On Agricultural Econometrics

David A. Bessler

However the agricultural economy is put to-gether, we observe it in two distinct ways: through the lens of experimental data or through the lens of observational data. The types of inferences we make under each are distinct, at least at first glance. Haavelmo (1944) made this point clear in Chapter 2 of his 1944 treatise, "The Probability Approach in Econometrics." He writes (page 14): "The experiments (economists) have in mind may be grouped into two different classes, namely, (1) experiments that we should like to make to see if certain real economic phenomena – when artificially isolated from 'other influences'—would verify certain hypotheses, and (2) the stream of experiments that Nature is steadily turning out from her enormous laboratory, and which we merely watch as passive observers."

Today it is common to label Haavelmo's first class as "econometrics on experimental data" and his second class as "econometrics on observational data." Even before Haavelmo's work, agricultural science was making progress with both experimental and observational data. Ronald Fisher at Rothamsted Experiment Station (Fisher, 1926) and Sewall Wright with the Bureau of Animal Industry of the U.S. Department of Agriculture (Wright, 1921) were developing (seemingly) distinct inference methods on experimental data and observational data, respectively. In this note I elucidate the differences and similarities between experimental and observational inference. Key to this articulation is an understanding of the causal structure that generates the data. In observational studies, one does as Haavelmo suggests and passively observes the consequences of that structure. Under experimental studies, one manipulates that structure. To obtain structure with observational data, one needs to find manipulations in nature that mimic the experiment. To this end, I conclude the article with Ronald Fisher's advice for finding causation in observational studies: "Make your theories elaborate" (Cochran, 1965). Perhaps not along the line Fisher had in mind, I suggest our theory elaboration needs to be about the "causes of our causes."

The article is organized as follows. First, I offer a simple graphical representation on the Inference Problem. Hopefully this representation will allow one to see clearly the fundamental difference between inference with observational data and inference with experimental data. I then consider suggested solutions with observational data motivated by work of Phillip and Sewall Wright over 80 years ago in their invention of instrumental variables. This instrumental variables solution is perhaps too simple in the sense that it assumes we know much about the world that may not actually be the case. I turn then to inference under a representation that more generally represents our human condition: where we do not know all of the variables that generate our observed data. This work on causal inference in the presence of latent variables builds on an instrumental variable presentation. Although in the

David A. Bessler is a professor at Texas A&M University, College Station, Texas.

simple case one (usually) secures consistent estimation by assumption, in the more general case, one can actually test for consistent estimation if instruments are of a particular type.[1]

The Inference Problem

In Figure 1, I present a simple two-panel diagram elucidating the problem of inference under observational and experimental data. In Figure 1A, I represent the causal flow from X to Y. However, both X and Y are influenced by (caused by) a set of unobserved variables, L. In this simple model, ordinary least squares regression of Y on X will yield biased and inconsistent parameter estimates of $\partial Y/\partial X$. Such regression will yield systematically too large or too small of partial derivative estimates; subsequent elasticity estimates (the bread and butter of applied economic policy analysis) will be either too high or too low. Figure 1A is what Pearl (2009) labels "the back door path problem." Now if L (the set of variables L) is observable, the estimation problem would be trivially solved. One need only condition on L (the variables in L) to block the back door path in an ordinary least squares regression of Y on X and L.[2] The estimated parameter associated with X will be an unbiased (and consistent) estimate of $\partial Y/\partial X$. Unfortunately, as the graph is drawn and as the world often presents itself, L is unobservable, either because we do not have theories sufficient to identify (specify) all of the causes of Y or, although theory is clear on all such causes, one or more are unmeasurable.

Consider the many applied economic studies in demand or supply. It is common to rely on *a priori* theory of maximizing behavior (Hicks, 1946; Samuelson, 1947) to derive first-order

conditions from a constrained optimization problem (the constraint being the budget constraint for demand and production technology of supply) where a choice vector of quantities consumed or produced are derived is a function of exogenous prices. This theory is then used to confront observational data on prices, quantities, and money income (sometimes expenditures). It is the general case that unobserved variables are not
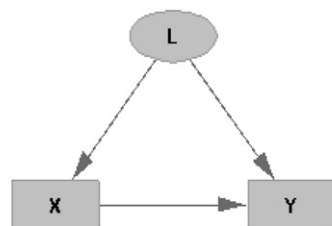
Panel A: The Observational World
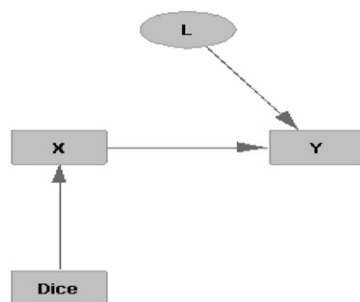
Panel B: The Experimental World

Figure 1. A Simple Representation of the Observational World (A) and an Experimental World (B) (Note: X and Y are observable in both panels. L is unobservable in both panels as well. The arrows from X to Y communicate that X causes Y (or the arrows from L to X and Y A or to Y in B indicate that the latent variable L causes X and/or Y). The graphic representations in this article ("rectangles and arrows") are drawn with TETRAD software developed by members of the Philosophy Department of Carnegie Mellon University found at: www.phil.cmu.edu/projects/tetrad/.)

---

[1] Of course, if one knows the true model that generates the observed data, she or he can dispense with much of what I cover here. I, perhaps dogmatically, insist that such knowledge (of the truth) is rarely, if ever, possible.

[2] The back door path is defined in Pearl (2009, p. 79): a set of variables, L, satisfies the back door criterion relative to an ordered pair of variables X and Y in a direct graph if: (1) no node in L is a descendent of X and (2) L blocks every path between X and Y that contains an arrow into X.

discussed in any detail. The theory used to derive our first-order conditions included no latent variables (unobserved variables), yet few would suggest latents are not present. Indeed, Malinvaud (1980) is quite upfront when he discusses estimation of the parameters of consumption: "Two households with exactly the same income, consumption, liquid assets, etc. will generally behave differently. The most reasonable solution is therefore to take account in our assumptions of the fact that the factors determining consumption are partly unknown to us" (Malinvaud, 1980, p. 50). A rather simple question one might ask is: "Do these unknown variables affect prices, quantities, or income?" Might they play the role of L in Figure 1A? Or might prices themselves be interrelated in a causal network?[3] Might a retailer of product 1 strategically set its price dependent on price of product 2? Or in agricultural supply, might input or output prices be related to one another or to other (unknown) variables?

Econometrics on Experimental Data

The experimentalist addresses this last series of questions with the answer, "We don't know." She or he enters the laboratory to measure the effect of price on quantity (and indeed estimates the parameters holding these two pillars of economics together) through a random assignment experiment.[4] That is, she or he follows Figure 1B. Metaphorically (Pearl, 2009), she or he "cuts" the edge between the latent (L) and X by assigning values of X through a random device (I label this as the dice in Figure 1B). Although in Figure 1A, we passively observe (we "see") X and Y, in Figure 1B, we "do X" by intervening. By construction, assuming no "fat-handed experiments" (Scheines, 2005), ordinary

least squares will provide us unbiased and consistent estimates of $\partial Y / \partial X$ for this experiment.[5]

The economics literature is now rich with experiments providing internally valid answers on demand, supply, information processing, price expectations, etc.[6] Smith (1989) provides a nice (if a bit dated) overview. In the agricultural/resource sector, economists have used the laboratory to study numerous topics, including demand for products (Lusk et al., 2006), food safety (Hayes et al., 1995), the willingness to pay for products (McAdams et al., 2013), the willingness to pay versus the willingness to accept (Horowitz and McConnell, 2002), price expectations (Nelson and Bessler, 1989), and now many other important topics.

Although the laboratory does provide a valid model for inferring causation and generating data for desirable parameter estimates on a particular experiment, extension of the results to other environments is problematic. Here one faces an "external validity" problem as did our predecessors in other agricultural sciences (Borlaug, 1954). In my opinion, based on my teaching of Campbell and Stanley (1966) for now over 30 years and even longer commitment to much of the writing of David Hume (1949), we will never adequately solve the external validity problem (as long as we admit we do not know the "true" model). Campbell and Stanley write, "Whereas the problems of internal validity are solvable within the limits of the logic of probability statistics, the problems of external validity are not logically solvable in any neat, conclusive way. Generalization always turns out to involve extrapolation into a realm not represented in one's sample. Such

---

[3] Wang and Bessler (2006) find retail beef prices cause retail pork and poultry prices in aggregate U.S. data.

[4] In an earlier paper, Bessler and Covey (1993) elucidate the experimental model in terms of the potential outcomes and average causal effect model of Rubin (1974) and its experimental application as offered in Holland (1986). The graphical presentation offered here is equivalent if a bit more transparent (Pearl, 2009).

[5] A "fat-handed" experiment is such that the intervention (random assignment) in Figure 1B also affects Y (directly, not just through X) or through a third variable that affects Y as well as X (Scheines, 2005).

[6] It is probably not necessary to point out to "older readers" there was a reluctance of many in the agricultural economics profession to adopt the experimental paradigm. Perhaps this was because most econometrics texts of the day, say between 1970 and 1990, did not admit discussions of the experiment in any positive light, if at all. For a more modern discussion, the reader is pointed to Angrist and Pischke (2009).

extrapolation is made by assuming one knows the relevant laws" (page 17). Indeed, Pearl and Bareinboim (2012) give graphical conditions for external validity (what they call transportability from the laboratory to the target population) as conditions on the underlying graph structures of the two sets (laboratory and target). The causal diagrams holding in the experiment and in the target population give license or not to transportability across studies. Their work requires what Campbell and Stanley required: "knowledge of the relevant laws."

Economists, in general, and agricultural economists in particular, have made significant strides in the direction of external validity. Lusk and Norwood (2009) and Roth (2002) are examples. It would be interesting to study the graph structure on laboratory experiments that informed the field applications that have actually been successful (and those that have failed, too). Roth (2002) offers us a glance of the complications (without formal graph structures) with respect to market designs for medical markets. My impression is details are important and substantial (as I suppose Borlaug found in his work on wheat a half century earlier).

Econometrics on Observational Data

Although random assignment represents what Angrist and Pischke (2009) call the "most credible and influential resign designs," there are many areas where experimental inquiry (or random assignment) is not possible. Some experiments are not ethical; who would consider random assignment of literacy rates to help us understand literacy's role in the distribution of poverty? Who would propose the study of the influence of conflict on poverty through random assignment of different levels of conflict among nations or neighborhoods? Therefore, we are forced to confront observational data to address many questions of the day. Figure 1A becomes the starting point. As stated previously, if L is observable, the estimation problem is solved by conditioning on the observable L (blocking the backdoor path). If L is unobservable, problems arise. Early work in economics and agricultural economics failed to recognize the problems that the hidden L variable (or set

of variables) had on our parameter estimates. Indeed, Henry Moore (1914) and his positive slope demand curve for pig iron remains today an excellent classroom example of inappropriate interpretation of results from ordinary least squares (Hamilton, 1994). Given one cannot cut the edge between L and X (of Figure 1A), he or she can follow Philip Wright (1928) and attempt to capture variation in X that is unrelated to variation in L through an additional variable Z, as represented in Figure 2A. Here we call Z an instrument, because it has (it is usually assumed) certain desirable properties. In particular, Z causes X but has no direct link (no arrow directly into either L or Y) to L or Y.[7] One can predict X (label this prediction $\hat{X}$) based on variation in Z alone (without the influence of L) in a first-stage regression. In a second stage, one can regress Y on $\hat{X}$.

Instrumental variable (IV) estimation was and continues to be a great idea, the problem being that we assume a lot with Figure 2A. In particular, how does one know that his or her instrument does not itself link to L as in Figure 2B? The conditional correlation between Z (our instrument) and Y given X is nonzero in both panels of Figure 2. If one cannot see or even know of L, how might he or she determine that Figure 2A generates the data rather than Figure 2B?[8] Thinking back to the experiment, might the instrument, Z, induce a "fat-handed" response in that it moves X, as the researcher wants to be the case, but the instrument as well affects L (or is affected by L)?[9] The response of critics to this last question may well be "Get real; of course we can find instruments that don't link to L." However, remember L is a set of variables and may go back

---

[7] See Stock and Trebbi (2003) on instrumental variables, Moore's problem, and the interesting detective work on authorship of what appears to be the first formal presentation of instrumental variables in economics.

[8] There may be special cases in which we are almost surely convinced that Z is not connected to L (Angrist and Krueger, 1991). These cases seem to justify choice of Z through intuition.

[9] The analogy between the random assignment experiment and instrumental variables is strong. With observational data, the instrument serves the same role as the dice in Figure 1B.

Panel A   Panel B

Figure 2. A Simple Representation of a Proper Instrumental Variable (A) and an Improper Instrumental Variable (B) (Note: Here we have the same graph on X, Y, and L as that given in Figure 1A. Only now we represent the existence of an observable variable Z (an instrument). In A, we see no edges connecting Z with L; we label the Z in this graph a proper instrument. In B, there is an edge between Z and L. Here instrumental variable techniques will not work, because any prediction of X (in a first stage) based on Z (call this prediction $\hat{X}$) will not purge X of the influence of L.)

in time before Z and, recalling the Malinvaud quote given previously, we may not know exactly what L consists of. Unless we embrace Laplace's Demon, caution would suggest that Figure 2B is a possibility worth consideration.[10,11]

Recently, strides have been made in inference in the presence of latent variables. Scheines (2005) in particular offers conditions for demonstration of noncausality and causality, both in the presence of latents. If the world is put together according to Figure 3, one can indeed say Y does not cause X.[12] Here

instrument Z influences Y or both are influenced by unobserved latents, L2. Similarly, either X causes Y or both are caused by an unseen set of latents, L1. However, X and Z are independent (correlation is zero). Variables Z, Y, and X are thus held together seeming as "causal inverted forks" (Pearl, 2009). Although such an inference cannot say that it is not latents that are responsible for the links between Z and Y and Y and X, we can say that Y does not cause X. If Y did cause X, the correlation between Z and X would be nonzero, which is not the case. Bryant, Bessler, and Haigh (2009) demonstrate the statistical size and power of such an inference under varying sample size and signal strength.[13]

Finally, Scheines' work does offer conditions for which one can conclude X causes Y. Figure 4 offers a simple graph depicting such. First we need independent instruments Z1 and Z2, which move X or are linked to latents (L1 and L2) that move Z1 and Z2 and X (in terms of correlations, we need $\rho_{Z1,Z2} = 0$, $\rho_{X,Z1} \neq 0$, and $\rho_{X,Z2} \neq 0$). Second we need X and Y to not be independent ($\rho_{X,Y} \neq 0$). Finally, we require the conditional correlation between each instrument and Y given knowledge of X to be zero ($\rho_{Z1,Y| X)} = 0$, $\rho_{Z2,Y| X)} = 0$). The three conditions, along with Gaussian data and faithfulness

---

[10] Much of econometrics before say 1980 seems to have lined up well with Laplace's Demon: "An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes" (Laplace, 1902).

[11] Of course, it goes without saying (perhaps) that the experimentalist does not require the Demon to achieve internal validity as he or she explicitly intervenes to "cut the edge" as in Figure 1B. Only on questions of external validity does he or she dance with the Demon (Campbell and Stanley, 1966).

[12] We need additional conditions of "faithfulness" and "Markov probability factorization" for this argument to hold. Faithfulness, in particular, requires that vanishing correlations are the result of the fact that there is no edge (causation) between Z and X and not the result of cancellations of deeper parameters (see Spirtes, Glymour, and Scheines, 2000)). Furthermore, we assume all variables are Gaussian. If one has heavy tails or excess peakedness, one might consult methods of Hoyer et al. (2008). However, I am unaware of the latter's use under the presence of latents.

[13] Haigh and Bessler (2004) and Yu, Bessler, and Fuller (2007) find inverted forks on innovations in hinterland (HP) and the Gulf of Mexico grain and soybean prices (GP) and associated barge rates (BR): BR → HP ← GP. Their work does not, however, preclude unknown variables lying between BR and HP and GP and HP.
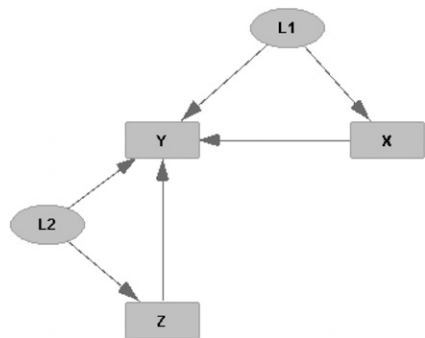
Figure 3. A Case Where We Can Conclude Y Does Not Cause X, Even in the Presence of Latents (Note: Here X, Y, and Z are measured variables. Variable L1 and L2 are unobserved latent variables. X and Z are independent (in a Gaussian world the correlation between Z and X is zero) and Z and Y and Y and X are not independent (in a Gaussian world respective correlations are not zero). Finally, we know that a latent does not lie between Z and X, because the associated correlation is zero.)

Figure 4. A Case Where We Can Conclude X Does Cause Y, Even in the Presence of Latents (Note: Here X, Y, Z1, and Z2 are measured variables. Variables L1 and L2 are unobserved latent variables. Z1 and Z2 are independent (in a Gaussian world the correlation between Z1 and Z2 is zero) and the Z_i and Y and Y and X are not independent (in a Gaussian world these correlations are not zero). The conditional correlations between Y and Z_i, i = 1,2 given knowledge of X are zero. These last two conditions on conditional correlations assure us that no latent lies between X and Y or between Y and L1 or L2.)

(that vanishing correlations are the result of independence and not cancellations of deep parameters), allow us to infer that X does cause Y. The last two conditions on conditional independence of Y and Z_i, i = 1,2, given knowledge of X, allows us to say that a latent (say L_3) does not lie between X and Y. Had such been the case, the conditional correlation between Z_i and Y, given knowledge of X, would be nonzero. Bryant and Bessler (2011) study systems such as that represented in Figure 4 under varying signal strength and sample size.

Concluding Comments

Some 60 years ago when asked what could be done in observational studies to go from association to causation in observational studies, Ronald Fisher suggested, "Make your theories elaborate" (as described and quoted in Cochran, 1965). My suggestion (not necessarily Fisher's) is that our theoretical elaboration needs to be with respect to the "causes of our causes." Furthermore, this elaboration in theories will be required in both experimental and observational studies (the former for external validity; see the
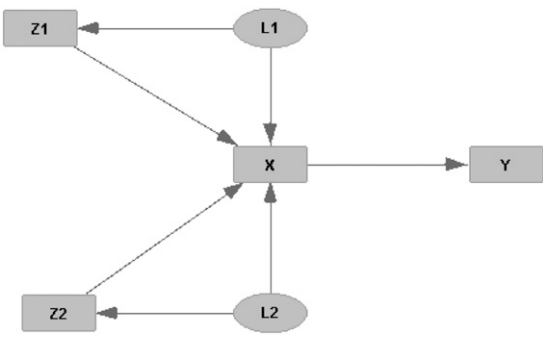
Campbell and Stanley quote previously). In observational studies, one needs a clear articulation on what moves our causes. If these ancestors are actually observable (and measurable), solutions are quite straightforward, as discussed previously. If we know there are ancestors (causes of our causes) but do not know what they are or by what tangled network they link to both our causes (X) and our effects (Y), our challenge is much greater. This is a challenge of theory and not statistics.

Yes, in the short run, many of us invoke (perhaps without explicit recognition) Laplace's Demon and either assume no "backdoor paths" or assume no linkages between our instruments and unseen omitted variables. Long-run progress will surely ask us to go further. Success will not be in terms of satisfying parameter estimates, good fit, correct signs, and agreement with previous studies. Rather, the metric of success will be defined in terms of forecasts of new data using the parameters generated from our econometric exercise. This brings me to my concluding remark, which

takes me back to the early days of my career at Purdue and my work with Jon Brandt (Brandt and Bessler, 1981). Whatever models we were considering then, univariate Autoregressive Integrated Moving Average (ARIMA) models, structural econometric models, or expert opinion, we withheld assignment of the descriptive "true model." I do not think much has changed on this issue or will ever change. As long as we admit that we do not know the true model, we will continuously have to re-evaluate fit models from either experimental or observational data in terms of their out-of-sample performance.

References

Angrist, J., and A. Krueger. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106(1991):976–1014.

Angrist, J., and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press, 2009.

Bessler, D.A., and T. Covey. "On the Search for Econometric Structure in Agriculture." *American Journal of Agricultural Economics* 75(1993):41–47.

Borlaug, N.E. "Mexican Wheat Production and Its Role on the Epidemiology of Stem Rust in North America." *Phytopathology* 44(1954):398–404.

Brandt, J.A., and D.A. Bessler. "Composite Forecasting: An Application with U.S. Hog Prices." *American Journal of Agricultural Economics* 63(1981):135–40.

Bryant, H.L., and D.A. Bessler. "Proving Causal Relationships with Observational Data." Unpublished manuscript, Texas A&M University, 2011.

Bryant, H.L., D.A. Bessler, and M.S. Haigh. "Disproving Causal Relationships Using Observational Data." *Oxford Bulletin of Economics and Statistics* 71(2009):357–74.

Campbell, O.T., and J.C. Stanley. *Experimental and Quasi-Experimental Designs for Research.* Boston, MA: Houghton Mifflin Company, 1966.

Cochran, W.G. "The Planning of Observational Studies of Human Populations." *Journal of the Royal Statistical Society A* 128(1965):134–55.

Fisher, R.A. "The Arrangement of Field Experiments," *Journal of the Ministry of Agriculture of Great Britain* 33(1926):503–13.

Haavelmo, T. "The Probability Approach in Econometrics." *Econometrica* 12,Suppl(1944):1–118.

Haigh, M., and D.A. Bessler. "Causality and Price Discovery: An Application of Directed Acyclic Graphs." *The Journal of Business* 77(2004):1097–21.

Hamilton, J. *Time Series Econometrics*. Princeton, NJ: Princeton University Press, 1994.

Hayes, D., J. Shogren, S. Shin, and J. Kliebenstein. "Valuing Food Safety in Experimental Auction Markets." *American Journal of Agricultural Economics* 77(1995):40–53.

Hicks, J.R. *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory.* 2nd ed. London, UK: Oxford University Press, 1946.

Holland, P.W. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(1986):945–60.

Horowitz, J., and K. McConnell. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44(2002):426–47.

Hoyer, P.O., A. Hyvärinen, R. Scheines, P. Spirtes, J. Ramsey, G. Lacerda, and S. Shimizu. "Causal Discovery of Linear Acyclic Models with Arbitrary Distributions." Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, Helsinki, Finland, 2008.

Hume, D. *An Enquiry Concerning Human Understanding. 1777.* LaSalle, IL: Reprinted Open Court Press, 1949.

Laplace, S. *A Philosophical Essay on Probabilities. Truscott and Emory Translation.* London: Chapman and Hall, Limited. 1902.

Lusk, J., and F. Norwood. "Bridging the Gap between Laboratory Experiments and Naturally Occurring Markets: An Inferred Valuation Method." *Journal of Environmental Economics and Management* 58(2009):236–50.

Lusk, J.L., W.B. Traill, L.O. House, C. Valli, S.R. Jaeger, M. Moore, and B. Morrow. "Comparative Advantage in Demand: Experimental Evidence of Preferences for Genetically Modified Food in the United States and European Union." *Journal of Agricultural Economics* 57(2006):1–21.

Malinvaud, E. *Statistical Methods of Econometrics.* Amsterdam, The Netherlands: North Holland, 1980.

McAdams, C., M. Palma, C. Hall, and A. Ishdorj. "A Nonhypothetical Ranking and Auction

Mechanism for Novel Products." *Journal of Agricultural and Applied Economics* 45(2013): 35–52.

Moore, H. *Economic Cycles: Their Law and Causes*. New York, NY: MacMillan, 1914.

Nelson, R.G., and D.A. Bessler. "Subjective Probabilities Elicited under Proper and Improper Scoring Rules: A Laboratory Test of Predicted Responses." *American Journal of Agricultural Economics* 71(1989):363–69.

Pearl, J. *Causality: Models, Reasoning and Inference*. New York, NY: Cambridge University Press, 2009.

Pearl, J., and E. Bareinboim. "External Validity: From Do-Calculus to Transportability Across Populations." Unpublished paper, University of California Computer Science Department, Technical Report R-400, 2012.

Roth, A. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70(2002):1341–78.

Rubin, D. "Estimating Causal Effects of Treatments from Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(1974):688–701.

Samuelson, P.A. *The Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press, 1947.

Scheines, R. "The Similarity of Causal Inference in Experimental and Non-Experimental Studies." *Philosophy of Science* 72(2005):927–40.

Smith, V. "Theory, Experiment and Economics." *The Journal of Economic Perspectives* 5(1989): 151–69.

Spirtes, P., C. Glymour, and R. Scheines. *Causation, Prediction and Search*. Cambridge, MA: MIT Press, 2000.

Stock, J., and F. Trebbi. "Who Invented Instrumental Variable Regression?" *The Journal of Economic Perspectives* 17(2003):177–94.

Wang, Z., and D.A. Bessler. "Price and Quantity Endogeneity in Demand Analysis: Evidence from Directed Acyclic Graphs." *Agricultural Economics* 34(2006):87–95.

Wright, P. *The Tariff on Animal and Vegetable Oils*. New York, NY: MacMillan, 1928.

Wright, S. "Correlation and Causation." *Journal of Agricultural Research* 20(1921):557–85.

Yu, T.-H, D.A. Bessler, and S.W. Fuller. "Price Dynamics in U.S. Grain and Freight Markets." *Canadian Journal of Agricultural Economics* 55(2007):381–97.