

Calibration of Models in Experiments

Marcel Boumans

*Department of Economics, University of Amsterdam, Amsterdam, Netherlands,
boumans@fee.uva.nl*

Abstract: The assessment of models in an experiment depends on their material nature and their function in the experiment. Models that are used to make the phenomenon under investigation visible - sensors - are assessed by calibration. However, calibration strategies assume material intervention. The experiment discussed in this paper is an experiment in economics to measure the influence of technology shocks on business cycles. It uses immaterial, mathematical instruments. It appears that calibration did not work for these kinds of models, it did not provide reliable evidence for the facts of the business cycle.

1. INTRODUCTION

The way in which models used in experiments can be assessed depends on at least two characteristics: their material nature and their function in the experiment. The kind of materiality not only determines the nature of control and inference in the experiment, but also the confidence one can have in the experiment's outcomes (Boumans and Morgan, 2001; Morgan, 2000). Traditionally, models are defined in terms of their logical and semantic connections with theories. So, usually no methodological distinction is made between the assessment of models and theories. However, by answering the question "What role do models play?" Morrison and Morgan (1999) showed that models function as autonomous agents, that is they are partially independent of both theories and the world, and therefore can be used as instruments of investigation in both domains. Hence, models should be assessed as

Model-Based Reasoning: Science, Technology, Values, edited by
L. Magnani and N.J. Nersessian, Kluwer Academic/Plenum Publishers, New York, 2002

instruments and not as theories. Moreover, the kind of assessment depends on the specific function the particular model has in the experiment.

This paper discusses two different models used in Kydland and Prescott's (1996) "Computational Experiment" to measure the percentage of the business cycle fluctuations caused by technology shocks. Borrowing Morgan's (2000) taxonomy, their experiment can be considered as a "virtual experiment": a non-material experiment on non-material objects. Both models are used as sensors to make facts about the business cycle phenomenon visible. One sensor functions as a filter, the other as a measuring instrument. Although business cycles are detected through the use of data, they are in general not directly observable, which in general applies to all phenomena, that is all potential objects of explanation by a theory (Bogen and Woodward, 1988; see below). In Kydland and Prescott's experiment, the phenomenological facts are unobservable components of the observed data. To display them we need a filter. The filter's role is to extract a specific signal from the available data by removing noise and other irrelevant components. The role of the measuring instrument is more specific, namely to generate quantitative facts about the phenomenon. Considering models as "mediating" instruments between theory and the world (Morrison and Morgan 1999), both sensors are located on the theory-world axis as "mediators" between facts about the phenomenon and data.

Theory \rightarrow Facts about Phenomenon $x \leftarrow$ SENSORS \leftarrow Data $y = x + \text{noise}$

To understand this position both instruments take on the theory-world axis, it is helpful to use the distinction Woodward (1989) makes between data and phenomena. Phenomena are relatively stable and general features of the world which are potential objects of explanation and prediction by general theory. Data, the observations, play the role of evidence for claims about phenomena. The contrast between data and phenomena can be characterized in three ways. In the first place, the difference between data and phenomena can be indicated in terms of the notions of error applicable to each. In the case of data the notion of error involves observational mistakes, while in the case of phenomena one worries whether one is detecting a real fact rather than an artifact produced by peculiarities of one's instruments or detection procedures. A second contrast between data and phenomena is that phenomena are more "widespread" and less idiosyncratic, less closely tied to the details of a particular instrument or detection procedure. A third way of thinking about the contrast between data and phenomena is that scientific investigation is typically carried on in a noisy environment, an environment in which the observations reflect the operation of many different causal factors.

The problem of detecting a phenomenon is the problem of detecting a signal in this sea of noise, of identifying a relatively stable and invariant pattern of some simplicity and generality with recurrent features - a pattern which is not just an artifact of the particular detection techniques we employ or the local environment in which we operate (Woodward, 1989, pp. 396-7).

Underlying the contrast between data and phenomena is the idea that theories do not explain data, which typically will reflect the presence of a great deal of noise. Rather, an investigator first subjects the data to analysis and processing, or alters the experimental design or detection technique, all in an effort to separate out the phenomenon of interest from extraneous background factors. "It is this extracted signal rather than the data itself which is then regarded as a potential object of explanation by theory" (p. 397).

Instruments located between data and phenomena on the theory-world axis are not assessed in the same way as theories are by confronting the outcome of the model with facts about the phenomenon. Confronting sensors with phenomenological facts rather means comparing these instruments with other instruments that generate these facts and are chosen to act as standards. In other words, assessment of filters and measuring instruments is done by calibration, that is standardization in the sense of comparing the instrument's performance with a standard. But - as will be shown - the validation problems of filters differ from those of measuring instruments. Both problems were caused by the lack of materiality, but in one case with respect to the standard, and in the other case with respect to the instrument itself.

2. ROBERT LUCAS'S PROGRAM

Kydland and Prescott's experiment was an implementation of Lucas's program for business-cycle research. This program not only framed their account of models and theories, but also advanced the view that business cycles should be considered as phenomena in the above-described meaning of the word.

Lucas's contribution to macroeconometric evaluation of economic policy - the "Lucas critique" - implied that estimated parameters which 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. The Lucas critique was an implicit call for a new research program. This program involves formulating and estimating macroeconometric models with parameters that are invariant under policy variations, so that they can be used

for evaluating alternative policies. According to Lucas, the model invariance is located on the level of the parameters describing tastes and technology. Lucas was the first who emphasized stable parameters instead of stable relationships.

Lucas characterized the business cycle by enumerating seven “qualitative features” of economic time series:

- i) Output movements across broadly defined sectors move together. [...]
- ii) Production of producer and consumer durables exhibits much greater amplitude than does the production of nondurables. iii) Production and prices of agricultural goods and natural resources have lower than average conformity. iv) Business profits show high conformity and much greater amplitude than other series. v. Prices generally are procyclical. vi. Short-term interest rates are procyclical; long-term rates are slightly so. vii). Monetary aggregates and velocity measures are procyclical (Lucas, 1977, p. 9).

By defining the business cycle in this way, Lucas indicated that he considered the business cycle as a general phenomenon of capitalist economies:

There is, as far as I know, no need to qualify these observations by restricting them to particular countries or time periods: they appear to be regularities common to all decentralized market economies. Though there is absolutely no theoretical reason to anticipate it, one is led by the facts to conclude that, with respect to the qualitative behavior of co-movements among series, business cycles are all alike. To theoretically inclined economists, this conclusion should be attractive and challenging, for it suggests the possibility of unified explanation of business cycles, grounded in the general laws governing market economies, rather than in political or institutional characteristics specific to particular countries or periods (Lucas, 1977, p. 10).

Borrowing this definition from Lucas, Prescott (1986) preferred to refer to business cycles as “business cycle phenomena”, “which are nothing more or less than a certain set of statistical properties of a certain set of important aggregate time series” (p. 10). By explicitly treating the business cycle as a general phenomenon not restricted to particular countries or time periods, the business cycle was stabilized and its “qualitative features” considered as “stylized facts”.

Lucas also introduced 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 be-

haves through time so as to imitate closely the time series behavior of actual economics (Lucas, 1977, p. 11).

In the “equilibrium business-cycle program” dominated by Lucas’ instructions, it became standard practice to run an experiment with an artificial economy:

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 costs (Lucas, 1980, p. 696).

But not all models will be equally useful, to test them as “useful imitations of reality” we should subject them to shocks “for which we are fairly certain how actual economies, or parts of economies, would react. The more dimensions on which the model mimics the answer actual economies give to simple questions, the more we trust its answer to harder questions” (Lucas, 1980, pp. 696-7).

3. CALIBRATION

If theories do not provide systematic explanations of facts about data, as Bogen and Woodward (1988) convincingly have argued, how do data constitute reliable evidence for some phenomenon? Because facts about phenomena are actually created by instruments for which data function as input, assessing reliability implies the investigation of these instruments. Beside this strategy, Woodward (1989) also mentions other possibilities to increase reliability, but these are not applicable in the case of business-cycle research: control of possible confounding effects and systematic error, replicability, and data reduction. Control and replication are impossible in a macroeconomic environment and data reduction is a luxury economists cannot afford. In the case of business cycle research, reliability can only be achieved by investigation of the equipment used. Because Woodward only discusses material instruments, he confines himself to the empirical investigation of the equipment, and calibration is seen as falling under this heading. But the experiment discussed here is “virtual” and thus empirical investigation is not possible. Fortunately, calibration is defined in such a broad sense that it does not exclude the investigation of nonmaterial instruments. This broader meaning will be used for the validation of the two sensors. But one should be warned, calibration does not guarantee a correct result; though its successful performance does argue for the validity of the result (Franklin, 1997, p. 76).



<http://www.springer.com/978-0-306-47244-2>

Model-Based Reasoning

Science, Technology, Values

Magnani, L.; Nersessian, N. (Eds.)

2002, XV, 404 p., Hardcover

ISBN: 978-0-306-47244-2