

Testing Calibrated General Equilibrium Models

Fabio Canova*

Department of Economics, Universitat Pompeu Fabra, Balmes 132,
E-08008 Barcelona, Spain;

Department of Economics, University of Modena, I-41100 Modena, Italy;
and CEPR

and

Eva Ortega

Department of Economics, European University Institute, Via dei Roccettini, 9
I-50016 San Domenico di Fiesole (FI), Italy

This draft: March 1996

Abstract

This paper illustrates the philosophy which forms the basis of calibration exercises in general equilibrium macroeconomic models and the details of the procedure, the advantages and the disadvantages of the approach, with particular reference to the issue of testing “false” economic models. We provide an overview of the most recent simulation-based approaches to the testing problem and compare them to standard econometric methods used to test the fit of non-linear dynamic general equilibrium models. We illustrate how simulation-based techniques can be used to formally evaluate the fit of a calibrated model to the data and obtain ideas on how to improve the model design using a standard problem in the international real business cycle literature, i.e. whether a model with complete financial markets and no restrictions to capital mobility is able to reproduce the second order properties of aggregate saving and aggregate investment in an open economy.

Key words: Calibration, Simulation, Evaluation, Saving and Investment Correlations.

JEL Classification numbers: C15, C52, D58.

*This paper has been written for the book “Simulation Based Inference in Econometrics: Methods and Applications”, edited by Robert Mariano, T. Schuermann and Mervin Weeks for Cambridge University Press. We have benefitted from the comments of John Geweke, Frank Diebold, Jane Marrinan and of the participants of the conference “Simulation Based Inference in Econometrics” held in Minneapolis, USA. The financial support of a DGICYT grant is gratefully acknowledged. Part of the work was conducted when the second author was visiting the IMF and the University of Pennsylvania

1 Introduction

Simulation techniques are now used in many fields of applied research. As shown elsewhere in this book, they have been employed to compute estimators in situations where standard methods are impractical or fail, to evaluate the properties of parametric and nonparametric econometric estimators, to provide a cheap way of evaluating posterior integrals in Bayesian analysis and to undertake linear and nonlinear filtering with a computationally simple approach.

The task of this chapter is to describe how simulation based methods can be used to evaluate the fit of dynamic general equilibrium models specified using a calibration methodology, to compare and contrast their usefulness relative to more standard econometric approaches and to provide an explicit example where the various features of the approach can be highlighted and discussed.

The structure of this chapter is as follows. First, we provide a definition of what we mean by calibrating a model and discuss the philosophy underlying the approach and how it differs from standard dynamic time series modelling. Second, we discuss various approaches to the selection of model parameters, how to choose the vector of statistics used to compare actual with simulated data and how simulations are performed. Third, we describe how to formally evaluate the model's approximation to the data and discuss alternative approaches to account for the uncertainty faced by a simulator in generating time paths for the relevant variables. Although we present a general overview of alternative evaluation techniques, the focus is on simulation based methods. Finally, we present an example, borrowed from Baxter and Crucini (1993), where the features of the various approaches to evaluation can be examined.

2 What is Calibration?

2.1 A Definition

Although it is more than a decade since calibration techniques emerged in the main stream of dynamic macroeconomics (see Kydland and Prescott (1982)), a precise statement of what it means to calibrate a model has yet to appear in the literature. In general, it is common to think of calibration as an unorthodox procedure to select the parameters of a model. This need not be the case since it is possible to view parameter calibration as a particular econometric technique where the parameters of the model are estimated using an "economic" instead of a "statistical" criteria (see e.g. Canova (1994)). On the other hand, one may want to calibrate a model because there is no available data to estimate its parameters, for example, if one is interested in studying the effect of certain taxes in a newly born country.

Alternatively, it is possible to view calibration as a cheap way to evaluate models. For example, calibration is considered by some a more formal version of the standard back-of-the-envelope calculations that theorists perform to judge the validity of their models (see e.g. Pesaran and Smith (1992)). According to others, calibration is a way to conduct quantitative experiments using models which are known to be "false", i.e. improper or simplified approximations of the true data generating processes of the actual data (see e.g. Kydland and Prescott (1991)).

Pagan (1994) stresses that the unique feature of calibration exercises does not lie so

much in the way parameters are estimated, as the literature has provided alternative ways of doing so, but in the particular collection of procedures used to test tightly specified (and false) theoretical models against particular empirical facts. Here we take a more general point of view and identify 6 steps which we believe capture the essence of the methodology. We call calibration a procedure which involves:

- (i) Formulating an economic question to be addressed.
- (ii) Selecting a model design which bears some relevance to the question asked.
- (iii) Choosing functional forms for the primitives of the model and finding a solution for the endogenous variables in terms of the exogenous variables and the parameters.
- (iv) Choosing parameters and stochastic processes for the exogenous variables and simulating paths for the endogenous variables of the model.
- (v) Selecting a metric and comparing the outcomes of the model relative to a set of “stylized facts”.
- (vi) Doing policy analyses if required.

By “stylized facts” the literature typically means a collection of sample statistics of the actual data such as means, variances, correlations, etc., which (a) do not involve estimation of parameters and (b) are self-evident. More recently, however, the first requirement has been waived and the parameters of a VAR (or the impulse responses) have also been taken as the relevant stylized facts to be matched by the model (see e.g. Smith (1993), Cogley and Nason (1994)).

The next two subsections describe in details both the philosophy behind the first four steps and the practicalities connected with their implementation.

2.2 Formulating a question and choosing a model

The first two steps of a calibration procedure, to formulate a question of interest and a model which bears relevance to the question, are self evident and require little discussion. In general, the questions posed display four types of structure (see e.g. Kollintzas (1992) and Kydland (1992)):

- Is it possible to generate Z using theory W ?
- How much of the fact X can be explained with impulses of type Y ?
- What happens to the endogenous variables of the model if the stochastic process for the control variable V is modified ?
- Is it possible to reduce the discrepancy D of the theory from the data by introducing feature F in the model?

Two economic questions which have received considerable attention in the literature in the last 10 years are the so-called equity premium puzzle, i.e. the inability of a general equilibrium model with complete financial markets to quantitatively replicate the excess returns of equities over bonds over the last hundred years (see e.g. Mehra and Prescott (1985)) and how much of the variability of GNP can be explained by a model whose only source of dynamics are technology disturbances (see e.g. Kydland and Prescott (1982)). As is clear from these two examples, the type of questions posed are very specific and the emphasis is on the numerical implications of the exercise. Generic questions with no numerical quantification are not usually studied in this literature.

For the second step, the choice of an economic model, there are essentially no rules except that it has to have some relevance with the question asked. For example, if one is interested in the equity premium puzzle, one can choose a model which is very simply specified on the international and the government side, but very well specified on the financial side so that it is possible to calculate the returns on various assets. Typically, one chooses dynamic general equilibrium models. However, several authors have used model designs coming from different paradigms (see e.g. the neo-keynesian model of Gali (1994), the non-walrasian models of Danthine and Donaldson (1992) or Gali (1995) and the model with union bargaining of Eberwin and Kollintzas (1995)). There is nothing in the procedure that restricts the class of model design to be used. The only requirement is that the question that the researcher formulates is quantifiable within the context of the model and that the theory, in the form of a model design, is fully specified.

It is important to stress that a model is chosen on the basis of the question asked and not on its being realistic or being able to best replicate the data (see Kydland and Prescott (1991) or Kydland (1992)). In other words, how well it captures reality is not a criteria to select models. What is important is not whether a model is realistic or not but whether it is able to provide a quantitative answer to the specific question the researcher poses.

This brings us to discuss an important philosophical aspect of the methodology. From the point of view of a calibrator all models are approximations to the DGP of the data and, as such, false. This aspect of the problem has been appreciated by several authors even before the appearance of the seminal article of Kydland and Prescott. For example, Hansen and Sargent (1979) also concede that an economic model is a false DGP for the data. Because of this and in order to test the validity of the model using standard statistical tools, they complete the probabilistic structure of the model by adding additional sources of variability, in the form of measurement errors or unobservable variables, to the fundamental forces of the economy.

For calibrators, the model is not a null hypothesis to be tested but an approximation of a few dimensions of the data. A calibrator is not interested in verifying whether the model is true (the answer is already known from the outstart), but in identifying which aspects of the data a false model can replicate and whether different models give different answers because they are false in different dimensions. A calibrator is satisfied with his effort if, through a process of theoretical respecification, a simple and highly stylized model captures an increasing number of features of the data (confront this activity with the so-called normal science of Kuhn (1970)).

Being more explicit, consider the realization of a vector of stochastic processes y_t (our data) and some well specified theoretical model $x_t = f(z_t, \gamma)$ which has something to say about y_t , where z_t are exogenous and predetermined variables and γ is a parameter vector. Because

the model does not provide a complete description of the phenomenon under investigation we write

$$y_t = x_t + u_t \tag{1}$$

where u_t is an error representing what is missing from $f(z_t, \gamma)$ to reproduce the stochastic process generating y_t and whose properties are, in general, unknown (it need not necessarily be mean zero, serially uncorrelated, uncorrelated with the x 's and so on). Let B_y and B_x be continuous and differentiable functions of actual and simulated data, respectively. Then standard econometric procedures judge the coherence of the model to the data by testing whether or not $B_x = B_y$, given that the difference between B_x and B_y and their estimated counterpart \hat{B}_x and \hat{B}_y arise entirely from sampling error. While this is a sensible procedure when the null hypothesis is expected to represent the data, it is less sensible when it is known that the model does not completely capture all aspects of the data.

The third step of a calibration exercise concerns the solution of the model. To be able to obtain quantitative answers from a model it is necessary to find an explicit solution for the endogenous variables of the model in terms of the exogenous and predetermined variables and the parameters. For this reason it is typical to parameterize the objective function of the agents so that manipulation of the first order conditions is analytically tractable. For example, in general equilibrium models, it is typical to choose Cobb-Douglas production functions and constant relative risk aversion utility functions. However, although the main objective is to select simple enough functional forms, it is well known that almost all general equilibrium models and many partial equilibrium models have exact analytical solutions only in very special situations.

For general equilibrium models, a solution exists if the objective function is quadratic and the constraints linear (see e.g. Hansen and Sargent (1979)) or when the objective function is log-linear and the constraints linear (see e.g. Sargent (1987, ch.2)). In the other cases, analytical expressions relating the endogenous variables of the model to the "states" of the problem does not exist and it is necessary to resort to numerical techniques to find solutions which approximate equilibrium functionals either locally or globally. There has been substantial theoretical development in this area in the last few years and several solution algorithms have appeared in the literature (see e.g. the special January 1990 issue of the JBES or Marcet (1994)).

The essence of the approximation process is very simple. The exact solution of a model is a relationship between the endogenous variables x_t , the exogenous and predetermined variables z_t and a set of "deep" parameters γ of the type $x_t = f(z_t, \gamma)$ where f is generally unknown. The approximation procedures generate a relationship of the type $x_t^* = g(z_t, \theta)$ where $\theta = h(\gamma)$ and where $\|f - g\| < \epsilon$ is minimal for some local or global metric. Examples of these types of procedures appear in Kydland and Prescott (1982), Coleman (1989), Tauchen and Hussey (1991), Novales (1990), Baxter (1992) and Marcet (1992), among others. The choice of a particular approximation procedure depends on the question asked. If one is concerned with the dynamics of the model around the steady state, local approximations are sufficient. On the other hand, if one is interested in comparing economic policies requiring drastic changes in the parameters of the control variables, global approximation methods must be preferred.

2.3 Selecting Parameters and Exogenous Processes

Once an approximate solution has been obtained, a calibrator needs to select the parameters γ and the exogenous stochastic process z_t to be fed into the model in order to generate time series for x_t^* . There are several approaches to the choice of these two features of the model. Consider first the question of selecting z_t . This choice is relatively uncontroversial. One either chooses it on the basis of tractability or to give the model some realistic connotation. For example, one can assume that z_t is an AR process with innovations which are transformations of a $N(0, 1)$ process and draw one or more realizations for z_t using standard random number generators. Alternatively, one can select the Solow residuals of the actual economy, the actual path of government expenditure or of the money supply. Obviously, the second alternative is typically preferred if policy analyses are undertaken. Note that while in both cases z_t is the realization of a stochastic process, in the first case the DGP is known while in the second it is not and this has implications for the way one measures the uncertainty in the outcomes of the model.

Next, consider the selection of the vector of parameters γ . Typically, they are chosen so that the model reproduces certain observations. Taking an example from physics, if one is interested in measuring water temperature in various situations it will be necessary to calibrate a thermometer for the experiments. For this purpose a researcher arbitrarily assigns the value 0 C to freezing water and the value 100 C to boiling water and interpolates values in the middle with, say, a linear scale. Given this calibration of the thermometer, one can then proceed to measure the results of the experiments: a value close to 100 C indicates “hot” water, a value close to 30 C indicates “tepid” water, and so on. To try to give answers to the economic question he poses, a calibrator must similarly select observations to be used to calibrate the model-thermometer. There are at least three approaches in the literature. One can follow the deterministic computable general equilibrium (CGE) tradition, summarized, e.g. in Showen and Walley (1984), the dynamic general equilibrium tradition pioneered by Kydland and Prescott (1982) or employ more standard econometric techniques. There are differences between the first two approaches. The first one was developed for deterministic models which do not necessarily possess a steady state while the second one has been applied to dynamic stochastic models whose steady state is unique. Kim and Pagan (1994) provide a detailed analysis of the differences between these two approaches. Gregory and Smith (1993) supplement the discussion by adding interesting insights in the comparison of the first two approaches with the third.

In CGE models a researcher solves the model linearizing the system of equations by determining the endogenous variables around a hypothetical equilibrium where prices and quantities are such that there is no excess demand or excess supply. It is not necessary that this equilibrium exists. However, because the coefficients of the linear equations determining endogenous variables are functions of these equilibrium values, it is necessary to measure this hypothetical equilibrium. The main problem for this literature is therefore to find a set of “benchmark data” and to calibrate the model so that it can reproduce this data. Finding this data set is the most complicated part of the approach since it requires a lot of judgement and ingenuity. The process of specification of this data set leaves some of the parameters of the model typically undetermined, for example, those that describe the utility function of agents. In this

situation a researcher either assigns arbitrary values or fixes them to values estimated in other studies in the literature. Although these choices are arbitrary, the procedure is coherent with the philosophy of the models: a researcher is interested in examining deviations of the model from a hypothetical equilibrium, not from an actual economy.

In stochastic general equilibrium models, the model is typically calibrated at the steady state: parameters are chosen so that the model, in the steady state, produces values for the endogenous variables which match corresponding long run averages of the actual data. In both this approach and the CGE approach point estimates of the parameters used to calibrate the model to the equilibrium are taken to be exact (no standard deviations are typically attached to these estimates). As in the previous setup, the steady state does not necessarily pin down all the parameters of the model. Canova (1994) and Gregory and Smith (1993) discuss various methods to select the remaining parameters. Briefly, a researcher can choose parameters a-priori, pin them down using values previously estimated in the literature, can informally estimate them using simple method of moment conditions or formally estimate them using procedures like GMM (see e.g. Christiano and Eichenbaum (1992)), SMM (see e.g. Duffie and Singleton (1993)) or maximum likelihood (see e.g. McGratten, Rogerson and Wright (1993)). As pointed out by Kydland and Prescott (1991), choosing parameters using the information contained in other studies imposes a coherence criteria among various branches of the profession. For example, in the business cycle literature one uses stochastic growth models to examine business cycle fluctuations and checks the implications of the model using parameters typically obtained in micro studies, which do not employ data having to do with aggregate business cycle fluctuations (e.g. micro studies of labor markets).

If one follows a standard econometric approach, all the parameters are chosen by minimizing the MSE of the error u_t in (1), arbitrarily assuming that the error and the model designs are orthogonal, or by minimizing the distance between moments of the actual data and the model or maximizing the likelihood function of the data given the model design. As we already pointed out, this last approach is the least appealing one from the point of view of a calibrator since it makes assumptions on the time series properties of u_t which are hard to justify from an economic point of view.

To clearly understand the merits of each of these procedures it is useful to discuss their advantages and their disadvantages. Both the CGE and the Kydland and Prescott approach where some of the parameters are chosen a-priori or obtained from a very select group of studies are problematic in several respects. First, there is a selectivity bias problem (see Canova (1995)). There exists a great variety of estimates of the parameters in the literature and different researchers may refer to different studies even when they are examining the same problem. Second, there is a statistical inconsistency problem which may generate very spurious and distorted inference. As Gregory and Smith (1989) have shown, if some parameters are set a-priori and others estimated by simulation, estimates of the latter may be biased and inconsistent unless the parameters of the former group are the true parameters of the DGP or consistent estimates of them. Third, since any particular choice is arbitrary, extensive sensitivity analysis is necessary to evaluate the quality of the results. To solve these problems Canova (1994)-(1995) suggests an approach for choosing parameters which allows, at a second stage, to draw inferences about the quality of the approximation of the model to the data. The idea is very simple. Instead of choosing one set of parameters over another he suggests calibrating

each parameter of the model to an interval, using the empirical information to construct a distribution over this interval (the likelihood of a parameter given existing estimates) and conducting simulation by drawing parameter vectors from the corresponding joint “empirical” distribution. An example may clarify the approach. If one of the parameters of interest is the coefficient of constant relative risk aversion of the representative agent, one typically chooses a value of 2 and tries a few values above and below this one to see if results change. Canova suggests taking a range of values, possibly dictated by economic theory, say $[0,20]$, and then over this range constructing a histogram using existing estimates of this parameter. Most of the estimates are in the range $[1,2]$ and in some asset pricing models researchers have tried values up to 10. Given this information, the resulting empirical distribution for this parameter can be very closely approximated by a $\chi^2(2)$, which has the mode at 2 and about 5% probability in the region above 6.

The selection of the parameters of theoretical models through statistical estimation has advantages and disadvantages. The main advantage is that these procedures avoid arbitrary choices and explicitly provide a measure of dispersion for the estimates which can be used at a second stage to evaluate the quality of the approximation of the model to the data. The disadvantages are of various kinds. First of all, to undertake a formal or informal estimation it is typically necessary to select the moments one wants to fit, and this choice is arbitrary. The standard approach suggested by Kydland and Prescott can indeed be thought of as a method of moment estimation where one chooses parameters so as to set only the discrepancy between the first moment of the model and the data (i.e. the long run averages) to zero. The formal approach suggested by Christiano and Eichenbaum (1992) or Langot and Fève (1994), on the other hand, can be thought of as a method of moment estimation where a researcher fits the discrepancies between model and data first and second moments to zero. The approach of choosing parameters by setting to zero the discrepancy between certain moments has the disadvantage of reducing the number of moments over which it will be possible to evaluate the quality of the model. Moreover, it is known that estimates obtained with the method of moments or GMM may be biased. Therefore, simulations and inference conducted with these estimates may lead to spurious inference (see e.g. Canova, Finn and Pagan (1994)). In addition, informal SMM may lead one to select parameters even though they are not identifiable (see Gregory and Smith (1989)). Finally, one should note that the type of uncertainty which is imposed on the model via an estimation process does not necessarily reflect the uncertainty a calibrator faces when choosing the parameter vector. As is clear from a decade of GMM estimation, once the moments are selected and the data given, sample uncertainty is pretty small. The true uncertainty is in the choice of moments and in the data set to be used to select parameters. This uncertainty is disregarded when parameters are chosen using extremum estimators like GMM.

Finally, it is useful to compare the parameter selection process used by a calibrator à-la Kydland and Prescott and the one used by a traditional econometric approach. In a traditional econometric approach parameters are chosen so as to minimize some **statistical** criteria, for example, the MSE. Such criteria do not have any economic content, impose stringent requirements on the structure of u_t and are used, primarily, because there exists a well established statistical and mathematical literature on the subject. In other words, the parameter selection criteria used by traditional econometricians does not have economic justification. On the

other hand, the parameter selection criteria used by followers of the Kydland and Prescott methodology can be thought of as being based on **economic** criteria. For example, if the model is calibrated so that, in the steady state, it matches the long run features of the actual economy, parameters are implicitly selected using the condition that the sum (over time) of the discrepancies between the model and the data is zero. In this sense there is an important difference between the two approaches which has to do with the assumptions that one is willing to make on the errors u_t . By calibrating the model to long run observations a researcher selects parameters assuming $E(u) = 0$, i.e. using a restriction which is identical to the one imposed by a GMM econometrician who chooses parameters using only first moment conditions. On the other hand, to conduct classical inference a researcher imposes restrictions on the first and second moments of u_t .

The comparison we have made so far concerns, obviously, only those parameters which enter the steady state conditions of the model. For the other parameters a direct comparison with standard econometric practice is not possible. However, if all parameters are calibrated to intervals with distributions which are empirically determined, the calibration procedure we have described shares a tight connection with Bayesian inferential methods such as Consensus Analysis or Meta-Analysis (see e.g. Genest and Zidak (1986) or Wolf (1986)).

Once the parameters and the stochastic processes for the exogenous variables are selected and an (approximate) solution to the model has been found, simulated paths for x_t^* can be generated using standard Monte Carlo simulation techniques.

3 Evaluating Calibrated Models

The questions of how well a model matches the data and how much confidence a researcher ought to give to the results constitute the most crucial steps in the calibration procedure. In fact, the most active methodological branch of this literature concerns methods to evaluate the fit of a model selected according to the procedures described in section 2. The evaluation of a model requires three steps: first, the selection of a set of stylized facts; second, the choice of a metric to compare functions of actual and simulated data and third, the (statistical) evaluation of the magnitude of the distance. Formally, let S_y be a set of statistics (stylized facts) of the actual data and let $S_{x^*}(z_t, \gamma)$ be a set of statistics of simulated data, given a vector of parameters γ and a vector of stochastic processes z_t . Then model evaluation consists of selecting a function $\psi(S_y, S_{x^*}(z_t, \gamma))$ measuring the distance between S_y and S_{x^*} and in assessing its magnitude.

The choice of which stylized facts one wants to match obviously depends on the question asked and on the type of model used. For example, if the question is what is the proportion of actual cyclical fluctuations in GNP and consumption explained by the model, one would choose stylized facts based on variances and covariances of the data. As an alternative to the examination of second moments, one could summarize the properties of actual data via a VAR and study the properties of simulated data, for example, by comparing the number of unit roots in the two sets of data (as in Canova, Finn and Pagan (1994)), the size of VAR coefficients (as in Smith (1993)) or the magnitude of certain impulse responses (as in Cogley and Nason (1994)). Also, it is possible to evaluate the discrepancy of a model to the data by choosing specific events that one wants the model to replicate e.g., business cycle turning points, (as in King and Plosser (1994) or Simkins (1994)) or variance bounds (as in Hansen and Jagannathan

(1991)).

Classical pieces in the calibration literature (see e.g. Kydland and Prescott (1982) or (1991)) are typically silent on the metric one should use to evaluate the quality of the approximation of the model to the data. The approach favored by most calibrators is to glare over the exact definition of the metric used and informally assess the properties of simulated data by comparing them to the set of stylized facts. In this way a researcher treats the computational experiment as a measurement exercise where the task is to gauge the proportion of some observed statistics reproduced by the theoretical model. This informal approach is also shared by cliometricians (see e.g. Summers (1991)) who believe that rough reproduction of simple sample statistics is all that is needed to evaluate the implications of the model (“either you see it with naked eyes or no fancy econometric procedure will find it”).

There are, however, alternatives to this informal approach. To gain some understanding of the differences among approaches, but at the cost of oversimplifying the matter, we divide evaluation approaches into five classes:

- Informal approaches.
- Approaches which do not consider sampling variability of actual or the uncertainty in simulated data, but instead use the statistical properties of u_t in (1) to impose restrictions on the time series properties of ψ . This allows them to provide an R^2 -type measure of fit between the model and the data (see Watson (1993)).
- Approaches which use the sampling variability of the **actual** data (affecting S_y and, in some cases, estimated γ) to provide a measure of the distance between the model and the data. Among these we list the GMM based approach of Christiano and Eichenbaum (1992), Cecchetti, Lam and Mark (1993) or Fève and Langot (1994), and the frequency domain approaches of Diebold, Ohanian and Berkowitz (1995) and Ortega (1995).
- Approaches which use the uncertainty of the **simulated** data to provide a measure of distance between the model and the data. Among these procedures we can distinguish those who take z_t as stochastic and γ as given, such as Gregory and Smith (1991), Söderlind (1994) or Cogley and Nason (1994) and those who take both z_t and γ as stochastic, such as Canova (1994) and (1995).
- Finally, approaches which consider the sampling variability of the **actual** data and the uncertainty in **simulated** data to evaluate the fit of the model. Once again we can distinguish approaches which, in addition to taking S_y as random, allow for variability in the parameters of the model (keeping z_t fixed) such as DeJong, Ingram and Whiteman (1995) from those which allow for both z_t and γ to vary such as Canova and De Nicoló (1995).

Because the emphasis of this book is on simulation techniques, we will only briefly examine the first three approaches and discuss in more detail the last two, which make extensive use of simulation techniques to conduct inference. Kim and Pagan (1994) provide a thorough critical review of several of these evaluation techniques and additional insights on the relationship among them.

The evaluation criteria that each of these approaches proposes is tightly linked to the parameter selection procedure we discussed in the previous section.

As mentioned the standard approach is to choose parameters using steady state conditions. Those parameters which do not appear in the steady state are selected a-priori or with reference to existing literature. Also, since S_y is chosen to be a vector of numbers and no uncertainty is allowed in the selected parameter vector, one is forced to use an informal metric to compare the model to the data. This is because, apart from the uncertainty present in the exogenous variables, the model links the endogenous variables to the parameters in a deterministic fashion. Therefore, once we have selected the parameters and we have a realization of S_y , it is not possible to measure the dispersion of the distance $\psi(S_y, S_{x^*}(z_t, \gamma))$. From the point of view of the majority of calibrators this is not a problem. As emphasized by Kydland and Prescott (1991) or Kydland (1992), the trust a researcher has in an answer given by the model does not depend on a statistical measure of discrepancy, but on how much he believes in the economic theory used and in the measurement undertaken.

Taking this as the starting point of the analysis Watson (1993) suggests an ingenious way to evaluate models which are known to be an incorrect DGP for the actual data. Watson asks how much error should be added to x_t^* so that its autocovariance function equals the autocovariance function of y_t . Writing $y_t = x_t^* + u_t^*$ where u_t^* includes the approximation error due to the use x_t^* in place of x_t , the autocovariance function of this error is given by

$$A_{u^*}(z) = A_y(z) + A_{x^*}(z) - A_{x^*y}(z) - A_{yx^*}(z) \quad (2)$$

To evaluate the last two terms in (2) we need a sample from the joint distribution of (x_t^*, y_t) which is not available. In these circumstances it is typical to assume that either u_t^* is a measurement error or a signal extraction noise (see e.g. Sargent (1989)), but in the present context neither of the two assumptions is very appealing. Watson suggests choosing $A_{x^*y}(z)$ so as to minimize the variance of u_t^* subject to the constraint that $A_{x^*}(z)$ and $A_y(z)$ are positive semidefinite. Intuitively, the idea is to select $A_{x^*y}(z)$ to give the best possible fit between the model and the data (i.e. the smallest possible variance of u_t^*). The exact choice of $A_{x^*y}(z)$ depends on the properties of x_t^* and y_t , i.e. whether they are serially correlated or not, scalar or vectors, full rank processes or not. In all cases, the selection criteria chosen imply that x_t^* and y_t are perfectly linearly correlated where the matrix linking the two vectors depends on their time series properties and on the number of shocks buffeting the model. Given this framework of analysis, Watson suggests two measures of fit, similar to a $1 - R^2$ from a regression, of the form

$$r_j(\omega) = \frac{A_{u^*}(\omega)_{jj}}{A_y(\omega)_{jj}} \quad (3)$$

$$R_j(\omega) = \frac{\int_{\omega \in Z} A_{u^*}(\omega)_{jj} d\omega}{\int_{\omega \in Z} A_y(\omega)_{jj} d\omega} \quad (4)$$

where the first statistic measures the variance of the j -th component of the error relative to the variance of the j -th component of the data for each frequency and the second statistic is the sum of the first over a set of frequencies. This last measure may be useful to evaluate the model, say, at business cycle frequencies. It should be stressed that (3) and (4) are lower

bounds. That is, when $r_j(\omega)$ or $R_j(\omega)$ are large, the model poorly fits the data. However, when they are small, it does not necessarily follow that the model is appropriate since it may still fit the data poorly if we change the assumptions about $A_{x^*y}(z)$.

To summarize, Watson chooses the autocovariance function of y as the set of stylized facts of the data to be matched by the model, the ψ function as the ratio of A_{u^*} to A_y and evaluates the size of ψ informally (i.e. if it is greater than one, between zero and one or close to zero). Note that in this approach, γ and z_t are fixed, and A_{x^*} and A_y are assumed to be measured without error.

When a calibrator is willing to assume that parameters are measured with error because, given an econometric technique and a sample, parameters are imprecisely estimated, then model evaluation can be conducted using measures of dispersion for simulated statistics which reflect parameter uncertainty. There are various versions of this approach. Christiano and Eichenbaum (1992), Cecchetti, Lam and Mark (1993) and Fève and Langot (1994) use a version of a J-test to evaluate the fit of a model. In this case S_y are moments of the data while ψ is a quadratic function of the type

$$\psi(S_y, S_{x^*}(z_t, \gamma)) = [S_y - S_{x^*}(\gamma)]V^{-1}[S_y - S_{x^*}(\gamma)]' \quad (5)$$

where V is a matrix which linearly weights the covariance matrix of S_{x^*} and S_y , and S_{x^*} is random because γ is random. Formal evaluation of this distance can be undertaken following Hansen (1982): under the null that $S_y = S_{x^*}(z_t, \gamma)$ the statistic defined in (5) is asymptotically distributed as a χ^2 with the number of degrees of freedom equal to the number of overidentifying restrictions, i.e. the dimension of S_y minus the dimension of the vector γ . Note that this procedure is correct asymptotically, that it implicitly assumes that $x_t = f(z_t, \gamma)$ (or its approximation x_t^*) is the correct DGP for the data and that the relevant loss function measuring the distance between actual and simulated data is quadratic.

The methods proposed by Diebold, Ohanian and Berkowitz (DOB) (1994) and Ortega (1995) are slightly different but can be broadly included into this class of approaches.

For DOB the statistic of interest is the spectral density matrix of y_t and, given a sample, this is assumed to be measured with error. They measure the uncertainty surrounding point estimates of the spectral density matrix employing (small sample) 90% confidence bands constructed using parametric and nonparametric bootstrap approaches and Bonferroni tunnels. On the other hand, they take calibrated parameters and the realization of z_t as given so that the spectral density matrix of simulated data can be estimated without error simply by simulating very long time series for x_t^* . Ortega (1995) also takes the spectral density matrix as the set of stylized facts of the data to be matched by the model. Unlike DOB, she considers the uncertainty in actual and simulated data by jointly estimating the spectral density matrix of actual and simulated data and constructs measures of uncertainty around point estimates of the spectral density matrix using asymptotic distribution theory.

In both cases, the measure of fit used is generically given by:

$$C(\gamma, z_t) = \int_0^\pi \psi(F_y(\omega), F_{x^*}(\omega, \gamma, z_t))W(\omega)d\omega \quad (6)$$

where $W(\omega)$ is a set of weights applied to different frequencies and F are the spectral density matrices of actual and simulated data.

DOB suggest various options for ψ (quadratic, ratio, likelihood type) but do not construct a direct test statistic to examine the magnitude of ψ . Instead, they compute a small sample distribution of the event that $C(\gamma, z_t)$ is close to a particular value (zero if ψ is quadratic, 1 if ψ is a ratio, etc.) Ortega, on the other hand, explicitly uses a quadratic expression for ψ and uses an asymptotic χ^2 test to assess whether the magnitude of the discrepancy between the model and the data is significant or not. The set of asymptotic tools she develops can also be used to compare the fit of two alternative models to the data and decide which one is more acceptable.

If a calibrator is willing to accept the idea that the stochastic process for the exogenous variables is not fixed, she can then compute measures of dispersion for simulated statistics by simply changing the realization of z_t while maintaining the parameters fixed. Such a methodology has its cornerstone in the fact that it is the uncertainty in the realization of the exogenous stochastic process (e.g. the technology shock), an uncertainty which one can call extrinsic, and not the uncertainty in the parameters, which one can call intrinsic, which determines possible variations in the statistics of simulated data. Once a measure of dispersion of simulated statistics is obtained, the sampling variability of simulated data can be used to evaluate the distance between statistics of actual and simulated data (as e.g. Gregory and Smith (1991) and (1993)).

If one uses such an approach, model evaluation can be undertaken with a probabilistic metric using well known Monte Carlo techniques. For example, one may be interested in finding out in what decile of the simulated distribution the actual value of a particular statistic lies, in practice, calculating the “size” of calibration tests. This approach requires two important assumptions: that the evaluator takes the model economy as the true DGP for the data and that differences between S_y and S_{x^*} occur only because of sampling variability. To be specific, Gregory and Smith take S_y be a set of moments of the data and assume that they can be measured without error. Then, they construct a distribution of $S_{x^*}(z_t, \gamma)$ by drawing realizations for the z_t process from a given distribution, given γ . The metric ψ used is probabilistic, i.e. they calculate the probability $Q = P(S_{x^*} \leq S_y)$, and judge the fit of the model informally, e.g. measuring how close Q is to 0.5.

An interesting variation on this setup is provided by Söderlind (1994) and Cogley and Nason (1994). Söderlind employs the spectral density matrix of the actual data while Cogley and Nason choose a “structural” impulse response function as the relevant statistics to be matched. Söderlind maintains a probabilistic metric and constructs the empirical rejection rate for the event that the actual spectral density matrix of y_t lies inside the asymptotic 90% confidence band for the spectral density matrix of the simulated data. Such an event is replicated by drawing vectors z_t for a given distribution. Cogley and Nason choose a quadratic measure of distance which, under the null that the model is the DGP for the data, has an asymptotic χ^2 distribution and then tabulate the empirical rejection rates of the test, by repeatedly constructing the statistic drawing realizations of the z_t vector. To be specific, the ψ function is in this case given by

$$\psi_{k,j}(\gamma) = [IRF_{x^*}^k(z_t^j, \gamma) - IRF_y^k]V^{-1}[IRF_{x^*}^k(z_t^j, \gamma) - IRF_y^k]' \quad (7)$$

where j indexes replications and k steps, IRF^k is the impulse response function and V is its asymptotic covariance matrix at step k . Because for every k and for fixed j $\psi_{k,j}(\gamma)$ is

asymptotically χ^2 , they can construct (a) the simulated distribution for $\psi_{k,j}$ and compare it with a χ^2 and (b) the rejection frequency for each model specification they examine.

In practice, all three approaches are computer intensive and rely on Monte Carlo methods to conduct inference. Also, it should be stressed that all three methods verify the validity of the model by computing the “size” of the calibration tests, i.e. assuming that the model is the correct DGP for y_t .

The approach of Canova (1994)-(1995) also belongs to this category of methods, but, in addition to allowing the realization of the stochastic process for the exogenous variables to vary, he also allows for parameter variability in measuring the dispersion of simulated statistics. The starting point, as discussed earlier, is that parameters are uncertain not so much because of sample variability, but because there are many estimates of the same parameter obtained in the literature, since estimation techniques, samples and frequency of the data tend to differ. If one calibrates the parameter vector to an interval, rather than to a particular value, and draws values for the parameters from the empirical distribution of parameter estimates, it is then possible to use the intrinsic uncertainty, in addition to or instead of the extrinsic one, to evaluate the fit of the model. The evaluation approach used is very similar to the one of Gregory and Smith: one simulates the model repeatedly by drawing parameter vectors from the empirical “prior” distribution and realizations of the exogenous stochastic process z_t from some given distribution. Once the empirical distribution of the statistics of interest is constructed, one can then compute either the size of calibration tests or the percentiles where the actual statistics lie.

The last set of approaches considers the uncertainty present in the statistics of both actual and simulated data to measure the fit of the model to the data. In essence what these approaches attempt to formally measure is the degree of overlap between the (possibly) multivariate distributions of S_y and S_x using Monte Carlo techniques. There are differences in the way these distributions have been constructed in the literature. Canova and De Nicoló (1995) use a parametric bootstrap algorithm to construct distributions for the statistics of the actual data. DeJong, Ingram and Whiteman (DIW) (1995), on the other hand, suggest representing the actual data with a VAR and computing posterior distribution estimates for the moments of interest by drawing VAR parameters from their posterior distribution and using the AR(1) companion matrix of the VAR at each replication. In constructing distributions of simulated statistics, Canova and De Nicoló take into account both the uncertainty in exogenous processes and parameters while DIW only consider parameter uncertainty. The two approaches also differ in the way the “prior” uncertainty in the parameters is introduced in the model. The former paper follows Canova (1995) and chooses empirical based distributions for the parameter vector. DIW use subjectively specified prior distributions (generally normal) whose location parameter is set at the value typically calibrated in the literature while the dispersion parameter is free. The authors use this parameter in order to (informally) minimize the distance between actual and simulated distributions of the statistics of interest. By enabling the specification of a sequence of increasingly diffuse priors over the parameter vector, such a procedure illustrates whether the uncertainty in the model’s parameters can mitigate differences between the model and the data.

Finally, there are differences in assessing the degree of overlap of the two distributions.

Canova and De Nicoló choose a particular contour probability for one of the two distributions and ask how much of the other distribution is inside the contour. In other words, the fit of the model is examined very much in the style of the Monte Carlo literature: a good fit is indicated by a high probability covering of the two regions. To describe the features of the two distributions, they also repeat the exercise varying the chosen contour probability, say, from 50% to 75%, 90%, 95% and 99%. The procedure allows them to detect anomalies in the shape of the two distributions due to clustering of observations in one area, skewness or leptokurtic behavior. In this approach actual data and simulated data are used symmetrically in the sense that one can either ask whether the actual data could be generated by the model, or viceversa, whether simulated data are consistent with the distribution of the empirical sample. This symmetry allows the researcher to understand much better the distributional properties of error u_t in (1). Moreover, the symmetry with which the two distributions are treated resembles very much the process of switching the null and the alternative in standard classical hypothesis testing.

DeJong, Ingram and Whiteman take the point of view that there are no well established criteria to judge the adequacy of a model's "approximation" to reality. For this reason they present two statistics aimed at synthetically measuring the degree of overlap among distributions. One, which they call Confidence Interval Criterion (CIC) is the univariate version of the contour probability criteria used by Canova and De Nicoló and is defined as

$$CIC_{ij} = \frac{1}{1-\alpha} \int_a^b P_j(s_i) ds_i \quad (8)$$

where s_i , $i = 1, \dots, n$ is a set of functions of interest, $a = \frac{\alpha}{2}$ and $b = 1 - a$ are the quantiles of $D(s_i)$, the distribution of the statistic in the actual data, $P_j(s_i)$ is the distribution of the simulated statistic where j is the diffusion index of the prior on the parameter vector and $1 - \alpha = \int_a^b D(s_i) ds_i$. Note that with this definition, CIC_{ij} ranges between 0 and $\frac{1}{1-\alpha}$. For CIC close to zero, the fit of the model is poor, either because the overlap is small or because P_j is very diffuse. For CIC close to $\frac{1}{1-\alpha}$ the two distributions overlap substantially. Finally, if $CIC > 1$, $D(s_i)$ is diffuse relative to $P_j(s_i)$, i.e. the data is found to be relatively uninformative regarding s_i .

To distinguish among the two possible interpretations when CIC is close to zero, DeJong, Ingram and Whiteman suggest a second summary measure analogous to a t-statistic for the mean of $P_j(s_i)$ in the $D(s_i)$ distribution, i.e.,

$$d_{ji} = \frac{EP_j(s_i) - ED(s_i)}{\sqrt{varD(s_i)}} \quad (9)$$

Large values of (9) indicate that the location of $P_j(s_i)$ is quite different from the location of $D(s_i)$.

The final problem of the DIW methodology is to choose α . DeJong, Ingram and Whiteman fix a particular value ($\alpha = 0.01$) but, as in Canova and De Nicoló, varying α for a given j is probably a good thing to do in order to describe the feature of the distributions. This is particularly useful when we are interested in partitions of the joint distributions of s_i because graphical methods or simple statistics are not particularly informative about distributions in high dimensional spaces.

4 Policy Analyses

Although it is not the purpose of this chapter to discuss in detail how calibrated models can be used for policy analyses, it is useful to describe the implications of the procedure for questions which have policy implications and how policy experiments can be undertaken. As we have already mentioned, a model is typically calibrated to provide a quantitative answer to very precise questions and some of these questions have potential policy implications. To forcefully argue the policy implications of the exercise one needs to be confident in the answer given by the model and to do this it is necessary to undertake extensive sensitivity analysis to check how results change when certain assumptions are modified.

As we have seen, the answers of the model come in the form of continuous functions $h(x_t^*) = h(g(z_t, \gamma))$ of simulated data. In theory, once g has been selected, the uncertainty in h is due to the uncertainty in γ and in z_t . Since in standard calibration exercises the γ vector is fixed, it is therefore typical to examine the sensitivity of the results in the neighborhood of the calibrated values for γ . Such experiments may be local, if the neighborhood is small, or global, in which case one measures the sensitivity of the results to perturbations of the parameters over the entire range. This type of exercise may provide two types of information. First, if results are robust to variations of a parameter in a particular range, its exact measurement is not crucial. In other words, the uncertainty present in the choice of such a parameter does not make the answers of the model tenuous and economic inference groundless. On the other hand, if results crucially depend on the exact selection of certain parameters, it is clearly necessary to improve upon existing measurement of these parameters.

A local sensitivity analysis can be undertaken informally, replicating the experiments for different values of the parameters (as in Kydland and Prescott (1982)) or more formally, calculating the elasticity of h with respect to γ (as in Pagan and Shannon (1985)). A global sensitivity analysis can be efficiently undertaken with Monte Carlo methods or numerical semi-deterministic techniques (see e.g. Niederreiter (1988)) if the function g is known and the distribution of the γ vector is specified. If g is only an approximation to the functional linking x to z and γ , one can use techniques like Importance Sampling (see Geweke (1989)) to take into account this additional source of uncertainty. Clearly the two types of sensitivity analysis are not incompatible and should both be undertaken to assess the degree of trust a researcher can attach to the answer given by the model. Finally, one should note that the type of sensitivity analysis one may want to undertake depends also on the way parameters are selected and models evaluated. For example, if one uses the approach of Canova (1994)-(1995) or DeJong, Ingram and Whiteman (1995), the evaluation procedure automatically and efficiently provides sensitivity analysis to global perturbations of the parameters within an economically reasonable range.

Once model answers to the question of interest have been shown to be robust to reasonable variations in the parameters, a researcher may undertake policy analyses by changing the realization of the stochastic process for z_t or varying a subset of the γ vector, which may be under the control of, say, the government. Analyses involving changes in the distribution of z_t in the g function are also possible, but care should be exercised in order to compare results across specifications.

5 An example

In the field of international economics, robust stylized facts are usually hard to obtain. One of the most stable regularities observed in the data is the high correlation of national saving and domestic investment, both in time series analysis of individual countries and in cross sections regressions where the average over time of these variables is treated as a single data point for each country. High saving and investment correlations are observed in small economies as well as large ones, although the correlation tends to be lower for smaller countries. These findings were originally interpreted as indicating that the world economy is characterized by a low degree of capital mobility. Yet most economists believe that the world is evolving toward an increasingly higher degree of international capital mobility. Baxter and Crucini (1993) forcefully turned this initial interpretation around by providing a model in which there is perfect international mobility of financial and physical capital but which generates high time series correlations of national saving and investment. Their evaluation of the model lies entirely within the standard Kydland and Prescott approach, i.e. parameters are fixed at some reasonably chosen values, no uncertainty is allowed in actual and simulated statistics and the metric used to compare actual and simulated data is informal.

The task of this section is three fold. First, we want to study whether the time series properties of simulated saving and investment do indeed reproduce those of the actual data when the model is formally examined with the tools described in this article. To this end we provide several measures of fit which can be used to gauge the closeness of the model to the data using variants of the simulation-based procedures described in the previous section. Second, we wish to contrast the outcomes obtained with various evaluation procedures and compare them with those obtained using more standard techniques. This will shed further light on the degree of approximation of the model to the data, and point out, when they emerge, unusual features of the model. Finally, we wish to provide a few suggestions on how to fine tune the model design so that undesirable features are eliminated while maintaining the basic bulk of the results.

5.1 The model

We consider a model with two countries and a single consumption good. Each country is populated by a large number of identical agents and labor is assumed to be immobile across countries and variables are measured in per-capita terms. Preferences of the representative agent of country $h = 1, 2$ are given by:

$$U \equiv E_0 \sum_{t=0}^{\infty} \frac{\beta^t}{1-\sigma} [C_{ht}^\mu L_{ht}^{(1-\mu)}]^{1-\sigma} \quad (10)$$

where C_{ht} is private consumption of the single composite good by the representative agent of country h and L_{ht} is leisure, β is the discount factor, σ the coefficient of relative risk aversion and μ the share of consumption in utility. Leisure choices are constrained by:

$$0 \leq L_{ht} + N_{ht} \leq 1 \quad \forall h \quad (11)$$

where the total endowment of time in each country is normalized to 1 and N_t represents the number of hours worked. The goods are produced with a Cobb-Douglas technology:

$$Y_{ht} = A_{ht}(K_{ht})^{1-\alpha}(X_{ht}N_{ht})^\alpha \quad h = 1, 2 \quad (12)$$

where K_t is the capital input, α is the share of labor in GDP, and where $X_{ht} = \theta_x X_{ht-1} \forall h$ with $\theta_x \geq 1$. X_{ht} represents labor-augmenting Harrod-neutral technological progress with deterministic growth rate equal to θ_x . Production requires domestic labor and capital inputs and is subject to a technological disturbance A_{ht} with the following properties:

$$\begin{bmatrix} A_{1t} \\ A_{2t} \end{bmatrix} = \begin{bmatrix} \bar{A}_1 \\ \bar{A}_2 \end{bmatrix} + \begin{bmatrix} \rho & \nu \\ \nu & \rho \end{bmatrix} \begin{bmatrix} A_{1t-1} \\ A_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

where $\epsilon_t = [\epsilon_{1t} \ \epsilon_{2t}]' \sim N(0, \begin{bmatrix} \sigma_\epsilon^2 & \psi \\ \psi & \sigma_\epsilon^2 \end{bmatrix})$ and $[\bar{A}_1, \bar{A}_2]'$ is a vector of constants. The parameter ψ controls the contemporaneous spillover while ν the lagged spillover of the shocks.

Capital goods are accumulated according to:

$$K_{ht+1} = (1 - \delta_h)K_{ht} + \phi(I_{ht}/K_{ht})K_{ht} \quad h = 1, 2 \quad (13)$$

where $\phi(\frac{I_{ht}}{K_{ht}}) > 0$ is concave and represents the costs of adjusting capital. As explained in Baxter and Crucini (1993), there is no need to choose a functional form for ϕ ; it is sufficient to describe its behavior near the steady state. We do this by specifying two parameters: $\frac{1}{\phi'}$, which corresponds to Tobin's Q, i.e. the price of existing capital in one location relative to the price of new capital and $\xi_{\phi'}$, the elasticity of the marginal adjustment cost function with respect to the investment-capital ratio.

Governments finance their consumption purchases, G_{ht} , by taxing national outputs with a distorting tax and transferring what remains back to domestic residents. For simplicity we assume that $G_{ht} = G_h, \forall t$. The government budget constraint is given by:

$$G_h = TR_{ht} + \tau_h Y_{ht} \quad \forall h \quad (14)$$

where τ_h are tax rates and TR_h are lump sum transfers in country h .

The economy wide resource constraint is given by:

$$\pi(Y_{1t} - G_{1t} - C_{1t} - I_{1t}) + (1 - \pi)(Y_{2t} - G_{2t} - C_{2t} - I_{2t}) \geq 0 \quad (15)$$

where π is the fraction of world population living in country 1.

Finally, following Baxter and Crucini (1993), we assume complete financial markets and free mobility of financial capital across countries so that agents can write and trade every kind of contingent security.

To find a solution to the model we first detrend those variables which drift over time by taking ratios of the original variables with respect to the labor augmenting technological progress, e.g. $y_{ht} = \frac{Y_{ht}}{X_{ht}}$, etc. Second, since there are distortionary taxes in the model, the competitive equilibrium is not Pareto optimal and the competitive solution differs from the social planner's solution. As in Baxter and Crucini (1993) we solve the problem faced by a

pseudo social planner, modifying the optimality conditions to take care of the distortions. The weights in the social planner problem are chosen to be proportional to the number of individuals living in each of the countries. The modified optimality conditions are approximated with a log-linear expansion around the steady state as in King, Plosser and Rebelo (1988). Time series for saving and investment in each of the two countries are computed analytically from the approximate optimality conditions. The second order properties of saving and investment of actual and simulated data are computed eliminating from the raw time series a linear trend.

The parameters of the model are $\gamma = [\beta, \sigma, \alpha, \theta_x, \delta, \rho, \nu, \sigma_\epsilon, \psi, \pi, \xi_{\phi'}, \phi', \tau,]$ plus steady state hours and the steady state Tobin's Q which we set equal to 1. The exogenous processes of the model are the two productivity disturbances so that $z_t = [A_{1t} A_{2t}]'$.

The actual data we use are per capita basic saving (i.e. computed as $S_t = Y_t - C_t - G_t$) and investment for the period 1970:1-1993:3 for the US and for Europe in real terms, seasonally adjusted and are from OECD Main Economic Indicators. Plots of the detrended series appear in figure 1.

The statistics we care about are the diagonal elements of the 4×4 spectral density matrix of the data and the coherences between saving and investment of the two "countries". Spectral density estimates at each frequency are computed smoothing with a flat window 13 periodogram ordinates. Figure 2 plots these statistics.

In the benchmark experiment the vector γ is the same as in Baxter and Crucini (1993) except for σ_ϵ which they normalize to 1, while we set it equal to the value used in Backus, Kehoe and Kydland (1995), and are reported in the first column of table 1. When we allow for parameters to be random we take two approaches: the one of Canova (1994) and the one of DeJong, Ingram and Whiteman (1995). In the first case empirical based distributions are constructed using existing estimates of these parameters or, when there are none, choosing a-priori an interval on the basis of theoretical considerations and imposing a uniform distribution on it. The distributions from which the parameters are drawn and their features are displayed in the second column of table 1. In the second case distributions for the parameters are assumed to be normal, with means equal to the basic calibrated parameters presented in column 1 while dispersions are a-priori chosen. The third column of table 1 reports these distributions.

We generate samples of 95 observations to match the sample size of actual data. Because the initial conditions for the capital stock are set arbitrarily, the first 50 observations for each replication of the model are discarded. The number of replications used for each exercise is 500.

5.2 The Results

Table 2 summarizes the results obtained using four different evaluation approaches. Each row reports how the model fares in reproducing the spectral densities of saving and investment and the saving-investment coherence for US and Europe on average at business cycle frequencies (cycles of 3-8 years).

As a reference for comparison, the two first rows report the average spectral densities and coherences at business cycle frequencies for actual and simulated data when parameters are fixed (Kydland and Prescott approach). National saving is highly correlated with domestic investment but the average coherence at business cycle frequencies is higher for Europe than for

the US. The variability of both US series is also higher and US investment are almost two times more volatile than European ones. This pattern does not depend on the averaging procedure we choose; in fact, it is present at every frequency within the range we examine.

Given the symmetry of the model specification, the variability of simulated saving and investment is similar in both continental blocks, it is somewhat lower than the actual data for Europe, but definitively too low relative to the actual US series. Moreover, as in the actual data, the variability is higher for national savings than for domestic investment. Consistent with Baxter and Crucini's claims, the model produces high national saving and investment correlations at business cycle frequencies. In fact, the model coherences for the US are higher than those found in the actual data.

The following rows of table 2 check whether the above results persist when the performance of the model is evaluated using some of the procedures described in this paper.

The first approach, which we use as a benchmark, is the one of Watson (1993). Given the spectral density matrix of the actual saving and investment for the two economic blocks, we calculate the spectral density matrix of the approximation error and compute the measure of fit (4) where Z includes frequencies corresponding to cycles of 3-8 years. Since in the model there are two technology disturbances, the spectral density matrix of simulated saving and investment for the two countries is singular and of rank equal to two. Therefore, to minimize the variance of the approximation error we consider two different identification schemes: in "identification 1" we jointly minimize the error term of the saving and investment of the first country (row 3 of table 2) and in "identification 2" we jointly minimize the saving and investment errors of the second country (row 4 of table 2). Note that to generate $R_j(\omega)$ we make two important assumptions: (i) that the spectral density matrix of the actual and simulated data can be measured without error and (ii) that the parameters of the model can be selected without error.

The results suggest that the fit of the model depends on the identification scheme used. On average, the size of the error at business cycle frequencies is between 2% and 5% of the spectral density of those variables whose variance is minimized and between 20% and 30% of the spectral density of other variables, suggesting that "some" error should be added to the model to capture the features of the spectral density matrix of the data. Overall, we find small differences in the fit for the two continental blocks, and within continental blocks between the two variables of interest. Changes in the coherences across identifications are somewhat relevant and the model appears to fit coherences much better when we minimize the variance of US variables.

To show how the Monte Carlo techniques discussed in this paper can be used to evaluate the quality of the model's approximation to the data we compute three types of statistics. First, we report how many times on average, at business cycle frequencies, the diagonal elements of the spectral density matrix and the coherences of model generated data lie within a 95% confidence band for the corresponding statistics of actual data. That is, we report $T_1 = \int_{\omega_1}^{\omega_2} \int_{S_1(\omega)}^{S_2(\omega)} p_\omega(x) dx d\omega$ where $S_1(\omega)$ and $S_2(\omega)$ are the lower and upper limits for the asymptotic 95% confidence band for the spectral density of actual data, ω_1 and ω_2 are the lower and upper limits for the business cycle frequencies and $p_\omega(x)$ is the empirical distribution of the simulated

spectral density matrix for the four series at frequency ω .

If the spectral density matrix of the actual data is taken to be the object of interest to be replicated, T_1 reports the *power* of a test which assumes that the model is the correct DGP for the actual data. If we are not willing to assume that the model is the correct DGP for the actual data, these numbers judge the quality of the approximation by informally examining the magnitude of the probability coverings. No matter which interpretation we take, a number close to 95% would indicate a “good” model performance at a particular frequency band.

We compute 95% confidence bands for the actual data in two ways: using asymptotic distribution theory (as in Ortega (1995)) and using a version of the parametric bootstrap procedure of Diebold, Ohanian and Berkowitz (1995). In this latter case, we run a four variable VAR with 6 lags and a constant, construct replications for saving and investment for the two countries by bootstrapping the residuals of the VAR model, estimate the spectral density matrix of the data for each replication and extract 95% confidence bands after ordering the replications, frequency by frequency.

Replications for the time series generated by the model are constructed using Monte Carlo techniques in three different ways. In the first case we simply randomize on the innovations of the technology shocks, keeping their distribution fixed (as in Gregory and Smith (1991)), and use the basic parameter setting displayed in the first column of table 1. In the second and third cases parameters are random and drawn from the distributions listed in the second and third columns of table 1. The results appear in rows 5 to 7 under the heading “Probability Covering”. To economize on space and because simulated results are similar when the 95% confidence bands for actual data are computed asymptotically or by bootstrap, row 5 presents the probability covering using an asymptotic 95% band when only the stochastic processes of the model are randomized, row 6 present the probability covering using an asymptotic 95% band when we randomize on the stochastic processes of the model and parameters are drawn from empirically based distributions, and row 7 when parameters are drawn from normal prior distribution.

The results obtained with this testing approach highlight interesting features of simulated data. With fixed parameters, the average percentage of times the model spectra is inside the 95% band of the actual spectra is, in general, much smaller than 95%, its magnitude depends on the series and it is highest for European saving. When we randomize the parameters using DIW approach, results are more uniform across series and the probability covering is always of the order of 30% while when we randomize using empirical based distributions, the average percentage of times model’s spectra are inside the 95% confidence band is somewhat lower. These results occur because with random parameters, simulated distributions are shifted and stretched: the model produces a wider range of variabilities than those possibly consistent with the data and this reduces the percentage of times simulated data are inside the asymptotic 95% band for each frequency. For coherences the results are very similar across the three rows: in this case, adding parameter variability does not change the outcomes. This is because parameter variability increases the volatility of saving and investment and their covariance by the same factor and this factor cancels out in the computation of coherences. In general, we find that the model slightly “overfits” US coherences, i.e. on average too many simulations fall inside the asymptotic 95% band, while the opposite is true for European coherences. However, with empirical based priors, the coverage in both cases is close to 95%.

In sum, this evaluation procedure confirms that the model is better suited in matching coherences than volatilities at business cycle frequencies and that the covering properties of the model do not improve when we allow the parameters to be random.

To gain further evidence on the properties of the simulated distributions of the data, we next compute a second statistic: the percentile of the simulated distribution of the spectral density matrix of saving and investment for the two countries, where the value of the spectral density matrix of actual data (taken here to be estimated without an error) lies, on average, at business cycle frequencies. Implicitly, this p-value reports, on average over the selected frequency band, the proportion of replications for which the simulated data is less than the historical value. In other words, if $\bar{S}_y(\omega)$ is the spectral density matrix of the actual data at frequency ω we report $T_2 = \int_{\omega_1}^{\omega_2} \int_{-\infty}^{\bar{S}_y(\omega)} p_\omega(x) dx d\omega$ where all variables have been previously defined. Seen through these lenses the spectral density matrix of the actual data is treated as a “critical value” in examining the validity of the theoretical model. Values close to 0% (100%) indicate that the actual spectral density matrix is in the left (right) tail of the simulated distribution of the spectral density matrix of simulated data at that particular frequency band, in which case the model is poor in reproducing the statistics of interest. Values close to 50%, on the other hand, suggest that the actual spectral density matrix at those frequencies is close to the median of the distribution of the spectral density matrix of simulated data so the model is appropriate at those frequencies. Note also that values of the statistic in the range $[\alpha, 100 - \alpha]$, where α is a chosen confidence percentage, would indicate that the model is not significantly at odds with the data. We report the results of this exercise in rows 8 to 10 of table 2 under the heading “Critical Value”. Row 8 presents results when only the innovations of the technology disturbances are randomized, row 9 displays results when the parameters are drawn for normal priors and row 10 when parameters are drawn from an empirical based distribution.

As expected, the model with fixed parameters is unable to match the variabilities of the four series at business cycle frequencies. For all variables the statistics of actual data are in the right tail of the simulated distribution of the statistics at each frequency, i.e., a large proportions of simulations generate average values for the spectral density at business cycle frequencies which are lower than those found in the actual data. For European variables however, the picture is less dramatic. With parameter variability the picture changes. For all variables it is still true that actual variability exceeds the median of the simulated distribution on average at business cycle frequencies, but, at least with empirical priors, it is now within the interquartile range of the simulated distribution for three of the four variables. This is because parameter variability pushes the median of the simulated distribution close to the actual values, shifting the location to the left (less variability is generated). In essence, with parameter variability the model generates two features which improve its overall distributional fit: a wider range of variabilities at business cycle frequencies (with a somewhat larger percentage of more extreme values) and a less concentrated and less skewed distribution.

For coherences the results are somewhat different. With fixed parameters the model generates average coherences at business cycle frequencies which are much higher than in the data for the US but close to the median for Europe (actual values are in the 15th and 50th percentile). With random parameters (and empirical based priors), the situation improves for the US (actual coherence moves up to the 33rd percentile) but worsens for Europe. Once

again, parameter variability enhances the range of possibilities of the model but it fails to tilt the distribution so as to more adequately reproduce the data.

Taken together, the results of these two exercises suggest that with fixed parameters the model generates a distribution for variability which is skewed to the left and only partially overlapping with a normal asymptotic range of variabilities for the data. For coherences the opposite occurs: the overlapping is high but also the skewness within the band is high. Parameter uncertainty, by tilting and stretching the shape of the simulated distribution, ameliorates the situation and in terms of the distributions of certain statistics used, actual and simulated data are almost indistinguishable.

To complete the picture, we finally compute the distributional properties of the approximation error by Monte Carlo methods, i.e. we compute the distribution of the error needed to match the spectral density matrix of the actual data given the model's simulated spectral density matrix. To compute the distributional properties of the log of the error, we draw, at each replication, parameters and innovations from the posterior distribution of the VAR representation of the actual data, construct time series of interest following the procedure of DeJong, Ingram and Whiteman (1995) and estimate the spectral density matrix of the four series. At each replication, we also draw parameters and innovations from the distributions presented in table 1, construct the spectral density matrix of simulated data and compute $S_u^i(\omega) = S_y^i(\omega) - S_x^i(\omega)$, i.e. the error in matching the spectral density matrix of the data, $S_y^i(\omega)$ at replication i . By drawing a large number of replications we can construct a nonparametric estimate of this distribution (using e.g. kernels) and compute moments and fractiles at each frequency. If the model is the correct DGP for the data, the distribution for this error would be degenerate at each frequency. Otherwise the features of this distribution (median value, skewness, kurtosis, etc.) may indicate what is missing from the model to capture the features of interest in the data. The last three rows in table 2 present the median (across replications) of the average error across business cycle frequencies for the six statistics of interest under the heading "Error". Once again, we performed the calculations randomizing both on the stochastic processes of the model and the parameters of the model. Row 11 reports the results when parameters are fixed and rows 12 and 13 when the simulated time series incorporate uncertainty in both stochastic processes and parameters.

The results are quite similar in the three cases for the diagonal elements of the spectral density matrix. The model fails to generate enough variability at business cycle frequencies for US investments while for the other three variables the error is much smaller. The magnitude of the difference is, however, significant. For example for US savings and keeping parameters fixed, the error is about one-third of the actual variability at business cycle frequencies. The results for coherences depend on which of the two countries we consider. For US variables, the model generates systematically higher coherences (negative spectral errors) while for Europe the opposite is true. Relatively speaking, the magnitude of these error are smaller than those obtained comparing spectra. Adding parameter variability as in DeJong, Ingram and Whiteman does not change the results too much. However, when parameters are drawn from empirical based priors, the model generates higher coherences in both cases.

5.3 What did we learn from the exercises?

Our exercise pointed out several important features of the model used by Baxter and Crucini (1993). As claimed by the authors, we find it generates high coherences between national saving and investment at business cycle frequencies which are of the same magnitude as the actual ones for European saving and investment. However, we also saw that the model tends to generate coherences which are uniformly higher than those observed in US data and this is true regardless of whether we used fixed or random parameters. In particular, we show that in only about 20% of the simulations is the simulated coherence smaller than the actual one and that there is tendency of the model to cluster saving and investment correlations in the vicinity of 1. Nevertheless, also in this case, the magnitude of the error is small. The model performance is worse when we try to account for the variability of saving and investment for the two continental blocks at business cycle frequencies. With fixed parameters, the simulated distribution at business cycle frequencies is skewed toward lower than actual values for all variables of interest and that the degree of overlap of simulated and actual distributions varies between 8 and 50%. Parameter variability helps but it does not represent a complete solution to the problem. This is clearly demonstrated by the size of the median value of the spectral error at business cycle frequencies which is sometimes larger than the error obtained with fixed parameters and always positive.

These results suggest that if one is interested in replicating the distributional properties of the statistics of the data (rather than their point estimates), it is necessary to respecify the model, at least for the US. What is primarily needed are two types of features. First, we need some real friction, maybe by adding a new sector (non-traded goods) which uses capital to produce goods; this modification is likely to reduce the median value of the distribution of correlation of saving and investment at business cycle frequencies. Second, we need an additional propagation or variability enhancing device, maybe in the form of a lower adjustment cost of capital or higher elasticity of investment to technology innovations. For the US this can bring simulated variabilities at business cycle frequencies more in the range of the values we found in the data.

6 Conclusions

The task of this chapter was to illustrate how simulation techniques can be used to evaluate the quality of a model's approximation to the data, where the basic theoretical model design is one which fits into what we call a calibration exercise. In section 2 we first provide a definition of what calibration is and then describe in detail the steps needed to generate time series from the model and to select relevant statistics of actual and simulated data. In section 3 we overview four different formal evaluation approaches recently suggested in the literature, comparing and contrasting them on the basis of what type of variability they use to judge the closeness of the model's approximation to the data. In section 4 we describe how to undertake policy analysis with models which have been calibrated and evaluated along the lines discussed in the previous two sections. Section 5 presents a concrete example, borrowed from Baxter and Crucini (1993), where we design four different simulation-based statistics which allow us to shed some light on the quality of the model approximation to the data, in particular, whether the model is able to

reproduce the main features of the spectral density matrix of saving and investment for the US and Europe at business cycle frequencies. We show that, consistent with Baxter and Crucini's claims, the model qualitatively produces a high coherence of saving and investment at business cycle frequencies in the two continental blocks but it also has the tendency to generate a highly skewed simulated distribution for the coherence of the two variables. We also show that the model is less successful in accounting for the volatility features of US and European saving and investment at business cycle frequencies and that taking into account parameter uncertainty helps in certain cases to bring the properties of simulated data closer to those of the actual data.

Overall, the example shows that simulation based evaluation techniques are very useful to judge the quality of the approximation of fully specified general equilibrium models to the data and may uncover features of the model which are left hidden by more simple but more standard informal evaluation techniques.

References

- [1] Backus, D., Kehoe, P. and Kydland, F. (1995), “International Business Cycles: Theory and Evidence”, in T. Cooley (ed.), *Frontiers of Business Cycle Analysis*, Princeton, NJ: Princeton University Press.
- [2] Baxter, M. (1991) “Approximating Suboptimal Dynamic Equilibria: An Euler Equation Approach”, *Journal of Monetary Economics*, 27, 173-200.
- [3] Baxter, M. and Crucini, M. (1993) “Explaining Saving-Investment Correlations”, *American Economic Review*, 83, 416-436.
- [4] Canova, F. (1994) “Statistical Inference in Calibrated Models”, *Journal of Applied Econometrics*, 9, S123-S144.
- [5] Canova, F. (1995) “Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies”, *International Economic Review*, 36, 477-501.
- [6] Canova, F. and De Nicoló, G. (1995), “The Equity Premium and the Risk Free Rate: A Cross Country, Cross Maturity Examination”, CEPR working paper 1119.
- [7] Canova, F., Finn, M. and Pagan, A. (1994), “Evaluating a Real Business Cycle Model”, in C. Hargreaves (ed.), *Nonstationary Time Series Analyses and Cointegration*, Oxford, UK: Oxford University Press.
- [8] Cecchetti, S.G., Lam, P. and N. Mark (1993), “The Equity Premium and the Risk Free Rate: Matching Moments”, *Journal of Monetary Economics*, 31, 21-45.
- [9] Christiano, L. and M. Eichenbaum (1992), “Current Business Cycle Theories and Aggregate Labor Market Fluctuations”, *American Economic Review*, 82, 430-450.
- [10] Cogley, T. and Nason, J.M. (1994), “Testing the Implications of Long-run Neutrality for Monetary Business Cycle Models”, *Journal of Applied Econometrics*, 9, S37-S70.
- [11] Coleman, W. (1989) “An Algorithm to Solve Dynamic Models”, Board of Governors of the Federal Reserve System, International Finance Division, Discussion Paper no. 351.
- [12] Danthine, J.P. and Donaldson, (1992), “Non-Walrasian Economies”, Cahiers de Recherche Economique, Université de Lausanne, No.9301.
- [13] DeJong, D., Ingram, B. and C. Whiteman, (1995), “A Bayesian Approach to Calibration”, forthcoming, *Journal of Business and Economic Statistics*.
- [14] Diebold, F., Ohanian, L. and J. Berkowitz, (1995), “Dynamic Equilibrium Economies: A Framework for Comparing Models and Data”, NBER Technical Working Paper No.174.
- [15] Duffie, D. and Singleton, K. (1993), “Simulated Moments Estimation of Markov Models of Asset Prices”, *Econometrica*, 61, 929-950.

- [16] Eberwein, C. and Kollintzas, T. (1995), “A Dynamic Model of Bargaining in a Unionized Firm with Irreversible Investment”, *Annales d’Economie et de Statistique*, 37-38, 91-116.
- [17] Fève, P. and Langot, F. (1994), “The RBC Models through Statistical Inference: An Application with French Data”, *Journal of Applied Econometrics*, 9, S11-S37.
- [18] Gali, J. (1994), “Monopolistic Competition, Business Cycles and the Composition of Aggregate Demand”, *Journal of Economic Theory*, 63, 73-96.
- [19] Gali, J. (1995), “Non Walrasian Unemployment Fluctuations”, NBER working paper 5337.
- [20] Geweke, J. (1989), “Bayesian Inference in Econometric Models Using Monte Carlo Integration”, *Econometrica*, 57, 1317-1339.
- [21] Genest, C. and Zidak, M. (1986) “Combining Probability Distributions: A Critique and an Annotated Bibliography”, *Statistical Science*, 1, 114-148.
- [22] Gregory, A. and Smith, G. (1989), “Calibration as Estimation”, *Econometric Reviews*, 9(1), 57-89.
- [23] Gregory, A. and Smith, G. (1991), “Calibration as Testing: Inference in Simulated Macro Models”, *Journal of Business and Economic Statistics*, 9(3), 293-303.
- [24] Gregory, A. and Smith, G. (1993), “Calibration in Macroeconomics”, in Maddala, G.S. (ed.), *Handbook of Statistics*, vol. 11, Amsterdam, North Holland.
- [25] Hansen, L. (1982), “Large Sample Properties of Generalized Method of Moment Estimators”, *Econometrica*, 50, 1029-1054.
- [26] Hansen, L. and Sargent, T. (1979), “Formulating and Estimating Dynamic Linear Rational Expectations Models”, *Journal of Economic Dynamic and Control*, 2, 7-46.
- [27] Hansen, L. and Jagannathan, R. (1991), “Implications of Security Market Data for Models of Dynamic Economies”, *Journal of Political Economy*, 99, 225-262.
- [28] *Journal of Business and Economic Statistics*, January 1990.
- [29] King, R., and Plosser, C. (1994), “Real Business Cycles and the Test of the Adelmans” , *Journal of Monetary Economics*, 33, 405-438.
- [30] King, R., Plosser, C. and Rebelo S. (1988), “Production, Growth and Business Cycles: I and II”, *Journal of Monetary Economics*, 21, 195-232 and 309-342.
- [31] Kim, K. and Pagan, A. (1994) “The Econometric Analysis of Calibrated Macroeconomic Models”, forthcoming, in Pesaran, H. and M. Wickens (eds.), *Handbook of Applied Econometrics*, Vol.I, London: Blackwell Press.
- [32] Kollintzas, T. (1992), “Comment to J.P. Danthine: Calibrated Macroeconomic Models: What and What for”, manuscript, Athens University.

- [33] Kuhn, T. (1970), *The Structure of Scientific Revolutions*, Chicago, IL: Chicago University Press.
- [34] Kydland, F. (1992) “ On the Econometrics of World Business Cycles”, *European Economic Review*, 36, 476-482.
- [35] Kydland, F. and Prescott, E. (1982), “Time To Build and Aggregate Fluctuations”, *Econometrica*, 50, 1345-1370 .
- [36] Kydland, F. and Prescott, E. (1991), “The Econometrics of the General Equilibrium Approach to Business Cycles”, *The Scandinavian Journal of Economics*, 93(2), 161-178.
- [37] Marcet, A. (1992) “Solving Nonlinear Stochastic Models by Parametrizing Expectations: An Application to Asset Pricing with Production”, Universitat Pompeu Fabra, working paper 5.
- [38] Marcet, A. (1994), “Simulation Analysis of Stochastic Dynamic Models: Applications to Theory and Econometrics” in Sims, C. (ed.), *Advances in Econometrics. Sixth World Congress of the Econometric Society*, Cambridge: Cambridge University Press.
- [39] Mehra, R. and Prescott, E. (1985), “The Equity Premium: A Puzzle”, *Journal of Monetary Economics*, 15, 145-162.
- [40] McGratten, E. , Rogerson, B. and Wright, R. (1993), “Estimating the Stochastic Growth Model with Household Production”, Federal Reserve Bank of Minneapolis, manuscript.
- [41] Niederreiter, H. (1988), “ Quasi Monte Carlo Methods for Multidimensional Numerical Integration”, *International Series of Numerical Mathematics*, 85, 157-171.
- [42] Novales, A. (1990), “Solving Nonlinear Rational Expectations Models: A Stochastic Equilibrium Model of Interest Rates”, *Econometrica*, 58, 93-111.
- [43] Ortega, E. (1995), “Assessing and Comparing Computable Dynamic General Equilibrium Models”, manuscript, European University Institute.
- [44] Pagan, A. (1994), “Calibration and Econometric Research: An Overview”, *Journal of Applied Econometrics*, 9, S1-S10.
- [45] Pagan, A. and Shannon, (1985), “Sensitivity Analysis for Linearized Computable General Equilibrium Models”, in J.Piggott and J. Whalley (eds.) *New Developments in Applied General Equilibrium Analysis*, Cambridge: Cambridge University Press.
- [46] Pesaran, H. and Smith, R. (1992), “The Interaction between Theory and Observation in Economics”, University of Cambridge, manuscript.
- [47] Sargent, T. (1987), *Dynamic Macroeconomic Theory*, Cambridge, Ma: Harvard University Press.

- [48] Showen, J. and Whalley, J. (1984), “Applied General Equilibrium Models of Taxation and International Trade: An Introduction and Survey”, *Journal of Economic Literature*, 22, 1007-1051.
- [49] Simkins, S.P. (1994), “Do Real Business Cycle Models Really Exhibit Business Cycle Behavior?”, *Journal of Monetary Economics*, 33, 381-404.
- [50] Smith, T. (1993) “Estimating Nonlinear Time Series Models Using Simulated VAR”, *Journal of Applied Econometrics*, 8, S63-S84.
- [51] Söderlind, P. (1994), “Cyclical Properties of a Real Business Cycle Model”, *Journal of Applied Econometrics*, 9, S113-S122.
- [52] Summers, L. (1991), “Scientific Illusion in Empirical Macroeconomics”, *Scandinavian Journal of Economics*, 93(2), 129-148.
- [53] Tauchen, G. and Hussey, R. (1991) “Quadrature Based Methods for Obtaining Approximate Solutions to Integral Equations of Nonlinear Asset Pricing Models”, *Econometrica*, 59, 371-397.
- [54] Watson, M. (1993) “Measures of Fit for Calibrated Models”, *Journal of Political Economy*, 101, 1011-1041.
- [55] Wolf, F. (1986) *Meta-Analysis: Quantitative Methods for Research Synthesis*, Beverly Hill, Ca.: Sage Publishers.

Table 1: **Parameter values used in the simulations**

Parameter	Basic	Empirical Density	Prior Normal
Steady State hours (H)	0.20	Uniform[0.2, 0.35]	Normal(0.2, 0.02)
Discount Factor (β)	0.9875	Trunc.Normal[0.9855, 1.002]	Normal(0.9875, 0.01)
Risk Aversion (σ)	2.00	Truncated $\chi^2(2)[0, 10]$	Normal(2, 1)
Share of Labor in Output (α)	0.58	Uniform[0.50, 0.75]	Normal(0.58, 0.05)
Growth rate (θ_x)	1.004	Normal(1.004, 0.001)	1.004
Depreciation Rate of Capital (δ)	0.025	Uniform[0.02, 0.03]	Normal(0.025, 0.01)
Persistence of Disturbances (ρ)	0.93	Normal(0.93, 0.02)	Normal(0.93, 0.025)
Lagged Spillover of Disturbances (ν)	0.05	Normal(0.05, 0.03)	Normal(0.05, 0.02)
Standard Deviation of Technology Innovations (σ_ϵ)	0.00852	Truncated $\chi^2(1) [0, 0.0202]$	Normal(0.00852, 0.004)
Contemporaneous Spillover (ψ)	0.40	Normal(0.35, 0.03)	Normal(0.4, 0.02)
Country Size (π)	0.50	Uniform[0.10, 0.50]	0.5
Elasticity of marginal adjustment cost function ($\xi_{\phi'}$)	-0.075	-0.075	-0.075
Steady State Tobin's Q ($\frac{1}{\phi r}$)	1.0	1.0	1.0
Tax Rate (τ)	0.0	0.0	0.0

Notes: “Empirical density” refers to distributions for the parameters constructed using either existing estimates or a-priori intervals as in Canova (1994). “Prior Normal” refers to distributions for the parameters which are a-priori chosen as in DeJong, Ingram and Whiteman (1995). The range for the parameter is reported inside the brackets. The mean and the standard deviation for the distribution are reported inside the parentheses.

Table 2: **Fit of the model at Business Cycle frequencies**

	US Spectra		Europe Spectra		US Coherence	Europe Coherence
	S	I	S	I	S-I	S-I
Actual data	0.75	0.88	0.68	0.49	85.41	93.14
Simulated data	0.36	0.18	0.35	0.18	94.04	93.00
Watson approach						
Identification 1	0.02	0.05	0.20	0.23	0.04	0.13
Identification 2	0.24	0.21	0.05	0.04	0.20	0.15
Probability Covering						
Fixed parameters	46.46	8.63	55.71	43.57	98.99	92.91
Normal distribution	35.30	23.40	32.89	37.00	98.17	90.34
Empirical distribution	19.63	18.60	21.11	20.20	94.71	95.69
Critical Value						
Fixed parameters	90.80	99.89	82.16	93.91	15.60	49.04
Normal Distribution	71.80	89.90	66.00	76.60	19.80	51.89
Empirical distributions	62.50	79.70	73.30	74.60	33.46	29.60
Error						
Fixed parameters	0.25	0.55	0.30	0.28	-9.17	0.37
Normal Distribution	0.19	0.56	0.29	0.28	-9.01	0.81
Normal distribution	0.13	0.58	0.42	0.35	-6.07	-2.86

Notes: Actual and simulated data are linearly detrended and logged, in real per capita terms. Simulations are undertaken using 500 draws. All rows except the third and the fourth report numbers in percentage terms. “Watson approach” reports the average statistic (4) at business cycle frequencies, “Probability covering” reports the average covering at business cycle frequencies of the theoretical 95% range, “Critical value” the percentile where the actual data lies on average at business cycle frequencies, and “Error” the median error across simulations on average at business cycle frequencies. S refers to saving and I to investment.