JOURNAL OF APPLIED ECONOMETRICS, VOL. 9, S123–S144 (1994)

STATISTICAL INFERENCE IN CALIBRATED MODELS

FABIO CANOVA

Department of Economics, Universitat Pompeu Fabra, Balmes 132, 08008 Barcelona, Spain and Department of Economics, Università di Catania, 95100 Catania, Italy, and CEPR

SUMMARY

This paper describes a Monte Carlo procedure to assess the performance of calibrated dynamic general equilibrium models. The procedure formalizes the choice of parameters and the evaluation of the model and provides an efficient way to conduct a sensitivity analysis for perturbations of the parameters within a reasonable range. As an illustration the methodology is applied to two problems: the equity premium puzzle and how much of the variance of actual US output is explained by a real business cycle model.

1. INTRODUCTION

The current macroeconometrics literature has proposed two ways to confront general equilibrium rational expectations models with data. The first, an estimation approach, is the direct descendant of the econometric methodology proposed 50 years ago by Haavelmo (1944). The second, a calibration approach, finds its justification in the work of Frisch (1933) and is closely linked to the computable general equilibrium literature surveyed e.g. in Shoven and Whalley (1984).

The two methodologies share the same strategy in terms of model specification and solution. Both approaches start from formulating a fully specified general equilibrium dynamic model and in selecting convenient functional forms for preferences, technology, and exogenous driving forces. They then proceed to find a decision rule for the endogenous variables in terms of the exogenous and predetermined variables (the states) and the parameters. When the model is nonlinear, closed-form expressions for the decision rules may not exist and both approaches rely on recent advantages in numerical methods to find an approximate solution which is valid either locally or globally (see e.g. the January 1990 issue of the *Journal of Business and Economic Statistics* for a survey of the methods and Christiano, 1990, and Dotsey and Mao, 1991, for a comparison of the accuracy of the approximations).

It is when it comes to choosing the parameters to be used in the simulations and in evaluating the performance of the model that several differences emerge. The first procedure attempts to find the parameters of the decision rule that best fit the data either by maximum likelihood (ML) (see e.g. Hansen and Sargent, 1979, or Altug, 1989) or generalized method of moments (GMM) (see e.g. Hansen and Singleton, 1983, or Burnside *et al.*, 1993). The validity of the specification is examined by testing restrictions, by general goodness of fit tests or by comparing the fit of two nested models. The second approach 'calibrates' parameters using a set of alternative rules which includes matching long-run averages, using previous microevidence or *a priori* selection, and assesses the fit of the model with an informal distance criterion.

These differences are tightly linked to the questions the two approaches ask. Roughly speaking, the estimation approach asks the question 'Given that the model is true, how false is

CCC 0883–7252/94/0S0S123–22

Received July 1992
Revised August 1994

it?' while the calibration approach asks 'Given that the model is false, how true is it?' Implicit in the process of estimation is in fact the belief that the probability structure of a model is sufficiently well specified to provide an accurate description of the data. Because economic models are built with an emphasis on tractability, they are often probabilistically underspecified so that measurement errors or unobservable shocks are added at the estimation stage to complete their probability structure (see e.g. Hansen and Sargent, 1980, or Altug, 1989). By testing the model, a researcher takes the model seriously as a data-generating process (DGP) and examines what features of the specification are at variance with the data. A calibrationist takes the opposite view: the model, as a DGP for the data, is false. That is, as the sample size grows, it is known that the data are generated by the model will be at increasingly greater variance with the observed time series. An economic model is seen, at best, as an approximation to the true DGP which need not be either accurate or realistic and, as such, should not be regarded as a null hypothesis to be statistically tested (see Prescott, 1991, p. 5). In confronting the model with the data, a calibrationist wants to indicate the dimensions where the approximation is poor and suggest modifications to the theoretical model in order to obtain a better approximation.

Both methodologies have weak points. Model estimation involves a degree of arbitrariness in specifying which variables are unobservable or measured with error. In the limit, since all variables are indeed measured with error, no estimation seems possible and fruitful. In addition, tests of the model's restrictions may fail to indicate how to alter the specification to obtain a better fit. The limitations of the calibration approach are also well known. First, the selection criterion for parameters which do not measure long-run averages is informally specified and may lead to contradictory choices. Information used in different studies may in fact be inconsistent (e.g. a parameter chosen to match labour payments from firms in national account data may not equal the value chosen to match the labour income received by households) and the range of estimates for certain parameters (e.g. risk aversion parameter) is so large that selection biases may be important. Second, the outcomes of the simulations typically depend on the choice of unmeasured parameters. However, although some authors (see e.g. Prescott, 1991, p. 7, or Kydland, 1992, p. 478) regard a calibration exercise as incomplete unless the sensitivity of the results to reasonable perturbations of the parameters selected *a priori* or not well tied to the data is reported, such an analysis is not often performed. Third, because the degree of confidence in the results depends on both the degree of confidence in the theory *and* in the underlying measurement of the parameters and because either parameter uncertainty is disregarded or, when a search is undertaken, the number of replications typically performed is small, we must resort to informal techniques to judge the relevance of the theory.

This paper attempts to eliminate some of the weaknesses of calibration procedures while maintaining the general analytical strategy employed in calibration exercises. The focus is on trying to formalize the selection of the parameters and the evaluation process and in designing procedures for meaningful robustness analysis on the outcomes of the simulations. The technique we propose shares features with those recently described by Gregory and Smith (1991) and Kwan (1990), has similarities with stochastic simulation techniques employed in dynamic nonlinear large scale macro models (see e.g. Fair, 1991), and generalizes techniques on randomized design for strata existing in the static computable general equilibrium literature (see e.g. Harrison and Vinod, 1989).

The idea of the technique is simple. We would like to reproduce features of actual data, which is taken to be the realization of an unknown vector stochastic process, with an 'artificial economy' which is almost surely the incorrect generating mechanism for the actual data. The features we may be interested in include conditional and unconditional moments (or densities), the autocovariance function of the data, functions of these quantities (e.g. measures of relative volatility), or specific events (e.g. a recession). A model is simulated repeatedly using a Monte

Carlo procedure which randomizes over *both* the exogenous stochastic processes *and* the parameters. Parameters are drawn from a data-based density which is consistent with the information available to a simulator (which may include both time-series and cross-sectional aspects). We judge the validity of a model on its ability to reproduce a number of 'stylized facts' of the actual economy (see Friedman, 1959). The metric used to evaluate the discrepancy of the model from the data is probabilistic. We construct the simulated distribution of the statistics of interest and, taking the actual realization of the statistic as a critical value, examine (1) in what percentile of the simulated distribution the actual value lies and (2) how much of the simulated distribution is within a $k\%$ region centred around the critical value. Extreme values for the percentile (say, below $\alpha\%$ or above $(1-\alpha)\%$) or a value smaller than k for the second probability indicates a poor fit in the dimensions examined.

The approach has several appealing features. First, it accounts for the uncertainty faced by a simulator in choosing the parameters of the model in a realistic way. Second, it has a built-in feature for global sensitivity analysis on the support of the parameter space and allows for other forms of conditional or local sensitivity analysis. Third, it provides general evaluation criteria and a simple and convenient framework to judge the relevance of the theory.

The paper is divided into six sections. The next section introduces the technique, provides a justification for the approach and describes the details involved in the implementation of the procedure. Section 3 deals with robustness analysis. Section 4 spells out the relationship with existing techniques. Two examples describing the potential of the technique for problems of practical interest appear in Section 5. Section 6 presents conclusions.

2. THE TECHNIQUE

A General Framework of Analysis

We assume that a researcher is faced with an $m \times 1$ vector of time series $\bar{x}_t$, which are the realizations of a vector stochastic process $\bar{X}_t$ and that she is interested in reproducing features of $\bar{X}_t$ using a dynamic general equilibrium model. $\bar{X}_t$ is assumed to have a continuous but unknown distribution and moments up to the nth. For the sake of presentation we assume that the unconditional distribution of $\bar{X}_t$ is independent of t but shifts in the unconditional distribution of $\bar{X}_t$ at known points can easily be handled. $\bar{X}_t$ may include macro variables like GNP, consumption, interest rates, etc. We also assume that dynamic economic theory gives us a model expressing the endogenous variables X_t as a function of exogenous and predetermined variables Z_t (the states of the problem) and of the parameters β. Z_t may include objects like the existing capital stock, exogenous fiscal, and monetary variables or shocks to technologies and preferences. We express the model's functional relation as $X_t = f(Z_t, \beta)$. Under specific assumptions about the structure of the economy (e.g. log or quadratic preferences, Cobb–Douglas production function, full depreciation of the capital stock in the one-sector growth model), f can be computed analytically either by value function iteration or by solving the Euler equations of the model subject to the transversality condition (see e.g. Hansen and Sargent, 1979). In general, however, f cannot be derived analytically from the primitives of the problem. A large body of current literature has concentrated on the problem of finding approximations which are either locally or globally close to f for a given metric.[1]

[1]These include linear or log-linear expansions of f around the steady state (Kydland and Prescott, 1982; and King *et al.*, 1988), backward-solving methods (Sims, 1984; Novales, 1990), global functional expansions in polynomials (Marcet, 1992; Judd, 1992), piecewise linear interpolation methods (Coleman, 1989; Baxter, 1991) and quadrature techniques (Tauchen and Hussey, 1991).

Here we assume that either f is available analytically or that one of the existing numerical procedures has been employed to obtain a functional $\mathcal{F}$ which approximates f in some sense, i.e. $\| \mathcal{F}(Z_t, \gamma) - f(Z_t, \beta) \| < \varepsilon$, where $\gamma = \iota(\beta)$ and $\| . \|$ is a given norm. Given the model f, an approximation procedure $\mathcal{F}$, a set of parameters β, and a probability distribution for Z_t (denoted by $\kappa(Z_t)$), we can infer the model-based probability distribution for X_t.

Let $\mathcal{G}(X_t \mid \beta, f)$ be the density of the X_t vector, conditional on the parameters β and the model f, let $\pi(\beta \mid \mathcal{I}, f)$ be the density of the parameters, conditional on the information set $\mathcal{I}$ available to the simulator and the model f, and let $\mathcal{H}(X_t, \beta \mid f, \mathcal{I})$ be the joint density of simulated data and of parameters. $\mathcal{G}(X_t \mid \beta, f)$ represents the probability that a particular path for the endogenous variables will be drawn given a parametric structure for the artificial economy and a set of parameters, while $\pi(\beta \mid \mathcal{I}, f)$ summarizes the information on the parameters available to a researcher. Note that $\mathcal{G}$ is assumed to be independent of $\mathcal{I}$ and π may depend on f, i.e. if we are using a GE model we may want to use only estimates obtained with similar GE models. For a given β, X_t is random because Z_t is random, i.e. $\mathcal{G}(X_t \mid \beta, f)$ is a deterministic transformation of $\kappa(Z_t)$.

Throughout this paper we are interested in studying the behaviour of functions of simulated data (denoted by $\mu(X_t)$) under the predictive density $p(X_t \mid \mathcal{I}, f) = \int \mathcal{H}(X_t, \beta \mid f, \mathcal{I})\, d\beta$, i.e. evaluating objects of the form:

$$E(\mu(X_t) \mid f, \mathcal{I}, \mathcal{A}, \mathcal{C}) = \int_{\mathcal{C}} \mu(X_t) p(X_t, \mathcal{I}, f)\, dX_t$$

$$= \int_{\mathcal{A}} \int_{\mathcal{C}} \mu(X_t) \mathcal{H}(X_t, \beta \mid \mathcal{I}, f)\, d\beta\, dX_t \tag{1}$$

where $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{B}$ is the parameter space and $\mathcal{C}$ is the support of the exogenous variables. Let $h(\tilde{x}_t)$ be the corresponding vector of functions of the actual data.

The problem of measuring the fit of the model can be summarized as follows. How likely is the model to generate $h(\tilde{x}_t)$? To answer note that from equation (1) we can compute probabilities of the form $P(\nu(X_t) \in D)$, where D is a bounded set and $\nu(X_t)$ includes moments and other statistics of the simulated data. To do this let $\mu(X_t) = \chi(X_t, [X_t : \nu(X_t) \in D])$ *where* $\chi(X_t, S)$ is the indicator function, i.e. $\chi(X_t, S) = 1$ if $\nu(X_t) \in S$ and zero otherwise. Similarly, we can construct quantiles $q(X_t)$ by appropriately choosing D (see e.g. Geweke, 1989). Finally, we can also find a $\tilde{h}$ satisfying $\mathrm{P}[\nu(X_t) \leqslant \tilde{h}] = \upsilon$ for any given vector υ, by appropriately selecting the indicator function.

Model evaluation then consists of several types of statements which are interchangeable and differ only in the criteria used to measure distance. First, we can compute $\mathrm{P}[\nu(X_t) \leqslant h(\tilde{x}_t)]$. In other words, we can examine the likelihood of an event (the observed realization of the summary statistics in the actual data) from the point of view of the model. Extreme values for this probability indicate a poor 'fit' in the dimensions examined. Alternatively, if we can measure the sampling variability of $h(\tilde{x}_t)$, we can then choose the set D to include the actual realization of $h(\tilde{x}_t)$ plus one or two standard deviations and either check whether $\tilde{h}$ is in D or calculate $\mathrm{P}[\nu(X_t) \in D]$.

Implementation

There are four technical implementation issues which deserve some discussion. The first concerns the computation of integrals like those in equation (1). When the (β, Z_t) vector is of high-dimension simple discrete grid approximations, spherical or quadrature rules quickly become infeasible since the number of function evaluations increases exponentially with the

dimension of β and Z_t. In addition, unless the contours of $\mathcal{H}(X_t, \beta \mid \mathcal{I}, f)$ (and of $p(X_t \mid \mathcal{I}, f)$) are of ellipsoidal forms, grid evaluations may explore this density inefficiently. There are several feasible alternatives: one is the Monte Carlo procedure described in Geweke (1989), another is the data-augmentation procedure of Tanner and Wong (1987), a third is the 'Gibbs sampler' discussed in Gelfand and Smith (1990). Finally, we could use one of the quasi-random procedures proposed by Niederreiter (1988).

In this paper we adopt a Monte Carlo approach. After drawing with replacement i.i.d. β vectors and Z_t paths, we substitute sums over realizations for the integrals appearing in equation (1) and appeal to the strong law of large numbers for functions of i.i.d. random variables to obtain

$$\frac{1}{N}\sum_{i=1}^{N}\mu_i(X_t) \xrightarrow{\text{a.s.}} E(\mu(X_t)) \tag{2}$$

where N is the number of replications. Note that, although $\mathcal{H}$ (and p) are, in general, unknown, sampling from them can be conveniently accomplished by simulating the model repeatedly for random (Z_t, β), i.e. randomly drawing exogenous forces and selecting a parameter vector and using the decision rule to compute time paths for X_t.

Second, since in most cases the function f is unknown, $\mathcal{G}$ itself becomes unknown and the direct computation of equation (1) is infeasible. If the approximation $\mathcal{F}$ to f is accurate, we could simply neglect the error and proceed using $\mathcal{J}(X_t, \beta \mid \mathcal{I}, \mathcal{F})$ in place of $\mathcal{H}(X_t, \beta \mid \mathcal{I}, f)$ where $\mathcal{J}$ is the joint density of simulated data and parameters using the information set $\mathcal{I}$ and the approximation rule $\mathcal{F}$. However, since only little is known about the properties ofapproximation procedures and some have only local validity (see e.g. Christiano, 1990; Dotsey and Mao, 1991), we may want to explicitly account for the existence of an approximation error in conducting inference. In this case, following Geweke (1989), we would replace equation (1) with:

$$E(\mu(X_t) \mid f, \mathcal{I}, \mathcal{A}, \mathcal{C}) = \int_{\mathcal{A}}\int_{\mathcal{C}} \mu(X_t)\mathcal{J}(X_t, \beta \mid \mathcal{I}, \mathcal{F})\mathcal{L}(\beta, f, \mathcal{F})\, d\beta\, dX_t \tag{3}$$

where $\mathcal{L}(\beta, f, \mathcal{F})$ are weights which depend on the 'true' density $\mathcal{H}(X_t, \beta \mid \mathcal{I}, f)$ and on the approximation density $\mathcal{J}(X_t, \beta \mid \mathcal{I}, \mathcal{F})$. For example, if a quadratic approximation around the steady state is used, the density $\mathcal{L}$ can be chosen so that draws of Z_t inducing paths of X_t which are in the tails of $\mathcal{J}$ (i.e. paths which are very far away from steady states) receive a very small weight in the calculation of the statistics of interest.

Third, we must specify a density for the parameters of the model. We could select it on the basis of one specific data set and specify $\pi(\beta \mid \mathcal{I}, f)$ to be the asymptotic distribution of a GMM estimator (as in Burnside *et al.*, 1993), of a simulated method of moments (SMM) estimator (as in Canova and Marrinan, 1993), or of a ML estimator of β (as in Phillips, 1991). The disadvantage of these approaches is that the resulting density measures the uncertainty surrounding β present in a particular data set and does not necessarily reflect the uncertainty faced by a researcher in choosing the parameters of the model. As Larry Christiano has pointed out to the author, once a researcher chooses the moments to match, the uncertainty surrounding estimates of β is small. The true uncertainty lies in the choice of moments to be matched and in the sources of data to be used to compute estimates.

A better approach would be to select $\pi(\beta \mid \mathcal{I}, f)$ so as to summarize efficiently all existing information, which may include point estimates of β obtained from different estimation techniques, data sets, or model specifications. El-Gamal (1993a,b) has formally solved the

problem of finding such a $\pi(\beta \mid \mathcal{I}, f)$ using Bayesian methods. The resulting $\pi(\beta \mid \mathcal{I}, f)$ is the least informative pseudo-posterior density on the parameter space which is consistent with a set of constraints describing the information contained in various estimation experiments. El-Gamal suggests a Gibbs sampler algorithm to compute this density but, in practice, there are simpler ways to construct empirical approximations to this type of density. One would be to count estimates of β previously obtained in the literature and construct $\pi(\beta \mid \mathcal{I}, f)$ by smoothing the resulting histogram. For example, if one of the elements of the β vector is the risk aversion parameter, counting estimates obtained over the last 15 years from fully specified general equilibrium models and smoothing the resulting histogram, we would obtain a truncated (below-zero) bell-shaped density, centred around two and very small massabove four. Alternatively, we could take what the profession regards as a reasonable range for β and assume more or less informative densities on the support depending on available estimates. If theoretical arguments suggest that the maximum range for e.g. the risk aversion parameter is $[0, 20]$, we can put higher weights on the interval $[1, 3]$ where most of the estimates lie. If for some parameters previous econometric evidence is scant, measurement is difficult, or there are no reasons to expect that one value is more likely than others, we could assume uniform densities on the chosen support.

Available estimates of β are not necessarily independent (the same data set is used in many cases) and some are less reliable than others. Non-independent estimates are legitimate candidates to enter the information set as long as they reflect sampling variability or different estimation techniques. The influence of less reliable estimates or of estimates obtained with different model specifications can be discounted by giving them a smaller weight in constructing histograms (see also El-Gamal, 1993a,b).

Finally, in many applications the joint density of the parameter vector can be factored into the product of lower-dimensional densities. If no relationship across estimates of the various parameters exists, $\pi(\beta \mid \mathcal{I})$ is the product of univariate densities. If estimates of certain parameters are related (e.g. in the case of the discount factor and the risk aversion parameter in asset pricing models), we can choose multivariate densities for these dimensions and maintain univariate specifications for the densities of the other parameters.

To summarize, to implement the procedure we need to do the following:

- Select a reasonable (data-based) density $\pi(\beta \mid \mathcal{I}, f)$, where $\mathcal{I}$ represents the information set available to a researcher, and a density $\kappa(Z_t)$ for the exogenous processes.
- Draw vectors β from $\pi(\beta \mid \mathcal{I}, f)$ and z_t from $\kappa(Z_t)$.
- For each draw of β and z_t, generate $\{x_t\}_{t=1}^{T}$ and compute $\mu(x_t)$ using the model $x_t = f(z_t, \beta)$ or the approximation $x_t = \mathcal{F}(z_t, \gamma)$.
- Repeat the two previous steps N times.
- Construct the frequency distribution of $\mu(x_t)$, compute probabilities, quantiles and other measures of interest.

An Interpretation

The proposed framework of analysis lends itself to a simple Bayesian interpretation. In this case we treat $\pi(\beta \mid \mathcal{I}, f)$ as the prior on the parameters. The function $\mathcal{G}$ is entirely analogous to a classical likelihood function for X_t in a standard regression model. The difference is that $\mathcal{G}$ need not be the correct likelihood for $\bar{X}_t$ and need not have a closed form. Then equation (1) is the conditional expectation of $\mu(X_t)$ under the predictive density of the model and probability statements based on equation (1) can be justified using the arguments contained in Box (1980).

There is also a less orthodox interpretation of the approach which exchanges the role of $\pi(\beta \mid \mathcal{I}, f)$ and $\mathcal{G}(X_t \mid \beta, f)$ and is nevertheless reasonable. In this case $\mathcal{G}(X_t \mid \beta, f)$ is the prior and represents the *a priori* degree of confidence posed by the researcher on the time path generated by the model given the parameters while $\pi(\beta \mid \mathcal{I}, f)$ summarizes the information contained in the data. Then equation (1) is a 'pseudo-posterior' statement about the model's validity once the empirical evidence on the parameters is taken into account.

It is useful to note that, if we follow the first approach, we can relate the proposed construction of $\pi(\beta \mid \mathcal{I}, f)$ to the data-based priors employed in Meta-Analysis (see Wolf, 1986) and in the 'consensus literature' (see e.g. Genest and Zideck, 1986). El-Gamal (1993a) spells out in detail the connection with these two strands of literature.

3. ROBUSTNESS ANALYSIS

If we adopt a Monte Carlo approach to compute simulated densities for the statistics of interest, an automatic and efficient global sensitivity analysis is performed on the support of the parameter space as a by-product of the simulations. Sensitivity analysis, however, can take other more specific forms. For example, we may be interested in examining how likely $\mu(X_t)$ is to be close to $h(\bar{x}_t)$ when β is fixed at some prespecified value $\hat{\beta}$. This would be the case, for example, if β includes parameters which can be controlled by the government and $h(\bar{x}_t)$ is e.g. the current account balance of that country. In this case we could choose a path for Z_t and analyse the conditional distribution of $\mu(X_t)$ for the selected value(s) of β. Alternatively, we might wish to assess the maximal variation in $\mu(X_t)$ consistent, say, with β being within two standard deviations of a particular value. Here we choose a path for Z_t and construct paths for $\mu(X_t)$ for draws of β in the range. Finally, we may be interested in knowing which dimensions of β are responsible for particular features of the distribution of $\mu(X_t)$. For example, if the simulated distribution of $\mu(X_t)$ has a large spread or fat tails, a researcher may be interested in knowing whether technology or preference parameters are responsible for this feature. In this case we would partition β into $[\beta_1, \beta_2]$ and compute the simulated distribution of $\mu(X_t)$ conditional on $\beta_2 = \hat{\beta}_2$, where $\hat{\beta}_2$ is a prespecified value (or set of values.

So far, we have examined the robustness of the results to variations of the parameters within their support. In some cases it is necessary to study the sensitivity of the results to local perturbations of the parameters. For example, we may be interested in determining how robust the simulation results are to changes of the parameters in a small neighbourhood of a particular vector of calibrated parameters. To undertake this type of analysis we can take a numerical version of the average derivative of $\mu(X_t)$ in the neighbourhood of a calibrated vector (see Pagan and Ullah, 1991). Because global and local analyses aim at examining the sensitivity of the outcomes to perturbations in the parameters of different size they provide complementary information and should both be used as specification diagnostics for models whose parameters are calibrated.

4. A COMPARISON WITH EXISTING METHODOLOGIES

The framework of analysis in Section 2 is general enough to include simulation undertaken after the parameters are both calibrated and estimated via method of moments as special cases. To show this it is convenient to recall that $\mathcal{H}(X_t, \beta \mid \mathcal{I}, f)$ is a deterministic transformation of $Q(Z_t, \beta \mid \mathcal{I}, f) = \pi(\beta \mid \mathcal{I}, f)\, \kappa(Z_t)$. The two procedures can then be recovered by imposing restrictions on the shape and the location of $\pi(\beta \mid \mathcal{I}, f)$ and, in some cases, also on the shape and the location of $\kappa(Z_t)$.

Calibration exercises impose a point mass for $\pi(\beta \mid \mathcal{I}, f)$ on a particular value of β (say, $\hat{\beta}$). One interpretation of this density selection is that a simulator is perfectly confident in the vector β used and does not worry about the cross-study or time-series uncertainty surrounding estimates of β. In certain situations a path for the vector of exogenous variables is also selected in advance either by drawing only one realization from their distribution or by choosing a z_t on the basis of extraneous information, e.g. inputting Solow's residuals in the model, so that $\kappa(Z_t)$ is also a singleton. In this instance, the density of $\mu(X_t)$ has a point mass and because the likelihood of the model to produce any event is either 0 or 1, we must resort to informal techniques to assess the discrepancy of simulated and actual data. In some studies the randomness in Z_t is explicitly considered, repeated draws for the exogenous variables are made for a fixed $\hat{\beta}$, and moments of the statistics of interest are computed by averaging the results over a number of simulations (see e.g. Backus *et al.*, 1989).

Simulation exercises undertaken with estimation of the parameters are also special cases of the above framework. Here $\pi(\beta \mid \mathcal{I})$ has a point mass at β^*, where β^* is either the GMM estimator, the SMM estimator (see Lee and Ingram, 1991), or the simulated quasi-maximum likelihood (SQML) estimator of β (see Smith, 1993). Simulations are typically performed by drawing one realization from $\mathcal{G}(X_t \mid \beta^*, f, \mathcal{I})$ and standard errors for $\mu(X_t)$ are computed using the asymptotic standard errors of β^* and the functional form for μ. In some cases, $\pi(\beta \mid \mathcal{I})$ is taken to be the asymptotic distribution of one of the above estimators (e.g. Canova and Marrinan, 1993). In this case, simulations are performed by drawing from $\mathcal{G}(X_t \mid \beta^*, f, \mathcal{I})\pi(\beta^* \mid \mathcal{I})$ and the distance of simulated and actual data is computed using measures of discrepancy like the ones proposed here.

In assessing the model's performance this last set of procedures has two advantages over calibration. First, they allow formal statements on the likelihood of selected parameter values to reproduce the features of interest. For example, if a four standard deviations range around the point estimate of the AR(1) parameter for the productivity disturbance is [0·84, 0·92], then it is highly unlikely (with a probability higher than 99%) that a unit root productivity disturbance is needed to match the data. Second, they provide a set-up where sensitivity analysis can easily be undertaken (although not often performed).

These procedures, however, have also two major shortcomings. First, they impose a strong form of ignorance which does not reflect available *a priori* information. The vector β may include meaningful economic parameters which can be bounded on the basis of theoretical arguments but the range of possible β with GMM, SMM, or SQML procedures is $[-\infty, \infty]$. By appropriately selecting a hypercube for their densities a researcher can make 'unreasonable' parameter values unlikely and avoid *a posteriori* adjustments. Second, simulations conducted after parameters are estimated may not constitute an independent way to validate the model because the parameter estimates are obtained from the same data set which is used later to compare results.

Simulation procedures where parameters are selected using a mixture of calibration and estimation strategies have recently been employed by e.g. Heaton (1993) and Burnside *et al.* (1993). Here some parameters are fixed using extraneous information while others are formally estimated using moment (or simulated moment) conditions. Although these strategies allow a more formal evaluation of the properties of the model than pure calibration procedures, they face two problems. First, as in the case when the parameters are all selected using GMM, SMM, and SQML procedures, the evaluation of the model is problematic because measures of dispersion for the statistics of interest are based on one data set and do not reflect the uncertainty faced by a simulator in choosing the unknown features of the model. Second, as Gregory and Smith (1989) have pointed out, the small-sample properties of estimators obtained from these

strategies may be far from reasonable unless calibrated parameters consistently estimate the true parameters. When this condition is not satisfied, estimates of the remaining parameters are sensitive to errors in pre-setting and results are misleading.

The Monte Carlo methodology we employ to evaluate the properties of the model is related to those of Kwan (1990) and Gregory and Smith (1991) but several differences need to be emphasized. First, Gregory and Smith take the model as a testable null hypothesis while this is not the case here. Second, they do not account for parameter uncertainty in evaluating the outcomes of the model. Third, because they take a calibrated version of the model as the 'truth', they conduct sensitivity analysis inefficiently, by replicating experiments for different calibrated values of the parameters. Kwan (1990) allows for parameter uncertainty in his simulation scheme, but, following an orthodox Bayesian approach, he chooses a subjective prior density for the parameters. In addition, he evaluates the outcomes of the model in relative terms by comparing two alternative specifications using a posterior-odds ratio: a model is preferred to another if it maximizes the probability that the simulated statistics are in a prespecified set.

The procedure for sensitivity analysis we proposed extends the approach that Harrison and Vinod (1989) used in deterministic computable general equilibrium models. The major difference is that in a stochastic framework parameter uncertainty is only a part of the randomness entering the model and the uncertainty characterizing the exogenous processes is important in determining the randomness of the outcomes.

To conclude, we should mention that, parallel to the literature employing Monte Carlo methods to evaluate calibrated models, there is also a branch of the literature which uses alternative tools to examine the fit of calibrated models. This is the case e.g. of Smith (1993), Watson (1993), and Canova *et al.* (1993) which assess the relevance of theoretical models with regression R^2's, tests based on restricted and unrestricted VARs, and encompassing procedures.

5. TWO EXAMPLES

The Equity Premium Puzzle

Mehra and Prescott (1985) suggest that an asset-pricing model featuring complete markets and pure exchange cannot simultaneously account for the average risk-free rate and the average equity premium experienced by the US economy over the sample 1889–1978 with reasonable values of the risk aversion parameter and of the discount factor.

The model they consider is a frictionless Arrow–Debreu economy with a single representative agent, one perishable consumption good produced by a single productive unit or a 'tree', and two assets, an equity share and a risk-free asset. The tree yields a random dividend each period and the equity share entitles its owner to that dividend in perpetuity. The risk-free asset entitles its owner to one unit of the consumption good in the next period only. The agent maximizes:

$$E_0 \sum_{t=0}^{\infty} \theta^t \left[\frac{c_t^{1-\omega} - 1}{1-\omega} \right] \tag{4}$$

subject to:

$$c_t = y_t e_{t-1} + p_t^e (e_{t-1} - e_t) + f_{t-1} - p_t^f f_t \tag{5}$$

where E_0 is the mathematical expectation operator conditional on information at time zero, θ is the subjective discount factor, ω is the risk aversion parameter, c_t is consumption, y_t is the

tree's dividend, p_t^e and p_t^f are the prices of the equity and the risk-free asset, and e_t and f_t are the agent's equity and risk-free asset holding at time t, Production evolves according to $y_{t+1} = x_{t+1} y_t$ where x_t, the gross growth rate, follows a two-state ergodic Markov chain with probability $P(x_{t+1} = x_j \mid x_t = x_i) = \phi_{ij}$. Defining the states of the problem as (c, i) where $y_t = c$ and $x_t = \lambda_i$, the period t equilibrium asset prices are

$$p^e(c, i) = \theta \sum_{j=1}^{2} \phi_{i,j} (\lambda_j c)^{-\omega} [p^e(\lambda_j c, j) + \lambda_j c] c^{\omega} \tag{6}$$

$$p^f(c, i) = \theta \sum_{j=1}^{2} \phi_{i,j} \lambda_j^{-\omega} \tag{7}$$

When the current state is (c, i), the expected equity return and the risk-free rate are:

$$R_i^e = \sum_{j=1}^{2} \phi_{i,j} \left(\frac{p^e(\lambda_j c, j) + \lambda_j c}{p^e(c, i)} - 1 \right) \tag{8}$$

$$R_i^f = \frac{1}{p^f(c, i)} - 1 \tag{9}$$

The unconditional (average) expected returns on the two assets are $R^e = \Sigma_{j=1}^2 \pi_i R_i^e$, $R^f = \Sigma_{j=1}^2 \pi_i R_i^f$, and the average equity premium is $EP = R^e - R^f$, where π_i are the Markov chain stationary probabilities, satisfying $\pi = \phi^T \pi$ and $\Sigma_i \pi_i = 1$, where $\phi^T = \phi_{j,i}$.

Mehra and Prescott specified the two states for consumption (output) to be $\lambda_1 = 1 + \mu + \nu$; $\lambda_2 = 1 + \mu - \nu$ and restricted $\phi_{1,1} = \phi_{2,2} = \phi$ and $\phi_{1,2} = \phi_{2,1} = 1 - \phi$. They calibrated the three technology parameters so that the mean, the standard deviation, and the AR(1) coefficient of the model's consumption match those of the growth rate of annual US consumption over the period 1889–1978 and searched for combinations of the preference parameters (θ, ω) in a prespecified interval to obtain values for the risk-free rate and the equity premium. Given that the average, the standard deviations, and the AR(1) coefficient of annual growth rate of US consumption are 1·018, 0·036, and −0·14, the implied values of μ, ν, and ϕ are 0·018, 0·036, and 0·43, respectively. The range for ω was selected to be [0, 10] and this choice was justified citing a number of empirical estimates of this parameter (see Mehra and Prescott, 1985, p. 154). The range for θ was chosen to be [0, 1], but simulations which produced a risk-free rate in excess of 4% were eliminated on the grounds that 4% constitutes an upper bound consistent with historical experience. The puzzle is that the *largest* equity premium generated by the model is 0·35%, which occurred in conjunction with a real risk-free rate of about 4%, while the US economy for the period 1889–1978 experienced an annual average equity premium of 6·18% and an average real risk-free rate of 0·80%.

Two hidden assumptions underlie Mehra and Prescott's procedure. First, they believe that technology parameters can be tightly estimated while the uncertainty surrounding the choice of preference parameters is substantial. Consequently, while the sensitivity of the results is explored to variations in θ and ω within the range, no robustness check is made for perturbations of the technology parameters. Second, by providing only the largest value generated, they believe that it is a sufficient statistic to characterize the puzzle.

Here we repeat their exercise with three tasks in mind: first, to study whether the uncertainty present in the selection of the technology parameters is important in determining the magnitude of the puzzle; second, to formally measure the discrepancy of the model from the data using a

variety of statistics based on the probability distribution of outcomes of the model; third, to evaluate the contribution of two alleged solutions to the equity premium puzzle proposed in the literature.

This first example is particularly simple since we have an exact solution for the endogenous variables of the model. In addition, because the model produces values for the mean of R^f and EP, variations in X_t are entirely determined by variations in β, so that $\mathcal{H}(X_t, \beta \mid \mathcal{I}, f)$ is proportional to $\pi(\beta \mid \mathcal{I}, f)$. Therefore, once we have selected $\pi(\beta \mid \mathcal{I}, f)$, we can immediately determine the distribution of simulated means of the R^f–EP pair.

To select the density for the five parameters of the model we proceed in two steps. First, we choose a maximum range for the support of β on the basis of theoretical considerations.

Table I. Equity premium puzzle

(A) Parameter values

Basic case

θ	Truncated normal, range [0·9523, 1·000], mode at 0·9708									
ω	$\chi^2(2)$ range [0, 10]									
μ	−0·0025	0·0044	0·0148	0·0183	0·0219	0·0230	0·0237	0·0241	0·0255	0·0300
ν Uniform	0·0528	0·0307	0·0100	0·0357	0·0252	0·0490	0·0100	0·0140	0·0531	0·0397
ϕ	−0·0700	−0·0500	−0·0100	0·1100	−0·1800	0·0600	−0·0400	0·0400	0·0700	0·1100

Experiment 1

β Truncated normal, range [0·9523, 1·040], mode at 0·9708

Experiment 2

ζ Exponential, range [0·0001, 0·2]

ξ $1 - \frac{\nu}{1+\mu}$ with prob ϕ; $\frac{1-\mu-\nu}{1+\mu}$ with prob $1 - \phi - \zeta$; $1 - \frac{\mu}{1+\mu}$ with prob ζ

(B) Statistics of the simulated distribution

	M-P case		Basic case		Alternative		Experiment 1		Experiment 2	
	EP	R	EP	R	EP	R	EP	R	EP	R
Mean	0·0094	0·0913	0·0035	0·0702	0·0015	0·0670	0·032	−0·061	0·1087	−0·1519
S.D.	0·0067	0·0379	0·0073	0·0470	0·0032	0·0433	0·0066	0·0419	0·1808	0·3327
Skewness	0·35	−0·15	2·66	1·15	3·32	1·18	2·68	1·14	1·58	−1·57
Kurtosis	−1·08	−0·86	6·67	0·42	12·95	0·42	6·83	0·39	1·03	0·86
Maximum	0·022	0·167	0·034	0·215	0·022	0·203	0·031	0·061	0·747	0·224
5%	0·0005	0·028	0·00001	0·023	0·00001	0·023	0·00001	−0·107	0·00002	−0·938
Median	0·0084	0·094	0·0002	0·051	0·0001	0·0499	0·0001	−0·008	0·0098	0·022
95%	0·021	0·150	0·022	0·170	0·0079	0·159	0·019	0·020	0·511	0·068
Mode	0·0094	0·110	0·0008	0·052	0·0001	0·0478	0·0001	−0·074	0·007	−0·018
Pr 1	0·736		0·817		0·803		0·855		0·727	
Pr 2	0·99		0·99		0·99		0·95		0·62	
Pr 3 Q1	0·994		0·994		1·000		0·927		0·577	
Pr 3 Q2	0·000		0·000		0·000		0·000		0·295	
Pr 3 Q3	0·006		0·006		0·000		0·073		0·126	
Pr 3 Q4	0·000		0·000		0·000		0·000		0·002	

Note:
Pr 1 refers to the frequency of simulations for which the pair (R^f, EP) is in a classical 95% region around the actual values. Pr 2 reports the percentile of the simulated distribution where the actual (R^f, EP) pair lies. Pr 3 reports the probability that the model generates values in each of the four quadrants delimited by the actual pair of (R^f, EP). Q1 is the region where $R^{fs} > R^f$ and $EP^s < EP$, Q2 is the region where $R^{fs} > R^f$ and $EP^s \geqslant EP$, Q3 is the region where $R^{fs} \leqslant R^f$ and $EP^s < EP$ and Q4 is the region where $R^{fs} \leqslant R^f$ and $EP^s \geqslant EP$.

Second, we specify the joint density to be the product of five univariate densities and select each univariate density to be a smoothed version of the frequency distribution of estimates existing in the literature. The densities and their support are in panel A of Table I. Therange for ω is the same as that of Mehra and Prescott and the chosen χ^2 density has a mode at 2, where most of the estimates of this parameter lie, and a low mass (smaller than 5%) for values exceeding 6. The range for θ reflects the results of several estimation studies which obtained values for the steady-state real interest rate in the range [−0·005, 0·04] (see e.g. Altug, 1989; Dunn and Singleton, 1986; or Hansen and Singleton, 1983) and of simulation exercises which have a steady-state real interest rate in the range [0, 0·05] (see e.g. Kandel and Stambaugh, 1990; or Mehra and Prescott, 1985). The density for θ is skewed to express the idea that a steady-state real interest rate of 2–3% or lower is more likely than a steady-state interest rate in excess of 4%. Note that although we assume that the densities of θ and ω are independent, many estimates of these two parameters are not. However, the rank correlation coefficient for the pairs of estimates is small and none of the results we present depends on this simplifying assumption.

To provide a density for μ, ν and ϕ we experimented with two procedures. The first, which is used in the basic experiment, involves taking the 10 sub-sample estimates of the mean, of the standard deviation, and of the AR(1) coefficient of the growth rate of consumption over 10-year samples contained in Mehra and Prescott (1985, p. 147) as characterizing reasonable consumption processes and then constructing a uniform discrete density over these triplets. The second involves dividing the growth rates of consumption over the 89 years of the sample into two states (above and below the mean), estimating a measure of dispersion for the first two moments and for the AR(1) coefficient of the growth rate of consumption in each state and directly inputting these estimates into the model. In this case simulations are performed by assuming a joint normal density for the mean, the standard deviation, and AR(1) coefficient in each state centred around the point estimate of the parameters and maximum support within two standard deviations of the estimate.

Figures 1–4 present scatterplots of the simulated pairs (R^f, EP) when 10,000 simulations are performed. We summarize the features of the joint distribution in panel B of Table I using a

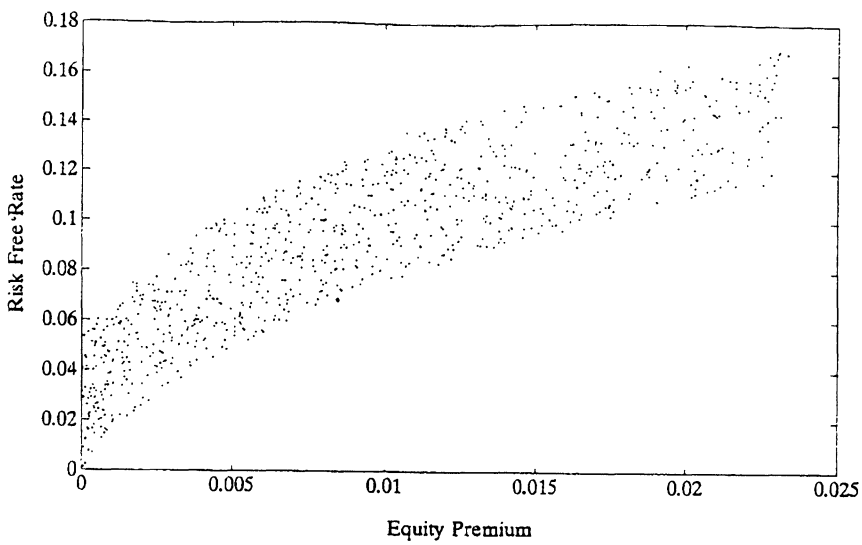

Figure 1. Scatterplot risk-free rate-equity premium: Mehra–Prescott case

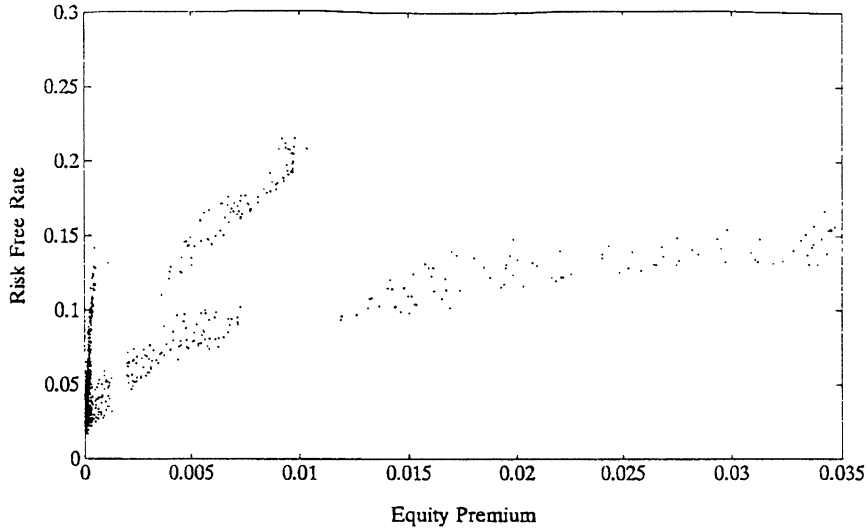

Figure 2. Scatterplot risk-free rate-equity premium: basic case

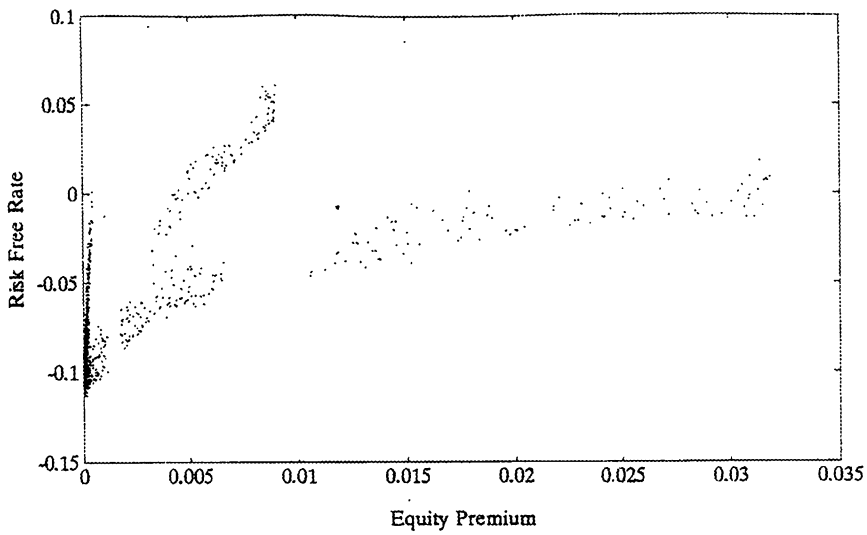

Figure 3. Scatterplot risk-free rate-equity premium: beta > 1 case

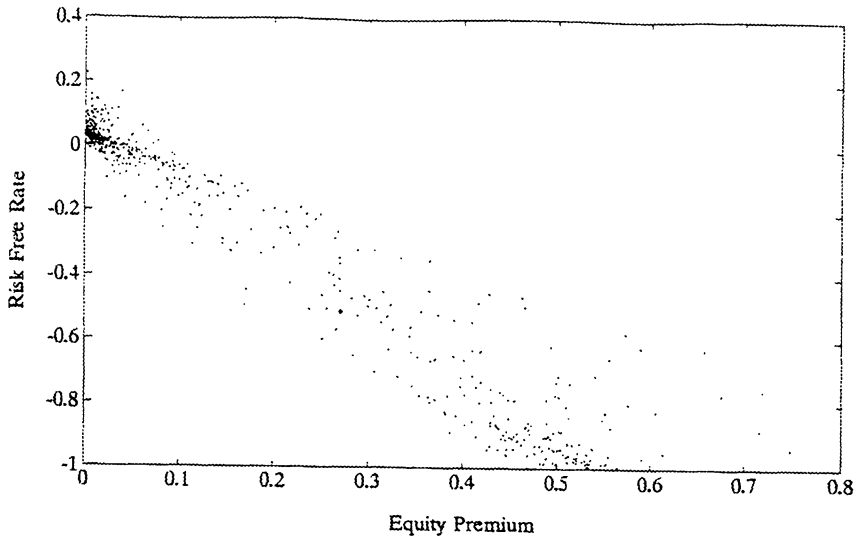

Figure 4. Scatterplot risk-free rate-equity premium: Reitz case

number of statistics. To evaluate the discrepancy of the model from the data we report (1) the probability that the model generates values for (R^f, EP) which fall within a two standard deviation band of the actual mean, (2) the percentile contour of the simulated distribution where the actual means of (R^f, EP) lies, and (3) the probability that the simulated pair is in each of the four quadrants of the space delimited by the actual means of (R^f, EP).

Figure 1 reports the scatterplot obtained with the Mehra and Prescott specification (i.e. when technology parameters are fixed and we draw replications from the densities of θ and ω only). It is necessary to check that the maximum value of the equity premium consistent with a risk free-rate not exceeding 4% is only 0·0030, confirming Mehra and Prescott's conclusion. Also for this specification,the distribution of the model's outcomes is uniform and the mode of the joint distribution (the most likely value from the point of view of the model) is at $R^f = 0{\cdot}110$, $\text{EP} = 0{\cdot}0094$. The probabilistic measures of discrepancy suggest that a large portion of the simulations are in the region where the simulated R^f exceeds the mean of R^f and the simulated EP is below the mean of EP we find in the data, that about 73% of the simulations produce pairs within a classical 95% ball around the actual means of (R^f, EP), and that the actual mean pair is outside the 99 percentile contour.

Figure 2 reports the scatterplot obtained with the basic specification of the model. Also in this case, the puzzle, as defined by Mehra and Prescott, is evident: if we set a 4% upper bound to the risk-free rate, the maximum equity premium generated is only 0·0038. However, with this specification, the distribution is bimodal and most of the simulated pairs lie on a ridge parallel to the R^f axis. The probability that the model generates values in a ball centred around the actual means of (R^f, EP) is now 81·4%. However, in 100% of the cases the simulated risk-free rate exceeds the actual mean and the simulated equity premium is below the actual mean and the actual pair still lies outside the 99 percentile contour of simulated distribution.

To examine whether the selection of the density for the technology parameters has effects on the results, we also conducted simulations using the alternative distribution for these parameters. No substantial changes emerge. For example, the probability that the model generates pairs in a

ball centred around the actual means of (R^f, EP) is 80·3% and the maximum value for EP compatible with a R^f not exceeding 4% is 0·0025.

Several conclusions can be drawn from this first set of exercises. First, even after taking into account the uncertainty surrounding estimates of the technology parameters, the puzzle remains regardless of the way it is defined (maximum values, modes, or contour probabilities): the model cannot generate (R^f, EP) pairs which match what we see in the data. Second, once the uncertainty surrounding estimates of the technology parameters is taken into account, the simulated distributions are bimodal, highly left skewed, and have a fat left tail, indicating that lower than average values are more probable and that very small values have nonnegligible probability. Third, the simulated risk-free rate is always in excess of the actual one, a result that Weil (1990) has termed the risk-free rate puzzle. Fourth, while the model fails to generate values for (R^f, EP) which replicate the historical experience, in more than 80% of the simulations it produces pairs which are within two standard deviations of the actual means.

Next, we conduct two exercises designed to examine the contribution of the modifications suggested by Kocherlakota (1990), Benninga and Protopapadakis (1990), and Rietz (1988) to the solution of the puzzle. The first experiment allows the discount factor θ to take on values greater than 1. The justification is that, in a growing economy, reasonable values for the steady-state real interest rate can be obtained even with θ greater than 1. In this experiment we still maintain the truncated normal density for θ used in the baseline case but increase the upper value for its range to 1·04 and allow about 10% of the density in the region above 1.0.

The second experiment assumes the presence of a third unlikely crash state where consumption falls substantially.[2] The justification for including a third state is that in the Great Depression consumption fell substantially and excluding such a state may have important implications on the results (a conclusion denied by Mehra and Prescott, 1988). With this specification there are two new parameters which cannot be measured from available data: ζ, the probability of a crash state and ξ, the percentage fall in consumption in the crash state. Rietz (1988) searched over the *a priori* ranges of $[0{\cdot}0001, 0{\cdot}2]$ and $[\mu/(1+\mu), 1-\nu/(1+\mu)]$ and examined the magnitude of the maximum simulated equity premium that the model consistent with a simulated risk-free rate below 4%. We maintain these ranges in our experiment and assume on these supports an exponential density for ζ and a three-point discrete density for ξ summarizing the three cases examined by Rietz.

Allowing the discount factor to take on values greater than 1 goes a long way towards reducing the discrepancy of the model from the data (see Figure 3) since it shifts the univariate distribution of R^f towards negative values (the minimum and maximum values of R^f are now $(-0{\cdot}084, 0{\cdot}0.092)$. For example, the probability that the model generates pairs in a ball centred around the actual means of (R^f, EP) is now 85·7% and in only 7·4% of the cases is the simulated risk-free rate in excess of the actual means. Because of this shift in the univariate distribution of R^f, the maximum value of EP consistent with a risk-free rate below 4% is now 0·031. Despite these differences, the location and the shape of the univariate distribution of EP are unaffected. Hence, although the equity premium puzzle is 'solved' when defined in terms of the maximum simulated EP consistent with a simulated R^f below 4%, it is still very evident when we look at the distributional properties of the simulated EP.

---

[2]The three consumption states are $\lambda_1 = 1+\mu+\nu$, $\lambda_2 = 1+\mu-\nu$, $\lambda_3 = \xi * (1+\mu)$ and the transition matrix has elements: $\phi_{1,1} = \phi_{2,2} = \phi$; $\phi_{1,2} = \phi_{2,1} = 1-\phi-\zeta$, $\phi_{1,3} = \phi_{2,3} = \zeta$, $\phi_{3,1} = \phi_{3,2} = 0.5$, $\phi_{3,3} = 0.0$. Note that the experiment is conceptually different from the previous ones since there are two extra degrees of freedom (the new parameters ξ and ζ) and no extra moments to be matched.

The second modification is much less successful (see Figure 4). It does shift the univariate distribution of EP to the right (the mode of 0·035) and increases the dispersion of simulated EPs but it achieves this at the cost of shifting the distribution of R^f towards unrealistic negative value (the mean is −0·15 and the 90% range is [−0·940, 0·068]) and of scattering the simulated (R^f, EP) pairs all over the place. For example, the probability that the simulated pair is in a ball centred around the actual means of (R^f, EP) decreases to 72·7% and the probabilities that the model generates values in each of thefour quadrants delimited by the actual means of (R^f, EP) are almost identical. Finally, the maximum EP consistent with a R^f below 4% is 0·747. Therefore, adding a crash state shifts the mode and stretches and tilts the shape of the joint simulated distribution. Roughly speaking, too many (R^f, EP) configurations now have equal probability, and this weakens the ability of the theory to provide a coherent answer to the question posed.

Technology Shocks and Cyclical Fluctuations in GNP

Kydland and Prescott (1982) showed that a one-sector growth model driven by technology shocks calibrated to reproduce the statistical properties of Solow residuals explains about 70% of the variance of per capita US output. This result has spurred much of the subsequent literature which tries to account for business cycle regularities in models where monetary impulses play no role (the so-called real business cycle literature). Kydland and Prescott's initial estimate has been refined by adding and subtracting features to the basic model (see Hansen, 1985) but the message of their experiment remains: a model where technology shocks are the only source of disturbance explains a large portion of the variability of per capita US output.

Recently, Eichenbaum (1991) has questioned this assertion because 'decisions based solely on the point estimate of λ_y are whimsical (where $\lambda_y = \text{var}(y_t^s)/\text{var}(y_t)$ and $\text{var}(y_t^s)$ and $\text{var}(y_t)$ are the variance of the cyclical component of simulated and actual output) and suggests that 'the model and the data, taken together, are almost completely uninformative about the role of technology shocks in generating fluctuations in US output' (pp. 614–615). Using an exactly identified GMM procedure to estimate the free parameters, he finds that the model explains anywhere between 5% and 200% of the variance of per capita US output.

In this section we repeated Eichenbaum's exercise with three goals in mind. First, we are interested in knowing that is the most likely value of λ_y from the point of view of the model (i.e. in locating the mode of the simulated distribution). Second, we want to provide confidence bands for λ_y which reflect the uncertainty faced by a researcher in choosing the parameters of the model (not the uncertainty present in the data, as in Eichenbaum). Third, we wish to verify whether normal confidence bands appropriately describes the uncertainty surrounding point estimates of λ_y and examine which feature of the model make deviations from normality more evident.

The model is the same as Eichenbaum's and is a simple variation of Hansen's (1985) model which allows for deterministic growth via labour-augmenting technological progress. The social planner of this economy maximizes

$$E_0 \sum_{t=0}^{\infty} \theta^t [\log(c_t) + \psi(T - h_t)] \tag{10}$$

subject to:

$$c_t + k_{t+1} - (1 - \delta)k_t \leqslant A_t k_t^{1-\alpha} (\gamma^t h_t)^{\alpha} \tag{11}$$

where c_t is per capital consumption, $T - h_t$ is leisure, and k_t the capital stock. When δ is different from 1, a closed-form stationary solution to the problem does not exist. Here we compute an approximate decision rule for the endogenous variables using a loglinear expansion around the steady state after variables have been linearly detrended as in King *et al.* (1988), but we neglect the approximation error in constructing probability statements on the outcomes of the model (i.e. we use $\mathcal{P}(X_t, \beta \mid \mathcal{I}, \mathcal{F})$ and no weighting).

There are seven parameters in the model, five deep (δ, the depreciation rate of capital; β, the subjective discount rate; ψ, leisure's weight in the utility function; α, labour's share in output; γ, the constant unconditional growth rate of technology) and two which appear only because of the auxiliary assumptions we made on the stochastic process for technology shocks (ρ, the AR parameter and σ the standard deviation of the shock). Hansen (1985) calibrated these seven parameters (the values are in the first column of panel A of Table II) and found that $\lambda_y \approx 1$. Eichenbaum (1991) estimated all parameters except β (which is calibrated) using a method of moments estimator (estimates and standard deviations are in the second column of panel A of Table II) and found (1) a point estimate of λ_y of 0·80, (2) a large standard deviation about the point estimate of λ_y due primarily to the uncertainty surrounding estimates of ρ and σ, and (3) a strong sensitivity of the point estimate of λ_y to small perturbations in the parameter vector used.

Table II. Technology shocks and cyclical fluctuations in GNP

(A) Parameter values

	Hansen (1985)	Eichenbaum (1991)	Canova (1994)
θ	0·99	0·9926	Truncated normal, range [0·9855; 1·002], mode 0·9926
ϕ	2·60	3·6779 (0·0003)	Endogenous
α	0·64	0·6553 (0·0570)	Uniform [0·50; 0·75]
γ	1·00	1·0041 (0·0003)	Normal (1·0002, 0·001)
δ	0·25	0·0209 (0·0003)	Uniform [0·02; 0·03]
ρ	0·95	0·9772 (0·0289)	Normal (0·95, 0·01)
σ	0·00712	0·0072 (0·0012)	Truncated χ^2, range [0; 0·0091], mean 0·0073

(B) Statistics of the simulated distribution

Mean	0·8775
S.D.	0·7635
Skewness	1·9802
Kurtosis	4·4083
Minimum	0·1566
Maximum	7·2355
5%	0·2261
Median	0·5949
95%	2·6018
Mode	0·9046
Pr 1	0·427
Pr 2	0·673

Note:
Estimated standard errors are in parentheses. Pr 1 refers to the frequency of simulations for which the variance of simulated output is in a classical 95% region around the actual value of the variance of detrended output. Pr 2 reports the percentile of the simulated distribution where the point estimate of the actual variance of output lies.

In the exercise we conduct, we assume that $\pi(\beta \mid \mathcal{I}, f)$ is the product of seven univariate densities. Their specification appear in the third column of panel A of Table II. The range for the quarterly discount factor corresponds to the one implied by the annual range used in the previous example and the density is the same. δ is chosen so that the annual depreciation rate of the capital stock is uniformly distributed between 8% and 12% per year. The range is selected because in simulation studies δ is commonly set to 0·025, which corresponds to a 10% annual depreciation rate, while estimates of this parameter lie around this value (e.g. McGratten *et al.*, 1991, have a quarterly value of 0·0310 and a standard deviation of 0·0046, while Burnside *et al.*, 1993, have a quarterly value of 0·0209 and a standard deviation of 0·0003). The range for α reflects calculations appearing in Christiano (1988) where, depending on how proprietors' income is treated, the share of total output paid to capital varies between 0·25 and 0·43, and the estimate obtained, among others, in McGratten *et al.* (1991). We chose the densities for ρ and σ as in Eichenbaum because the econometric evidence on these two parameters is scant andthe values used in most simulation studies fall within a one standard deviation band around the mean of the assumed density (see e.g. Kydland and Prescott, 1982; Hansen, 1985). Finally, T is fixed at 1369 hours per quarter, the density for γ matches the quarterly distribution of unconditional quarterly growth rates of US output for the period 1950–1990, and ψ is endogenously chosen so that the representative household spends between one sixth and one third of its time working in the steady state.

We performed 1000 simulations with time series of length $T = 124$ and filtered both simulated and actual GNP data with the Hodrick and Prescott filter.[3] The results appear in panel B of Table II and in Figure 5, where we present a smoothed version of the simulated distribution of λ_y. The distribution is scaled so that with the point estimates of the parameters used by Eichenbaum $\lambda_y = 0{\cdot}80$. The implied value of λ_y using Hansen's parameters is 0·84.

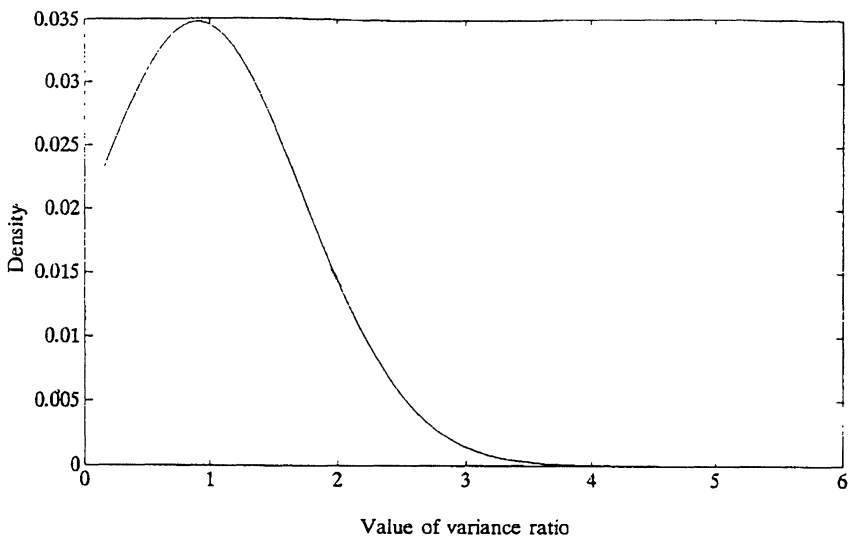

Figure 5. Density of variance ratio: HP filtered data

[3]We use the Hodrick and Prescott filter to maintain comparability with previous work. The results obtained when the data are linearly detrended or first-order differenced are not substantially different.

The mode of the distribution of λ_y is at 0·9046, the mean at 0·8775, and the median at 0·5949. The dispersion around these measures of location is very large. For example, the standard deviation is 0·7635 and the 90% range of the distribution is [0·2261, 2·6018]. The simulated distribution is far from normal and its right tail tends to be very long. Hence the range of reasonable values of λ_y is very large, and, as in Eichenbaum, small perturbations in the parameter vector induce large variations in the variance ratio. In addition, normal confidence bands do not appropriately characterize the uncertainty surrounding the outcomes of the model.

Several other features of the simulated distribution are worth mentioning. First, in 67·3% of the cases the variance of simulated output is smaller than the variance of actual output. Second, in 42·7% of the simulations the variance of simulated output is within a 95% confidence interval centred around the estimate of the variance of actual output. Third, if we select $v = 0{\cdot}5$ and look for the $\hat{\lambda}$ satisfying $\Pr(\lambda_{y_i} \leqslant \hat{\lambda}) = 0{\cdot}5$, i.e. $\hat{\lambda}$ is the median of the simulated distribution, we find that the median value of the variance of simulated GNP is outside the 95% normal confidence interval for the variance of actual GNP.

When we ask which parameter is responsible for the wide dispersion in the estimates of λ_y, we find that it is the location and width of the support of ρ which induce this feature in the distribution of λ_y. For example, assuming that the density of ρ has a point mass at 0·94 and maintaining the same densities for the other parameters, we find that location measures of the simulated distribution of λ_y decrease (the mode is now at 0·792) and the standard deviation drops to 0·529. Similar conclusions are obtained by shifting the range of ρ towards 0·90 or by cutting the range of possible ρ in half without changing the mean value. Hence, as in Eichenbaum, we find that it is the uncertainty present in the choice of the parameters of the exogenous processes rather than the uncertainty present in the selection of the deep parameters of the model that is responsible for the large spread in the distribution of λ_y.

6. CONCLUSIONS

This paper describes a Monte Carlo procedure to evaluate the properties of calibrated general equilibrium models. The procedure formalizes the choice of the parameters and the evaluation of the properties of the model while maintaining the basic approach used in calibration exercises. It also realistically accounts for the uncertainty faced by a simulator in choosing the parameters of the model. The methodology allows for global sensitivity analysis for parameters chosen within the range of existing estimates and evaluates the discrepancy of the model from the data by attaching probabilities to events a simulator is interested in characterizing. The approach is easy to implement and includes calibration and simulation exercises conducted after the parameters are estimated by simulation and GMM techniques as special cases. We illustrate the usefulness of the approach as a tool to evaluate the performance of theoretical models with two examples which have received much attention in the recent macroeconomic literature: the equity premium puzzle and the ability of a real business cycle model to reproduce the variance of actual US output. Finally, it is worth noting that for problems of moderate size, the computational complexity of the procedure is limited. For both examples presented the entire Monte Carlo routine required about a minute on a 486-33 MHz machine using RATS386 programs.

ACKNOWLEDGEMENTS

Part of this research was undertaken while the author was also associated with the European University Institute, Florence. The author has benefited from the comments and the suggestions

of a large number of colleagues, including two anonymous referees, David Backus, Larry Christiano, Frank Diebold, Javier Diaz, Mahmoud El-Gamal, John Geweke, Eric Ghysels, Bruce E. Hansen, Gary Hansen, Jane Marrinan, Yaw Nyarko, Adrian Pagan, Franco Peracchi, Victor Rios, Gregor Smith, Herman Van Dijk, and Randall Wright. He would also like to thank the participants of seminars at Brown, European University Institute, NYU, Montreal, Rochester, Penn, University of Rome, Carlos III Madrid, Free University Bruxelles, CERGE, University of Minnesota, University of Maryland, Summer Meetings of the North American Econometric Society and the Conference on 'Econometric Inference Using Simulation Techniques' held in Rotterdam on 5–6 June 1992 for useful discussions.

REFERENCES

Altug, S. (1989), 'Time to build and aggregate fluctuations: some new evidence', *International Economic Review*, **30**, 889–920.

Backus, D., A. Gregory and A. Zin (1989), 'Risk premiums in the terms structure: evidence from artificial economies', *Journal of Monetary Economics*, **24**, 371–399.

Baxter, M. (1991), 'Approximating suboptimal dynamic equilibria: an Euler equation approach', *Journal of Monetary Economics*, **27**, 173–200.

Benninga, S. and A. Protopapadakis (1990), 'Leverage, time preference, and the equity premium puzzle', *Journal of Monetary Economics*, **25**, 49–58.

Box, G. (1980), 'Sampling and Bayes inference in scientific modelling and robustness', *Journal of the Royal Statistical Society*, Ser. A, **143**, 383–430.

Burnside, C., M. Eichenbaum and S. Rebelo (1993), 'Labor hoarding and the business cycle', *Journal of Political Economy*, **101**, 245–273.

Canova, F. and J. Marrinan (1993), 'Profits, risk and uncertainty in foreign exchange markets', *Journal of Monetary Economics*, **32**, 259–286.

Canova, F., M. Finn and A. Pagan (1993), 'Evaluating a real business cycle model', forthcoming, in C. Hargreaves (ed.), *Nonstationary Time Series Analyses and Cointegration*, Oxford: Oxford University Press.

Christiano, L. (1988), 'Why does inventory investment fluctuate so much?' *Journal of Monetary Economics*, **21**, 247–280.

Christiano, L. (1990), 'Solving the stochastic growth model by linear quadratic approximation and by value function iteration', *Journal of Business and Economic Statistics*, **8**, 23–26.

Coleman, W. (1989), 'An algorithm to solve dynamic models', Board of Governors of the Federal Reserve System, International Finance Division, Discussion Paper No. 351.

Dotsey, M. and C.S. Mao (1991), 'How well do linear approximation methods work? Results for suboptimal dynamic equilibria', *Journal of Monetary Economics*, **29**, 25–58.

Dunn, D. and K. Singleton (1986), 'Modelling the term structure of interest rates under non-separable utility and durability of goods', *Journal of Financial Economics*, **17**, 27–55.

Eichenbaum, M. (1991), 'Real business cycle theory: wisdom or whimsy?' *Journal of Economic Dynamic and Control*, **15**, 607–621.

El-Gamal, M. (1993a), 'The extraction of information from multiple point estimates', forthcoming in *Journal of Nonparametric Statistics*.

El-Gamal, M. (1993b), 'A Bayesian interpretation of extremum estimators', manuscript, California Institute of Technology.

Fair, R. (1991), 'Estimating event probabilities from macroeconomic models using stochastic simulations', Yale University, manuscript.

Friedman, M. (1959), *Essays in Positive Economics*, New York: Aldine Press.

Frisch, R. (1933), 'Propagation problems and impulse problems in dynamics economies', in *Economic Essays in Honor of Gustav Cassel*, London.

Geweke, J. (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica*, **57**, 1317–1339.

Gelfand, A. and A. Smith (1990), 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409.

Genest, C. and M. Zidak (1986), 'Combining probability distributions: a critique and an annotated bibliography', *Statistical Science*, **1**, 114–148.
Gregory, A. and G. Smith (1989), 'Calibration as estimation', *Econometric Reviews*, **9**(1), 57–89.
Gregory, A. and G. Smith (1991), 'Calibration as testing: inference in simulated macro models', *Journal of Business and Economic Statistics*, **9**(3), 293–303.
Haavelmo, G. (1944), 'The probability approach in econometrics', *Econometrica*, **12**, Supplement.
Hansen, G. (1985), 'Indivisible labor and the business cycle', *Journal of Monetary Economics*, **16**, 309–328.
Hansen, L. and T. Sargent (1979), 'Formulating and estimating dynamic linear rational expectations models', *Journal of Economic Dynamic and Control*, **2**, 7–46.
Hansen, L. and K. Singleton (1983), 'Stochastic consumption, risk aversion and temporal behavior of asset returns', *Journal of Political Economy*, **91**, 249–265.
Harrison, G. and H. D. Vinod (1989), 'Sensitivity analysis of applied general equilibrium models: completely randomized factorial sampling designs', University of New Mexico, manuscript.
Heaton, J. (1993), 'The interaction between time nonseparable preferences and time aggregation', *Econometrica*, **61**, 353–381.
Journal of Business and Economic Statistics, January 1990.
Judd, K. (1992), 'Projection methods for solving aggregate growth models', *Journal of Economic Theory*, **58**, 410–452.
Kandel, S. and R. Stambaugh (1990), 'Expectations and volatility of consumption and asset returns', *Review of Financial Studies*, **3**, 207–232.
King, R., C. Plosser and S. Rebelo (1988), 'Production, growth and business cycles: I and II', *Journal of Monetary Economics*, **21**, 195–232 and 309–342.
Kocherlakota, N. (1990), 'On the discount factor in growth economies', *Journal of Monetary Economics*, **25**, 45–48.
Kydland, F. (1992), 'On the econometrics of world business cycles', *European Economic Review*, **36**, 476–482.
Kydland, F. and E. Prescott (1982), 'Time to build and aggregate fluctuations', *Econometrica*, **50**, 1345–1370.
Kydland, F. and E. Prescott (1991), 'The econometrics of the general equilibrium approach to business cycles', *The Scandinavian Journal of Economics*, **93**(2), 161–178.
Kwan, Y. K. (1990), 'Bayesian calibration with an application to a non-linear rational expectation two country model', mimeo, University of Chicago Business School.
Lee, B.S. and B. Ingram (1991), 'Simulation estimators of time series models', *Journal of Econometrics*, **47**(2/3), 197–206.
Marcet, A. (1992), 'Solving nonlinear stochastic models by parametrizing expectations: an application to asset pricing with production', Universitat Pompeu Fabra, Working Paper 5.
Mehra, R. and E. Prescott (1985), 'The equity premium: a puzzle', *Journal of Monetary Economics*, **15**, 145–162.
Mehra, R. and E. Prescott (1988), 'The equity risk premium: a solution?' *Journal of Monetary Economics*, **22**, 133–136.
McGratten, E., R. Rogerson and R. Wright (1991), 'Estimating the stochastic growth model with household production', Federal Reserve Bank of Minneapolis, manuscript.
Niederreiter, H. (1988), 'Quasi Monte Carlo methods for multidimensional numerical integration', *International Series of Numerical Mathematics*, **85**, 157–171.
Novales, A. (1990), 'Solving nonlinear rational expectations models: a stochastic equilibrium model of interest rates', *Econometrica*, **58**, 93–111.
Pagan, A. and A. Ullah (1991), 'Nonparametric estimation', University of Rochester, manuscript.
Phillips, P.C.B. (1991), 'To criticize the critics: an objective Bayesian analysis of stochastic trends', *Journal of Applied Econometrics*, **6**(4), 333–354.
Prescott, E. (1991), 'Real business cycle theories: what have we learned?' Federal Research of Minneapolis, Working Paper 486.
Reitz, T. (1988), 'The equity risk premium: a solution', *Journal of Monetary Economics*, **22**, 117–132.
Shoven, J. and J. Whalley (1984), 'Applied general equilibrium models of taxation and international trade: an introduction and survey', *Journal of Economic Literature*, **22**, 1007–1051.
Sims, C. (1984), 'Solving nonlinear stochastic equilibrium models backward', University of Minnesota, manuscript.

Smith, T. (1993), 'Estimating nonlinear time series models using simulated VAR', *Journal of Applied Econometrics*, **8**, s63–s84.
Tanner, M. and W. Wong (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association*, **87**, 528–550.
Tauchen, G. and R. Hussey (1991), 'Quadrature based methods for obtaining approximate solutions to integral equations of nonlinear asset pricing models', *Econometrica*, 59, 371–397.
Watson, M. (1993), 'Measures of fit for calibrated models', *Journal of Political Economy*, **101**, 1011–1041.
Weil, P. (1990), 'The equity premium puzzle and the risk free puzzle', *Journal of Monetary Economics*, **24**, 401–421.
Wolf, F. (1986), *Meta-Analysis: Quantitative Methods for Research Synthesis*, Beverly Hills, CA: Sage.