Measure for Measure: How Economists Model the World into Numbers

BY MARCEL BOUMANS

Only a further development of the engineering skill of econometrics will help in this respect (Tinbergen, 1959 [1936]: 84).

But technique is interesting to technicians (which is what we are, if we are to be of any use to anyone). . . (Lucas, 1987: 35).

1. Introduction

The practice of economic science is dominated by model building. To evaluate economic policy, models are built and used to produce numbers to inform us about economic phenomena. Although phenomena are detected through the use of observed data, they are in general not directly observable. To “see” them we need instruments, and to obtain numerical facts of the phenomena in particular we need measuring instruments. This paper will argue that in economics, models function as such instruments of observation—more specifically, as measuring instruments. In measurement theory, measurement is the mapping of a property of the empirical world into a set of numbers. This paper’s view is that economic modeling is a specific kind of mapping to which the standard account on how models are obtained and assessed

*I thank Mark Blaug and Harro Maas for their valuable remarks.
does not apply. Models are not easily or simply derived from theories and subsequently tested against empirical data. Instruments are constructed by integrating theoretical and empirical ideas and requirements in such a way that their performance meets a previously chosen standard. The empirical requirement is that the model should take account of the phenomenological facts so that the reliability of the model is not assessed by post-model testing but obtained by calibration (see Boumans, 1999).

It will be argued that the practice of treating models as measuring instruments is typical of twentieth-century economics and is most evident in the birth of two new branches—econometrics and macroeconomics—during that century. It is here, particularly where both branches interact and overlap, that the discussion of the construction of reliable instruments emerges most clearly and where the solutions to this problem have been worked out in practical and intellectual responses. These new practices can be associated with Nobel laureates such as Ragnar Frisch and Jan Tinbergen (who received the award in 1969), Tjalling Koopmans (in 1975), Milton Friedman (in 1976), Herbert Simon (in 1978), Lawrence Klein (in 1980), Trygve Haavelmo (in 1989), and Robert Lucas (in 1995) who in their work treated these problems explicitly. Their contributions to macroeconomics and econometrics will be discussed in this paper, which will focus on attempts to model business-cycle phenomena.

The model account explored in this paper is closely related to Hoover’s and Morgan’s discussions of modeling in econometrics. According to Hoover (1994), econometrics is an observational science analogous to astronomy; therefore its models should not be assessed as to whether they are valid or not, but as to whether they are useful or not. And the standard by which the usefulness of an instrument for observation or measurement is to be judged varies with what one seeks to observe or to measure. Morgan (1988) showed that the econometricians of the 1930s were not trying to prove theories to be true or untrue, but were primarily concerned with finding “satisfactory” empirical models. Testing
involved an idea of quality control and quality ranking. And more recently Morrison and Morgan (1999) have shown that models function as "instruments of investigation." We can learn from them about the world because of their representative power. The second section of this paper will discuss, using Galileo's model of intelligibility based on Archimedean simple mechanisms, how this representative power of models can be attained.

The particular case examined in section 2 is the problem that economic theories do not provide the mathematics needed to represent the quantitative facts of the phenomenon. At the same time, the phenomena are not very helpful, for they also do not prescribe any particular kind of formalism. It appears that the choice of the mathematics for constructing a model comes from simple mechanisms that are used as recipes for building new models. These simple mechanisms function as analogies and can in principle be picked up from anywhere, though it seems to be that the favorite sources are hydraulics and mechanics. These mechanisms not only contribute to the explanation of the phenomenon but also provide the accompanying tool kit, the mathematical techniques. When the construction of a new model succeeds, it can function as recipe for another related model.

Why models cannot simply be derived from theories stems from the fact that theories are incomplete. Theories are incomplete with respect to data. The main message of Bogen's and Woodward's intriguing paper, "Saving the Phenomena" (1988), is that the "facts for which typical scientific theories are expected to account are not, by and large, facts about observables"; on the contrary, "scientific theories do predict and explain facts about phenomena" (305-6). The reason is that observations—data—are typically the result of complex interactions among a large number of disparate causal factors that are idiosyncratic to a particular experimental design, detection device, or data-gathering procedure an investigator employs. No theory can accurately predict or explain an outcome that depends on the confluence of so many variable and transient factors. But phenomena are expected to have stable, repeatable
characteristics that will be detectable by means of a variety of different procedures, which may yield quite different kinds of data. In other words, phenomena refer to general features, while observations are always idiosyncratic to a particular context. Theories are not about observations—particulars—but about facts of phenomena—universals. Because facts about phenomena are not directly measured but must be inferred from the observed data, we need to consider the reliability of the data. These considerations cannot be derived from theory but are based on a closer investigation of the experimental design, the equipment used, and need a statistical underpinning. This message was well laid out for econometrics by Haavelmo (1944).

The mechanisms used to construct models do not have to be adequate descriptions of real "mechanisms" in the outer world. To use models as instruments for evaluating policy measures it suffices that they mimic the relevant facts of the phenomena. This antirealistic approach in economics is called instrumentalism and is explored in section 3.

But a comparison of the facts generated by a measurement model with those about phenomena, as can be done in the case of theories, cannot assess the quality of measuring instruments. There are no other facts with which the measurements can be assessed, unless one has at one’s disposal facts from another measuring instrument, which in economics is rarely the case. Another problem in observing phenomena is how to distinguish between facts about the phenomena and the artifacts created by the instrument. A third problem is the requirement for sufficient stability: although data are idiosyncratic to the specific, partially uncontrollable circumstances, one nevertheless aims at a relation between phenomenological facts and data that is as much as possible insensitive to the environment—a feature better known in econometrics by the term “autonomy” (see Frisch, 1995 [1938]; Haavelmo, 1944). Allan Franklin (1986) discusses nine epistemological strategies to improve the reliability of an instrument. One of these strategies is calibration. It will be shown in section 4 that
this specific strategy can be used to deal with the aforementioned problems of empirical assessment. Section 5 concludes this discussion by stating that, at the end of the twentieth century, the prevailing view regarding economic modeling is that economists should aim at simple but calibrated models.

2. The Instruments of an Economist

One usually associates instruments with physical devices, like a thermometer or a ruler, but in economics they are immaterial. Although the term “model” originally referred to a material object—that is, a three-dimensional representation of a structure on a different scale, such as a figure in clay or wax—models have lost their physical substance in economics. Nevertheless, Morrison and Morgan (1999) have shown that models in economics still function as if they were physical instruments. They can function as such because they involve some form of representation. This representative power enables us to learn something about the thing it represents. But,

we do not learn much from looking at a model—we learn more from building the model and manipulating it. Just as one needs to use or observe the use of a hammer in order to really understand its function, similarly, models have to be used before they will give up their secrets. In this sense, they have the quality of a technology—the power of the model only becomes apparent in the context of its use (Morrison and Morgan, 1999: 12).

In other words, Morrison and Morgan treat models as quasi-empirical objects.

Morrison’s and Morgan’s account of understanding that is gained by building and using models fits into a longer tradition that started with what Galileo took to be intelligible and the model of intelligibility that he developed. Machamer (1998) shows that
the Archimedean simple machines, such as the balance, the inclined plane, and the screw, and the experience related to them, became Galileo's model for both theory and experiment.

Intelligibility or having a true explanation for Galileo had to include having a mechanical model or representation of the phenomenon. In this sense, Galileo added something to the traditional criteria of mathematical description (from the mixed sciences) and observation (from astronomy) for constructing scientific objects (as some would say) or for having adequate explanations of the phenomena observed (as I would say)... To get at the true cause, you must replicate or reproduce the effects by constructing an artificial device, so that the effects can be seen (Machamer, 1998: 69).

This mode of scientific understanding was also emphasized by William Thomson (Lord Kelvin): "It seems to me that the test of 'Do we or do we not understand a particular subject in physics?' is 'Can we make a mechanical model of it?"' (quoted in Duhem, 1981 [1954]: 71). In this tradition, understanding a phenomenon became the same thing as "designing a model imitating the phenomenon; whence the nature of material things is to be understood by imagining a mechanism whose performance will represent and simulate the properties of the bodies" (Duhem 1981 [1954]: 72). (For a more recent, philosophically related discussion, see Cartwright's "simulacrum" account of models: "The success of a model depends on how much and how precisely it can replicate what goes on" [Cartwright, 1983: 153]).

It seems that twentieth-century mathematical modeling took place according to this Galilean model of intelligibility: understanding a phenomenon is provided by Archimedean simple machines, which not only provide the mechanism that explains the phenomenon but also the mathematics (geometry) to describe it. The exemplar of a simple machine used in economics in the 1930s to understand business cycles was the pendulum. The two leading figures of the Econometric Society in the interbellum (and the first Nobel laureates in economics) were Jan Tin-
bergen and Ragnar Frisch. Both took harmonic oscillation—the mathematical representation of the pendulum—as the starting point for modeling the business cycle. During the 1933 Leiden meeting of the Econometric Society, Tinbergen discussed the question, “Is the theory of harmonic oscillation useful in the study of business cycles?” (Marschak, 1934: 187). And Frisch’s classic Rocking Horse model of the business cycle (1933b) was a pendulum (or rocking horse) hampered by friction but frequently hit by a stick to maintain the cycle.

Nevertheless, besides the choice of an adequate form, mathematical shaping also implies tuning: a choice of certain parameter values such that the mechanism reproduces certain facts of the phenomenon. Parameter values are derived from observations, generally by estimation. Therefore they include various unwanted kinds of background noise and confounding factors. Tuning is a necessary part of the mathematical integration of theoretical notions with the facts of the phenomenon. It restricts the range of parameter values in the hope of reducing the role of observational errors. A nice but not well-known example of tuning can be found in Tinbergen’s “Suggestions on Quantitative Business Cycle Theory” in Econometrica (Tinbergen, 1935).

Tinbergen systematically outlined the requirements for an adequate business-cycle theory. The first requirement was that the theory should be “dynamic”: the “mechanism,” that is, the system of relations between the variables, should contain at least one dynamic relation. A relation was defined to be dynamic when “variables relating to different moments [of time] appear in one equation” (1935: 241). Dynamic relations were obtained by introducing integrals, \( \int p(t)dt \), derivatives, \( \dot{p}(t) \), and lag terms, \( p(t-t_i) \), into the relations, so that in general the reduced form equation would appear as:

\[
\sum_i^n a_i p(t-t_i) + \sum_i^n b_i \dot{p}(t-t_i) + \sum_i^n c_i \int_{t_i}^{t} p(\tau) d\tau = 0
\]

The second requirement was that the parameters satisfy the “wave condition” and the “long wave condition.” The “wave con-
dition" indicated that the solution to the above reduced form equation should consist of a sine function, \( p(t) = C\lambda \sin(\omega t) \), so that the time shape of \( p(t) \) is cyclic. The "long wave condition" prescribed that the cycle period should be long compared with the "time units" and that the cycle should not differ "too much from an undamped one" (280). According to Tinbergen, "These conditions will be a guide in a statistical test of the different schemes as to their accord with reality" (280). As a first approximation to these conditions, Tinbergen put \( \lambda = 1 \) and \( \omega = 0 \). Then the period of the cycle, \( 2\pi/\omega \), goes to infinity. A consequence of these conditions was that:

\[
\sum_{i} c_i = 0
\]

In other words, mechanisms "only then lead to long, not too much damped waves when the integral terms are of small importance" (281).

Tinbergen also considered a second approximation of the long wave conditions by assuming that \( \lambda = 1 + \delta \) and \( \omega = \epsilon \), where both \( \epsilon \) and \( \delta \) are very small. Again this resulted in restrictions on the parameters of the possible mechanisms. Tinbergen considered several mechanisms as possible explanations of the business cycle. The wave conditions were used to detect the correct mechanism by comparing the order of magnitude required by the conditions with the estimated parameter values.

Tinbergen was the first to succeed in modeling a real economy based on the method he had worked out in his 1935 "Quantitative Business Cycle Theory" (see Boumans, 1992). In 1936 he presented a model of the Dutch economy to the Dutch Economic Association. The paper was read and published in Dutch, but in the same year Tinbergen was commissioned by the Economic Intelligence Service of the League of Nations to undertake statistical tests of the business-cycle theories examined by Haberler for the league. Tinbergen worked at this task for two years and reported his results in a two-volume work, *Statistical Testing of Business-Cycle Theories*, published in 1939. The first contained an expla-
nation of a method of econometric testing and a demonstration in three case studies of what could be achieved. The second volume developed a model of the United States. It should be noted that Tinbergen did not build his models within one general (macro-) economic theoretical framework. This was still the pre-Keynesian era.

3. Instrumentalism

Tinbergen's first models were awesome examples of what could be achieved by the program of the Econometric Society in its early stage. The Econometric Society was founded in 1930. Its scope was "the advancement of economic theory in its relation to statistics and mathematics" and its main object "to promote studies that aim at a unification of the theoretical-quantitative and empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking similar to that which has come to dominate in the natural sciences" ("Constitution," 1933: 106). One of the founders, Ragnar Frisch, and editor of the society's own journal, *Econometrica*, emphasized the aim of econometrics:

*Experience has shown that each of these three viewpoints, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. And it is this unification that constitutes econometrics* (Frisch, 1933a: 2).

A group of scholars closely linked to the Econometric Society was the Cowles Commission for Research in Economics, founded in 1932. It was "dedicated to research in economic theory and measurement. It seeks to make additions to fundamental knowledge about society, through theory construction, through mea-
surement for the testing of theory, through development of methods of measurement, and through application of results in specific areas" (Cowles Commission, 1952: 2). The Cowles Commission (later Cowles Foundation) became an important financial supporter of the Econometric Society and the journal *Econometrica* was published by it.

Bogen's and Woodward's conclusion that theories are incomplete with respect to data had a predecessor in Haavelmo's seminal paper, "The Probability Approach in Econometrics": "The data [the economist] actually obtains are, first of all, nearly always blurred by some plain errors of measurement, that is, by certain extra 'facts' which he did not intend to 'explain' by his theory" (Haavelmo, 1944: 7). From this insight he draw the conclusion that to infer facts from the observations we need to take account of the error term: "one should study very carefully the actual series considered and the conditions under which they were produced, before identifying them with the variables of a particular theoretical model" (7).

The paper became the paradigm for the research pursued at the Cowles Commission. Its aim was to "bridge" economic theory involving exact functional relationships and actual measurement with a stochastic scheme, but it in fact shifted the focus from identifying "mechanisms" with the aid of phenomenological facts to identifying the real economic "structure" with the use of data. It was the Haavelmo-Cowles approach to macroeconometric modeling that became dominant for the next 40 years and can be summarized as follows:

(a) that the economy may be characterized as a set of autonomous and simultaneous behavioural (causal) relations with structural features captured by the parameters of these relations; and

(b) that these relations are essentially stochastic (De Marchi and Gilbert, 1989: 5).
In the 1940s, Lawrence Klein was commissioned to build models of the United States in the tradition of Tinbergen’s macro-econometric modeling. The program’s aim was to build increasingly comprehensive models to improve their predictability so they could be used as reliable instruments for economic policy. The implications of a policy change could then be forecasted with the aid of these models.

An important critique to this approach came from Milton Friedman (1951), who strongly doubted the validity of these econometric models on the basis of the poor results of postmodel forecasting tests. It appeared that Klein’s models performed much worse in comparison with simple extrapolation models, the so-called naive models. The assertion that the econometric model can predict the consequences of policy changes was, according to Friedman, a “pure act of faith” (Friedman, 1951: 111). Greater comprehensiveness did not lead to better policy instruments. Friedman’s lack of faith in the macroeconometric program sent him in another research direction, namely that of partitioning: “The direction of work that seems to me to offer most hope for laying a foundation for a workable theory of change is the analysis of parts of the economy in the hope that we can find bits of order here and there and gradually combine these bits into a systematic picture of the whole” (114). This opinion became increasingly shared in the applied circle. Many applied modelers shifted their interest in macro-modeling away from a whole economy to parts of economic activities in which economic theories were relatively well developed (Qin, 1993: 138-139).

For Friedman the forecasting abilities were the qualities of the model that should be evaluated, not its realism. This methodological standpoint was spelled out in his well-known article “The Methodology of Positive Economics” (1953) and later labeled by Boland (1979) as “instrumentalism”: “For some policy-oriented economists, the intended job is the generation of true or successful predictions. In this case a theory’s predictive success is always a sufficient argument in its favor. This view of the role of
theories is called 'instrumentalism.' It says that theories are convenient and useful ways of (logically) generating what have turned out to be true (or successful) predictions or conclusions" (Boland, 1979:508).

Of course, it is not necessary to lose faith in the Cowles program, as Friedman did, because of the poor forecasting abilities of the early macroeconometric models. In the first place, one can be confident that the forecasting abilities will steadily improve—and, in fact, they did. In the second place, one may wonder whether a test of success in extrapolation and prediction is the most suitable standard by which to judge these models. When such models are built to gain insight into the mechanism of business fluctuations, a more adequate postmodel test might be to verify whether the model mimics the relevant characteristics of these fluctuations. For example, Tinbergen’s last stage of valuing the United States model (1939) was to check whether the model as a whole would represent the cycles adequately (Morgan, 1990: 116). Indeed, it was this kind of postmodel test that provided the Cowles program with strong support. Irma and Frank Adelman’s (1959) computer simulation of the Klein-Goldberger (1955) model of the United States economy—at that time the most advanced macroeconometric model—showed that this model, when shocked by disturbances, could generate cycles with the same characteristics as those of the United States economy. Indeed, the Klein-Goldberger model cycles were remarkably similar to those described by the National Bureau of Economic Research (NBER) as characterizing the American economy. From this the Adelmans concluded that the Klein-Goldberger model was “not very far wrong” (1959: 621).

To Robert Lucas the Adelmans’ achievement also signaled a new standard for what it means to understand business cycles: “One exhibits understanding of business cycles by constructing a model in the most literal sense: a fully articulated artificial economy which behaves through time so as to imitate closely the times series behavior of actual economics” (Lucas, 1977: 11). However,
in Lucas’s view, the ability of models to imitate actual behavior assessed by “characteristics testing” (see Kim, De Marchi, and Morgan, 1995) is a necessary but not sufficient condition to use these kinds of econometric models for policy evaluation: “simulations using these models can, in principle, provide no useful information as to the actual consequences of alternative economic policies” (Lucas, 1976: 20). The underlying idea, known as the Lucas Critique, is that estimated parameters that were previously regarded as “structural” in econometric analysis of economic policy actually depend on the economic policy pursued during the estimation period. Hence, the parameters may change with shifts in the policy regime (Lucas, 1976).

Lucas’s 1976 paper is perhaps the most influential and cited paper in macroeconomics in the last 25 years (Hoover, 1995b) and contributed to the decline in popularity of the Cowles approach. The Lucas Critique was an implicit call for a new research program. This alternative to the Cowles program involved formulating and estimating macroeconometric models with parameters that are invariant under policy variations and can thus be used to evaluate alternative policies. And the only parameters Lucas “hopes” to be invariant under policy changes are those describing “tastes and technology” (Lucas, 1981: 11-12; 1977: 12).

Lucas’s approach was not to aim for more realism in the models but, on the contrary, to advocate “superficiality”:

A “theory” is not a collection of assertions about the behavior of the actual economy but rather an explicit set of instructions for building a parallel or analogue system—a mechanical, imitation economy. A “good” model, from this point of view, will not be exactly more “real” than a poor one, but will provide better imitations. Of course, what one means by a “better imitation” will depend on the particular questions to which one wishes answers (Lucas 1980: 697).
4. Calibration

The problem of requiring conditions of invariance in an economic system is related to the problem of autonomy in econometrics (see Aldrich, 1989 and Hoover, 1995a). Ragnar Frisch coined this term in his 1938 memorandum discussing Tinbergen's results for the League of Nations. He defined the autonomous features of a system as those that "could be maintained unaltered while other features of the structure were changed" (Frisch, 1946 [1938]: 417). Trygve Haavelmo characterized autonomous relations as those that "describe the functioning of some parts of the mechanism irrespective of what happens in some other parts" (Haavelmo, 1944: 28). Both econometricians were concerned with modeling and statistical procedures that would allow them to identify stable relations in the economy.

Thus, autonomy is connected to Lucas's critique of macroeconometric models. Hoover (1995a) shows that Herbert Simon's The Sciences of the Artificial (1969) provided the materials to construct a methodological foundation for Lucas's approach to modeling in which the problem of autonomy is taken into account in relation to "superficiality." For Simon, an artifact can be thought of as a meeting point—an "interface" in today's terms—between an "inner" environment, the substance and organization of the artifact itself, and an "outer" environment, the surroundings in which it operates. If the inner environment is appropriate to the outer environment, or vice versa, the artifact will serve its intended purpose (1969: 7).

To clarify this definition of an artifact, Simon uses the example of a clock. The purpose of a clock is to measure time. The inner environment of the clock is its internal construction. Simon emphasizes that whether a clock will in fact tell time is also dependent on where it is placed. The artifact is molded by the environ-
ment: a sundial performs as a clock in sunny climates, but to devise a clock that would tell time on a rolling and pitching ship it has to be endowed with "many delicate properties, some of them largely or totally irrelevant to the performance of a landlubber's clock" (6).

The designer insulates the inner system from the environment, so that an invariant relation is maintained between inner system and goal, independent of variations over a wide range in most parameters that characterize the outer environment (9).

In contrast to physics, in which one is able to create stable environments for measurements, in economics one has often to take measurements in a constantly changing environment. Unable to command the environment, stability has to be built into the instrument. Simon's clock example is a material instrument within which one can build stabilizing mechanisms. But economic models are not material, so that stability has to be built-in differently. Hoover (1995a: 36) showed that Simon's notion of artifact justified Lucas's insistence that only through constructing the model from invariants "can the model secure benefits of a useful abstraction and generality," namely an enlargement of its applicability.

Although Hoover's article, "Facts and Artifacts: Calibration and the Empirical Assessment of Real-Business-Cycle Models" (1995a), discusses Lucas's program of modeling from invariants and calibration techniques in one paper, it does not explore the link between both, mainly because its aim is to compare and evaluate estimation versus calibration. However, it will be shown below that calibration fits perfectly in Lucas's program and is in fact modeling from invariants.

Calibration techniques were imported by Kydland and Prescott (1982) as a means of quantifying their new-classical-equilibrium, real-business-cycle model. Since its introduction to economics, calibration has been controversial. One reason for this is its
ambiguous meaning. Kydland and Prescott introduced calibration as "the selection of parameter values for which the model steady-states values are near average values for the American economy during the period being explained" (1360). The selection itself was not further explicated, leaving the door open for different views.

One view sees calibration as method of estimation: "The method involves simulating a model with ranges of parameters and selecting elements from these ranges which best match properties of the simulated data with those of historical data" (Gregory and Smith, 1990: 57). Calibration considered as such is in fact "tuning" (see above). An alternative interpretation is that calibration is a test of a model. If there are no free parameters, then the comparison of a model's population moments (or perhaps some other population measure) with those of historical time-series can be thought of a test, namely a specific type of the previously mentioned "characteristics testing." If the correspondence between some aspect of the model and the historical record is deemed to be reasonable close, then the model is viewed as satisfactory. If the distance between population and historical moments is viewed as too great, then the model is rejected (see Gregory and Smith, 1991).

Both views of calibration are based on methodologies in which models are considered as simulation devices and other instruments of observation are available producing facts with which the quality of the model can be assessed. But if one considers models as an idiosyncratic measuring instrument, a third interpretation is needed—the one used in metrology. To understand what calibration means in the context of measuring instruments we first have to give an account of measurement itself. A "popular definition" from the *New Caxton Encyclopedia* states that measurement is:

the quantitative determination of a physical magnitude by comparison with a fixed magnitude adopted as the standard, or by mean of a calibrated instrument. The result of
measurement is thus a numerical value expressing the ratio between the magnitude under examination and a standard magnitude regarded as a unit . . . (quoted in Sydenham, 1979: 10).

Measurements are comparisons of the unknown with a standard. The standard is the means by which the unit of measurement is defined. A standard is the "material measure, measuring instrument, reference material or measuring system intended to define, realise, conserve or reproduce a unit or one or more values of a quantity to serve as a reference" (VIM, 6.1).4 The general philosophy adopted for the creation of standards is that they should be based upon some principle that is known to be as invariant as possible. Standards are entirely a human choice. Nothing about the natural world defines them; they are a contrivance of man. However, they are often based upon naturally occurring phenomena when these possess the required degree of definition.

The reliability of a measuring instrument is verified by comparing it against a common standard, a process called calibration. The VIM definition of calibration is the "set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material, and the corresponding values realized by standards" (6.11). In other words, it is a way of checking the accuracy of the measuring instrument.

An important problem that crops up in observing or measuring a phenomenon is how to distinguish between the facts of the phenomenon and the artifacts created by the instrument. Allan Franklin (1986) discusses nine epistemological strategies to distinguish between a valid observation and an artifact. One of these strategies is calibration, "the use of a surrogate signal to standardize an instrument." "If an apparatus reproduces known phenomena, then we legitimately strengthen our belief that the
apparatus is working properly and that the experimental results produced with that apparatus are reliable” (Franklin, 1997: 31).

The ambiguity surrounding calibration methods in economics was probably the reason the *Journal of Economic Perspectives* published a symposium on this subject in the winter 1996 issue. The journal gave Kydland and Prescott the opportunity to explicate the “tool” they used in their “Time to Build” paper (1982), which they now called “The Computational Experiment.” According to both authors, “any economic computational experiment involves five major steps: pose a question; use a well-tested theory; construct a model economy; calibrate the model economy; and run the experiment” (Kydland and Prescott, 1996: 70). The afterward question they seem to have posed in 1982 was, “What is the quantitative nature of fluctuations induced by technology shocks?” (1996: 71).

In explaining what they meant by calibration, the authors referred to the graduation of measuring instruments. “For example, a Celsius thermometer is calibrated to register zero degrees when immersed in water that contains ice and 100 degrees when immersed in boiling water. A thermometer relies on the theory that mercury expands (approximately) linearly within this range of temperatures” (1996: 74). Discussing business-cycle research, Prescott explicitly specified a model as “a measurement instrument used to deduce the implication of theory” (1998: 2). Quoting Lucas, he defined a theory as “an implicit set of instructions for constructing a model economy for the purpose of answering a question” so that the “quantitative answer to the question is deduced from the model economy” (2, 3). Comparing economic models with measuring instruments, Kydland and Prescott arrive at an interpretation of calibration that comes very close to the one given by Franklin above: “Generally, some economic questions have known answers, and the model should give an approximately correct answer to them if we are to have any confidence in the answer given to the question with unknown answer” (1996: 74). The answer to this latter question
was that “the model economy displays business cycle fluctuations 70 percent as large as did the U.S. economy” (74). In other words, the answer is supposed to be the measurement result carried out with a calibrated instrument.

But what are the economic questions for which we have known answers? Or: What are the facts with which the model is calibrated? The answer is explicitly given by Cooley and Prescott (1995). They described calibration as the following three-step process:

The first step is to restrict [equilibrium] processes to a parametric class. We stress the idea of using a model that is consistent with growth observations to study fluctuations. . . . The second step in this process is to construct a set of measurements that are consistent with the parametric class of models. . . . The third step is to assign values to the parameters of our models. This involves setting parameter values so that the behavior of the model economy matches features of the measured data in as many dimensions as there are unknown parameters. We observe over time that certain ratios in actual economies are more or less constant. We choose parameters for our model economy so that it mimics the actual economy on dimensions associated with long-term growth (15).

These “certain ratios” were the so-called stylized facts of economic growth, “striking empirical regularities both over time and across countries,” the “benchmarks of the theory of economic growth” (Cooley and Prescott, 1995: 3). Originally they were Kaldor’s “stylised facts” of growth (1978 [1958]), but the ones that were used in the real-business-cycle literature are those as characterized by Solow (1970).

Although we have seen that equilibrium business-cycle modelers aim to model from invariants, the choice to take these stylized facts as empirical facts of growth is dubious. Solow already remarked that “there is no doubt that they are stylized, though it
is possible to question whether they are facts" (1970: 2). The danger is that stylized facts may turn out to be more stylized than factual. Hacche provided an account of the British-American evidence relating to Kaldor’s six stylized facts and showed inconsistencies between economic history and Kaldor’s stylized facts:

the data for the United Kingdom provide little support for the hypothesis that there is some “steady trend” or “normal” growth rate of capital or output or both running through economic history—which is what Kaldor’s stylised facts suggest—unless the interpretation of the hypothesis is so liberal as to bear little meaning (1979: 278).

Calibration is thus understood in the real-business-cycle literature as a strategy to make the experimental result valid by taking care that it reproduces stable growth facts. But the instrument in the computational experiment was not reliable, because it was calibrated with dubious invariants.

5. Conclusions

The issue at stake here is whether we can build reliable instruments in economics—in other words, whether we can find invariants with which we can calibrate our models. But what are these “invariants,” or where can one localize stability? To real-business-cycle economists they must be found among the stylized growth facts. According to Lucas, the model invariance is located at the level describing “tastes and technology.” Frisch (1995 [1938]), in discussing the autonomy of Tinbergen’s macrodynamic model, suggested the interview method as a substitute for experimentation, a suggestion repeated 10 years later:

It is very seldom indeed that we have a clear case where the statistical data can actually determine numerically an autonomous structural equation. In most cases we only get
a covariational equation with a low degree of autonomy. . . .

We must look for some other means of getting information about the numerical character of our structural equations. The only possible way seems to be to utilize to a much larger extent than we have done so far the interview method, *i.e.*, we must ask persons or groups what they would do under such and such circumstances (Frisch, 1948: 370).

Frisch tended to localize stability on the microlevel. If economics is the study of a specific kind of behavior, then it may be worthwhile to localize stability in human nature. Stable characteristics are, for example, our need for food, the finiteness of our lives, etc.

On the other hand, it could be more fruitful to localize stability on the macrolevel, avoiding aggregation problems and, in contrast to the Cowles approach, to follow Zellner's advice (1997) to KISS: Keep It Sophistically Simple. Start with models as simple as possible and improve the model each time in the direction indicated with diagnostic checks of the properties of the model (Zellner, 1994). Stability is then found in simplicity, based on the Jeffreys-Wrinch simplicity postulate, which "suggest that simpler models will probably work better than complicated models" (Zellner, 1994: 218-219). But Simon's notion of artifacts also justifies simple mechanisms:

the first advantage of dividing outer from inner environment in studying an adaptive or artificial system is that we can often predict behavior from knowledge of the system's goals and its outer environment, with only minimal assumptions about the outer environment (Simon, 1969: 8).

An economy is a complex system, in the sense that it is a system made up of a large number of parts interacting in a nonsimple way. The Cowles approach tried to solve this problem of complexity by setting up a mathematical interdependent system of an entire economy without empirical content and then using econo-
metric techniques to give empirical flesh and blood to this system. Complexity was treated by building larger and more comprehensive models. One of the major efforts of the 1960s in this respect, building on the earlier work of Klein and Goldberger (1955), was the Brookings model (Dusenberry, Fromm, Klein, and Kuh, 1965). “This model was a joint effort of many individuals, and at its peak it contained nearly 400 equations. Although much was learned from this exercise, the model never achieved the success that was initially expected, and it was laid to rest around 1972” (Fair, 1992: 2). One can only wonder what “understanding” means in relation to such an enormous model, of which the builders only survey the part they have contributed and of which the overall dynamic behavior only can be revealed by computer simulations.

The appeal of the simple, elegant models Lucas started to develop in the 1970s was that they in their simplicity provided a kind of understanding one at least could communicate to students or colleagues. Lucas preferred to call them “mechanics,” “a system of differential equations the solution to which imitates some of the main features of the economic behavior we observe in the world economy” (Lucas, 1988: 39). In my view it would be better to call them “mechanisms,” defined by Machamer, Darden, and Craver as “entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (2000: 3). The understanding provided by a mechanistic explanation arises not from its correctness, but rather from an elucidative relation between the setup conditions and intermediate entities and activities and the termination condition or the phenomenon to be explained. “Mechanism descriptions show how possibly, how plausibly, or how actually things work” (Machamer, Darden, and Craver, 2000: 21). A mechanism can be communicated to others without the need to mention or explain matters from outside the mechanism, even more so when they are simple machines that “can be drawn or reproduced in a picture
or recipe book. Such things can be seen and made by everyone and anyone" (Machamer, 1998: 70).

To come back to the title of this paper, now put as a question—"How Do Economists Model the World into Numbers?"—my answer is that economists, after a century of mathematical modeling, now prefer very simple mechanisms with the faith that they will be calibrated in the future.

Notes

1A nice example of how the balance functioned as such in the work of William Stanley Jevons (1835-1882), founder of modern economics, is explored by Maas (2001). Morgan (1999) and Boumans (2001) show how in the work of Irving Fisher (1867-1947) the balance provided the appropriate mechanism for developing the Quantity Theory of Money.

2There is apparently a misprint in the original text (Tinbergen, 1935: 279). In the text I have reproduced the corrected version. It should be noted that an equation in which both the integral and the derivative term appears is equivalent to the equation describing harmonic oscillation:

\[ a\ddot{y}(t) + b\dot{y}(t) + c\int_{0}^{t} y(\tau) d\tau = 0 \]

\[ \rightarrow \quad a\ddot{y}(t) + b\dot{y}(t) + cy(t) = 0 \]

3For valuable histories of econometrics, see Morgan (1990) and Qin (1993).

4VIM is an abbreviation of the “International Vocabulary of Basic and General Terms in Metrology.” This document was prepared and published by the ISO.

References


Boland, Lawrence A. “A Critique of Friedman’s Critics.” *Journal of Economic Literature* 17 (June 1979): 503-522.


