Kevin D. Hoover

QUANTITATIVE EVALUATION OF IDEALIZED MODELS IN THE NEW CLASSICAL MACROECONOMICS

The new classical macroeconomics is today certainly the most coherent, if not the dominant, school of macroeconomic thought. The pivotal document in its two decades of development is Robert Lucas's 1976 paper, "Econometric Policy Evaluation: A Critique."[1] Lucas argued against the then reigning methods of evaluating the quantitative effects of economic policies on the grounds that the models used to conduct policy evaluation were not themselves invariant with respect to changes in policy.[2] In the face of the Lucas critique, the new classical economics is divided in its view of how to conduct quantitative policy analysis between those who take the critique as a call for better methods of employing theoretical knowledge in the direct empirical estimation of macroeconomic models, and those who believe that it shows that estimation is hopeless and that quantitative assessment must be conducted using idealized models. Assessing the soundness of the views of the latter camp is the main focus of this essay.

1. The Lucas Critique

The Lucas critique is usually seen as a pressing problem for macroeconomics, the study of aggregate economic activity (GNP, inflation, unemployment, interest rates, etc.). Economists typically envisage individual consumers, workers, or firms as choosing the best arrangement of their affairs, given their preferences, subject to constraints imposed by limited resources. Moving up from an account of individual behavior to an account of economy-wide aggregates is problematic. Typical macroeconometric models before Lucas's

---

[1] See Hoover (1988), (1991) and (1992) for accounts of the development of new classical thinking.
[2] The "Lucas critique," as it is always known, is older than Lucas's paper, going back at least to the work of Frisch and Haavelmo in the 1930s; see Morgan (1990), Hoover (1994).

In: Martin R. Jones and Nancy Cartwright (eds.), *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences* (*Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 86), pp. 15-33. Amsterdam/New York, NY: Rodopi, 2005.

paper consisted of a set of aggregate variables to be explained (endogenous variables). Each was expressed as a (typically, but not necessarily, linear) function of other endogenous variables and of variables taken to be given from outside the model (exogenous variables). The functional forms were generally suggested by economic theory. The functions usually contained free parameters; thus, in practice, no one believed that these functions held exactly. Consequently, the parameters were usually estimated using statistical techniques such as multivariate regression.

The consumption function provides a typical example of a function of a macromodel. The permanent income hypothesis (Friedman 1957) suggests that consumption at time t (C_t) should be related to income at t (Y_t) and consumption at t-1 as follows:

$$C_t = k(1 - \mu)Y_t + \mu C_{t-1} + e \tag{1}$$

where k is a parameter based on people's tastes for saving, μ is a parameter based on the process by which they form their expectations of future income, and e is an unobservable random error indicating that the relationship is not exact.

One issue with a long pedigree in econometrics is whether the underlying parameters can be recovered from an econometric estimation.[3] This is called the *identification problem*. For example, suppose that what we want to know is the true relationship between savings and permanent income, k. If the theory that generated equation (1) is correct, then we would estimate a linear regression of the form

$$C_t = \pi_1 Y_t + \pi_2 C_{t-1} + w \tag{2}$$

where w is a random error and π_1 and π_2 are free parameters to be estimated. We may then calculate $k = \pi_1 / (1 - \pi_2)$.

The Lucas critique suggested that it was wrong to take an equation such as (1) in isolation. Rather, it had to be considered in the context of the whole system of related equations. What is more, given that economic agents were optimizers, it was wrong to treat μ as a fixed parameter, because people's estimates of future income would depend in part on government policies that alter the course of the economy and, therefore, on GNP, unemployment, and other influences on personal income. The parameter itself would have to be modeled something like this:

$$\mu = f(Y_t, Y_{t-1}, C_{t-1}, G) \tag{3}$$

[3] See Morgan (1990), ch. 6, for a history of this problem.

where G is a variable indicating the stance of government policy. Every time government policy changes, μ changes. If equation (1) were a true representation of economic reality, the estimates of π_1 and π_2 in equation (2) would alter with every change in μ: they would not be *autonomous*.[4] The absence of autonomy, in turn, plays havoc with identification: the true k can be recovered from π_1 and π_2 only if they are stable. Identification and autonomy are distinct streams in the history of econometrics; the Lucas critique stands at their confluence. Autonomy had been a neglected issue in econometrics for over forty years. Identification, in contrast, is a central part of every econometrics course.

2. Calibration versus Estimation

New classical macroeconomists are radical advocates of microfoundations for macroeconomics. Explanations of all aggregate economic phenomena must be based upon, or at least consistent with, the microeconomics of rational economic agents. The new classicals stress constrained optimization and general equilibrium. A characteristic feature of new classical models is the rational expectations hypothesis. Expectations are said to be formed rationally if they are consistent with the forecasts of the model itself. This is usually seen to be a requirement of rationality, since, if the model is correct, expectations that are not rational would be systematically incorrect, and rational agents would not persist in forming them in such an inferior manner.[5] All new classicals share these theoretical commitments, yet they do not all draw the same implications for quantitative policy evaluation.

One approach stresses the identification aspects of the Lucas critique. Hansen and Sargent (1980), for instance, recommend starting with a model of individual preferences and constraints (i.e., "taking only tastes and technology as given," to use the jargon). From this they derive the analogs of equations (1)-(3). In particular, using the rational expectations hypothesis, they derive the correctly parameterized analog to (3). They then estimate *the reduced forms*, the analogs to equation (2) (which may, in fact, be a system of equations). Statistical tests are used to determine whether the restrictions implied by the theory and represented in the analogs to equations (1) and (3) hold. In the current example, since we have not specified (3) more definitely, there is nothing to test in equation (1). This is because equation (1) *is just identified*; i.e., there is only one way to calculate its parameters from the estimated coefficients of equation (2). Nonetheless, in many cases, there is more than one way to

[4] For the history of autonomy, see Morgan (1990), chs. 4 and 8.

[5] See Sheffrin (1983) for a general account of the rational expectations hypothesis, and Hoover (1988), ch. 1, for a discussion of the weaknesses of the hypothesis.

calculate these parameters; the equation is then said to be *overidentified*. A statistical test of whether, within some acceptable margin, all of these ways yield the same estimated parameters is a test of overidentifying restrictions and is the standard method of judging the success of the theoretical model in matching the actual data. Hansen and Sargent's approach assumes that the underlying theory is (or could be or should be) adequate, and that the important problem is to obtain accurate estimates of the underlying free parameters. Once these are known, they can be used to conduct policy analysis.

Lucas (1987, p. 45) and Kydland and Prescott (1991) argue that Hansen and Sargent's approach is inappropriate for macroeconomics. They do not dissent from the underlying theory used to derive the overidentifying restrictions. Instead, they argue that reality is sufficiently complex that no tractable theory can preserve a close mapping with all of the nuances of the data that might be reflected in the reduced forms. Sufficiently general reduced forms will pick up many aspects of the data that are not captured in the theory, and the overidentifying restrictions will almost surely be rejected.

Lucas and Prescott counsel not attempting to estimate or test macroeconomic theories directly. Instead, they argue that macroeconomists should build idealized models that are consistent with microeconomic theory and that mimic certain key aspects of aggregate economic behavior. These may then be used to simulate the effects of policy. Of course, theory leaves free parameters undetermined for Lucas and Prescott, just as it does for Hansen and Sargent. Lucas and Prescott suggest that these may be supplied either from independently conducted *microeconomic* econometric studies, which do not suffer from the aggregation problems of macroeconomic estimation, or from searches over the range of possible parameter values for a combination of values that well matches the features of the economy most important for the problem at hand. Their procedure is known as *calibration*.[6] In a sense, advocacy of calibration downplays the identification problem in the Lucas critique and emphasizes autonomy.

Estimation in the manner of Hansen and Sargent is an extension of standard and well-established practices in econometrics. When the type of models routinely advocated by Prescott (e.g., Kydland and Prescott 1982) are estimated, they are rejected statistically (e.g., Altug 1989). Prescott simply dismisses such rejections as applying an inappropriate standard. Lucas and Prescott argue that models that are idealized to the point of being incapable of passing such statistical tests are nonetheless the preferred method for generating *quantitative*

[6] Calibration is not unique to the new classical macroeconomics, but is well established in the context of "computable general equilibrium" models common in the analysis of taxation and international trade; see Shoven and Whalley (1984). All the methodological issues that arise over calibration of new classical macromodels must arise with respect to computable general equilibrium models as well.

evaluations of policies. The central issue now before us is: can models which clearly do not fit the data be useful as quantitative guides to policy? Who is right – Lucas and Prescott or Hansen and Sargent?

3. Models and Artifacts

"Model" is a ubiquitous term in economics, and a term with a variety of meanings. One commonly speaks of an econometric model. Here one means the concrete specification of functional forms for estimation: equation (2) is a typical example. I call these *observational models*. The second main class of models are *evaluative* or *interpretive models*. An obvious subclass of interpretive/evaluative models are *toy models*.

A toy model exists merely to illustrate or to check the coherence of principles or their interaction. An example of such a model is the simple exchange model with two goods and two agents (people or countries). Adam Smith's famous "invisible hand" suggests that a price system can coordinate trade and improve welfare. In this simple model, agents are characterized by functions that rank their preferences over different combinations of goods and by initial endowments of the goods. One can check, first, whether there is a relative price between the two goods and a set of trades at that price that makes both agents satisfied with the arrangement; and, second, given such a price, whether agents are better off in their own estimation than they would have been in the absence of trade. No one would think of drawing quantitative conclusions about the working of the economy from this model. Instead, one uses it to verify in a tractable and transparent case that certain qualitative results obtain. Such models may also suggest other qualitative features that may not have been known or sufficiently appreciated.[7] The utter lack of descriptive realism of such models is no reason to abandon them as test beds for general principles.

Is there another subclass of interpretive/evaluative models – one that involves quantification? Lucas seems to think so:

> One of the functions of theoretical economics is to provide fully articulated, artificial economic systems that can serve as laboratories in which policies that would be prohibitively expensive to experiment with in actual economies can be tested out at much lower cost (Lucas 1980, p. 271).

Let us call such models *benchmark models*. Benchmark models must be abstract enough and precise enough to permit incontrovertible answers to the questions put to them. Therefore,

---

[7] Cf. Diamond (1984), p. 47.

> . . . insistence on the "realism" of an economic model subverts its potential usefulness in thinking about reality. Any model that is well enough articulated to give clear answers to the questions we put to it will necessarily be artificial, abstract, patently unreal (Lucas 1980, p. 271).

On the other hand, only models that mimic reality in important respects will be useful in analyzing actual economies.

> The more dimensions in which the model mimics the answers actual economies give to simple questions, the more we trust its answers to harder questions. This is the sense in which more "realism" in a model is clearly preferred to less (Lucas 1980, p. 272).

Later in the same essay, Lucas emphasizes the quantitative nature of such model building:

> Our task . . . is to write a FORTRAN program that will accept specific economic policy rules as "input" and will generate as "output" statistics describing the operating characteristics of time series we care about, which are predicted to result from these policies (Lucas 1980, p. 288).

For Lucas, Kydland and Prescott's model is precisely such a program.[8]

One might interpret Lucas's remarks as making a superficial contribution to the debate over Milton Friedman's "Methodology of Positive Economics" (1953): must the assumptions on which a theory is constructed be true or realistic or is it enough that the theory predicts "as if" they were true? But this would be a mistake. Lucas is making a point about the architecture of models and not about the foundations of secure prediction.

To make this clear, consider Lucas's (1987, pp. 20-31) cost-benefit analysis of the policies to raise GNP growth and to damp the business cycle. Lucas's model considers a single representative consumer with a constant-relative-risk-aversion utility function facing an exogenous consumption stream. The model is calibrated by picking reasonable values for the mean and variance of consumption, the subjective rate of discount, and the constant coefficient of relative risk aversion. Lucas then calculates how much extra consumption consumers would require to compensate them in terms of utility for a cut in the growth rate of consumption and how much consumption they would be willing to give up to secure smoother consumption streams. Although the answers that Lucas seeks are quantitative, the model is not used to make predictions that might be subjected to statistical tests. Rather, it is used to set upper bounds to the benefits that might conceivably be gained in the real world. Its parameters must reflect some truth about the world if it is to be useful, but they could not be easily directly estimated. In that sense, the model is unrealistic.

---

[8] They do not say, however, whether it is actually written in FORTRAN.

In a footnote, Lucas (1980, p. 272, fn. 1) cites Herbert Simon's *Sciences of the Artificial* (1969) as an "immediate ancestor" of his "condensed" account. To uncover a more fully articulated argument for Lucas's approach to modeling, it is worth following up the reference.

For Simon, human artifacts, among which he must count economic models,

> can be thought of as a meeting point – an "interface" . . . – between an "inner" environment, the substance and organization of the artifact itself, and an "outer" environment, the surroundings in which it operates (Simon 1969, pp. 6, 7).

An artifact is useful, it achieves its goals, if its inner environment is appropriate to its outer environment.

Simon's distinction can be illustrated in the exceedingly simplified macroeconomic model represented by equations (1) and (3). The model converts inputs (Y_t and G) into an output C_t. Both the inputs and outputs, as well as the entire context in which such a model might be useful, can be considered parts of the outer environment. The inner environment includes the structure of the model – i.e., the particular functional forms – and the parameters, μ and k.[9] The inner environment controls the manner in which the inputs are processed into outputs.

The distinction between the outer and inner environments is important because there is some degree of independence between them. Clocks tell time for the outer environment. Although they may indicate the time in precisely the same way, say with identical hands on identical faces, the mechanisms of different clocks, their inner environments, may be constructed very differently. For determining when to leave to catch a plane, such differences are irrelevant. Equally, the inner environments may be isolated from all but a few key features of the outer environment. Only light entering through the lens for the short time that its shutter is open impinges on the inner environment of the camera. The remaining light is screened out by the opaque body of the camera, which is an essential part of its design.

Our simple economic model demonstrates the same properties. If the goal is to predict the level of consumption within some statistical margin of error, then like the clock, other models with quite different functional forms may be (approximately) observationally equivalent. Equally, part of the point of the functional forms of the model is to isolate features which are relevant to achieving the goal of predicting C_t – namely, Y_t and G – and screening out irrelevant features of the outer environment, impounding any relevant but ignorable influences in the error term. Just as the camera is an opaque barrier to ambient light, the functional form of the consumption model is an opaque barrier to economic influences other than income and government policy.

---

[9] One might also include C_{t-1}, because, even though it is the lagged value of C_t (the output), it may be thought of as being stored "within" the model as time progresses.

Simon factors adaptive systems into goals, outer environments, and inner environments. The relative independence of the outer and inner environments means that

> [w]e might hope to characterize the main properties of the system and its behavior without elaborating the detail of *either* the outer or the inner environments. We might look toward a science of the artificial that would depend on the relative simplicity of the interface as its primary source of abstraction and generality (Simon 1969, p. 9).

Simon's views reinforce Lucas's discussion of models. A model is useful only if it foregoes descriptive realism and selects limited features of reality to reproduce. The assumptions upon which the model is based do not matter, so long as the model succeeds in reproducing the selected features. Friedman's "as if" methodology appears vindicated.

But this is to move too fast. The inner environment is only *relatively* independent of the outer environment. Adaptation has its limits.

> In a benign environment we would learn from the motor only what it had been called upon to do; in a taxing environment we would learn something about its internal structure – specifically, about those aspects of the internal structure that were chiefly instrumental in limiting performance (Simon 1969, p. 13).

This is a more general statement of principles underlying Lucas's (1976) critique of macroeconometric models. A benign outer environment for econometric models is one in which policy does not change. Changes of policy produce structural breaks in estimated equations: disintegration of the inner environment of the models. Economic models must be constructed like a ship's chronometer, insulated from the outer environment so that

> . . . it reacts to the pitching of the ship only in the negative sense of maintaining an invariant relation of the hands on its dial to real time, independently of the ship's motions (Simon 1969, p. 9).

Insulation in economic models is achieved by specifying functions whose parameters are invariant to policy.

Again, this is easily clarified with the simple consumption model. If μ were a fixed parameter, as it might be in a stable environment in which government policy never changed, then equation (1) might yield an acceptable model of consumption. But in a world in which government policy changes, μ will also change constantly (the ship pitches, tilting the compass that is fastened securely to the deck). The role of equation (3) is precisely to isolate the model from changes in the outer environment by rendering μ a stable function of changing policy; μ changes, but in predictable and accountable ways (the compass mounted in a gimbal continually turns relative to the deck in just such a way as to maintain its orientation with the earth's magnetic field).

The independence of the inner and outer environments is not something that is true of arbitrary models; rather it must be built into models. While it may be enough in hostile environments for models to reproduce key features of the outer environment "as if" reality were described by their inner environments, it is not enough if they can do this only in benign environments. Thus, for Lucas, the "as if" methodology interpreted as an excuse for complacency with respect to modeling assumptions must be rejected.

New classical economists argue that only through carefully constructing the model from invariants – tastes and technology, in Lucas's usual phrase – can the model secure the benefits of a useful abstraction and generality. This is again an appeal to found macroeconomics in standard microeconomics. Here preferences and the production possibilities (tastes and technology) are presumed to be fixed, and the economic agent's problem is to select the optimal combination of inputs and outputs. Tastes and technology are regarded as invariant partly because economists regard their formation as largely outside the domain of economics: *de gustibus non est disputandum*. Not all economists, however, would rule out modeling the formation of tastes or technological change. But for such models to be useful, they would themselves have to have parameters to govern the selection among possible preference orderings or the evolution of technology. These parameters would be the ultimate invariants from which a model immune to the Lucas critique would have to be constructed.

4. Quantification and Idealization

Economic models are idealizations of the economy. The issue at hand, is whether, as idealizations, models can be held to a quantitative standard. Can idealized models convey useful quantitative information?

The reason why one is inclined to answer these questions negatively is that models patently leave things out. Simon's analysis, however, suggests that even on a *quantitative* standard that may be their principal advantage. To see this better, consider the analogy of physical laws.

Nancy Cartwright (1983, esp. essays 3 and 6) argues that physical laws are instruments, human artifacts in precisely the sense of Simon and Lucas, the principal use of which is to permit calculations that would otherwise be impossible. The power of laws to rationalize the world and to permit complex calculation comes from their abstractness, definitiveness, and tractability. Lucas (1980, pp. 271, 272) says essentially the same thing about the power of economic models.

The problem with laws for Cartwright (1989), however, is that they work only in ideal circumstances. Scientists must experiment to identify the actual working of a physical law: the book of nature is written in code or, more aptly perhaps, covered in rubbish. To break the code or clear the rubbish, the experimenter must either insulate the experiment from countervailing effects or must account for and, in essence, subtract away the influence of countervailing effects.[10] What is left at the end of the well-run experiment is a measurement supporting the law – a quantification of the law. Despite its tenuousness in laboratory practice, the quantified law remains of the utmost importance.

To illustrate, consider an experiment in introductory physics. To investigate the behavior of falling objects, a metal weight is dropped down a vertical track lined with a paper tape. An electric current is periodically sent through the track. The arcing of the electricity from the track to the weight burns a sequence of holes in the tape. These mark out equal times, and the experimenter measures the distances between the holes to determine the relation between time and distance.

This experiment is crude. When, as a college freshman, I performed it, I proceeded as a purely empirically – minded economist might in what I thought was a true scientific spirit: I fitted the best line to the data. I tried linear, log-linear, exponential and quadratic forms. Linear fit best, and I got a *C* – on the experiment. The problem was not just that I did not get the right answer, although I felt it unjust at the time that scientific honesty was not rewarded. The problem was that many factors unrelated to the law of gravity combined to mask its operation – conditions were far from ideal. A truer scientific method would attempt to minimize or take account of those factors. Thus, had I not regarded the experiment as a naive attempt to infer the law of gravity empirically, but as an attempt to quantify a parameter in a model, I would have gotten a better grade. The model says that distance under the uniform acceleration of gravity is $gt^2/2$. I suspect that given calculable margins of error, not only would my experiment have assigned a value to g, but that textbook values of g would have fallen within the range of experimental error – despite the apparent better fit of the linear curve.[11]

Even though the data had to be fudged and discounted to force it into the mold of the gravitational model, it would have been sensible to do so, because we know from unrelated experiments – the data from which also had to be fudged and discounted – that the quadratic law is more general. The law is right, and must be quantified, even though it is an idealization.

Confirmation by practical application is important, although sometimes the confirmation is exceedingly indirect. Engineers would often not know where to begin if they did not have quantified physical laws to work with. But laws as

---

[10] Cartwright (1989, secs. 2.3, 2.4), discusses the logic and methods of accounting for such countervailing effects.

[11] An explicit analog to this problem is found in Sargent (1989), in which he shows that the presence of measurement error can make an investment-accelerator model of investment, which is incompatible with new classical theory, fit the data better, even when the data were in fact generated according to Tobin's q-theory, which is in perfect harmony with new classical theory.

idealizations leave important factors out. Thus, an engineer uses elementary physics to *calculate* loadings for an ideal bridge. Knowing that these laws fail to account for many critical factors, the engineer then specifies material strengths two, four, or ten times what the calculations showed was needed. Error is inevitable, so err on the side of caution. Better models reduce the margins of error. It is said that the Golden Gate Bridge could easily take a second deck because it was designed with pencil, paper, and slide rule, with simple models, and consequently greatly overbuilt. England's Humber River Bridge, designed some forty years later with digital computers and more complex models is built to much closer tolerances and could not take a similar addition. The only confirmation that the bridges give of the models underlying their design is that they do not fall down.

Overbuilding a bridge is one thing, overbuilding a fine watch is quite another: here close tolerances are essential. An engineer still begins with an idealized model, but the effects ignored by the model now have to be accounted for. Specific, unidealized knowledge of materials and their behavior is needed to correct the idealized model for departures from ideal conditions. Such knowledge is often based on the same sort of extrapolations that I rejected in the gravity experiment. Cartwright (1983, essay 6) refers to these extrapolations as "phenomenal laws." So long as these extrapolations are restricted to the range of actual experience, they often prove useful in refining the fudge factors. Even though the phenomenal laws prove essential in design, the generality of idealized laws is a great source of efficiency: it is easier to use phenomenal laws to calculate departures from the ideal, than to attempt to work up from highly specific phenomenal laws. Again, the ideal laws find their only confirmation in the watch working as designed.

Laws do not constitute all of physics:

> It is hard to find them in nature and we are always having to make excuses for them: why they have exceptions – big or little; why they only work for models in the head; why it takes an engineer with a special knowledge of materials and a not too literal mind to apply physics to reality (Cartwright 1989, p. 8).

Neither do formal models constitute all of economics. Yet despite the short-comings of idealized laws, we know from practical applications, such as shooting artillery or sending rockets to the moon, that calculations based on the law of gravity get it nearly right and calculations based on linear extrapolation go hopelessly wrong.

Cartwright (1989, ch. 4) argues that "capacities" are more fundamental than laws. *Capacities* are the invariant dispositions of the components of reality. Something like the notion of capacities must lie behind the proposal to set up "elasticity banks" to which researchers could turn when calibrating computable general equilibrium models (Shoven and Whalley 1984, p. 1047). An elasticity is the proportionate change of one variable with respect to another.[12] Estimated elasticities vary according to the methods of estimation employed (e.g., functional forms, other controlling variables, and estimation procedures such as ordinary least squares regression or maximum likelihood estimation) and the set of data used. To "bank" disparate estimates is to assume that such different measures of elasticities somehow bracket or concentrate on a "true" value that is independent of the context of estimation.

Laws, at best, describe how capacities compose under ideal circumstances. That models should represent the ways in which invariant capacities compose is, of course, the essence of the Lucas critique. Recognizing that models must be constructed from invariants does not itself tell us how to measure the strengths of the component capacities.

5. Aggregation and General Equilibrium

Whether calibrated or estimated, real-business-cycle models are idealizations along many dimensions. The most important dimension of idealization is that the models deal in aggregates while the economy is composed of individuals. After all, the distinction between microeconomics and macroeconomics is the distinction between the individual actor and the economy as a whole. All new classical economists believe that one understands macroeconomic behavior only as an outcome of individual rationality. Lucas (1987, p. 57) comes close to adopting the *Verstehen* approach of the Austrians.[13] The difficulty with this approach is that there are millions of people in an economy and it is not practical – nor is it ever likely to become practical – to model the behavior of each of them.[14] Universally, new classical economists adopt *representative-agent models*, in which one agent or a few types of agents, stand in for the behavior of all agents.[15] The conditions under which a single agent's behavior can accurately represent the behavior of an entire class are onerous. Strict

---

[12] In a regression of the logarithm of one variable on the logarithms of others, the elasticities can be read directly as the value of the estimated coefficients.

[13] For a full discussion of the relationship between new classical and Austrian economics see Hoover (1988), ch. 10.

[14] In (Hoover 1984, pp. 64-66), and (Hoover 1988, pp. 218-220), I refer to this as the "Cournot problem," since it was first articulated by Augustine Cournot (1927, p. 127).

[15] Some economists reserve the term "representative-agent models" for models with a single, infinitely-lived agent. In a typical overlapping-generations model the new young are born at the start of every period, and the old die at the end of every period, and the model has infinitely many periods; so there are infinitely many agents. On this view, the overlapping-generations model is not a representative-agent model. I, however, regard it as one, because within any period one type of young agent and one type of old agent stand in for the enormous variety of people, and the same types are repeated period after period.

aggregation requires not only that every economic agent have identical preferences but that these preferences are such that any individual agent would like to consume goods in the same ratios whatever their levels of wealth. The reason is straightforward: if agents with the same wealth have different preferences, then a transfer from one to the other will leave aggregate wealth unchanged but will change the pattern of consumption and possibly aggregate consumption as well; if all agents have identical preferences but prefer different combinations of goods when rich than when poor, transfers that make some richer and some poorer will again change the pattern of consumption and possibly aggregate consumption as well (Gorman 1953). The slightest reflection confirms that such conditions are never fulfilled in an actual economy.

New classical macroeconomists insist on general equilibrium models. A fully elaborated general equilibrium model would represent each producer and each consumer and the whole range of goods and financial assets available in the economy. Agents would be modeled as making their decisions jointly so that, in the final equilibrium, production and consumption plans are individually optimal and jointly feasible. Such a detailed model is completely intractable. The new classicals usually obtain tractability by repairing to representative-agent models, modeling a single worker/consumer, who supplies labor in exchange for wages, and a single firm, which uses this labor to produce a single good that may be used indifferently for consumption or as a capital input into the production process. Labor, consumption, and capital are associated empirically with their aggregate counterparts. Although these models omit most of the details of the fully elaborated general equilibrium model, they nonetheless model firms and worker/consumers as making individually optimally and jointly consistent decisions about the demands for and supplies of labor and goods. They remain stripped-down general equilibrium models.

One interpretation of the use of calibration methods in macroeconomics is that the practitioners recognize that highly aggregated, theoretical representative-agent models must be descriptively false, so that estimates of them are bound to fit badly in comparison to atheoretical (phenomenal) econometric models. The theoretical models are nonetheless to be preferred because useful policy evaluation is possible only within tractable models. In this, they are exactly like Lucas's benchmark consumption model (see section III above). Calibrators appeal to microeconomic estimates of key parameters because information about individual agents is lost in the aggregation process. In general, these microeconomic estimates are not obtained using methods that impose the discipline of individual optimality *and* joint feasibility implicit in the general equilibrium model. Lucas (1987, pp. 46, 47) and Prescott (1986, p. 15) argue that the strength of calibration is that it uses multiple sources of information, supporting the belief that it is structured around true invariants. Again this comes close to endorsing a view of capacities as invariant dispositions independent of context.

In contrast, advocates of direct estimation could argue that the idealized representative-agent model permits better use of other information not employed in microeconomic studies. Hansen and Sargent (1980, pp. 91, 92), for example, argue that the strength of their estimation method is that it accounts consistently for the interrelationships between constituent parts of the model; i.e., it enforces the discipline of the general equilibrium method – individual optimality and especially joint feasibility. The tradeoff between these gains and losses is not clear cut.

Since both approaches share the representative-agent model, they also share a common disability: using the representative-agent model in any form begs the question by assuming that aggregation does not fundamentally alter the structure of the aggregate model. Physics may again provide a useful analogy. The laws that relate the pressures, temperatures and volumes of gases are macrophysics. The ideal laws can be derived from a micromodel: gas molecules are assumed to be point masses, subject to conservation of momentum, with a distribution of velocities. An aggregation assumption is also needed: the probability of the gas molecules moving in any direction is taken to be equal.

Direct estimation of the ideal gas laws shows that they tend to break down – and must be corrected with fudge factors – when pushed to extremes. For example, under high pressures or low temperatures the ideal laws must be corrected according to van der Waals's equation. This phenomenal law, a law in macrophysics, is used to justify alterations of the micromodel: when pressures are high one must recognize that forces operate between individual molecules.

Despite some examples of macro-to-micro inferences analogous to the gas laws, Lucas's (1980, p. 291) more typical view is that we must build our models up from the microeconomic to the macroeconomic. Unlike gases, human society does not comprise homogeneous molecules, but rational people, each choosing continually. To understand (*verstehen*) their behavior, one must model the individual and his situation. This insight is clearly correct. It is not clear in the least that it is adequately captured in the heroic aggregation assumptions of the representative-agent model. The analog for physics would be to model the behavior of gases at the macrophysical level, not as derived from the aggregation of molecules of randomly distributed momenta, but as a single molecule scaled up to observable volume – a thing corresponding to nothing ever known to nature.[16]

[16] A notable, non-new classical attempt to derive macroeconomic behavior from microeconomic behavior with appropriate aggregation assumptions is (Durlauf 1989).

6. Lessons for Econometrics

Does the calibration methodology amount to a repudiation of econometrics? Clearly not. At some level, econometrics still helps to supply the values of the parameters of the models. Beyond that, whatever I have said in favor of calibration methods notwithstanding, the misgivings of econometricians such as Sargent are genuine. The calibration methodology, to date, lacks any discipline as stern as that imposed by econometric methods. For Lucas (1980, p. 288) and Prescott (1983, p. 11), the discipline of the calibration method comes from the paucity of free parameters. But one should note that theory places only loose restrictions on the values of key parameters. In the practice of the new classical calibrators, they are actually pinned down from econometric estimation at the microeconomic level or accounting considerations. In some sense, then, the calibration method would appear to be a kind of indirect estimation. Thus, it would be a mistake to treat calibration as simply an alternative form of estimation, although it is easy to understand why some critics interpret it that way. Even were there less flexibility in the parameterizations, the properties ascribed to the underlying components of the idealized representative-agent models – the agents, their utility functions, production functions, and constraints – are not subject to as convincing cross-checking as the analogous components in physical models usually are. The fudge factors that account for the discrepancies between the ideal model and the data look less like van der Waals's equation, less like phenomenal laws, than like special pleading. Above all, it is not clear on what standards competing but contradictory models are to be compared and adjudicated.[17] Some such standards are essential if any objective progress is to be made in economics.[18]

The calibration methodology is not, then, a repudiation of econometrics; yet it does contain some lessons for econometrics.

In (Hoover 1994), I distinguish between two types of econometrics. *Econometrics as observation* treats econometric procedures as filters that process raw data into statistics. On this view, econometric calculations are not valid or

[17] Prescott (1983, p. 12), seems, oddly, to claim that the inability of a model to account for some real events is a positive virtue – in particular, that the inability of real-business-cycle models to account for the Great Depression is a point in their favor. He writes: "If any observation can be rationalized with some approach, then that approach is not scientific." This seems to be a confused rendition of the respectable Popperian notion that a theory is more powerful the more things it rules out. But one must not mistake the power of a theory with its truth. Aside from issues of tractability, a theory that rationalizes only and exactly those events that actually occur, while ruling out exactly those events that do not occur is the perfect theory. In contrast, Prescott seems inadvertently to support the view that the more exceptions, the better the rule.

[18] Watson (1993) develops a goodness-of-fit measure for calibrated models. It takes into account that, since idealization implies differences between model and reality that may be systematic, the errors-in-variables and errors-in-equations statistical models are probably not appropriate.

invalid, but useful if they reveal theoretically interpretable facts about the world and not useful if they do not. *Econometrics as measurement* treats econometric procedures as direct measurements of theoretically articulated structures. This view is the classic Cowles Commission approach to structural estimation that concentrates on testing identified models specified from *a priori* theory.[19]

Many new classicals, such as Cooley and LeRoy (1985) and Sargent (1989), advocate econometrics as measurement. From a fundamental new classical perspective, they seem to have drawn the wrong lesson from the Lucas critique. Recall that the Lucas critique links traditional econometric concerns about identification and autonomy. New classical advocates of economics as observation overemphasize identification. Identification is achieved through prior theoretical commitment. The only meaning they allow for "theory" is general equilibrium microeconomics. Because such theory is intractable, they repair to the representative-agent model. Unfortunately, because of the failure of the conditions for exact aggregation to obtain, the representative-agent model does not represent the actual choices of *any* individual agent. The representative-agent model applies the mathematics of microeconomics, but in the context of econometrics as measurement it is only a simulacrum of microeconomics. The representative-agent model does not solve the aggregation problem; it ignores it. There is no reason to think that direct estimation will capture an accurate measurement of even the average behavior of the individuals who make up the economy. In contrast, calibrators use the representative-agent model precisely to represent average or typical behavior, but quantify that behavior independently of the representative-agent model. Thus, while it is problematic at the aggregate level, calibration can use econometrics as measurement, when it is truly microeconometric – the estimation of fundamental parameters from cross-section or panel data sets.

Calibrators want their models to mimic the behavior of the economy; but they do not expect economic data to parameterize those models directly. Instead, they are likely to use various atheoretical statistical techniques to establish facts about the economy that they hope their models will ultimately imitate. Kydland and Prescott (1990, pp. 3, 4) self-consciously advocate a modern version of Burns and Mitchell's "measurement without theory" – i.e., econometrics as observation. Econometrics as observation does not attempt to quantify fundamental invariants. Instead it repackages the facts already present in the data in a manner that a well calibrated model may successfully explain.

---

[19] For a general history of the Cowles Commission approach, see Epstein (1987), ch. 2.

7. Conclusion

The calibration methodology has both a wide following and a substantial opposition within the new classical school. I have attempted to give it a sympathetic reading. I have concentrated on Prescott and Lucas, as its most articulate advocates. Calibration is consistent with appealing accounts of the nature and role of models in science and economics, of quantification and idealization. The practical implementation of calibration methods typical of new classical representative-agent models is less convincing.

The calibration methodology stresses that one might not wish to apply standard measures of goodness-of-fit (e.g., R^2 or tests of overidentifying restrictions) such as are commonly applied with the usual econometric estimation techniques, because it is along only selected dimensions that one cares about the model's performance at all. This is completely consistent with Simon's account of artifacts. New classical economics has traditionally been skeptical about discretionary economic policies. They are therefore more concerned to evaluate the operating characteristics of policy rules. For this, the fit of the model to a particular historical realization is largely irrelevant, unless it assures that the model will also characterize the future distribution of outcomes. The implicit claim of most econometrics is that it does assure a good characterization. Probably most econometricians would reject calibration methods as coming nowhere close to providing such assurance. Substantial work remains to be done in establishing objective, comparative standards for judging competing models.

Fortunately, even those converted to the method need not become Lucasians: methodology underdetermines substance. Simon, while providing Lucas with a foundation for his views on modeling, nonetheless prefers a notion of "bounded rationality" that is inconsistent with the rational expectations hypothesis or Lucas's general view of humans as efficient optimizers.[20] Favero and Hendry (1989) agree with Lucas over the importance of invariance, but seek to show that not only can invariance be found at the level of aggregate econometric relations (e.g., in the demand-for-money function), but that this rules out rational expectations as a source of noninvariance.[21]

Finally, to return to a physical analog, economic modeling is like the study of cosmology. Substantial empirical work helps to determine the values of key constants; their true values nonetheless remain doubtful. Different values within the margins of error, even given similarly structured models, may result in very

[20] E.g., Simon (1969, p. 33) writes: "What do these experiments tell us? First, they tell us that human beings do not always discover for themselves clever strategies that they could readily be taught (watching a chess master play a duffer should also convince us of that)."

[21] Favero and Hendry (1989) reject the practical applicability of the Lucas critique for the demand for money in the U.K.; Campos and Ericsson (1988) reject it for the consumption function in Venezuela.

different conclusions (e.g., that the universe expands forever or that it expands and then collapses). Equally, the same values, given the range of competing models, may result in very different conclusions. Nevertheless, we may all agree on the form that answers to cosmological or economic questions must take, without agreeing on the answers themselves.*

Kevin D. Hoover
Department of Economics
University of California, Davis
kdhoover@ucdavis.edu

REFERENCES

Altug, S. (1989). Time-to-Build and Aggregate Fluctuations: Some New Evidence. *International Economic Review* **30**, 889-920.

Campos, J. and Ericsson, N. R. (1988). Econometric Modeling of Consumers' Expenditure in Venezuela. Board of Governors of the Federal Reserve System International Finance Discussion Paper, no. 325.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.

Cooley, T. F. and LeRoy, S. F. (1985). Atheoretical Macroeconometrics: A Critique. *Journal of Monetary Economics* **16**, 283-308.

Cournot, A. ([1838] 1927). *Researches into the Mathematical Principles of the Theory of Wealth*. Translated by Nathaniel T. Bacon. New York: Macmillan.

Diamond, P. A. (1984). *A Search-Equilibrium Approach to the Micro Foundations of Macroeconomics: The Wicksell Lectures, 1982*. Cambridge, Mass.: MIT Press.

Durlauf, S. N. (1989). Locally Interacting Systems, Coordination Failure, and the Behavior of Aggregate Activity. Unpublished typescript, November 5th.

Epstein, R. J. (1987). *A History of Econometrics*. Amsterdam: North-Holland.

Favero, C. and Hendry, D. F. (1989). Testing the Lucas Critique: A Review. Unpublished typescript.

Friedman, M. (1953). The Methodology of Positive Economics. In: *Essays in Positive Economics*. Chicago: Chicago University Press.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Gorman, W. M. (1953). Community Preference Fields. *Econometrica* **21**, 63-80.

Hansen, L. P. and Sargent, T. J. (1980). Formulating and Estimating Dynamic Linear Rational Expectations Models. In R. E. Lucas, Jr. and T. J. Sargent (eds.), *Rational Expectations and Econometric Practice*. London: George Allen & Unwin.

Hoover, K. D. (1984). Two Types of Monetarism. *Journal of Economic Literature* **22**, 58-76.

Hoover, K. D. (1988). *The New Classical Macroeconomics: A Skeptical Inquiry*. Oxford: Blackwell.

* I thank Thomas Mayer, Kevin Salyer, Judy Klein, Roy Epstein, Nancy Cartwright, and Steven Sheffrin for many helpful comments on an earlier version of this paper.

Hoover, K. D. (1991). Scientific Research Program or Tribe? A Joint Appraisal of Lakatos and the New Classical Macroeconomics. In M. Blaug and N. de Marchi (eds.), *Appraising Economic Theories: Studies in the Application of the Methodology of Research Programs*. Aldershot: Edward Elgar.

Hoover, K. D. (1992). Reflections on the Rational Expectations Revolution in Macroeconomics. *Cato Journal* **12**, 81-96.

Hoover, K. D. (1994). Econometrics as Observation: The Lucas Critique and the Nature of Econometric Inference. *Journal of Economic Methodology* **1**, 65-80.

Kydland, F. E. and Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica* **50**, 1345-1370.

Kydland, F. E. and Prescott, E. C. (1990). Business Cycles: Real Facts and a Monetary Myth. *Federal Reserve Bank of Minneapolis Quarterly Review* **14**, 3-18.

Kydland, F. E. and Prescott, E. C. (1991). The Econometrics of the General Equilibrium Approach to Business Cycles. *Scandinavian Journal of Economics* **93**, 161-78.

Lucas, R. E., Jr. (1976). Econometric Policy Evaluation: A Critique. Reprinted in Lucas (1981).

Lucas, R. E., Jr. (1980). Methods and Problems in Business Cycle Theory. Reprinted in Lucas (1981).

Lucas, R. E., Jr. (1981). *Studies in Business-Cycle Theory*. Oxford: Blackwell.

Lucas, R. E., Jr. (1987). *Models of Business Cycles*. Oxford: Blackwell.

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

Prescott, E. C. (1983). 'Can the Cycle be Reconciled with a Consistent Theory of Expectations?' or A Progress Report on Business Cycle Theory. Federal Reserve Bank of Minneapolis Research Department Working Paper, No. 239.

Prescott, E. C Prescott, E. C. (1986). Theory Ahead of Business Cycle Measurement. *Federal Reserve Bank of Minneapolis Quarterly Review* **10**, 9-22.

Sargent, T. J. (1989). Two Models of Measurements and the Investment Accelerator. *Journal of Political Economy* **97**, 251-287.

Sheffrin, S. M. (1983). *Rational Expectations*. Cambridge: Cambridge University Press.

Shoven, J. B. and Whalley, J. (1984). Applied General-equilibrium Models of Taxation and International Trade. *Journal of Economic Literature* **22**, 1007-1051.

Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, Mass.: The MIT Press.

Watson, M. W. (1993). Measures of Fit for Calibrated Models. *Journal of Political Economy* **101**, 1011-41.