Journal of Economic Perspectives—Volume 24, Number 3—Summer 2010—Pages 1–6

Online Appendix: The Role of Theory in Instrument-Variables Strategies

In this appendix, I illustrate the role of theory further using the same example from Acemoglu and Johnson (2007) used in the text. The estimating equation for income per capita can be alternatively written as

$$\Delta y_i = \pi \Delta x_i + \Delta \mu + \Delta \varepsilon_i,$$

where Δ denotes the difference between dates t_0 and t_1 (for example, in Acemoglu and Johnson's estimation 1940 and 1980 or 2000). Estimating this differenced equation would (mechanically) lead to identical results to those obtained from the estimation of the level equation for income per capita presented in the text. However, the literature on the cross-country relationship between health and income sometimes estimates equations of the form

$$\Delta y_i = \alpha y_{it_0} + \beta x_{it_0} + \pi \Delta x_i + \Delta \mu + \Delta \varepsilon_i,$$

where x_{it_0} is the initial (1940) level of log life expectancy and y_{it_0} is the initial (1940) level of income per capita. It may then be tempting to estimate this modified model using a similar IV strategy, in particular, using the same predicted mortality variable (either by treating the initial levels x_{it_0} and y_{it_0} as exogenous or by instrumenting for them using their lagged values or geographic controls). The reasoning would be that if predicted mortality is a good instrument for our original estimation exercise, then it must be a good instrument for estimating this modifed model. But this reasoning is incorrect.

As already noted in the text, we need both the second and the first stages to be derived from appropriate (and logically consistent) economic models. The second stage in Acemoglu and Johnson was derived from the neoclassical growth model. A second stage equation such as the one in this modified model, with initial life expectancy on the right-hand side, could also be derived from theory, for example, from one where life expectancy in 1940 would have had a direct effect on productivity in 1980 or 2000 (40 or 60 years thereafter). However, the estimation strategy additionally requires a theoretical justification for the first stage, or the exclusion restriction (embedded in the assumption above that $\mathrm{Cov}(M_{it}, \varepsilon_{it}) = 0$). As highlighted above, this exclusion restriction could only make sense if the baseline level of mortality does not have a direct effect on future growth. If it did, the assumption that $\text{Cov}(M_{it}, \varepsilon_{it}) = 0$ would be directly violated.[1] But this implies that the theoretical argument underlying the exclusion restriction cannot be *logically* combined with a model that takes the form of this modified equation, even though it is entirely consistent with the original model we started with. We thus have a simple example where one needs to consider the theoretical foundations of the entire set of economic relations (or more explicitly, the first and second stages) together in order not to make logical errors.

The broader point is that one cannot think of "instruments" without theory. What makes a particular variable a valid instrument is a robust theoretical justification for the entire set of economic relationships being estimated, that is, both the specification of the structural parameters and the corresponding first stages and exclusion restrictions. When either of these changes, the validity of the instrument may be jeopardized. This, of course, should not be surprising in view of the discussion we started with: the plausibility of structural parameters, and thus their estimation, crucially depends on using economic theory to derive an "appropriate" model as the building block for estimation. In this light, it should be clear that there cannot be "instruments" without theory.

Conversation between Theory and Econometrics

The text emphasized the importance of economic theory in helping us specify empirical models with some degree of counterfactual validity. In this Appendix, I briefly discuss some issues that arise in attempts to implement this. Economic theory is based on mathematical models that are abstractions of reality. Our models involve several assumptions, many of which are adopted for convenience or as "simplifying" assumptions. With all of its assumptions, a model implies a set of relationships between different variables. But we may not necessarily wish to take all of these relationships as empirical predictions to be tested or used for counterfactual analysis. I now suggest that we ought to distinguish between the key implications and auxiliary implications of models, though in practice we often fail to do so.

Let us think of a model as a mapping from assumptions into empirical relationships. For concreteness, let us think of these empirical relationships as moments in the data (e.g., the conditional covariance of one variable with another), and focus on a specific problem: the relationship between income distribution, credit markets and occupational choice, for example, studied by Banerjee and Newman (1993). Let $\mathcal{A}$ denote the set of assumptions that one could make in modeling this problem. An element $A \in \mathcal{A}$ corresponds to a set of assumptions, that is, such

[1] This may or may not be a valid assumption in general. As noted in the text, Acemoglu and Johnson provide evidence to substantiate this assumption and increase the plausibility of this exclusion restriction.

things as the exact specification of production functions, the parameterization of the joint distribution of talent and initial wealth, the intertemporal preferences of agents, assumptions concerning how the credit market works, and assumptions on conjectures, expectations and the equilibrium concept. Since A is an instance of a complete set of assumptions for the problem at hand, we can think of this as a "model." This model A then generates some empirical implications. I will summarize these by a set of moments, denoted by $M \in \mathcal{M}$, which could include the correlations between the interest rate, the occupational distribution, productivity and initial wealth. Economic theory then amounts to using these assumptions in order to derive empirical relationships, or sets of moment conditions. Thus we can think of economic theory as a mapping (correspondence) $f: \mathcal{A} \rightrightarrows \mathcal{M}$, specifying which set of moments we should expect in the data for a set of assumptions.

The difficulty here is twofold. First, in writing down a model $A \in \mathcal{A}$, most theorists, rightly, will go for a minimalist structure. In many instances, A will not even contain any stochastic elements. For example, Banerjee and Newman's model leads to a unique nonstochastic equilibrium for most parameter values, where the initial distribution of wealth together with an individual's wealth determines his and his dynasty's occupational choices. To generate moment conditions that would correspond to correlations in the data, one would then have to add an unmodeled error term. Although one could interpret this error term as coming from "measurement error," this is clearly not a satisfactory interpretation, since there are many other factors relevant for occupational choices not captured by the model at hand, and they will all be subsumed into this error term, though they are not in reality related to measurement error. Yet this is not a shortcoming, but rather a strength of the model. Banerjee and Newman's model is successful largely because it abstracts from several features of the world. Second, for the same reasons of parsimony and simplicity, a model typically involves assumptions that are "auxiliary," meaning that they are made for convenience, and in the hope that they are not the source of its "main" conclusions. Naturally, what these main conclusions are is not always a simple matter to determine.

This issue notwithstanding, the central difficulty here is that the set of moment implications M depends on the entire set of assumptions, A. Suppose, for example, that $A = A' \cup A''$, where A' and A'' are two disjoint sets of assumptions, and those in the set A' correspond to the "key" assumptions, while A'' contains the auxiliary and simplifying assumptions. For example, in Banerjee and Newman, the assumption that all individuals have the same ability in all occupations, that there is no intensive margin of production, and that dynastic saving decisions are "myopic" are auxiliary assumptions. These assumptions, taken together, lead to a set of predictions. For example, taken naïvely, the model implies that there will exist a threshold level of wealth, such that all dynasties with initial wealth below this level will remain in subsistence or become workers, whereas those just above this threshold will become entrepreneurs. If we were to take such an assumption seriously, it would lead to the rejection of the model, but this would not be an insightful rejection. Instead, the Banerjee and Newman model is insightful because it highlights how the credit market problems create a link between wealth and occupational choice, and how this link depends on factor prices, which are themselves endogenously determined by the entire distribution of income.

This discussion highlights two problems. First, the insights from certain models may be "conceptual" and thus difficult to translate into moment conditions. For example, the insight that income distribution matters for occupational choices of an individual (with a given income level) is a conceptual point, even though one could devise tests by comparing different economies or the same economy over time in order to investigate the degree to which such a link is present. Second, not all of the implications of the model should be taken seriously. The second problem suggests that we may wish to separate the set of moment conditions, M, into two disjoint sets, $M = M' \cup M''$, so that M' corresponds to the set of "robust" moment predictions, which we should test or use as guidance for empirical work, whereas M'' corresponds to the moment conditions generated by the "auxiliary" assumptions. However, such a separation is not typically possible, since each moment implication of the model is potentially generated by all of the assumptions taken together. In the Banerjee and Newman model, for example, we cannot simply remove the assumptions regarding the form of the production function and still obtain moment conditions about the relationship between occupation and wealth.

This discussion suggests that in formulating economic theories which we wish to apply to data (either by ourselves or by others), we should pay special attention to which dimensions of the model are introduced just for achieving tractability and parsimony (the so-called "auxiliary") assumptions, and which assumptions and implications of the model are "robust" and should be relied upon and used empirically (or conceptually). Unfortunately, we do not have the theoretical and econometric tools to achieve this,[2] and developing such tools, or at the very least, trying to emphasize in specific instances which predictions are more robust, would be a useful direction for future research. In addition, even though such tools are not currently available, in specific instances, considerable progress is possible. I will now illustrate this using a recent paper by Weese (2009).

Weese (2009) studies the mergers across Japanese municipalities. Changes in Japanese government policy on municipality finance in 1995 led to major changes in municipality structure. Many small Japanese municipalities that did not previously have incentives to merge, because this would have reduced the transfers they received from the central government, were induced to merge after this change in policy and the number of municipalities declined from 3,232 to 1,800. Mergers across municipalities are important for public finance (because they determine the type and amount of local public goods), for development economics (since there are marked inequalities across municipalities in terms of income and provision of

[2] On the theory side, the literature on robust comparative statics, which provides qualitative predictions for a range of models, might be one useful direction. Such robust comparative statics can be obtained for environments that can be represented as supermodular games (Milgrom and Roberts, 1994; Vives, 1990) or for those that can be represented as aggregative games (Acemoglu and Jensen, 2009).

public goods, e.g., Acemoglu and Dell, 2010), and for political economy (as they are a major example of endogenous coalition formation). Weese is interested in estimating the "preferences" of different municipalities concerning mergers, and whether given these preferences, a better policy could have been devised. This type of counterfactual exercise clearly requires structural parameters in which we can have some confidence. Thus this exercise must start with a theoretical model, which will then be estimated to obtain structural parameters to use for the counterfactual and policy analysis.

One line of attack would be to specify a dynamic or static game of coalition formation, with specific assumptions on the game form. But these specific assumptions will translate into different predictions on which coalitions will form (which mergers will take place). Thus using a specific model, one could typically obtain significantly different predictions than using another related model, and structural estimation of each of these models is likely to lead to very different conclusions because auxiliary assumptions, such as those related to the order in which offers are made and how different equilibria are selected, will impact implications and inference.[3] Instead, Weese adopts a different approach, more in line with the type of conversation between theory and empirics suggested here: he specifies a general hedonic coalitional game, where municipality preferences depend on a few characteristics (average income, distance, etc.), and given these preferences he focuses on the Von Neumann-Morgenstern stable set.[4] This set is not a singleton, thus the model, equipped with this equilibrium/solution concept, does not make a unique prediction. Nevertheless, it rules out a large set of mergers given underlying preferences, and thus specifies a set of moment conditions that can be used for estimation. Crucially, these are not all of the moment conditions that will follow from a model that would make additional auxiliary assumptions to specify, say, a unique equilibrium merger structure for every value of the underlying parameter vector. Interestingly, in this case, Weese is able to estimate the underlying preferences and conduct counterfactual policy analysis.[5] His estimates show that a different government policy would have led to better (merger) outcomes, and that, somewhat surprisingly, allowing side transfers would have disadvantaged poor municipalities (because their willingness to merge with richer municipalities would have given the latter significant bargaining power).

Overall, one important direction for applied theory work (in economics in general and in development economics) would be to carefully delineate which sets

[3] Acemoglu, Egorov, and Sonin (2009) consider a class of dynamic coalition formation games, where "auxiliary" assumptions, for example, those concerning the order in which offers are made and acceptance and voting procedures, do not affect the set of predictions. This provides another example of a strategy to obtain "robust implications," even though such results can only be obtained under certain, somewhat restrictive, assumptions.

[4] See, among others, Pakes (2008) and Tamer (2003), for different approaches to the estimation of models with multiple equilibria.

[5] For example, one estimation strategy is to assign uniform probabilities to all coalition structures in the Von Neumann-Morgernstern set and then to estimate the underlying parameters by quasi-maximum likelihood.

of predictions are more robust, and thus (policy) invariant to auxiliary assumptions, and develop empirical strategies and methods of conducting counterfactual experiments that exploit these more robust implications. In the meantime, this discussion also highlights that if structural estimation relies on all of the moment conditions implied by a (simple) model, this may lead to misleading results. Making good use of theory does not mean taking all of the predictions of a model seriously, but to make use of the key and robust implications from theoretical models to specify and estimate structural parameters. It thus also requires us to be cognizant of which dimensions of a model are adopted just for simplicity, tractability and convenience.

Additional References

Acemoglu, Daron, and Melissa Dell. 2010. "Productivity Differences between and within Countries." Forthcoming. *American Economic Journal: Macroeconomics.*

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2009. "Dynamics and Stability of Constitutions, Coalitions and Clubs." NBER working paper # 14239.

Acemoglu, Daron, and Martin Kaee Jensen. 2009. "Aggregate Comparative Statics," MIT, working paper.

Banerjee, Abhijit V., and Andrew Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy,* 101 (February): 274–298.

Milgrom, Paul, and John Roberts. 1994. "Comparing Equilibria." *American Economic Review,* 84 (March): 441–459.

Tamer, Elie. 2003. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *Review of Economic Studies,* 70 (January): 147–165.

Townsend, Robert. 1994. "Risk and Insurance in Village India." *Econometrica,* 62 (May): 539–591.

Townsend, Robert. 2009. "Financial Structure and Economic Welfare: Applied General Equilibrium Development Economics." MIT mimeo.

Weese, Eric. 2009. "Political Mergers As Coalition Formation: Evidence from Japanese Municipal Mergers." Yale University mimeo.

Vives, Xavier. 1990. "Nash Equilibrium with Strategic Complementarities" *Journal of Mathematical Economics,* 19: 305–321.