relations in a distinctive way, that is, using static simultaneous equations.

This paper attempts to shed light on the meaning of structural models, first, by bringing out the significance of the functional form of structural equations as a bearer of structural content, and second, by presenting a brief outline of a semantics for the functional form based on Herbert Simon (1953). In the first part, a commitment by econometricians to a significant functional form, one which is used to signify the structural meaning of the model, is made explicit by showing that this is necessary for their explicitly adopted identification conditions to have substance and be useful for their epistemic purposes. The second part of the paper complements this by setting out, in outline, a semantics for the functional form of structural equations.

2. The Identification Conditions. To clarify structural models, it helps to look at the methods econometricians use, because these methods provide hints as to how their structural equations models should be interpreted. Particularly revealing method is the treatment of 'the problem of observational equivalence'. This is the problem of how to distinguish between two or more different structures that are consistent with the same set of observations.[3] The econometricians' solution to this problem is captured in 'the identification conditions'. The problem of observational equivalence is important, because careful scrutiny of the associated identification conditions reveals the necessity for a nonmathematical interpretation of the functional form of structural equations models.

The most influential problem of observational equivalence in economics

2. Though this is true of econometrics during the second half of the twentieth century, the founders of econometrics did use causal language. See, e.g., Haavelmo (1944); Frisch ([1938] 1995); and Simon (1953).

3. Philosophically, this is a problem of the underdetermination of hypotheses by observation.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

is that of measuring supply and demand curves from observations of prices and quantities transacted for a good. As Mary Morgan recounts, in early work on supply and demand concerns were raised that measurements did not measure what was claimed (1990, 163–168). The difficulty was and remains that observed prices and quantities of goods transacted result from supply and demand structures acting together. This presents a challenge of how to identify properties of the demand curve without mistakenly mixing in properties of the supply curve (or vice versa). In response to this problem, conditions were developed to make clear when one measured properties of the individual demand and supply curves. These 'identification conditions' are the econometricians' solution to the problem of observational equivalence.

To clarify this, consider the following example of a supply and demand model based on Koopmans (1949, 127). In the model q and p respectively denote the equilibrium quantity and equilibrium price of a good, both of which are observable. There are two error terms u_1 and u_2 covering the factors not explicitly represented in the demand and supply equations; these are unobservable. Taken together, the supply and demand equations determine the values of q and p given values of the error terms. The structural equations are

$$q = \alpha_1 p + u_1 \cdots \text{demand}$$

$$q = \alpha_2 p + u_2 \cdots \text{supply.}$$

Note that the functional form of the equations is taken as known, but the coefficients, α_1 and α_2, are unknown and it is these that the econometrician wants to estimate in order to infer the precise demand and supply equations.

Suppose that there are two price and quantity observations known to be generated by the structure represented by the two equations above in which the observed (q_2, p_2) lies 'northeast' of (q_1, p_1), as in Figures 1 and 2 below. Given the unobservability of the error terms, one does not know whether the change between observations was due to a shift in u_1, a shift in u_2 or both.[4] Then there are two possibilities to explain the change:[5] (1) a shift in both equations (u_1 and u_2) and (2) a shift in just the demand equation (in u_1). The two possibilities are represented in Figures 1 and 2.

4. Since q and p are assumed to be determined by the model, changes in q and p are assumed to follow from changes in the error terms, u.

5. I have not included a case where only the supply equation shift occurs as possible, because, given that (q_2, p_2) lies northeast of (q_1, p_1), this would require that the demand equation were upward sloping which I assume not to be the case in line with conventional assumptions about demand.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

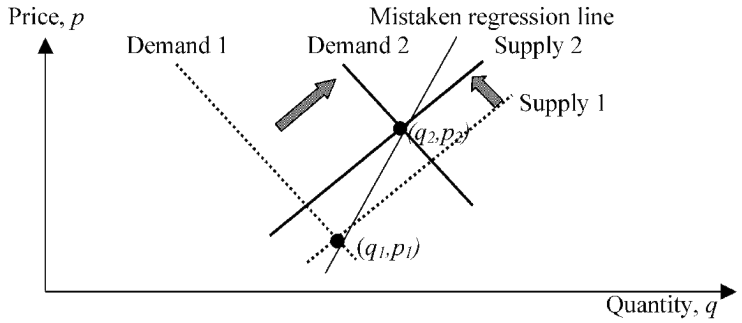


Figure 1. Possibility 1: Shift in both supply and demand (in u_1 and u_2).

Figure 1 shows how the two observations would have been generated if both the supply and demand equations had shifted, while the second shows how they would have been generated if the change was due to a shift only in the demand equation. Since changes in the u's are not observable, there is no way of knowing which of the two possibilities it is. In short, regressing a line through the observed points for price and quantity does not reliably measure either the supply or the demand equation. Regressing in this way will almost certainly yield a measurement of a mixture (i.e., linear combination) of the two equations, as shown in Figure 1. This is the problem of observational equivalence, one cannot distinguish between the underlying structure and those that mix up the demand and supply equations.

Yet, if one somehow knew that only the demand equation shifted, as shown in Figure 2, then one *could* regress on the observations to measure the supply equation because, as the Figure 2 illustrates, since in this case the regression line fits 'on' the supply equation. This second possibility suggests a way out of the problem: attributing shifts to particular equations.

How can this be done? The trick is to refine the model so that a factor that was previously covered in the error term is now explicitly modelled. Specifically, suppose in the example that an observable factor, income (say), i, previously impounded in the error term, u_1, figures in the demand equation but not the supply equation. The refined structural equations are

$$q = \alpha_1 p + \alpha_3 i + u_1' \cdots \text{demand}$$

$$q = \alpha_2 p + u_2 \cdots \text{supply.}$$

https://doi.org/10.1086/525642 Published online by Cambridge University Press

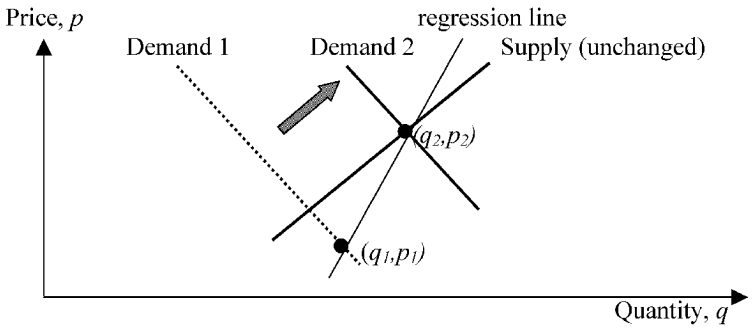


Figure 2. Possibility 2: Shift only in demand (in u_1).

Provided that the u's are not correlated with income,[6] then identification of the supply equation is possible. To see this, consider the case where the error terms are uncorrelated with one another,[7] and imagine that income changes to six different levels while the error terms change in an uncorrelated way with each other and income. Figure 3 shows how observations might be generated given shifts in income and the error terms. The graph shows the supply and demand curves at different instances when observations were made. The different positions of the demand curves are due to changes in income and the error term (u_1') while differences in the supply curves are due to changes in the error term (u_2).

Also shown in Figure 3 is how a regression line fitted through the observations should have a slope that is a good fit to the slope of the supply equation.

This approach to identification has been generalised by econometricians to give a necessary and sufficient condition for identifiability of an equation in a structural equations model where all variables are observable. The condition is:[8]

> **The Rank Condition**. Given a system of n linear structural equations for determining n variables, in which all variables are observable, a necessary and sufficient condition to identify all the coefficients in an equation is that the submatrix of coefficients formed from the

6. Stephen LeRoy (2004, 16–17) gives a more general and causal interpretation of this requirement in 'the uncorrelatedness assumption'.

7. This assumption is not necessary for identification. However, it allows a straightforward regression between quantity and price. If the errors are correlated the structural coefficient inferred directly here would be calculated from the reduced form coefficients.

8. For more, see Fisher (1966, 39–41); Gujarati (1995, 657–669); and Maddala (2001, 348–352).

https://doi.org/10.1086/525642 Published online by Cambridge University Press

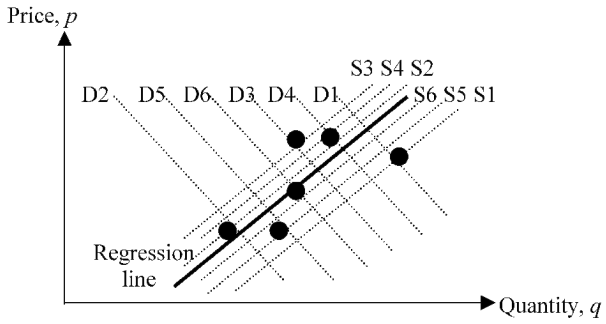


Figure 3. Multiple observations for income changes with error term changes.

columns of the coefficients of the variables excluded from that equation has rank $n - 1$.[9]

This condition appears to solve the problem of observational equivalence, because when it is met one can infer a unique structural equation from observation. However, as shown in the next section, there is an important caveat to this result, which when made explicit shows that identification conditions do not suffice as a solution to the problem of observational equivalence. This caveat in turn shows the need for a commitment by econometricians to a substantive interpretation for the functional form of structural equations.

3. A Caveat: Identifiability Is Relative to Functional Form. The discussion above is a simplified version of the accounts of identification typically found in introductory econometric textbooks. However, these discussions are incomplete. To see why, first note that the solution to the problem of observational equivalence in the example above relied on a variable (income) not appearing in the equation being identified (the supply equation). Likewise, whether the rank condition is met or not depends on the variables excluded from an equation. Therefore, whether or not a structural equation is identifiable depends on where variables appear in the equations. Identifiability of structural equations is relative to the functional form of the equations.

This alone may not raise worries with the econometrician's solution. But concerns arise once one considers that attributing significance to where variables appear in equations marks a departure from the mathematical interpretation of the set of equations. If one reads a set of struc-

9. The rank of a matrix is the dimension of the space spanned by its rows or columns.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

tural equations in a purely mathematical way, then any mathematically equivalent set of equations bears exactly the same content. This means that one can linearly combine the equations in the set to create new sets of equations with the same mathematical content. However, such manipulations change where variables appear in the equations and thus change identifiability properties. Therefore, identifiability of an equation is not preserved under all transformations that preserve mathematical content. Thus, identifiability of equations only makes sense, has meaning, if one attributes content to the equations beyond the standard mathematical interpretation. So, identification piggybacks on an implicit attribution of meaning to the equations that is nonmathematical, that is, an attribution of significance to the mathematically arbitrary functional form of the structural equations.

Moreover, it is easily seen that the nonmathematical interpretation of the functional form is part of what makes the equations structural. Reconsider the income example:

$$q = \alpha_1 p + \alpha_3 i + u_1' \cdots \text{demand}$$

$$q = \alpha_2 p + u_2 \cdots \text{supply}$$

If one transforms the system, replacing the original second equation with the sum of the two equations, one gets a new mathematically equivalent set of equations:

$$q = \alpha_1 p + \alpha_3 i + u_1' \cdots \text{demand}$$

$$q = \frac{1}{2}(\alpha_1 + \alpha_2)p + \frac{1}{2}\alpha_3 i + \frac{1}{2}(u_1' + u_2) \cdots \text{`shmupply'}.$$

In changing the functional form of the equations, not only is the supply equation no longer identifiable, but the nonmathematical interpretation of the equations has been changed. In the original system the supply equation represents structure that determines supply of a good. It says something about firms, and what firms do when faced with different market situations. In contrast, the new 'shmupply' equation tells us no such thing since it mixes up demand and supply structure. In the original set of equations, there was substantive content in the claim that income appeared in the demand equation but not in the supply equation, namely that income plays a role in how much is consumed, but not in how much the supplier decides to produce or decides to charge and so on. Income appearing in one equation but not another is not an accident of the way the equations were written. The functional form plays a key role in sig-

https://doi.org/10.1086/525642 Published online by Cambridge University Press

nifying the structural or causal content of the equations by setting out which factors play role in which part of the structure.

Unfortunately, the role of the functional form in bearing structural content of the model is left implicit in conventional discussions in econometrics. Yet, this implicit content is crucial, not only because the content attributed to it is part of what makes the equations structural, but because this content is necessary for the identification conditions to be properly solutions to the problem of observational equivalence. This is because, by definition of the problem of observational equivalence, to have a solution to the problem one must infer a unique model from observation. But the relativity of identifiability to functional form implies that identifiability ensures a unique model can be inferred from observation provided one limits the candidate models to those that have the same (identifiable) functional form. If one is free to change functional form, then any uniquely inferred model (of any particular identifiable functional form) will be changeable into another model of distinct functional form. In this case, one will have two or more distinct models that are equally good at describing the data, that is, the problem of observational equivalence returns. Therefore, to use the identification conditions to solve the problem of observational equivalence requires knowing the 'right' functional form. Or, more precisely, one must know that the structural features, which the functional form represents, hold.

In practice, econometricians implicitly assume this by taking the functional form as a given in their treatment of the problem of observational equivalence and in their inference of coefficients from data. Given the functional form must be attributed some significance for the identification conditions to be meaningful, it follows that economists assume that something of significance about the structure of the economy is assumed to be known when inferring coefficients for their structural models. But what is the important assumed background knowledge? To answer, it is necessary to investigate what the functional form means.

The earlier shmupply example gives a preliminary answer. It shows that the functional form intuitively carries information about what factors play what role in the underlying structure. Therefore, knowing the functional form amounts to knowing the roles different factors play in the structure. If one adopts this crude answer, then in claiming that the identification conditions are a solution to the problem of observational equivalence, econometricians assume what factors play what roles where in the equation are known.

In summary, the problem with standard econometric discussions of identification is that they do not explicitly discuss the role of the functional form of the structural equations in signifying structural content. Not only does this leave ambiguity over just what the 'structure' in structural equa-

https://doi.org/10.1086/525642 Published online by Cambridge University Press

tions is all about, it also leaves unquestioned, the arguably much more difficult epistemic problem of finding out the right functional form, that is, the right structure that is taken as given when inferring coefficients from observations. Since their solution to the problem of observational equivalence is hostage to this larger epistemic problem, the silence on the issue in their discussions of identification is worrying. To remedy this, one would like to know what the functional form tells us about the structure the equations are meant to represent. Then one can begin to ask how claims to know the functional form of a structural equations model are to be justified. The first of these tasks is attempted next.

4. Revealing Structure—What the Functional Form Represents. Filling in an interpretation of the functional form of structural equations is not too difficult, one need only look at the work of the founders of econometrics, who were more aware about the need to be explicit about what their mathematical models mean. Both Ragnar Frisch (1995) and Trygve Haavelmo (1944), central figures in the development of modern econometrics, wrote about the content of structural equations. Yet, perhaps the most explicit discussion is Herbert Simon's (1953) who defines a causal order for a set of structural equations.

In his paper, Simon formally defines the causal order for a set of equations. To obtain it, one first distinguishes between two types of variables in the model, internal and external.[10] The internal variables are those that are determined by the model (e.g., q and p above), while the external variables (e.g., income, i) and the error terms have values that are taken as given, from outside the model. One then solves for the internal variables one by one using the fewest equations required to solve for them; this stipulates an order for the solution of the internal variables. Any variable used to solve for and solved for prior to another variable causally precedes it. One variable directly causally precedes another if it causally precedes it and if it appears in the same equation as the other variable.[11] The resulting ordering is the causal order. Note that Simon's causal order depends on the functional form since the order of solution depends on the equations in which the variables appear.

10. Strictly speaking Simon uses a distinction between coefficients and variables. I prefer the approach here since it is more in line with econometric practice. Also, the difference is not significant for the purposes of this paper.

11. This is a simplification to keep the discussion manageable, the correct requirement is the causally precedent variable appears in the minimal set of equations by which the causally subsequent variable is solved. See Simon (1953) for details.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

Consider the income example, where one categorizes q and p as internal and i as external:[12]

$$q = \alpha_1 p + \alpha_3 i + u_1' \cdots \text{demand}$$

$$q = \alpha_2 p + u_2 \cdots \text{supply.}$$

Here q and p can only be solved for together in terms of i and the error terms, using both equations. Moreover, since i appears in an equation with q and p, i is a direct cause of p and q. Also, since p and q are both determined together (in the same minimal set of equations) they are 'co-determined'.[13] Thus, the causal order is $i \rightarrow \{p, q\}$. This yields an intuitive result for the example since it makes explicit that income is a cause of equilibrium price and quantity, while equilibrium price and quantity are codetermined, just what one would expect for an equilibrium model of supply and demand in which income has an impact (via demand).

Although Simon's causal order helps to make explicit the content of the functional form of structural equations, it is limited progress because his causal order is merely a formal relation among the variables defined from the functional form of the equations. Despite its 'causal' label, as it stands it says nothing about the content of 'cause'.

Luckily, Simon helps by briefly discussing how variables and equations should be interpreted. External variables are taken to denote factors that are directly controllable by an experimenter or nature, and internal variables taken to denote factors that are indirectly controllable. Equations are taken to denote mechanisms and error terms are taken to denote the joint role of omitted directly controllable factors in a mechanism. The idea is that the experimenter or nature has direct access to the directly controllable factors and is free to change them. Changing these then has an impact on the other indirectly controllable factors, in virtue of the mechanisms that connect the indirectly controllable factors to the directly controllable factors. Under this interpretation, the causal order arises from the joint action of mechanisms, and maps out how changes in a factor will 'in general'[14] lead to changes in other factors. It sets out that changing

12. This categorization is justified by the model being one of supply and demand but *not* income.

13. I leave aside for now the interesting question of how these codetermining relations should be understood. There is an important question as to whether or not this relationship should be considered as one of mutual causation.

14. The 'in general' caveat stems from Simon (1953, 25). It is important since it is possible if coefficients have 'just so' values that one factor may cause another while all its changes cancel out, i.e., that a change in a cause does not lead to a change in its effect.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

a cause 'in general' changes its effects, whereas it is possible to change an effect without changing one of its causes.[15]

Though this gives us some idea of how to interpret the functional form, there is much which is not discussed by Simon. For example, it is also important in his interpretation that mechanisms be invariant to changes brought about by the experimenter or nature. Otherwise, the equations expressing the mechanisms could be completely changed upon intervention and would tell us nothing about what happens to indirectly controllable factors as a result of changes to the directly controllable factors. Also, it is important that the directly controllable factors be independent of each other in the sense that the experimenter must be 'free' to change them.[16] Fleshing out these issues is an important step—one not attempted here—in setting out a clear interpretation of the functional form.

Nevertheless, one can return to the spurious demand-shmupply model to show how Simon's causal order brings out the significance of the functional form. The two models were

$$q = \alpha_1 p + \alpha_3 i + u_1' \cdots \text{demand}$$

$$q = \alpha_2 p + u_2 \cdots \text{supply}$$

and

$$q = \alpha_1 p + \alpha_3 i + u_1' \cdots \text{demand}$$

$$q = \frac{1}{2}(\alpha_1 + \alpha_2)p + \frac{1}{2}\alpha_3 i + \frac{1}{2}(u_1' + u_2) \cdots \text{shmupply.}$$

Reading the first set of equations following Simon, the indirectly controllable equilibrium price and quantity are constrained to take values, by the demand mechanism, in line with the value of the directly controllable income. Similarly, the supply mechanism constrains the indirectly controllable equilibrium price and quantity to take values in line with the values of other factors in the mechanism that are not explicitly modelled. In contrast, the shmupply model has a second mechanism which involves directly controllable income. In this way it is distinct from the supply mechanism, simply because income is not a factor in the supply mechanism. More generally, using the Simon interpretation, changing the functional form changes which factors are in which mechanisms, thus changing

15. Provided the effect has more than one cause.

16. How one fleshes out this concept of independence is debatable. One interpretation, adopted by Hoover (2001), is to assume that independence requires that the directly controllable factors have values that are variation-free; i.e., they can take any value as a group as they can individually.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

the underlying structure modelled. Thus, one cannot change the functional form without changing the structural interpretation. So Simon's interpretation makes explicit what is important about the functional form of the structural equations.

Simon's semantics are a useful beginning for shedding light on what the functional form and, more generally, what structural equations in econometrics mean. Nevertheless, a lot remains to be done. To complete the semantic analysis requires further details, for instance, about mechanisms. What are these? What invariance relations do they satisfy? What counterfactuals do they support? Answering these is important to clarify further what structural models mean.

Finally, I have not discussed the suitability or attractiveness of the Simon semantics for structural models. This is also important to debate. Providing a semantics is not the same as providing a semantics that fits with econometricians' use of their models. Ideally, any proposed semantics should be tested to see if it can clarify the actual use of structural models in economics. I do not attempt this here. Instead, the point of presenting the Simon semantics, is to show how progress can be made in shedding light on the functional form of structural equations models.

5. Conclusion. This paper has attempted to show that the identification conditions in econometrics only makes sense if one attributes a nonmathematical significance to the functional form of the structural equations, that is, to which variables appear in which equations. It has presented work by Herbert Simon that provides a mechanistic interpretation from which one can begin to understand what is meant by the structural equations. In this, Simon's causal order provides a possible missing piece in understanding the functional form of structural equations.

To finish on why work like this matters, it is worth returning to the problem the paper started on, the problem of observational equivalence. Ultimately, this is a problem of inference, how one learns about the structures that govern the economy. In their solution of the problem, econometricians rely on the functional form being known for the model they parameterise. This is an important assumption, and an advantage of the Simon mechanistic interpretation is that it provides the beginnings of a way to make explicit just what the claim to know the functional form amounts to. It amounts to a claim to know what the mechanisms governing the economy are. These claims are central to the econometrician's method. If we are to fully scrutinise their method, it is crucial that we be able to judge the strengths and limits of this knowledge claim. Providing an explicit mechanistic account provides a ground from which to begin to consider epistemically the knowledge claim that is central to econometric method.

https://doi.org/10.1086/525642 Published online by Cambridge University Press

REFERENCES

Fisher, Franklyn (1966), *The Identification Problem in Econometrics*. Huntington, NY: Krieger.
Frisch, Ragnar (1995), "Autonomy of Economic Relations", in David Hendry and Mary Morgan (eds.), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 271–273.
Gujarati, Damodar (1995), *Basic Econometrics*. New York: McGraw-Hill.
Haavelmo, Trygve (1944), "The Probability Approach in Econometrics", *Econometrica* 12: 1–115.
Hoover, Kevin (2001), *Causality in Macroeconomics*. Cambridge: Cambridge University Press.
——— (2004), "Lost Causes", *Journal of the History of Economic Thought* 26: 149–164.
Koopmans, Tjalling (1949), "Identification Problems in Economic Model Construction", *Econometrica* 17: 125–144.
LeRoy, Stephen (2004), "Causality in Economics", in *Causality: Metaphysics and Methods*. Technical Report CTR 20/04, Centre for the Philosophy of the Natural and Social Sciences, London: London School of Economics.
Maddala, Kameswari (2001), *Introduction to Econometrics*. 3rd ed. New York: Wiley.
Morgan, Mary (1990), *The History of Econometric Ideas*. Cambridge: Cambridge University Press.
Simon, Herbert (1953), "Causal Ordering and Identifiability", in *Models of Man*. New York: Wiley, 10–35.

https://doi.org/10.1086/525642 Published online by Cambridge University Press