Federal Reserve Bank of Minneapolis
Research Department Staff Report 178

August 1994

The Computational Experiment: An Econometric Tool

Finn E. Kydland*

University of Texas, Austin

Edward C. Prescott*

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

An economic experiment consists of the act of placing people in an environment desired by the experimenter, who then records the time paths of their economic behavior. Performing experiments that use actual people at the level of national economies is obviously not practical, but constructing a model economy and computing the economic behavior of the model's people is. We refer to such experiments as *computational experiments* because the economic behavior of the model's people is computed. In this essay, we specify the steps in designing a computational experiment to address some well posed quantitative question. We emphasize that the computational experiment is an econometric tool used in the task of deriving the quantitative implications of theory.

*We thank Graham Candler, Tryphon Kollintzas, and Jim Schmitz for helpful discussions and the National Science Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Introduction

An economic experiment consists of the act of placing people in the environment desired by the experimenter, who then records the time paths of their economic behavior. Performing experiments that use actual people at the level of national economies is obviously not practical, but constructing a model economy and computing the economic behavior of the model's people is. We refer to such experiments as *computational experiments* because the economic behavior of the model's people is computed.[1] The computational experiment has become invaluable in quantitative aggregate economic theory. It is being used, for example, to estimate the quantitative effects of trade liberalization policies, the welfare consequences of changes in the tax system, and the magnitude and nature of business cycle fluctuations induced by different types of shocks.

One question that has arisen is whether or not the computational experiment is actually an econometric tool.[2] It is not an econometric tool in the modern (narrow) sense of the term in that it is not used in the "measurement of economic relations" (Marshack 1948, p. 1). Yet it is an econometric tool in the original sense of the term (which we prefer) in that it is used to derive the quantitative implications of economic theory (Frisch 1933a, p. 1). We do not enter into this semantic debate here.[3] Instead, we review the use of the computational experiment in economics, noting that the task of deriving the quantitative implications of theory differs from that of measuring economic parameters.

Computational experiments are not unique to economic science—they are heavily used in the physical sciences. Even though computational experiments in economics are similar to those in the physical sciences, they differ in one crucial respect. Unlike theory in the physical sciences, theory in economics does not provide a law of motion or, in the case of uncertainty, a Markov process governing the evolution of the system. Rather, economic theory provides a specification of people's ability and willingness to substitute among commodities. Consequently, computational experiments in economics include the additional step of computing the allocation process in which all the model's people behave in a way that is in each person's best interest—that is, economists must *compute* the equilibrium law of motion or process of the model economy. Given the process governing the system, the next and final step in both economic and physical science is to use the computer to compute realizations of this process. If the model is deterministic, only one possible equilibrium realization exists for the path of the model economy. If the model economy has aggregate uncertainty (as it must if the phenomena of interest are business cycle fluctuations), the realization is random. In the case of uncertainty, the computer can be used to generate any number of independent realizations of the equilibrium stochastic process, and these realizations along with statistical estimation theory are used to measure the sampling distribution of any finite set of statistics to any degree of desired accuracy.

Several theoretical developments over the past 30 or 40 years have been crucial in making the economic computational experiment feasible in cases which involve uncertain intertemporal behavior. Among these developments is statistical decision theory, which provides a consistent way for people to make decisions under uncertainty. Another important development is the Arrow-Debreu general equilibrium theory, which extends equilibrium theory to uncertain environments. Also important is the development of recursive methods for the study of economic dynamics because these methods allow the economist to use the computational experiment to generate time series disciplined by factual studies. (See Stokey and Lucas 1989.) With these methods, the elements being computed are decision or policy rules that describe, for a given environment, the decisions made by rational individuals as functions of a suitably defined state of the economy. Typical elements of the state vector, in addition to sufficient statistics for forecasting shocks to the economy, are stocks of various sorts, such as productive capital in the business sector, human capital, consumer durables, and inventories. Once the equilibrium decision rules have been computed, the equilibrium aggregate behavior of the model economy is fully described; the experimenter can then generate as many equilibrium realizations, in the form of model time series, as are needed to answer the posed question to the desired accuracy. This methodological framework facilitates bringing to bear factual knowledge about the actual economy and enables the researcher to produce time series that correspond to reported statistics, such as those reported in national income and product accounts (NIPA).

Steps in an Economic Computational Experiment

Pose a Question

The purpose of a computational experiment is to derive a quantitative answer to some well-posed question. Judging whether an experimenter's model economy is a good abstraction can only be done relative to the posed question. Examples of the types of questions that computational experiments address are as follows:

(i) What are the welfare consequences of policy A relative to those of policy B? (For example, researchers have explored the quantitative welfare consequences of alternative monetary policies.)

(ii) How much of fact X is accounted for by factor Y? (For example, researchers might examine the contribution of different sources of impulse to business cycle fluctuations.)

(iii) Does established theory display quantitative feature Z? [For example, researchers might ask whether standard theory, when extended to more than one country, displays the J-curve pattern of covariance between the terms of trade and the trade balance (Backus, Kehoe, and Kydland 1994).]

(iv) Does the introduction of feature F into a standard model for a particular class of phenomena account for part of deviation D from standard theory and, if so, for how much of this deviation? [For example, many researchers have tried out alternative features in an attempt to account for the equity premium puzzle demonstrated by Mehra and Prescott (1985).]

Use Well-Tested Theory

To carry out a computational experiment, a researcher needs some strong theory—that is, some theory that has been tested through use and found to provide reliable answers to a class of questions. Modern business cycle theory builds upon the neoclassical growth framework. This framework has served well when dealing with growth within reasonably stable economic institutions. It has been used to address public finance as well as business cycle questions. We emphasize that it has not been successful in addressing *all* aggregate issues. In particular, it fails spectacularly when used to address economic development issues.

Neoclassical growth theory represents a good example of the importance of interaction between factual studies and theory development. Solow (1970) lists several growth facts that influenced the development of neoclassical growth theory. Once the main ingredients of the theory, such as the production function, were established, new light was thrown upon the data. Business cycle models are stochastic versions of neoclassical growth theory. This theory's implication that the economy should display business cycle fluctuations of the quantitative nature observed in response to technology, public finance, and terms-of-trade shocks dramatically adds to our confidence in the answers it provides to public finance questions.

One definition of *theory* is "a formulation of apparent relationships or underlying principles of certain observed phenomena which has been verified to some degree" (Guralnik 1978, p. 775). Neoclassical growth theory certainly satisfies that criterion. Central to this theory is its description of aggregate production possibilities, with the output of goods resulting from the input of labor and capital. With an explicit description of the household sector (including its focus on the time-allocation decision), the neoclassical growth model becomes an internally consistent framework for addressing business cycle questions, as well as other questions of interest to macroeconomists.

Construct a Model Economy

The amount of detail included in a model economy depends on the question being addressed as well as the feasibility of computing the equilibrium process. Often the experimenters are constrained to deal with a much simpler model economy than they would like because computing the equilibrium is impossible, given currently available tools. This situation is no different from that in the physical sciences, where, as in economics, the computational experiment has become accepted as an invaluable scientific tool. In his overview of climate modeling, Schneider (1987, p. 72) states

> Although all climate models consist of mathematical representations of physical processes, the precise composition of a model and its complexity depend on the problem it is designed to address.

And later (p. 72):

> Often it makes sense to attack a problem first with a simple model and then employ the results to guide research at higher resolution.

In the physical sciences, as in economics, confidence in a particular framework or approach is gained through successful use.

So far, most of the model environments that economists have used share certain characteristics. The environments are inhabited by a large number of people whose decision problems are described explicitly. Both the household sector and business sector play a central role. For some questions, government or foreign sectors must be included as well. That everyone is alike is a reasonable abstraction for some purposes but not for others. Some questions (such as those for which demographic changes are important) dictate that abstractions with heterogeneous people be used. For example, heterogeneity is crucial in the Auerbach and Kotlikoff (1987) model to predict the consequences of the population's changing age structure on savings. At the same time, as Ríos-Rull (1993) demonstrates, these same features, even when combined with elements of market incompleteness, are not quantitatively important to business cycle findings regarding issues such as the contribution of technology shocks to business cycle fluctuations. We reemphasize that an abstraction can only be judged relative to some given question. To criticize or reject a model because it is an abstraction is foolish. All models are necessarily abstractions and therefore false.

While it obviously must be computable, a model environment must be selected based on the question being addressed. Model-economy selection should not depend on the answer provided. Moreover, searching within some parametric class of economies for the one that best fits a set of aggregate time series makes little sense. Thinking of interesting questions for which such a practice would provide an answer is difficult. For example, if the question is of the type, How much of fact X is accounted for by Y, then choosing the parameter values in such a way as to make the amount accounted for as large as possible according to some metric makes no sense. A model economy is obviously an abstraction and, by definition, false. With enough data, statistical hypothesis-testing almost surely will reject any model along some dimension. A model is useful insofar as it provides a quantitative answer to an interesting question. A given model may be appropriate for some question (or class of questions) but not for others. Consequently, a model economy can only be judged relative to the question it is being used to answer.

We will not debate the legitimacy of these methods. Such debates generally serve to define schools rather than produce agreement. They are almost nonexistent during normal science but tend to recur during scientific revolutions. As stated by Kuhn (1962, p. 145), "Few philosophers of science still seek absolute criteria for the verification of scientific theories." Using probabilistic verification theories that ask us to compare a given scientific theory with all others that might fit the same data is a futile effort. We agree with Kuhn (p. 146) that "probabilistic theories disguise the verification situation as much as they illuminate it." All historically significant theories have agreed with the facts, but only more or less. No more precise answer can be found to the question of how well an individual theory fits the facts.

Quantitative economic theory uses theory and measurement to estimate how big something is. For this purpose, a researcher needs an instrument or apparatus. In our case, the instrument is a computer program that computes the equilibrium process of the model economy and generates realizations of the equilibrium process. The computational experiment, then, is the act of using this instrument, usually for the purpose of finding a quantitative answer to some specific question.

Calibrate the Model Economy

Before the computational experiment can be executed, the model must be calibrated. Note that calibration is not estimation. *Estimation* is the determination of the approximate quantity of something. Quantitative theory, therefore, is estimation in the sense that the quantitative answer to a posed question is an estimate. For example, quantitative theory is used to measure the welfare implications of alternative tax policies. A related, but fundamentally different, activity is using statistical decision theory to estimate the magnitude of some economic parameter that is important in an established economic theory.

Estimation . . .

To estimate a parameter, a researcher looks for a situation in which the signal-to-noise ratio is high. Using the existing data and some theory, the researcher constructs a probability model. An estimator is developed which, relative to the parameter that is to be estimated, is robust to the questionable features of the maintained hypothesis. Good estimates of key parameters are used when constructing a model economy to yield a quantitative answer to a posed question.

. . . *Versus Calibration*

Calibration is a very different activity. Originally, in the physical sciences, *calibration* referred to the graduation of measuring instruments. For example, a thermometer is calibrated to register 0 when immersed in water that contains ice and 100 when immersed in boiling water. The following theory is used: Mercury expands approximately linearly within this range of temperatures. This theory also tells us how to recalibrate the thermometer if the measurements are made in Denver or in Lima rather than at sea level. In a sense, model economies, like thermometers, are measuring devices. In physics, they are artificial physical systems or models that are used to estimate quantitatively what will happen under different contingencies. Generally, some questions have known answers, and the model should give an approximately correct answer if we are to have any confidence it. The model systems are calibrated so that this happens. Some of the model's parameters may have to be varied until the model system mimics reality on some key dimensions. In the physical sciences, this activity has come to be called *calibration.* Since this task is not an attempt at assessing the size of something, it is not estimation.

Note that the goal of a computational experiment is not to try to match correlations. In other words, the criterion for choosing parameter values is not how close model correlations are to those in the data. In some cases, a discrepancy between a correlation in the data and the corresponding one in the model provides additional support for the answer. One example is the cyclical hours-productivity correlation. A model economy with only technology shocks as an impulse will display a high correlation between these two variables. If the question is, What fraction of the cycle has been accounted for by such shocks, and the answer is that the fraction is substantially less than one, then a low correlation in the data is crucial in confirming the answer. All other sources of shocks will lead to movements along declining marginal-product-of-labor schedules. Because the capital stock varies little over the cycle, such sources of impulse induce labor input and productivity movements which are in the opposite direction. Thus, the empirical correlation being different in a particular way from that in the model economy provides additional support for the quantitative answer to the question about the role of technology shocks.

Run the Experiment

To place the model's people in the desired experimental environment, we describe the economy in the form of a computer program. Under the neoclassical framework, the parameters are those describing preferences, technology, information sets, and institutional arrangements, including policy rules.

The Computational Experiment in Business Cycle Research

Business Cycle Questions

We follow Lucas (1977) in regarding business cycles as movements about trend in gross national product and business cycle regularities as comovements about trend in different aggregative time series with gross national product. Business cycle theory is largely concerned with estimating the contributions of various factors to these fluctuations.

Obviously, we can find many ways to quantitatively characterize the cyclical properties of economies. A method that has proven useful is one in which researchers fit a smooth curve through the time series and then examine the second moments of the time series' deviations from its smooth component. The view in the 1970s was that one set of factors (most likely monetary shocks) were behind the cyclical component and that an entirely different set of factors accounted for the movement of the growth component. This view motivated Hodrick and Prescott (1980) to use standard curve-fitting techniques to define a growth component as being the curve that best fits a time series in a least-square sense, subject to a penalty on the sum of the second differences squared. The larger this penalty parameter, the smoother the fitted curve. For quarterly series, they found that a penalty parameter of 1600 made the fitted curve mimic well the one that business cycle analysts would draw. But given the unanticipated finding that these features of the data, which we label *business cycle fluctuations,* are quantitatively just what neoclassical growth theory predicts, these deviations are nothing more than well-defined statistics. We emphasize that given the way the theory has developed, these statistics measure nothing. As is clear from the above discussion, business cycle theory treats growth and cycles as being integrated, not as a sum of two components driven by different factors. For that reason, talking about the resulting statistics as imposing spurious cycles makes no sense. The Hodrick-Prescott filter is simply a statistical decomposition that summarizes in a reasonable way what happens at business cycle frequencies. It has been used as a way of presenting the findings and judging the reliability of the answer, as well as a way of demonstrating remaining puzzles or anomalies relative to theory.

The Theory Used in Model Selection

The basic theory used is the neoclassical growth model. A key construct in this theory is the aggregate production function F, with inputs of capital K and labor H. This function specifies the maximum aggregate output, which is divided between consumption C and investment I. This constraint is

(1) $$C_t + I_t \leq A_t F(K_t, H_t)$$

where A_t is the technology parameter that grows at random rates. The neoclassical aggregate production function displays constant returns to scale, and under the assumption that factors are paid their marginal product, we obtain the NIPA identity that gross national product and income are equal: $C + I = wH + rK$, where w and r are factor rental prices. In the model economy the depreciation of capital is proportional to the capital stock with proportionality constant δ. Thus

(2) $$K_{t+1} = (1-\delta)K_t + I_t.$$

To complete the specification of technology, the Markov process generating the technology parameter must be specified. Given that a Markov structure which displays persistence is needed, we assume that

$$(3) \qquad A_{t+1} = \rho A_t + \epsilon_{t+1}$$

where ρ is large but less then one and the shocks ϵ_{t+1} are identically and independently distributed. The technology described by equations (1)–(3) specifies people's ability to substitute.

Also needed for a fully specified economy is a specification of people's willingness to substitute between consumption and leisure, both intertemporally and intratemporally. For this purpose our model economy has a stand-in household with utility function

$$(4) \qquad E \sum_{t=0}^{\infty} \beta^t U(C_t, 1-H_t)$$

where we normalize so that market and nonmarket productive time add to one. For simplicity, we assume that the household owns the capital stock directly. For a complete specification of the economy, values of the parameters β, δ, and ρ are needed as well as the explicit utility function U, the aggregate production function F, and the distribution of the shocks to technology ϵ_{t+1}. The final required element is an equilibrium concept. The one used is the *competitive equilibrium,* which equates marginal rates of substitution and transformations to price ratios.

Through this theory, business cycle theorists make contact with other fields of economics. Macroeconomics is no longer largely separate from the rest of economics. The utility and production functions that are used by public finance researchers (see, for example, Auerbach and Kotlikoff 1987) are almost the same as those used by business cycle theorists. The introduction of household production (see, for example, Benhabib, Rogerson, and Wright 1991 and Greenwood and Hercowitz 1991) illustrates the close connection with the work of labor economists. The connection with international trade (see, for example, Backus, Kehoe, and Kydland 1994) is another example.

For some questions, the set of techniques or methods that are currently compatible with computable models severely restricts the creation of useful model environments. The development of appropriate methods must therefore be given high priority. For example, a difficult methodological problem in aggregate economics is to analyze dynamic equilibriums in which the people are heterogeneous along dimensions that are key for the issue studied. For business cycle questions that have been addressed to date, little evidence exists that demographic factors play much of a role. For some other questions, however, demographic movements are at the heart of the issue. Thus, a whole new class of models must be created in which heterogeneity is incorporated. The set of computable general equilibrium models in this category has expanded dramatically over the past few years. (See Ríos-Rull 1994 for an overview.)

Calibration

Growth Facts

Often calibration involves the simple task of computing a few averages. For example, if the standard Cobb-Douglas production function is used [that is, we let $F(K,H) = K^{1-\theta}H^{\theta}$], then a numerical value for the parameter θ can be obtained by computing the average labor share of total output over a period of years. Several other growth relations map more or less directly into parameter values for typical models within the neoclassical growth framework, at least if they have been formulated with calibration in mind. As a consequence, computational experiments replicate the key long-term or growth relations among model aggregates.

Most growth relations have not changed much, on average, from one cycle to the next for several decades. Exceptions exist, however. The inventory stock as a fraction of gross national product (GNP) has declined steadily. Durables expenditures as a fraction of total output have risen. Depending on the associated pattern in the corresponding relative price, such information often enables the researcher to obtain a quite precise estimate of some elasticity of substitution. At the same time, abstracting from the difference in growth rates of the same two quantities may be acceptable if that feature is not likely to play a role for the answer to the question posed.

A good example is the fact that per household hours of work is about the same now as it was four decades ago in spite of a large rise in the real wage rate over the same period. This fact indicates that the elasticity of substitution between consumption and nonmarket time is near one. Still, many business cycle models abstract from the long-run productivity growth that is required to imply this sort of wage growth. The reason they do this is that the answer to the questions addressed in those studies would have been the same. For example, Hansen (1986) compares otherwise identical model economies and permits growth in one version and not in the other. The model without growth needs a slight adjustment in the capital depreciation rate in order to be calibrated to the investment share of output and the observed capital/output ratio. With this adjustment, both models estimate the same role of Solow residuals for cyclical fluctuations.

Panel Averages

Because these model economies are populated by people, other data used in calibration are averages across large numbers of the relevant population in the actual economy. For example, some model environments employ a utility function in consumption and leisure that, like the production function above, has a share parameter. The approximate empirical counterpart turns out to be the average fraction of time spent in market activity. This fraction, in principle, can be obtained from panel data for large samples of individuals. An example of a careful measurement study is Ghez and Becker (1975). To carry out such a study, a researcher needs to make choices about a variety of issues. What should be the upper and lower age limits for the people to be included? What is a reasonable definition of the total time allocated to market and nonmarket activities? For business cycle models, at least, the choice by Ghez and Becker to exclude time for both sleep and personal care is a reasonable one.

Definition of Variables

Even in the computations of growth relations, the empirical definition of particular variables in relation to the model economy may depend on the question. For example, both Benhabib, Rogerson, and Wright (1991) and Greenwood and Hercowitz (1991) consider household production in addition to market production, but the two studies are motivated by somewhat different questions. Both use capital and labor as inputs in nonmarket production. Benhabib, Rogerson, and Wright divide the time allocation into three uses: market and nonmarket production time and leisure time. The model is designed to capture the household decision to combine its labor with machines, such as stoves and washing machines, to produce household consumption services. They argue that houses do not need to be combined with labor, at least not to the same extent that household machines do. Consequently, they exclude housing capital from their concept of household capital. Greenwood and Hercowitz, on the other hand, distinguish only between market and nonmarket time and include the stock of housing, along with consumer durables, in their concept of household capital. To be consistent, they then subtract gross housing product (the measure of the service flow from the economy's housing stock) from GNP and add it to the consumer durables component of personal consumption expenditures.

Purposeful Inconsistencies

In calibration we sometimes make the model economy inconsistent with the data on one dimension so that it will be consistent on another. For example, İmrohoroğlu (1992) explores the welfare consequences of alternative monetary arrangements in worlds in which agents are liquidity-constrained. Cooley and Hansen (1989) explore the welfare consequences in worlds in which people use money for transaction purposes. These are two very different environments, each of which abstracts from the main feature of the other. İmrohoroğlu calibrates her model economy to yield a stock of money held per household in line with U.S. observations. In her model, however, people hold money only because they do not have access to an insurance technology to insure against randomness in the market value of their time. Equivalently, if they do have access to such an insurance technology, they find it so costly that, in equilibrium, they do not employ it. This is the only reason, in her model, for people to hold money; if she had calibrated the model to the amount of variation in individual income found in panel data, the model would have implied that average household holdings of liquid assets were about half of those actually held.

Of course, households have other reasons for holding liquid assets that earn much less than the average return on physical capital. Households accumulate savings so that they will be in a position to make a down payment on a house at some future date. They hold liquid assets as a substitute for insurance against sickness and accidents. They hold some for transaction purposes, as in the Cooley-Hansen environment. These and other factors are abstracted from in the İmrohoroğlu world, which led her to introduce greater variation in the market value of households' time so as to make per capita holdings of money in the model match actual holdings. This calibration is reasonable, given the question she addresses. Her implicit assumption is that it is unimportant which liquidity factor gives rise to these holdings. Subsequent research will either support this working hypothesis or disprove it and, in the process, lead to better model economies for evaluating monetary and credit policy arrangements. This sequence is how economic science progresses.

Running Experiments

With explicit functional forms for the production and utility functions in relations (1) and (4), with values assigned to the parameters, and with a probability distribution for the shocks, a researcher can use this economy to perform computational experiments. The objects that need to be computed first are the aggregate equilibrium decision functions $C(K_t,A_t)$, $I(K_t,A_t)$, and $H(K_t,A_t)$. In other words, the decisions are viewed as functions of the list of state variables that provide sufficient information about the position of the economy. The computer needs these three decision functions along with the two laws of motion for the state variables and the probability distribution for the shocks to generate time series for this model economy. For each t, given K_t and A_t inherited from period $t - 1$, the values are computed for C_t, I_t, and H_t from the decision functions, and the computer's random number generator makes a draw from the distribution for ε and updates the two state variables for period $t + 1$ using the laws of motion.

For a given model environment, these steps can be repeated for the desired number of periods. By making the time series long enough, a researcher could determine, with any degree of accuracy, the long-run probability distribution of the model's decision and state variables. A more useful approach, however, when making a comparison with actual data over a time period of a particular length (say, T periods) is to determine from the model economy the sampling distribution of T-period samples. In other words, each model time series is T periods long, and the computer produces multiple independent samples of that same length. A researcher can then determine the sampling distribution of the model statistics that characterize its cyclical properties. This information helps to assess the reliability of the quantitative answer obtained from a particular set of experiments. Sometimes we may say that the model mimics well on some dimension and point out that the value of some statistic for the actual economy is not far from the center of support of the sampling distribution of the corresponding statistic for the model economy.

We usually have some idea of what we think is a large deviation relative to some use of the model economy. If the deviation is quantitatively big, and this assessment takes into consideration the fact that the model statistic has a sampling distribution, then the answer provided by the model economy is less trustworthy.

Business Cycle Applications

Contribution of Technology Shocks[4]

A source of shocks suggested as far back as in work by Wicksell (1907) is fluctuation in technological growth. In the 1960s and 1970s, this source was dismissed by many as being unlikely to play much of a role in the aggregate. Most researchers accepted that considerable variation could exist in productivity at the industry level, but they believed that industry-level shocks would average out in the aggregate. During the 1980s, however, technology shocks gained renewed interest as a major source of fluctuations, supported largely by quantitative economic theory. So the question addressed was, How much would the U.S. postwar economy have fluctuated if technology shocks had been the only source of fluctuations?

Our selection of a model economy to address this question follows (see Kydland and Prescott 1982). We began by extending the neoclassical growth model to include leisure as an argument of the stand-in household's utility function. Given that more than half of business cycle fluctuations are accounted for by variations in the labor input, introducing this element was crucial. We then calibrated the deterministic version of the model so that its consumption-investment shares, factor income shares, capital/output ratios, leisure/market time shares, and depreciation shares matched the average values for the U.S. economy in the postwar period. Throughout this analysis, constant elasticity structures were used. Since uncertainty is crucial to the question, computational considerations led us to select a linear-quadratic economy whose average behavior is the same as the calibrated, deterministic, constant elasticity of the substitution economy.

We abstracted from public finance considerations and consolidated the public and private sectors. We introduced Frisch's (1933b) assumption of time to build new productive capital. The construction period considered was four quarters, with new capital becoming productive only upon completion, but with resources being used up throughout its construction. Given the high volatility of inventory investment, inventory stocks were included as a factor of production. We found, using the variance of Solow residuals estimated by Prescott (1986), that the model economy's output variance was 55 percent as large as the corresponding variance for the U.S. economy in the postwar period.

In the early 1980s, much discussion occurred in the profession about the degree of aggregate intertemporal substitution of leisure. Many felt that this elasticity had to be quite high in order for a market-clearing model to account for the highly volatile and procyclical movements in hours. The discussion may have started with the famous paper by Lucas and Rapping (1969). Realizing that the standard utility function implied a rather small elasticity of substitution, they suggested that past leisure choices might directly affect current utility. Being sympathetic to that view, we also considered a non-time-separable utility function as a tractable way of introducing this feature. When lags on leisure were considered, the estimate of how volatile the economy would have been if technology shocks were the only disturbance increased from 55 to near 70 percent. But until more support exists for this alternative preference structure, we rely on estimates obtained using the economy with a time-separable utility function. Unlike the system-of-equations approach, here the model economy that better fits the data is not the one used. Rather, currently established theory dictates which one is used.

Probably the most questionable assumption of this theory is that of homogeneous workers, with the additional implication that all variation in hours occurs in the form of changes in hours per worker. According to aggregate data for the U.S. economy, only about one-third of the quarterly fluctuations in market hours are of this form, while the remaining two-thirds arise from changes in the number of workers (see Kydland and Prescott 1990, Table 1).

This observation led Hansen (1985) to introduce the Rogerson (1988) labor indivisibility construct into a business cycle model. In the Hansen world, all fluctuations in hours are in the form of employment variation. To deal with the apparent nonconvexity arising from the assumption of indivisible labor, Hansen makes the problem convex by assuming that the commodity points are contracts in which every agent is paid the same amount whether that agent works or not and that a lottery randomly chooses who actually works in every period. He finds that with this labor indivisibility, his model economy fluctuates as much as did the U.S. economy. Our view is that with the extreme assumption of only fluctuations in employment, Hansen overestimates the amount of aggregate fluctuations accounted for by Solow residuals in the same way that the equally extreme assumption of only fluctuations in hours per worker led to an underestimation.

In Kydland and Prescott (1991b), the major improvement on the 1982 version of the model economy is that variation is permitted in both the number of workers and the number of hours per worker. The number of hours in which a plant is operated in any given period is endogenous.

Because the cost of moving workers in and out of the labor force is not included, a property of the equilibrium is that all the hours variation is in the form of employment change and none in hours per worker. In that respect, the Kydland and Prescott (1991b) model is identical to Hansen's (1985) model. For the economy with no moving costs, the estimate is that Solow residuals account for about 90 percent of the aggregate output variance. For the economy with moving costs, we calibrated it so that the relative variations in hours per worker and the number of workers matched

U.S. data. The estimate of the fraction of the cycle accounted for by Solow residuals is then reduced to 70 percent.

A widespread and misguided criticism of our econometric studies (for example, McCallum 1989) is that the correlation between labor productivity and labor input is almost one for our model economy while it is approximately zero for the U.S. postwar economy. If we had found that technology shocks account for nearly all fluctuations and that other factors were unimportant, the failure of the model economy to mimic the data in this respect would cast serious doubt on our findings. But we did not find that the Solow technology shocks are all-important. We estimate that these technology shocks account for about 70 percent of business cycle fluctuations. If technology shocks account for 70 percent, and some other shocks that are orthogonal to technology shocks account for 30 percent, then the theory implies a correlation between labor productivity and labor input near zero—just as in the data. Christiano and Eichenbaum (1992a) have established this possibility formally in the case that the other shock is variations in public consumption. But the result holds for any shock that is orthogonal to the Solow technology shocks (see Aiyagari 1994). The fact that this correlation for our model economy and the actual data differ as they do adds to our confidence in our findings.

The estimate of the contribution of technology shocks to aggregate fluctuations has been found to be robust in several modifications of the model economy. Greenwood, Hercowitz, and Huffman (1988) permit the utilization rate of capital to vary and affect its depreciation rate, while all technology change is embodied in new capital. Danthine and Donaldson (1990) introduce an efficiency-wage construct, while Cho and Cooley (forthcoming) permit nominal-wage contracting. Ríos-Rull (1993) uses a model calibrated to life-cycle earnings and consumption patterns. Gomme and Greenwood (forthcoming) have heterogeneous agents with recursive preferences and equilibrium risk allocations. In none of these cases is the estimate of the contribution of technology shocks to aggregate fluctuations significantly altered.

Contribution of Monetary Shocks

One interesting question is, How important a contributor to business cycle fluctuations are monetary shocks? Cooley and Hansen (1989, 1992) have addressed this issue using the Lucas and Stokey (1987) cash-credit-good construct.[5] The beauty of this construct is that it permits the introduction of money into the neoclassical growth model in a computationally tractable way. Models of this type have been used to evaluate monetary policy. Unlike the case of fiscal policy evaluation, however, we have little confidence in using this construct to evaluate monetary stabilization policy.

Our lack of confidence stems from three related reasons. The first is that, unlike the actual economies, these model economies fail to display the sluggishness of the inflation rate's response to changes in the growth rate of money.[6] The second is that households hold large quantities of liquid assets that earn low and, for extended periods, even negative returns. In the United States during the postwar period, households' holdings of M2 were more than half of annual GNP. The magnitude of these assets seems much larger than that needed for transaction purposes. The third reason is that the evaluation of monetary policy appears to be sensitive to the reason people hold these liquid assets. İmrohoroğlu (1992) has constructed a model economy in which people vary their holdings of liquid assets as their income varies in order to smooth their consumptions.[7] She finds that if a transaction cost model is calibrated to data generated by her economy and the calibrated economy used to estimate the cost of inflation, this estimate would grossly underestimate the true cost of inflation for her model world. This result is surprising and bothersome. Typically, how some feature is introduced is unimportant as long as the aggregate substitution elasticities and quantities match.

Given that the answer to monetary policy questions depends upon whether money is held for transaction or precautionary purposes, analytic tractability cannot dictate the way money is introduced. Besides matching better with micro observations, model economies in which the principal reason that people hold money is precautionary display considerable sluggishness in the inflation response to changes in the growth rate of the money supply. We currently do not have the tools for computing equilibriums of models with both the features of the neoclassical growth model and the idiosyncratic shocks that result in people holding money for precautionary reasons. That is why we say we need stronger theory when it comes to evaluating non-steady-state monetary policy and determining the contribution of monetary policy shocks to business cycle fluctuations.

Summary

With the general equilibrium approach, empirical knowledge is organized around preferences and technologies. Given the question and given existing economic theory and measurement, a researcher creates a model economy. This researcher then determines a quantitative answer to the posed question for the model economy. If the theory is strong and the measurements good, we have confidence that the answer for the model economy will be essentially the same as for the actual economy.

Sometimes, however, measurement is not very good, and the results of the computational experiments are that different plausible values of some parameter give very different answers to the posed question. If so, this parameter, which measures some aspect of people's willingness and ability to substitute, must be more accurately measured before theory can provide an answer in which we have confidence. Sometimes the theory relative to the question is weak, and the answer depends upon which of the currently competing theories is used to construct the model economy. If so, these competing theories must be subjected to further tests before there is a good basis for choosing among them. At other times, the theory relative to the question is not only weak but nonexistent. No theory passes all of the key tests. At still other times, the computational tools needed to derive the implications of the theory do not exist, so better computational methods or more powerful computers are needed.

A key issue is, How do we test a theory? Any model of a national economy is necessarily an abstraction and therefore false. Consequently, statistical hypothesis-testing is not a useful tool for testing theory. One useful way to test a theory is to note whether its model economy mimics certain aspects of reality. If a theory passes these tests, then the theory is tested further through challenges to the findings. The standard challenge is to conduct a computational experiment that includes some feature of reality not previously included in other computational experiments. More often than not, introducing this feature of reality does not change the answers, and currently established theory becomes stronger. Occasionally, however, this feature turns out to be important, and established theory is improved. In this way, economic science progresses. A theory is tested through successful use. Perhaps the ultimate test of a theory is whether or not the predictions of the theory are confirmed—that is, Did the economy behave as predicted, given the policy rule selected?

When controlled experiments are not feasible, the computational experiment is the tool of quantitative research. This is true in both the physical and economic sciences.

Footnotes

[1] Lucas (1980), in his paper on methods and problems in business cycle theory, explains the need for computational experiments in business cycle research.

[2] See, for example, Gregory and Smith (1993).

[3] We develop the position that the computational experiment is an econometric tool in Kydland and Prescott (1991a).

[4] This subsection, with minor modifications, is taken from Kydland and Prescott (1991a).

[5] Kydland (1989) also introduces money into the business cycle. People hold real cash balances in his world because this economizes on their time.

[6] Christiano and Eichenbaum (1992b) make this point.

[7] İmrohoroğlu and Prescott (1991) introduce a banking technology to intermediate government debt.

References

Aiyagari, S. Rao. 1994. On the contribution of technology shocks to business cycles. *Federal Reserve Bank of Minneapolis Quarterly Review* 18 (Winter), 22–34.

Auerbach, Alan J., and Kotlikoff, Laurence J. 1987. *Dynamic fiscal policy.* Cambridge: Cambridge University Press.

Backus, David K.; Kehoe, Patrick J.; and Kydland, Finn E. 1994. Dynamics of the trade balance and the terms of trade: The J-curve? *American Economic Review* 84, 84–103.

Benhabib, Jess; Rogerson, Richard; and Wright, Randall. 1991. Homework in macroeconomics: Household production and aggregate fluctuations. *Journal of Political Economy* 99, 1166–87.

Cho, Jong Ok, and Cooley, Thomas F. Forthcoming. The business cycle with nominal contracts. *Journal of Economic Theory.*

Christiano, Lawrence J., and Eichenbaum, Martin. 1992a. Current real-business-cycle theories and aggregate labor-market fluctuations. *American Economic Review* 82, 430–50.

Christiano, Lawrence J., and Eichenbaum, Martin. 1992b. Liquidity effects, monetary policy, and the business cycle. Working Paper 4129. National Bureau of Economic Research.

Cooley, Thomas F., and Hansen, Gary D. 1989. The inflation tax in a real business cycle model. *American Economic Review* 79, 733–48.

Cooley, Thomas F., and Hansen, Gary D. 1992. Tax distortions in a neoclassical monetary economy. *Journal of Economic Theory* 58, 290–316.

Danthine, Jean-Pierre, and Donaldson, John B. 1990. Efficiency wages and the business cycle puzzle. *European Economic Review* 34, 1275–1301.

Frisch, Ragnar. 1933a. Editorial. *Econometrica* 1, 1–5.

Frisch, Ragnar. 1933b. Propagation problems and impulse problems in dynamic economics. In *Economic essays in honour of Gustav Cassel,* pp. 171–205. London: G. Allen & Unwin ltd.

Ghez, Gilbert R., and Becker, Gary S. 1975. *The allocation of time and goods over the life cycle.* New York: Columbia University Press.

Gomme, Paul, and Greenwood, Jeremy. Forthcoming. On the cyclical allocation of risk. *Journal of Economic Dynamics and Control.*

Greenwood, Jeremy, and Hercowitz, Zvi. 1991. The allocation of capital and time over the business cycle. *Journal of Political Economy* 99, 1188–1214.

Greenwood, Jeremy; Hercowitz, Zvi; and Huffman, Gregory W. 1988. Investment, capacity utilization, and the real business cycle. *American Economic Review* 78, 402–18.

Gregory, Allan W., and Smith, Gregory W. 1993. Statistical aspects of calibration in macroeconomics. In *Handbook of statistics* 2, ed. G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 703–19. New York: Elsevier Science.

Guralnik, David B., ed. 1978. *Webster's New World dictionary of the American language.* 2d ed. New York: Avenel Books.

Hansen, Gary D. 1985. Indivisible labor and the business cycle. *Journal of Monetary Economics* 16, 309–27.

Hansen, Gary D. 1986. Growth and fluctuations. In *Three essays on labor indivisibility and the business cycle.* Ph.D. Dissertation. University of Minnesota.

Hodrick, Robert J., and Prescott, Edward C. 1980. Post-war U.S. business cycles: An empirical investigation. Discussion Paper 451. Carnegie Mellon University.

İmrohoroğlu, Ayşe. 1992. The welfare cost of inflation under imperfect insurance. *Journal of Economic Dynamics and Control* 16, 79–91.

İmrohoroğlu, Ayşe, and Prescott, Edward C. 1991. Seigniorage as a tax: A quantitative evaluation. *Journal of Money, Credit and Banking* 23, 462–75.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Kydland, Finn E. 1989. The role of money in a business cycle model. Discussion Paper 23. Institute for Empirical Macroeconomics. Federal Reserve Bank of Minneapolis.

Kydland, Finn E., and Prescott, Edward C. 1982. Time to build and aggregate fluctuations. *Econometrica* 50, 1345–70.

Kydland, Finn E., and Prescott, Edward C. 1990. Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review* 14 (Spring), 3–18.

Kydland, Finn E., and Prescott, Edward C. 1991a. The econometrics of the general equilibrium approach to business cycles. *Scandinavian Journal of Economics* 93, 161–78.

Kydland, Finn E., and Prescott, Edward C. 1991b. Hours and employment variation in business cycle theory. *Economic Theory* 1, 63–81.

Lucas, Robert E., Jr. 1977. Understanding business cycles. In *Stabilization of the domestic and international economy,* ed. Karl Brunner and Allan H. Meltzer, pp. 7–22. *Carnegie-Rochester Conference Series on Public Policy* 5. Amsterdam: North-Holland.

Lucas, Robert E., Jr. 1980. Methods and problems in business cycle theory. *Journal of Money, Credit, and Banking* 12, 696–715.

Lucas, Robert E., Jr., and Rapping, Leonard A. 1969. Real wages, employment, and inflation. *Journal of Political Economy* 77, 721–54.

Lucas, Robert E., Jr., and Stokey, Nancy L. 1987. Money and interest in a cash-in-advance economy. *Econometrica* 55, 491–513.

Marshack, Jacob. 1948. *Introduction to econometrics.* Hectographed lecture notes. Buffalo: University of Buffalo.

McCallum, Bennett T. 1989. Real business cycle models. In *Modern business cycle theory,* ed. Robert J. Barro, pp. 16–50. Cambridge, Mass.: Harvard University Press.

Mehra, Rajnish, and Prescott, Edward C. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145–61.

Prescott, Edward C. 1986. Theory ahead of business cycle measurement. *Federal Reserve Bank of Minneapolis Quarterly Review* 10 (Fall), 9–22. Also, 1986 in *Real business cycles, real exchange rates and actual policies,* ed. Karl Brunner and Allan H. Meltzer. *Carnegie-Rochester Conference Series on Public Policy* 25, 11–44. Amsterdam: North-Holland.

Ríos-Rull, José-Víctor. 1993. On the quantitative importance of market completeness. Manuscript. University of Pennsylvania.

Ríos-Rull, José-Víctor. 1994. Models with heterogeneous agents. In *Frontiers of business cycle research,* ed. Thomas F. Cooley. Princeton, N.J.: Princeton University Press.

Rogerson, Richard. 1988. Indivisible labor, lotteries and equilibrium. *Journal of Monetary Economics* 21, 3–16.

Schneider, Stephen H. 1987. Climate modeling. *Scientific American* 256 (May), 72–80.

Solow, Robert M. 1970. *Growth theory: An exposition.* Oxford: Clarendon Press.

Stokey, Nancy L., and Lucas, Robert E. Jr. (with Edward C. Prescott). 1989. *Recursive methods in economic dynamics.* Cambridge, Mass.: Harvard University Press.

Wicksell, Knut. 1907. Krisernas gåta. *Statsøkonomisk Tidsskrift* 21, 255–70.