INTERNATIONAL ECONOMIC REVIEW
Vol. 36, No. 2, May 1995

SENSITIVITY ANALYSIS AND MODEL EVALUATION IN SIMULATED DYNAMIC GENERAL EQUILIBRIUM ECONOMIES*

BY FABIO CANOVA[1]

This paper describes a Monte Carlo procedure to evaluate dynamic nonlinear general equilibrium macro models. The procedure makes the choice of parameters and the evaluation of the model less subjective than standard calibration techniques, it provides more general restrictions than estimation by simulation approaches and provides a way to conduct global sensitivity analysis for reasonable perturbations of the parameters. As an illustration the technique is applied to three examples involving different models and statistics.

1. INTRODUCTION

A growing body of research in the applied macroeconomic literature uses simulation techniques to derive the time series properties of nonlinear stochastic general equilibrium models, to compare them to real world data and to evaluate policy options (see e.g. King, Plosser, and Rebelo 1988, or Cooley and Hansen 1990). In implementing numerical analyses of general equilibrium models, one has to overcome four hurdles. First, an economy must be specified and functional forms for its primitives selected. Second, a decision rule for the endogenous variables in terms of the exogenous (and predetermined) variables and of the parameters must be computed. Third, given the probability structure of the economy, values for the parameters must be chosen. Fourth, the closeness of functions of simulated and the actual data must be assessed in a metric which is relevant to the problem and policy conclusions, if any, should be drawn.

While models are often specified with an eye to analytical tractability and there has been progress in developing techniques to numerically approximate unknown decision rules for the endogenous variables (see e.g. Sims 1984, Coleman 1989, Novales 1990, Baxter 1991, Tauchen and Hussey 1991, Judd 1992, Marcet 1992 and the January 1990 issue of the *Journal of Business and Economic Statistics*), surprisingly little attention has been paid to the problems connected with the other two steps of the simulations. In particular, the selection of the parameters and the evaluation of the simulation results have been undertaken using procedures which

* Manuscript received October 1991; revised November 1994.

[1] I would like to thank David Backus, Javier Diaz, Frank Diebold, John Geweke, Eric Ghysels, Gary Hansen, Bruce E. Hansen, Jane Marrinan, Yaw Nyarko, Adrian Pagan, Franco Peracchi, Victor Rios-Rull, Gregor Smith, Herman van Dijk, Randall Wright, two anonymous referees, and the participants of seminars at Brown University, European University Institute, New York University, University of Montreal, University of Rochester, University of Pennsylvania, University of Rome and Carlos III Madrid for comments and suggestions.

lack statistical foundations (exceptions include Smith 1993, Burnside, Eichenbaum, and Rebelo 1993).

Starting with Kydland and Prescott (1982) it has been common to use a "calibration" methodology which typically consists of three steps: (i) select the parameters of the model using an array of criteria which range from matching long run averages, consistency with estimates obtained in the micro literature and a priori settings, (ii) represent the properties of actual data with simple statistics (the so-called "stylized facts") which are relatively insensitive to approximation and measurement errors, (iii) evaluate the quantitative properties of the model informally using a metric which is specific to the researcher and the question being asked (see Gregory and Smith 1993 and Kim and Pagan 1993 for detailed surveys of the methodology). Calibration procedures have been justified in different ways. Shoven and Whalley (1984) argue that calibration is a tractable procedure to convert general equilibrium structures from abstract representations into realistic models of actual economies. Jorgenson (1984) indicates that it is the only feasible alternative when it is impossible to simultaneously estimate the parameters without requiring an unrealistically large number of observations or overly severe identifying restrictions. Kydland and Prescott (1982) suggest that it is a reasonable way to assess the quality of a model when measurement errors are present and when its abstract nature is likely to result in a sure rejection when formally tested. Finally, Kydland and Prescott (1991) suggest that calibration is the natural heir of the original quantitative approach advocated by Frisch (1933).

Although popular among theorists, calibration procedures have always made econometricians uneasy. With the advent of modern computer technology, tractability and size limitations are no longer a stumbling block to the use of estimation methods. But apart from the issue of estimation (thoroughly discussed in Gregory and Smith 1989), there are other compelling reasons for considering the conclusions obtained with calibration procedures debatable.

The parameters used in simulations are typically chosen either to reproduce the long run properties of a particular data set or from existing econometric evidence. The former procedure is problematic since information used in different studies may be inconsistent (e.g. a parameter chosen to match average labor payments from firms in national account data may not equal the value chosen to match the average labor income received by households). The latter is dubious because existing evidence is contradictory and because the range of estimates for certain parameters (e.g. risk aversion) is so large that selection biases may be important. In addition, the micro studies that are cited to support particular parameter choices, may have obtained estimates using model specifications which are inconsistent with those imposed in the model under consideration (such as completeness versus incompleteness of markets or general versus partial equilibrium).

Because no uncertainty is typically allowed in the selection of the parameters and the number of replications typically performed is small, the results of the simulations can not be corroborated with formal statements on the range of possible outcomes of the model. Therefore, one must resort to informal techniques to judge the closeness of simulated and actual data and to evaluate policy alternatives. Moreover, although certain parameters are crucial in determining the conclusions

of the study, results are often reported without any sensitivity analysis on how summary statistics change with reasonable perturbations of the parameters of interest. While these problems are well known in the static computable general equilibrium (CGE) literature (see e.g. Pagan and Shannon 1985) and partially recognized by Kydland and Prescott (1982), they were neglected by most of the subsequent literature.

The purpose of this paper is methodological. I propose a simulation methodology which formalizes the evaluation of properties of the model and allows for meaningful sensitivity analysis on the outcomes of the simulations. The methodology shares features with those recently proposed by Gregory and Smith (1991) and Kwan (1991b), can be justified using simple Bayesian tools (see Box 1980 and El-Gamal 1993) and has similarities with stochastic simulation techniques employed in dynamic nonlinear large scale macro models (see e.g. Fair 1991). Sims (1989), Smith (1992), Watson (1993) and Canova, Finn, and Pagan (1993) have suggested alternative procedures to formally measure the fit of calibrated models.

The idea of the procedure is simple. The model is recognized to be a *false* data generating process for the observed time series and the task is to know in which dimensions it is most at odds with actual data. The metric I use to evaluate the fit is probabilistic. I simulate the model repeatedly using a Monte Carlo procedure which randomizes over *both* the exogenous stochastic processes *and* the parameters. Parameters are drawn from a density consistent with the frequency distribution of estimates existing in the literature. I then construct the frequency distribution for the realizations of the statistics of interest and examine either in what percentile of the distribution of the simulated statistics the actual value lies or the percentage of simulated values which lie in a ball around the actual realization or both. Extreme values for the first percentile (say, below α percent or above $(1 - \alpha)$ percent) or a low value for the second percentile indicates that the model is particularly poor in the dimensions examined.

The approach I propose has several appealing features. First, it accounts for the uncertainty faced by a simulator in choosing the parameters of the model with a "realistic" Bayesian prior. This prior can be formally obtained using information theoretical measures and the outcomes of point estimation experiments (see El-Gamal 1993). Second, it has a built-in feature which allows for global sensitivity analysis on the support of the parameter space and generalizes techniques on randomized design for strata existing in the static CGE literature (see e.g. Harrison and Vinod 1989). Third, it provides a general evaluation criteria which attaches probabilities to events we are interested in characterizing (as in Box 1980). Finally, it provides a simple and convenient metric to judge the closeness of the simulated and the actual data.

The paper is divided in 6 sections. The next section reviews model building procedures and the criteria employed to examine the empirical relevance of dynamic economic models. Section 3 introduces the technique and describes the details involved in the implementation of the procedure. Section 4 spells out the relationship between the approach and existing techniques. Section 5 presents some examples. Section 6 concludes.

2. ON EVALUATING THE EMPIRICAL RELEVANCE OF ECONOMIC MODELS

The formulation, estimation and evaluation of dynamic general equilibrium macro models is a relatively recent undertaking. Hansen and Sargent (1979) pose the foundations for a maximum likelihood estimation of the "deep" (preference, technological) parameters of these models and for testing their validity. Hansen and Sargent face two basic problems. First, since closed form solutions for the endogenous variables in terms of exogenous variables and parameters do not always exist, they concentrate on parametric structures which deliver closed form solutions (linear-quadratic specifications for the primitives of the model and linear processes for the exogenous variables). Second, since many economic models do not provide a "realistic" statistical specification for the endogenous variables and will be discarded as empirically irrelevant in formal testing, Hansen and Sargent augment their models with additional random components (measurement errors, error in variables or unobserved components). Once a closed (final) form solution is obtained and there are enough sources of randomness in the economy to make the model "complete" in a probabilistic sense (see Haavelmo 1944), one proceeds to identify and estimate the parameters. The empirical relevance of the model is then judged by performing statistical goodness of fit tests or likelihood ratio tests for hypotheses concerning the parameters of interest.

Given the intrinsic limitations existing in the choice of linear quadratic specifications, Hansen (1982) proposed to estimate and test hypotheses concerning "deep" parameters directly from the Euler equations using simple moment conditions. Hansen's GMM-IV approach does not require a closed (or a final) form solution for the endogenous variables, is robust to any failure of the econometrician to have the same information set as agents (see Pagan and Ullah 1988) but still requires a fully specified probability structure for the model. The validity of the model is examined using standard goodness of fit tests (the J-tests).

Contemporaneously with the work of Hansen, Kydland, and Prescott (1982) suggested an alternative procedure to tackle the problem of the probabilistic underspecification of the model. Rather than augmenting an artificial economy with extraneous random components to obtain a richer statistical structure, they start from the observation that the model, as a data generating mechanism (DGP), is false. That is, it is known that, as the sample size grows, the data generated by the model will be at greater and greater variance with the observed time series. For Kydland and Prescott an economic model is neither an accurate nor a realistic description of the actual data but only an approximation to the stochastic process generating it. The task of an applied researcher is to indicate in what dimensions the approximation is poor and suggest ways to modify the artificial economy to obtain a better fit.

There are several logical consequences of this point of view. First, because the model is a false DGP for the actual data, classical estimation of the parameters is meaningless. In addition, classical hypothesis testing is inappropriate because a false model can not be regarded as a null hypothesis to be statistically examined (it can be rejected even before the test is undertaken). Similarly, standard Bayesian analysis is inapplicable because the (simulated) likelihood need not be the correct

one, so that posterior statements for the parameters are worthless. In response to these deficiencies, researchers working in this area have adopted a two step approach which chooses the parameters so that the model replicates the data in some basic dimension of interest and evaluates the model on its ability to reproduce "stylized facts." For example, in an aggregate model of the business cycle, parameters are chosen so that the behavior of the endogenous variables in the steady state coincides with the long run behavior of the corresponding variables in the actual economy, and the model is evaluated on its ability to replicate variances and covariances of the cyclical component of macro variables. This first step of the approach is typically justified as a way to unify observations from different fields of economics but it should be noted that it is entirely analogous to the procedure employed in experimental sciences where the physical instrument used to measure some phenomenon is "calibrated" so as to reproduce some known result. For example, to measure the temperature of water a thermometer is calibrated so that in freezing water it gives a value of 32F and in boiling water it gives a value of 212F.

However, because the economy is not fully specified in a probabilistic sense and no measure of dispersion is attached to "calibrated" parameters, the metric employed to determine the quality of the approximation is left unspecified, inferential procedures are subjective to the researcher and, in general, lack statistical foundations. Continuing with the analogy with experimental sciences, if a measurement of 65F is reported it is hard to say if it is different from any value observed in real life or in any other experimental situation.

To overcome these problems Watson (1993) develops a classical procedure which makes evaluation less subjective in situations where the model is known to be a "false" description of the actual economy and the parameters are calibrated. The metric Watson uses is the relative contribution of the second order properties of the model to the second order properties of the actual data. A model fits the data well if the correlation between summary statistics of simulated and actual data in a particular range of frequencies is large (in a R^2 sense).

Sims (1989) and Smith (1992) have suggested a VAR metric to judge the fit of the model. Their approach applies to both situations where the parameters are calibrated or estimated. A VAR is a window which may only partially capture aspects of the data. A model is regarded as appropriate if the "distance" between the unrestricted VAR representations of simulated and actual data is small either in absolute terms or relative to the distance of other models to the actual data.

Finally, Canova, Finn, and Pagan (1993) use the restrictions implied by a calibrated model on the long and short run dynamics of the actual data to provide several general goodness of fit tests and an encompassing procedure to discriminate among models which pass the first round of goodness of fit tests. The procedure has some analogy to the one of Hansen and Sargent (1979) since it employs the restricted VAR representation implied by the model to examine exclusion restrictions for the actual data, and has the advantage of providing the information necessary to modify a model in response to its failure to pass the tests.

In developing an alternative framework of inference I follow Kydland and Prescott's philosophy very closely. I take the actual data to be the realization of an unknown underlying vector stochastic process. The task here is to reproduce

features of the data with an "artificial economy," which is known to be almost surely a false generating mechanism for the actual data. The features of the actual data we may be interested in include conditional and unconditional moments (or the entire densities), the autocovariance function of the data and various functions of these quantities (e.g. measures of persistence or of relative volatility) and specific events (e.g. a recession or an average upward sloping term structure of interest rates). I recognize that "calibrating" the model to the actual data involves sampling error and, more importantly, that some judgmental decisions need to be made which lead to a whole range of calibrated values indexed by data sets, measurement techniques, model specifications and evaluation procedures. The presence of this cross sectional variability is the crucial ingredient to construct numerical measures of discrepancy between simulated and actual data. The inferential procedure adopted here follows Friedman (1959) and judges the validity of a model on its ability to reproduce, in a probabilistic sense, a selected number of facts of the actual economy. If the model is regarded to be a good approximation to the actual data generating process, it can be fruitfully used to evaluate policy options.

3. MODEL EVALUATION AND SENSITIVITY ANALYSIS

I assume that a simulator 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. The analysis of policy options will be discussed later on in this section. $\bar{X}_t$ is assumed to be a Markov process with absolutely continuous but unknown distribution and moments up to the nth. For the sake of presentation, I assume that the unconditional distribution of $\bar{X}_t$ is independent of t but there is nothing in the framework that prevents shifts in the unconditional distribution of $\bar{X}_t$ at known points. $\bar{X}_t$ may include variables like GNP, consumption, interest rates, exchange rates, etc. I 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 technology and preferences. I express the model's functional relation as $X_t = f(Z_t, \beta)$ where f is, in general, an unknown function. 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), f can be computed analytically either by value function iteration (see e.g. Bertsekas 1976) or by solving the Euler equations of the model subject to the transversality condition (see e.g. Hansen and Sargent 1979). Under general specifications, however, f can not 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 in a given metric.[2]

[2] Kydland and Prescott (1982), King, Plosser, and Rebelo (1988) for example, locally approximate the function f by linear or log-linear expansions of f around the steady state of the model. Sims (1984) and Novales (1990) employ a backward solution to recover the function f. Their idea is that although f is

Here I assume that either f is available analytically or that one of the existing numerical procedures has been employed so that a simulator has a functional $\mathcal{F}$ which approximates f in some sense, i.e. $\|\mathcal{F}(Z_t, \gamma) - f(Z_t, \beta)\| < \varepsilon$, where γ are functions of the parameters β and $\|\cdot\|$ 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, one can infer the probability distribution of X_t from the model.

Let $\mathcal{G}(X_t|\beta, f)$ be the density of the X_t vector, conditional on the parameters β and the model f. $\mathcal{G}(X_t|\beta, f)$ represents the probability that a particular path for the endogenous variables will be drawn given a parametric model structure for the artificial economy and a set of parameters and is a deterministic (nonlinear) transformation of $\kappa(Z_t)$, the probability density of the exogenous variables. In other words, X_t is random because Z_t is random. The vector β is, in general, unknown. Let $\pi(\beta|\mathcal{I})$ be the density of the parameters of the model, conditional on the information set $\mathcal{I}$. $\pi(\beta|\mathcal{I})$ represents the information available to a simulator on the parameters of the model. Let $\mathcal{H}(X_t, \beta|f, \mathcal{I})$ be the joint density of simulated data and of parameters and let $p(X_t|f, \mathcal{I}, \mathcal{A}) = \int_{\mathcal{A}} \mathcal{H}(X_t, \beta|f, \mathcal{I})\, d\beta$ be the simulated predictive density of X_t where $\mathcal{A} \subset \mathcal{B}$ is the parameter space.

A generic formulation for the problem we are interested in is to compute functions of simulated data under $p(X_t|f, \mathcal{I}, \mathcal{A})$, i.e. evaluating objects of the form:

$$(1) \qquad E(\mu(X_t)|f, \mathcal{I}, \mathcal{A}, \mathcal{C}) = \int_{\mathcal{C}} \mu(X_t) p(X_t|f, \mathcal{I}, \mathcal{A})\, dX_t$$

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

where $\mu(X_t)$ is the vector of functions of simulated data and $\mathcal{C}$ is the support of the exogenous variables. Let $h(\bar{x}_t)$ be the corresponding vector of functions of the actual data.

The problem of examining the fit of the model can be summarized with the following question: how likely is the model to generate $h(\bar{x}_t)$? To answer note that (1) allows us to compute probabilities of the form $P(\mu(X_t) \in D)$, where D is a bounded set. To do this choose, for example, the mth component of μ to be $\mu_m(X_t) = \chi(X_t : \mu(X_t) \in D)$ where χ is the indicator function, i.e. $\chi(\mu(X_t); D) = 1$ if $\mu(X_t) \in D$ and zero otherwise. From (1) one can also compute a value $\tilde{h}$ satisfying $P[\mu(X_t) \leq \tilde{h}] = \alpha$ for any given α, by appropriately selecting the indicator function.

---

impossible to compute, f^{-1} may be easier to find. In their approach, a process for the endogenous variables is selected and one seeks processes for the exogenous variables which may have generated them under f. Marcet's (1992) method of parametrizing expectations and Judd's (1992) minimum weighted method can be seen as choosing a set of known functions which globally approximate f in a given norm. Baxter's (1991) and Coleman's (1989) methods are grid procedures which obtain the function f by piecewise linear interpolation. Finally, Tauchen and Hussey's (1991) quadrature method is a grid approximation procedure which is appropriate for integral equations which are of Fredholm's second type.

Model evaluation then consists of several types of statements which are complementary and differ only in the criteria used to measure distance. For example, one can compute the probability that the model can generate a $\mu_m(X_t)$ less than or equal to $h_m(\bar{x}_t)$. 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 would indicate a poor "fit" in the dimensions examined. Alternatively, if one can measure the sampling variability of $h_m(\bar{x}_t)$, one can chose an α and compute the implied $\tilde{h}$. Then, by choosing the set D to include the actual realization of $h_m(\bar{x}_t)$ plus one or two standard deviations around the point estimate, one can either see if $\tilde{h}$ lies inside D or calculate the probability that the model generates functions $\mu_m(X_t)$ in the chosen set. If evaluation needs to be done in several dimensions of the $\mu(X_t)$ vector simultaneously, one can partition the simulated distribution of $\mu(X_t)$ into hypercubes and check the likelihood of the event $h(\bar{x}_t)$ from the point of view of the model. For example, if the task is to study the equity premium-risk free rate (EP-R) puzzle (see Merha and Prescott 1985), one could partition the space of simulated EP-R pairs into 4 quadrants with the origin on the actual mean values of the EP-R pair and check the proportion of simulated pairs which falls in each quadrant.

3.1. *Implementation.* There are four technical issues regarding the implementation of the procedure that deserve some discussion. The first concerns the computation of integrals like those appearing in (1). If the (β, Z_t) vector is of high dimension simple discrete grid approximations, spherical or quadrature rules quickly become unfeasible since the number of function evaluations increases exponentially with the dimension of β and Z_t. In addition, unless the contours of $\mathcal{H}(X_t, \beta|\mathcal{I}, f)$ are of ellipsoidal forms, grid evaluations may miss most of the action of this density. There are several feasible alternatives available: one is the Monte Carlo procedure described in, e.g., Geweke (1989), another is the data augmentation procedure of Tanner and Wong (1987) or the "Gibbs sampler" discussed, e.g., in Gelfand and Smith (1990). Finally, one could use one of the quasi-random procedures presented in Niederreiter (1988).

In the examples of Section 5, I adopt a Monte Carlo approach. After drawing with replacement iid β vectors from $\pi(\beta|\mathcal{I})$ and Z_t paths from $\kappa(Z_t)$, I substitute sums over realizations for the integrals in (1) and appeal to the law of large numbers for functions of iid variables to show that

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

where N is the number of replications. Note that, although $\mathcal{H}$ is unknown, sampling from $\mathcal{H}$ (or p) can be conveniently accomplished by simulating the model repeatedly for random (β, Z_t).

Second, since in most cases the function f is unknown, $\mathcal{G}$ itself becomes unknown and the direct computation of integrals like (1) is not feasible. If the approximation $\mathcal{F}$ to f is accurate, one could simply neglect the approximation error and proceed using $\mathcal{J}(X_t|\beta, \mathcal{F})$ in place of $\mathcal{G}(X_t|\beta, f)$. However, since only very little is known

about the properties of any of the approximation procedures and some have only local validity (see e.g. Christiano 1990 and Dotsey and Mao 1991), one may want to condition explicitly on the existence of an approximation error in conducting inference (as e.g. in Geweke 1989). In this case one would replace (1) with

$$(3)\quad E(\mu(X_t)|f, \mathcal{I}, \mathcal{A}, \mathcal{C}) = \int_{\mathcal{A}}\int_{\mathcal{C}} \mu(X_t)\mathcal{J}(X_t,\beta|\mathcal{I}, \mathcal{F})\omega(\beta, f, \mathcal{F})\, d\beta\, dX_t$$

where ω are weights which depend on the "true" density $\mathcal{G}(X_t, \beta|\mathcal{I}, f)$ and on the approximation density $\mathcal{J}(X_t, \beta|\mathcal{I}, \mathcal{F})$. Thus, the approximation problem can be posed in terms of choosing a procedure which makes the weights in (3) as close as possible to 1. In the example of Section 5 where the function f is unknown I simply neglect the approximation error.

Third, one must specify a density $\pi(\beta|I)$ for the parameters. One could choose this density to reflect the asymptotic distribution of a GMM estimator of β (as in Burnside, Eichenbaum, and Rebelo 1993), of a simulated method of moments (SMM) estimator of β (as in Canova and Marrinan 1993), or of a maximum likelihood (ML) estimator of β (as in Phillips 1991). Two disadvantages of this approach need to be noted: first, because the density of β is selected on the basis of one data set, it does not reflect all the information available to a simulator which includes estimates of β indexed by data sets, estimation procedures or model specifications. Second, and as a consequence of the above, the dispersion associated with the density may have little relationship with the true uncertainty faced by a simulator in choosing the parameters of a model.

The idea of the paper is to choose $\pi(\beta|\mathcal{I})$ so as to reflect all available cross-sectional information. El Gamal (1993) has shown how to do this formally, using information theoretical measures. The resulting $\pi(\beta|\mathcal{I})$ is the least informative (Bayesian) density consistent with available cross sectional information. Roughly speaking, the procedure amounts to counting estimates of β previously obtained in the literature and constructing $\pi(\beta|\mathcal{I})$ by smoothing the resulting histogram. For example, if one of the elements of the β vector is the risk aversion parameter, and one counts estimates over the last 15 years obtained from fully specified GE models and smooths the resulting histogram, one would obtain a truncated (below zero) bell-shaped density, centered around two with a small mass above four. If for some parameters previous econometric evidence is scant or there is no theoretical reason to expect that one value is more likely to occur than others, one could assume uniform densities on the chosen support.

Estimates of β available in the literature are not necessarily independent (the same data set is used in some cases), some are less reliable than others and many may be noncompatible as different definition of variables and model specification are used. Nonindependent estimates are legitimate candidates to enter into 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 models which very are different from the theoretical framework used can be discounted by giving them a smaller weight in constructing histograms. In the examples of Section 5 I will choose a reasonable range for β based on theoretical

considerations and current simulation practices and impose informative densities only in those dimensions where econometric evidence is rich.

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 parameters exists, $\pi(\beta|\mathcal{I})$ is simply the product of univariate densities. If estimates of certain parameters are related (e.g. in the case of parameters describing the share of various intermediate goods in a production function), one can choose bivariate or trivariate densities for these dimensions and maintain univariate specifications for the densities of the other parameters.

3.2. *Sensitivity Analysis.* If one adopts a Monte Carlo approach to compute (1), an automatic global and efficient sensitivity analysis is performed on the entire support of the parameter space as a by-product of the simulations. Sensitivity analysis, however, can take other more specific forms. For example, one may be interested in examining how likely $\mu(X_t)$ is to be close to $h(\bar{x}_t)$ when $\beta = \hat{\beta}$ is a "cocktail party" estimate of β. In this case one could choose a path for Z_t and analyze the conditional distribution of X_t for the selected value(s) of β. Alternatively, one might wish to assess what is the maximal variation in $\mu(X_t)$ which is consistent, say, with β being in a two standard error band of a particular value. To answer this question one chooses a path for Z_t and constructs paths for x_t for draws of β in a particular range.

3.3. *Analysis of Policy Options.* Once a model has been validated, one can proceed to analyze policy options. The issue of policy analyses is subtle to deal with for two reasons. First, one has to decide how to model an intervention. Second, since many approximations to the decision rule are appropriate only locally, policy changes must be designed so that they belong to the region where approximations are valid.

The easiest case to analyze is when the component of Z_t we are interested in changing is deterministic (e.g. tax or tariff rates). In that case $\mathcal{H}(X_t, \beta|f, \mathcal{I}) \propto \pi(\beta|\mathcal{I})$ and only the randomness in the parameters affects the outcome of the experiment.

If the component of Z_t we are interested in is stochastic, but policy options do not involve changes in the distribution of the Z's, one can undertake an analysis of different policy options by simply comparing a band for $\mu(X_t)$ obtained by randomizing the β vector under the two policies. Finally, if a policy experiment involves changes in the entire distribution for Z_t one may want to compare $\mathcal{H}(X_t, \beta|f_1, \mathcal{I})$ with $\mathcal{H}(X_t, \beta|f_2, \mathcal{I})$ where f_1 and f_2 now represent two different specifications for Z_t. Differences in the outcomes can be examined using nonparametric methods as discussed in Pagan and Ullah (1991). In the example of Section 5 dealing with the evaluation of policy options, I will only consider deterministic policy changes.

4. A COMPARISON WITH EXISTING PROCEDURES

The approach we have described in the previous section lends itself to a simple Bayesian interpretation and shares features with several existing Bayesian techniques.

We have already mentioned that our "prior" on the parameters can be justified formally as the least informative density which is consistent with the information contained in a variety of estimation experiments. The procedure we employ to construct this density is also tightly linked to the selection procedure used in the so-called "consensus literature" (see Genest and Zidek 1986), where the problem is to combine different subjective Bayesian priors into an overall (more objective) one, and to the one employed in "meta-analysis" (see Wolf 1986), where the outcomes of a number of hypothesis testing experiments are combined to reach a decision (accept or reject) based on the collection of experimental p-values.

Because $G(x_t|\beta, f)$ is not necessarily the correct "likelihood" of the data, our procedure shares features also with the limited information approach of Kwan (1991a), where an unknown density $\bar{G}$ is approximated with G on the basis of one or more statistics. As Kwan shows, the approximation is appropriate if and only if the statistic on which the approximated likelihood function is built is consistent and uniformly asymptotically normal. He uses these properties to construct a diagnostic check for the quality of the approximation. Because in our setup the statistic $\mu(X_t)$ need not be consistent, Kwan's diagnostic check is inapplicable.

Our inferential approach has direct connections with the one pioneered by Box (1980). Box used predictive densities and functions of the data to provide a diagnostic check for model adequacy, which may be used to either discredit or support posterior statements about the parameters. There are two major difference between that approach and ours: first, the predictive density need not be the correct predictive density for the actual data and second, it need not have a closed form expression. The lack of closed form expression for the predictive density prevents us from computing the probability of paths analytically, like Box does.

Our model evaluation procedure is also related to the ones proposed by Gregory and Smith (1991) and Kwan (1991b). However, a few differences need to be emphasized. Gregory and Smith take the model as a testable null hypothesis and compute the probability of type I error by simulation. Although the Monte Carlo methodology underlying their procedure is identical to ours, the interpretation of the results is different for three reasons. First, Gregory and Smith assume that the mod is the true DGP for the actual data while this is not the case here. Second, they do not account for parameter uncertainty in evaluating the outcomes of the model. Finally, because they take a calibrated version of the model as the "truth," they conduct sensitivity analysis inefficiently, by replicating the experiments for different calibrated values. Kwan, on the other hand, allows for parameter uncertainty in his simulation scheme. However, he chooses a subjective "prior" density for the parameters. In addition, he evaluates the outcomes in relative terms, by comparing two alternative model specifications using a posterior-odds ratio. A model is preferred to another if it maximizes the probability that the simulated statistics are in a given set (typically chosen to be of two standard deviations width around the point estimate of the actual statistics).

Finally, the procedure for sensitivity analysis proposed here extends the approach that Harrison and Vinod (1989) used in deterministic CGE models and is complementary to the local analysis of Canova, Finn, and Pagan (1993). To determine how robust simulation results are to "small" perturbations of the

parameters around the calibrated values, they examine the magnitude of the local derivative of the statistic in the direction of interest. Because the two procedures measure the sensitivity of the results to perturbations in the parameters of different size and because they take a different point of view regarding the reliability of calibrated parameters, they provide complementary information and should both be used as specification diagnostics for simulated models.

It is simple to show that both "calibration" and "estimation by simulation" exercises appear as special cases of our simulation procedure. Calibration exercises can be seen as imposing a point mass for $\pi(\beta|\mathcal{I})$ on a particular value of β and, in certain cases, also selecting a particular path for the exogenous processes. One interpretation of this choice is that a simulator is perfectly confident that the vector β used is correct and does not worry about the cross-study or time series uncertainty surrounding estimates of β. Note that when the density of β is a singleton the marginal and the conditional density of X_t are identical. In addition, when a path for the vector of exogenous variables is selected in advance, either by drawing only one realization from their distribution or by choosing a z_t on the basis of extraneous information (for example, inputting Solow residuals in the model), the marginal for X_t has a point mass. In this last instance the likelihood of the model to produce any particular event is either 0 or 1 and one must resort to informal techniques to compare the closeness of functions of simulated and actual data. In some studies the randomness in Z_t is explicitly taken into account and repeated draws for the exogenous variables are made for a fixed value of β. In this case one computes moments of the statistics of interest by averaging the results over a small number of simulations (see, e.g., Backus, Gregory, and Zin 1989).

Simulation exercises conducted after parameters have been selected using a SMM or a GMM technique are also special cases of the proposed framework of analysis. Here $\pi(\beta|\mathcal{I})$ has a point mass at β^*, where β^* is either the SMM estimator of β (see Lee and Ingram 1990) or the SQML estimator of β (see Smith 1992) or the GMM estimator of β. In some cases, $\pi(\beta|\mathcal{I})$ is taken to be the asymptotic distribution of one of these estimators (see Canova and Marrinan 1993). Simulations are performed by drawing one or more realizations from $\mathcal{G}(X_t|\beta^*, f, \mathcal{I})$ (or from $\mathcal{H}(X_t, \beta|f, \mathcal{I})$, if the asymptotic distribution of β^* is used) and standard errors of $\mu(X_t)$ are computed using the asymptotic standard error of β^* and a linear approximation to μ.

In evaluating the model's performance these last procedures have two advantages over calibration. First, they allow formal statements on the likelihood of certain parameter values to reproduce the features of interest. For example, if the four standard error range around the point estimate of the AR(1) parameter for the productivity disturbance is [.84, .92], then it is highly unlikely (with probability higher than 99 percent) that persistent (in the sense of unit root) productivity disturbances are needed to match the data. Second, they provide a setup where sensitivity analysis to a reasonable perturbation of the parameters can easily be undertaken (although not often done).

Estimation procedures however, have two major shortcomings. First, they impose a strong form of ignorance on the simulator which does not reflect the available a priori information. The vector β may include meaningful economic

parameters which can be bounded on the basis of theoretical arguments. For example, a priori it is known that a risk aversion parameter which is negative or in excess of, say 30, is very unlikely. With SMM, GMM or SQML procedures the range of possible β 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 may not constitute an independent way to cross validate the model because the parameters used are obtained from the same data set which later will be used to compare results.

Procedures mixing calibration and GMM estimation and calibration and estimation by simulation recently employed by, e.g., Burnside, Eichenbaum, and Rebelo (1993), are also special cases of our approach. In this approach some parameters are fixed using extraneous information, while others are formally estimated using moment conditions. Although these strategies allow a more formal evaluation of the properties of the model than pure calibration procedures, they face several problems. First, parameters may be estimated regardless of their identifiability. Second, as for generic estimation procedures, the evaluation of the model is problematic because standard errors for the statistic of interest do not reflect the uncertainty faced by a simulator in choosing parameter values. Finally, as Gregory and Smith (1989) have pointed out, the small sample properties of estimators obtained from these procedures may be far from reasonable unless the parameters which are fixed in advance are consistent estimators of the true parameters. When this condition is not met, estimates of the remaining parameters may be sensitive to errors in pre-setting and close matching of simulated and actual moments may yield misleading inference.

5. SOME EXAMPLES

5.1. *A One Sector Growth Model.* The first example I consider is the Brock-Mirman one sector growth model. Here a social planner maximizes the discounted sum of utilities of the representative consumer subject to an economy wide resource constraint. The problem is of the form

$$\max_{c_t} E_0 \sum_{t=0}^{\infty} \theta^t U(c_t) \tag{4}$$

subject to

$$c_t + K_t - (1 - \delta)K_{t-1} \leq f(K_{t-1}, \varepsilon_t) \tag{5}$$

where $I_t = K_t - (1 - \delta)K_{t-1}$ is investment at t, K_t is the capital stock at t, δ is the depreciation rate, ε_t is a productivity shock and E_0 is the expectation operator. For computational convenience, I assume that the production function has the form $Y_t = f(K_{t-1}, \varepsilon_t) = K_{t-1}^{\xi}\varepsilon_t$ and that the instantaneous utility function has the form $U(c_t) = \ln(c_t)$.

For $\delta = 1$ a solution for consumption and investment in terms of the states of the problem (K_{t-1}, ε_t) exists and it is given by (see Sargent 1987)

$$c_t = (1 - \xi\theta)K_{t-1}^{\xi}\varepsilon_t \tag{6}$$

$$I_t = \xi\theta K_{t-1}^{\xi}\varepsilon_t.$$

When $\delta \neq 1$ a closed form solution for c_t and I_t does not exist and numerical techniques must be used to compute an approximation to it. To avoid the issue of numerical approximations in this first example I will set $\delta = 1$. Therefore, there are two parameters in the model $\beta = (\theta, \xi)$ and one driving process $Z_t = \varepsilon_t$. Since there need not be any relationship between the range of possible θ and ξ, I assume that the density of β is the product of the densities of the two parameters.

Several studies have estimated the discount factor θ to be, for monthly data, in the neighborhood of 0.996 (see e.g. Hansen and Singleton 1983). Estimates of this parameter range across studies from 0.990 to 1.0022. Theoretically, it is known that in the steady state, the discount factor determines the real interest rate and that a reasonable range for the annualized real interest rate is [−0.005, 0.05]. In simulation studies various authors have used values of θ in the range [0.9951, 0.9992] (see e.g., Cooley and Hansen 1990 or Backus, Gregory, and Zin 1989). I capture these observations by choosing the density for θ to be truncated normal (with truncation on both sides) centered around 0.997 and with range [0.990, 1.0022]. Note that this distribution is skewed to express the idea that a real interest rate of 2 to 3 percent or lower is more likely than an interest rate in excess of 5 percent. I assume that ξ has a uniform distribution in the range [0, 1]. This range is consistent with either decreasing or constant returns to scale. Finally, to make simulations operative I draw ε_t as iid from a $\mathcal{N}(0, 1)$. Although the iid assumption is clearly unrealistic, it avoids the introduction of a nuisance (AR) parameter in the problem.

Since the model generates stationary paths for the endogenous variables, I take stationary inducing transformations of the real data. Suppose we are interested in the relative volatility of consumption to output as in, e.g. Deaton (1987). Using U.S. monthly data on the growth rate of personal consumption expenditure and income (as proxied by an index of industrial production) for the sample 1955–1985, I obtain a value of 0.56 for this ratio with a standard deviation of 0.21. Using (6) and the production function, I generate time series for c_t and Y_t using the level of the capital stock in the U.S. in 1954, 12 for k_0.

Figure 1 presents the estimated density of the statistic when 10000 random replications for the β, $\{\varepsilon_t\}_{t=1}^{T}$ pair are drawn. Estimates of the density are obtained nonparametrically using a kernel estimator with variable width as in Pagan and Ullah (1991).[3] A value of 0.56 lies in the 98th percentile of the density of the simulated statistics and only 25 percent of the simulated density mass lies within one standard deviation band around 0.56. Moreover, the mean value for the simulated density is 0.26, the median is 0.24, the mode is 0.02, the standard deviation is 0.12, the 90 percent range is [0.7, 0.49] and the minimum and maximum are 0.001 and 0.71. Therefore, it is unlikely that this parameterization of the one

[3] The use of the nonparametric density estimate in place of the empirical frequency density is of no consequence for the results obtained here and in the next two examples (see also Gregory and Smith 1991).

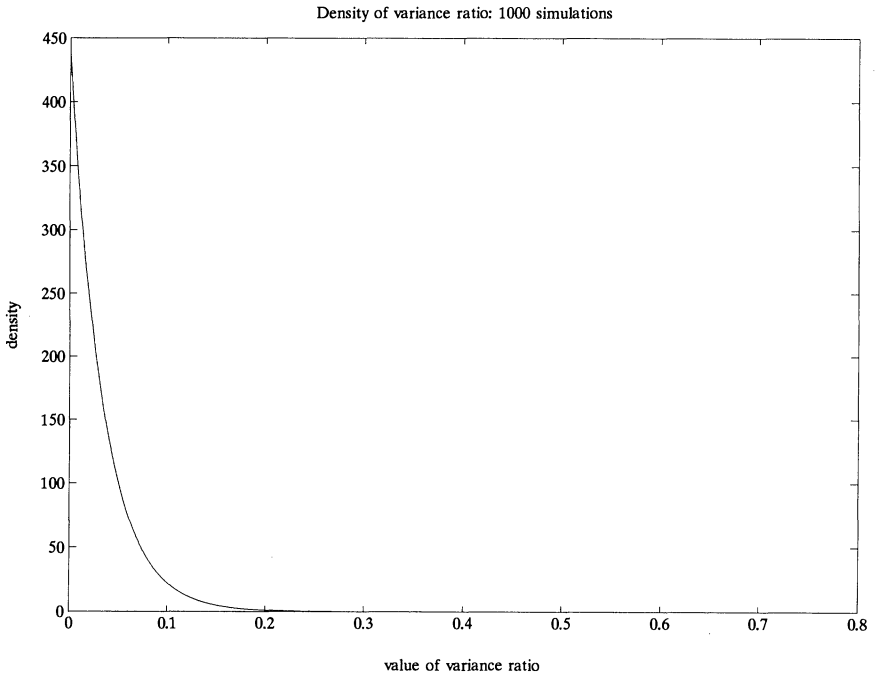

FIGURE 1

sector growth model is able to generate the relative variability of consumption to income we see in the U.S. data.

5.2. *Profits from Uncovered Speculative Strategies.* The second example draws from Canova and Marrinan (1991) and (1993) who attempt to reproduce features of profits from uncovered speculative strategies in several foreign exchange markets and of holding premiums in the term structure of U.S. interest rates.

The economy they consider is characterized by two countries. Every period, each country i is endowed with Y_{it}, $i = 1, 2$ units of a nonstorable consumption good. There are two governments which consume G_{it} units of their own country's good. To finance these consumption requirements each government issues a country specific money, M_{it}, collects real lump sum taxes, T_{it}, levied equally on agents from both countries, and issues debt to finance any purchases in excess of money creation and tax collections. This debt is in the form of state contingent nominal bills of maturity k, $k = 1, 2, \ldots, K$, denominated in their own country's currency. Endowments, government consumption requirements and money supplies are exogenous and follow independent first order Markov processes with a stationary and ergodic transition function.

Countries are each populated by a representative household maximizing a time separable utility function defined over the two goods. Households are subject to both a wealth constraint and a liquidity constraint which compels them to purchase goods with cash. The timing of the model is such that asset markets open first and

goods markets follow. At the beginning of each period the consumer enters the asset market and decides how to allocate her wealth among the productive assets of the 2 countries, currencies, and the state contingent nominal bonds issued by the 2 governments. After the asset market closes, the consumer enters the goods market and makes her consumption purchases with previously accumulated currency.

In equilibrium the expected nominal profits from holding a bond to maturity k for h periods relative to holding an h-period bond to maturity, defined as $HP_{i,t}^{k,h} = (k/h)r_{it,k} - ((k-h)/h)r_{it+h,k-h} - r_{it,h}$, are

$$(7)\quad HP_{i,t}^{k,h} = h^{-1}\left(\ln\left[\frac{E_t\theta^{k-h}Y_{it+k}(M_{it+k})^{-1}U_{it+k}}{Y_{it+h}(M_{it+h})^{-1}U_{it+h}}\right] - \ln\left[\frac{E_t\theta^{k}Y_{it+k}(M_{it+k})^{-1}U_{it+k}}{Y_{it}(M_{it})^{-1}U_{it}}\right] + \ln\left[\frac{E_t\theta^{h}Y_{it+h}(M_{it+h})^{-1}U_{it+h}}{Y_{it}(M_{it})^{-1}U_{it}}\right]\right).$$

The approximate annualized percentage expected nominal profits from speculating in foreign exchange markets defined as $EP_{t,h} = (\log E_t\{S_{t+h}\} - \log\{F_{t,h}\})$, where $F_{t,k} = S_{t,k}e^{r_{1t,k}-r_{2t,k}}$ are

$$(8)\quad EP_{t,h} = h^{-1}*\left(\log E_t\left\{\left[\frac{Y_{2t+h}(M_{2t+h})^{-1}U_{2t+h}}{Y_{1t+h}(M_{1t+h})^{-1}U_{1t+h}}\right]\right\} - \log\left\{\frac{E_t[Y_{2t+h}(M_{2t+h})^{-1}U_{2t+h}]}{E_t[Y_{1t+h}(M_{1t+h})^{-1}U_{1t+h}]}\right\}\right).$$

Canova and Marrinan examine a wide array of functions of (7) and (8) for different h. Here I confine attention to the variability and first order autocorrelation of three month holding premium and three months profits from forward speculation. The reason is that the second order properties induced by similar general equilibrium models are, in general, so different from those of actual data that some authors (e.g. Campbell and Shiller 1987, Frankel and Froot 1987) have concluded that the simple version of the rational expectations-efficient market hypothesis is severely flawed. By describing the distribution of the outcomes of these second moments from the point of view of the model we can shed light on this issue. The standard deviation and the AR(1) coefficient for profits from holding 3 months a 6 month T-bill as compared to holding a 3 month T-bill to maturity for the period 1960–1988 are .221 and .792 respectively. The standard deviation and the AR(1) coefficient for profits from 3 month forward speculation on the dollar in the dollar/mark market for the period 1979–1987 are .042 and .785 respectively.

To obtain closed form solutions for (7) and (8), I take a second order Taylor expansion around $\eta_t = (\Delta \ln(Y_{1t}), \Delta \ln(Y_{2t}), \Delta \ln(M_{1t}), \Delta \ln(M_{2t}), \ln(1-\psi_{1t}), \ln(1-\psi_{2t})$ where $\psi_{it} = G_{it}/Y_{it}$. I assume that the conditional mean and conditional variance of η_t evolve according to

$$(9)\quad E_t(\eta_{jt}) = A_{0j} + A_{1j}\eta_{jt-1} \quad j = 1, \ldots, 6$$

$$(10)\quad E_t(\eta_{jt} - E_t(\eta_{jt}))^2 \equiv \sigma_{jt}^2 = a_{0j} + a_{1j}\sigma_{jt-1}^2 + a_{2j}\varepsilon_{jt-1}^2,\ j = 1, \ldots, 6$$

where $\varepsilon_{jt} = \eta_{jt} - E(\eta_{jt}) \sim iid(0, \sigma_{jt}^2)$. Finally, I assume a utility function of the form $U(c_{1t}, c_{2t}) = (c_{1t}^{\pi} c_{2t}^{1-\pi})^{1-\gamma}/(1 - \gamma)$ where π is share of domestic goods in total consumption and γ is risk aversion parameter.

The problem under consideration is much more complex that the previous one since there are 6 exogenous sources of shocks (Y_{it}, G_{it}, M_{it}, $i = 1, 2$), 27 parameters to select and an approximation error to consider. Note that there are two types of parameters: preference parameters (π, γ) and "auxiliary" parameters, which enter expressions (7) and (8) only because of the particular time series model selected for the exogenous variables. For preferences parameters existing evidence is sufficiently rich to construct informative densities, while for the others the evidence is very slim and this forces us to choose uninformative densities for these parameters.

The density for $\beta = (\pi, \gamma)$ is assumed to be the product of univariate densities and it is selected as follows.[4] Since little is known about the mean value of π (the share of domestic good in total consumption) and different studies have used different values, I assume a uniform density over the range [0.5, 1.0]. I take the density for the risk aversion parameter to be truncated $\chi^2(4)$ and range [0, 20]. The rationale for this wide range is the large differences across studies for estimates of γ, which have obtained values between 0.5 and 3.0 (see e.g. Hansen and Singleton 1983 and Canova and Marrinan 1991) and recent simulation studies analyzing properties of financial data, which experimented with values ranging from 0.5 to 55 (see e.g. Merha and Prescott 1985, Backus, Gregory, and Zin 1989 and Kandel and Stambaugh 1990). Despite this diversity, there is enough consensus in the profession that $\gamma = 2$ is the most likely value. I capture this belief by selecting the mode of the density to take on this value. Finally, since the 95 percent range for a $\chi^2(4)$ is approximately [0.7, 10], no more than 1.5 percent of the mass of the distribution is in the region where $\gamma > 15$.

The remaining 25 parameters describe the conditional mean and variances of the exogenous processes. Based on the estimates of Stock and Watson (1989) and my own calculations I chose the ranges for A_{11}, A_{12} and A_{13} to be [−.10, .00], [−.06, .00], [.45, 60] respectively. On the ranges for A_{11} and A_{12} I assume that 50 percent of the density mass is uniformly distributed below 0 and 50 percent is lumped at 0 (see Sims 1988 for a rationale for this choice) while on the range for A_{13} I assume a uniform density. Based on arguments provided in Canova and Marrinan (1993) and estimates of Hodrick (1989) I assume a uniform density for all GARCH parameters. a_{11}, a_{12}, a_{13} have support on [−0.37, 0.13], [−0.41, 0.19] and [−0.14, −0.04], respectively while a_{21} and a_{22} both have support on [0.00, 0.50] and a_{23} has support on [0.17, 0.27].[5] The ranges for A_{01}, A_{02}, A_{03} and a_{01}, a_{02}, a_{03} are chosen endogenously so that the unconditional mean and the variance of the processes match the unconditional mean and variance of the growth rates of U.S. and OECD industrial production and of the monetary base in the U.S. For the remaining 10 parameters characterizing the behavior of government expenditure no

[4] Canova and Marrinan (1991) consider the case where the information about some of the parameters of the model is correlated. I will not examine this case here.

[5] The ranges for all the parameters are constructed from the point estimate for the period 64–88 plus or minus one standard error.

evidence exists because data on government expenditure shares in total output is not available at monthly frequency. Based on the quarterly estimates of Canova and Marrinan (1993) I assume that (A_{05}, A_{06}) have uniform densities in [0.05, 0.30], the densities of (a_{05}, a_{06}) are uniform in [0.05, 0.15]. Finally, the densities of $(A_{15}, A_{16}, a_{15}, a_{16}, a_{25}, a_{26})$ are chosen to be uniform [0, 1.0] but I eliminate all the paths for ψ_{it} which are too volatile or have a mean that lies outside of the cross sectional range of estimates of OECD countries.

For this problem I drew 5000 iid $(\beta, \{\eta_t\}_{t=1}^{T})$ vectors and neglect the approximation error due to the Taylor expansion by drawing η_t from a lognormal distribution. Estimates of the densities for the simulated variance and the simulated first autocovariance of the two series appear in Figure 2. For the holding profits series, the actual value of the standard deviation lies in the 28th percentile of the estimated density, while the actual value of the AR(1) coefficient is in 53th percentile. The means (modes) of the univariate distributions are 0.60, and 0.73 (0.18 and 0.35), the standard deviations are 0.53 and 0.22, 90 percent ranges are [0.18, 1.74] for the standard deviation and [0.35, 0.90] for the first order serial correlation. In 28 percent of the simulated values both the standard deviation and the AR(1) coefficient are below actual values, in 25 percent of the cases the variance is below the actual value but the AR(1) is greater than the actual value and in 47 percent of the cases both are greater than the actual values. For the risk profits series, the actual value of the standard error lies below all the simulated values, while the actual value of the AR(1) coefficient is in the 87th percentile. The means (modes) of the univariate distributions are 2.29 and 0.28 (1.14 and 0.57), the standard deviations are 1.36 and 0.32 and the 90 percent ranges are [1.37, 5.37] for the standard deviation and [0.07, 0.86] for the first order serial correlation. Finally, in 90 percent of the cases the standard deviation is above the actual value but the AR(1) is below than the actual value and in 10 percent of the cases they are both greater than the actual values.

In conclusion, the current model specification can generate on average more variability than what is found in the data but there are many reasonable parameter configurations for which the first order serial correlation coefficient is lower than what we see in the actual data.

5.3. *Optimal Taxation.* The final example considers the model employed by Cooley and Hansen (1990) and is chosen to illustrate how the procedure for sensitivity analysis outlined in the paper can be used to examine the consequences of policy options. The problem they examine is whether there is a combination of three different taxes which is less distorting than the actual U.S. taxation system.

The framework of analysis they employ is a closed economy model with production and two goods (cash and credit). To simplify the analysis, and because none of the conclusions depend on this, I assume that all goods are cash goods. The representative consumer maximizes lifetime utility given by

$$E_0 \sum_{t=0}^{\infty} \theta^t[\log (c_t) + B*(1 - h_t)]. \tag{11}$$

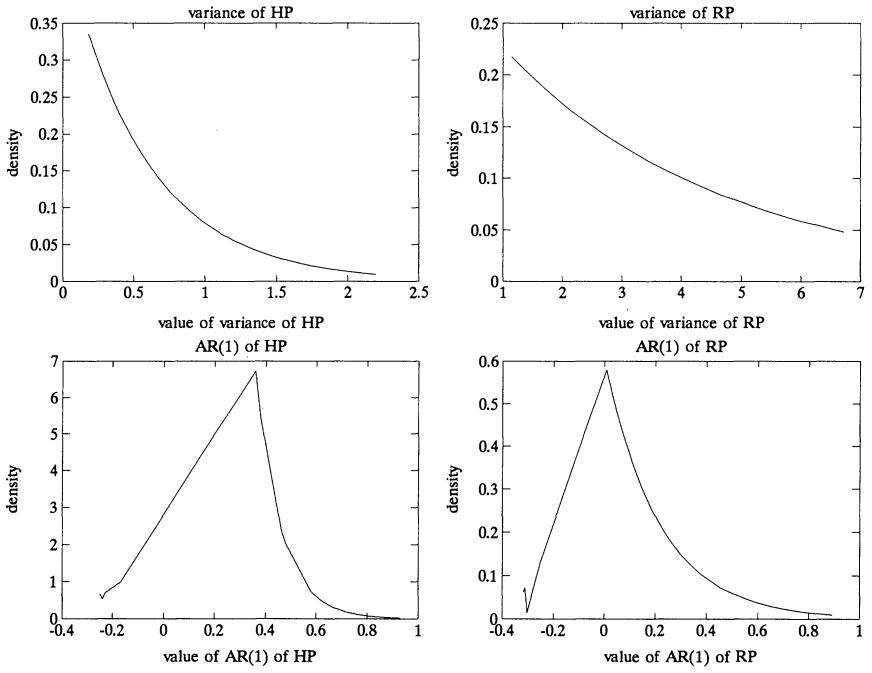

FIGURE 2

The cash-in-advance constraint and the wealth constraint are

$$p_t c_t \leq m_{t-1} + (\xi_t - 1)m_{t-1} \tag{12}$$

$$T_{ct}c_t + I_t + \frac{m_t}{p_t} \leq (1 - T_{ht})w_t h_t + (1 - T_{kt})r_t K_t + T_{kt}\delta K_t + \frac{m_{t-1}}{p_t} + TR_t \tag{13}$$

where $\xi_t > \theta$ is the gross growth rate of the money supply, I_t is investment, h_t is hours, $w_t h_t$ is labor income, $r_t K_t$ is capital income, T_{ct}, T_{ht}, T_{kt} and TR_t are the consumption tax, labor income tax, capital income tax and net transfers at t, respectively. Capital is accumulated according to

$$K_t = (1 - \delta)K_{t-1} + I_t. \tag{14}$$

There is a representative firm in the economy, owned by the consumer, maximizing profits

$$PR_t = K_t^{\alpha} h_t^{1-\alpha} - w_t h_t - r_t K_t. \tag{15}$$

Finally, there is a government which taxes agents using four distortionary taxes (inflation, consumption, labor income and capital tax) and transfers the total back to agents in a lump sum fashion. The government budget constraint is

$$(16) \qquad (\xi_t - 1)\frac{M_{t-1}}{p_t} + T_{ht} w_t h_t + T_{kt}(r_t - \delta)K_t + T_c c_t = TR_t.$$

The task here is to determine how consumer's welfare varies in the steady state with various levels and forms of taxes and to provide a upper and lower bound to the welfare costs of the current U.S. tax system. Because the analysis is conducted in the steady states, only the randomness in selecting β vector affects the outcomes of the exercise. As a welfare measure, Cooley and Hansen use the change in consumption as a percentage of steady state GNP which is required to restore an individual to the level of utility achieved with Pareto optimal allocations. Therefore Δc solves: $\log(\hat{c} + \Delta c) - \log \bar{c} - B(\hat{h} - \bar{h}) = 0$ where $\bar{c}$ and $\bar{h}$ are the steady states values of consumption and hours when all taxes are zero.

In this model there are four parameters $\beta = (\theta, B, \alpha, \delta)$ and three policy instruments (T_c, T_k, T_h). Note that because the CIA constraint is always binding, the inflation tax does not affect real allocations in the steady state it will not be considered here. I compute the welfare losses associated with each tax instrument separately using a 10 point grid on (0, 1) with 0.1 increments.

The density for the four parameters are as follows: θ is truncated normal, centered at 0.997 with range [0.990, 1.0022], α is uniform [0.25, 0.50], δ is uniform [0.006, 0.01] and B is endogenously chosen so that in the steady state agents spend between one-third and one-sixth of their time working. Choices for the range of θ has already been described. δ chosen so that the annual depreciation rate of the capital stock is between 8 and 12 percent per year. In simulation studies δ is commonly set to 0.025 which corresponds to a 10 percent annual depreciation rate. Cooley and Hansen chose a slightly lower value to match the value of the postwar investment-output ratio. In calculations I performed when quantities are not measured in per capita terms, I came up with a quarterly value for δ of 0.03. Finally, McGratten (1990) estimates δ to be 0.0226 with a standard error of 0.001. The range for α reflects calculations appearing in Christiano (1988) where, depending on how proprietors income is treated, the share of total output that is payment to capital varies between 0.25 and 0.43 and estimates obtained by, e.g., McGratten (1990).

Figure 3 plots the 90 percent bands for the welfare costs associated with each tax instrument when 10000 β vectors are drawn together with the median value of the distribution. The bands are, in general, large and nonmonotone, for a substantial portion of the grid the welfare costs of capital taxation include negative and zero values and the costs of consumption and income taxation are high for moderate tax rates. Note also that, in relative terms, the costs of capital taxation are smaller than with the other two taxes. The nonmonotonicity of the bands is due to the strong nonlinearities of Δc in the various tax rates. The fact that a low level of capital taxation yields negative welfare losses and that the costs of capital taxation appear to be smaller than with the other two taxes is related to the disincentive to work that capital taxation induces on agents. Therefore, the lower disutility of working is compensated by a lower level of consumption which needed to restore the agents to the nondistorted steady state level of utility.

To examine how far the U.S. economy is from an optimum, I compute the welfare losses using the values of the average tax rates on labor and capital obtained

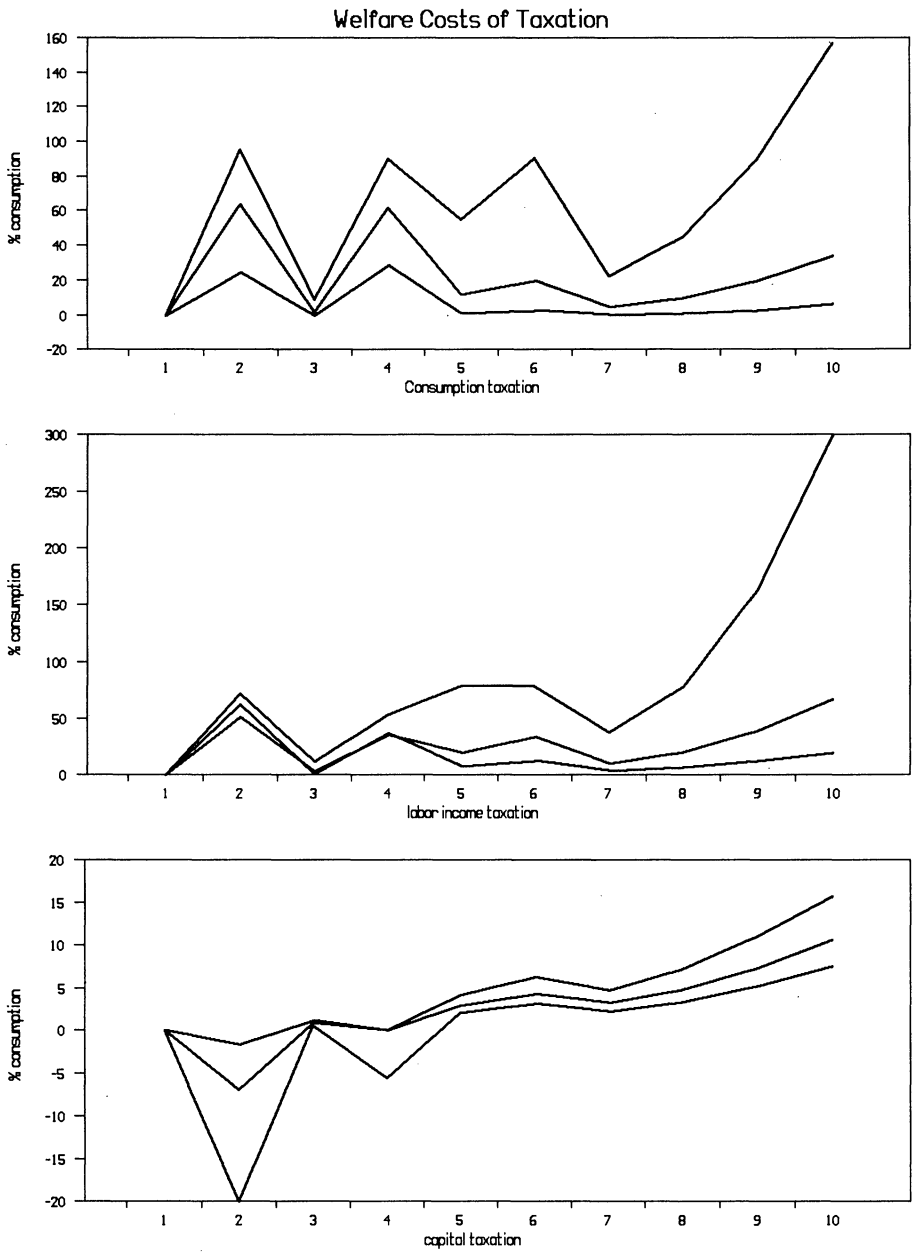

FIGURE 3

from Joines (1981) (0.23 and 0.50) and a consumption tax rate of zero. The 90 percent band for welfare losses corresponding to this tax vector is [0.12, 0.71], the median value of 0.25 and a mode of 0.52 (compared with a value of 0.31 obtained by Cooley and Hansen). Hence, if one takes the model seriously and believes

Joines' estimates, one can bound the distortions caused by the current U.S. tax system between 12 and 71 percent of U.S. steady state consumption with the most likely value at 52 percent.[6]

6. CONCLUSIONS

This paper describes a Monte Carlo approach to evaluate the properties of dynamic general equilibrium models. The starting point of the approach is the assumption that the model as DGP for the actual data is false. Given this point of view, standard classical inferential procedures are inappropriate and a new methodology for model evaluation is called for. The technique suggested in this paper can cope with the deficiencies of standard statistical analysis and provides formal foundations for the evaluation of the model via "stylized facts," which has been typical in the real business cycle literature. The procedure accounts for the uncertainty faced by a simulator in choosing the parameters of the model in a realistic way. The presence of this uncertainty becomes the key to provide a measure of dispersion for simulated statistics, a probabilistic metric to judge the closeness of the simulated and the actual data and an evaluation criteria for the model. The approach has a built-in feature which allows for global sensitivity analysis and several forms of conditional or local sensitivity analysis and evaluates the range of possibilities of the model by attaching probability statements to events a simulator may be interested in characterizing. Finally, the approach is easy to implement and includes existing calibration, estimation by simulation and GMM techniques as special cases.

The technique is applied to three examples involving different objectives, level of knowledge about the "deep" parameters and complexity and shows how to provide realistic conclusions to policy questions. Note also that computation considerations are not a major issue for problems of moderate size. For all the examples presented in this paper densities for the objects of interest were computed in a matter of minutes.

Universitat Pompeu Fabra, Spain

REFERENCES

BERTSEKAS, D., *Dynamic Programming and Stochastic Control* (New York: Academic Press, 1976).

BACKUS, D., A. GREGORY, AND S. ZIN, "Risk Premiums in the Terms Structure: Evidence from Artificial Economies," *Journal of Monetary Economics* 24 (1989), 371–399.

BAXTER, M., "Approximating Suboptimal Dynamic Equilibria: An Euler Equation Approach," *Journal of Monetary Economics* 27 (1991), 173–200.

BOX, G., "Sampling and Bayes Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society* Series A, 143 (1980), 383–430.

BURNSIDE, C., M. EICHENBAUM, AND S. REBELO, "Labor Hoarding and the Business Cycle," *Journal of Political Economy* 101 (1993), 245–273.

[6] McGratten (1990) shows that there are large differences in the time series properties for tax rates in the existing literature. Therefore both the magnitude of the band and its most likely value should be viewed with caution.

CANOVA, F., M. FINN, AND A. PAGAN, "Evaluating a Real Business Cycle Model," in C. Hargreaves, ed., *Nonstationary Time Series Analyses and Co-integration* (Oxford: Oxford University Press, 1993).

——— AND J. MARRINAN, "Reconciling the Term Structure of Interest Rates with the Consumption Based ICAP Model," *Journal of Economic Dynamic and Control* (forthcoming).

——— AND ———, "Profits, Risk and Uncertainty in Foreign Exchange Markets," *Journal of Monetary Economics* 32 (1993), 259–286.

CAMPBELL, J. AND R. SHILLER, "A Simple Account of the Behavior of Long Term Interest Rates," *American Economic Review, Papers and Proceedings* 74 (1987), 44–48.

CHRISTIANO, L., "Why Does Inventory Fluctuate so Much?" *Journal of Monetary Economics* 21 (1988), 247–280.

———, "Solving the Stochastic Growth Model by Linear Quadratic Approximation and by Value Function Iteration," *Journal of Business and Economic Statistics* 8 (1990), 23–26.

COLEMAN, W., "An Algorithm to Solve Dynamic Models," Discussion Paper No. 351, Board of Governors of the Federal Reserve System, International Finance Division, 1989.

COOLEY, T. AND G. HANSEN, "Tax Distortions in a Neoclassical Monetary Economy," Discussion Paper No. 38, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, 1990.

DEATON, A., "Life Cycle Models of Consumption: Is the Evidence Consistent with Theory?" in T. Bewley, ed., *Advance in Econometrics: Fifth World Congress*, Vol. 2 (New York: Cambridge University Press, 1987).

DOTSEY, M. AND C. S. MAO, "How Well Do Linear Approximation Methods Work? Results for Suboptimal Dynamic Equilibria," *Journal of Monetary Economics* 29 (1991), 25–58.

EL-GAMAL, M., "The Extraction of Information from Multiple Point Estimates," *Nonparametric Statistics* (forthcoming 1993).

FAIR, R., "Estimating Event Probabilities from Macroeconomic Models Using Stochastic Simulations," manuscript, Yale University, 1991.

FRANKEL, J. AND K. FROOT, "Using Survey Data to Tests Standard Propositions Regarding Exchange Rate Expectations," *American Economic Review* 77 (1987), 133–159.

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

FRISCH, R., "Propagation Problems and Impulse Problems in Dynamics Economics," in *Economic Essays in Honor of Gustav Cassel* (London: 1933).

GELFAND, A. AND A. SMITH, "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85 (1990), 398–409.

GENEST, G. AND J. ZIDEK, "Combining Probability Distributions: A Critique and an Annotated Bibliography," *Statistical Science* 1 (1986), 114–148.

GEWEKE, J., "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* 57 (1989), 1317–1339.

GREGORY, A. AND G. SMITH, "Calibration as Estimation," *Econometric Reviews* 9 (1989), 57–89.

——— AND ———, "Calibration as Testing: Inference in Simulated Macro Models," *Journal of Business and Economic Statistics* 9 (1991), 297–303.

——— AND ———, "Calibration in Macroeconomics," in G. S. Maddala, ed., *Handbook of Statistics*, Vol. 10 (Amsterdam: North Holland, 1993).

HAAVELMO, A., "The Probability Approach in Econometrics," *Econometrica* 12 (1944), Supplement.

HANSEN, L., "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica* 50 (1982), 1029–1054.

——— AND T. SARGENT, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamic and Control* 2 (1979), 7–46.

——— AND K. SINGLETON, "Stochastic Consumption, Risk Aversion and Temporal Behavior of Asset Returns," *Journal of Political Economy* 91 (1983), 249–265.

HARRISON, G. AND H. D. VINOD, "Sensitivity Analysis of Applied General Equilibrium Models: Completely Randomized Factorial Sampling Designs," manuscript, University of New Mexico, 1989.

HODRICK, R., "Risk, Uncertainty and Foreign Exchange Markets," *Journal of Monetary Economics* 23 (1989), 433–459.

JOINES, D., "Estimates of Effective Marginal Tax Rates on Factor Incomes," *Journal of Business* 54 (1981), 191–226.

JORGENSON, D., "Econometric Methods for Applied General Equilibrium Analysis," in H. Scarf and J. Whalley, eds., *Applied General Equilibrium Analysis* (New York: Cambridge University Press, 1984).

JOURNAL OF BUSINESS AND ECONOMIC STATISTICS (January 1990).

JUDD, K., "Projection Methods for Solving Aggregate Growth Models," *Journal of Economic Theory* 58 (1992), 410–452.

KANDEL, S. AND R. STAMBAUGH, "Expectations and Volatility of Consumption and Asset Returns," *Review of Financial Studies* 3 (1990), 207–232.

KIM, K. AND A. PAGAN, "The Econometric Analysis of Calibrated Macroeconomic Models," in H. Pesaran and M. Wickens, eds., *Handbook of Applied Econometrics*, Vol. 1 (London: Blackwell Press, 1993).

KING, R., C. PLOSSER, AND S. REBELO, "Production, Growth and Business Cycles: I," *Journal of Monetary Economics* 21 (1988), 195–232.

KWAN, Y. K., "Bayesian Analysis with an Unknown Likelihood Function: A Limited Information Approach," manuscript, University of Chicago Business School, 1991a.

———, "Bayesian Calibration with an Application to a Non-linear Rational Expectation Two Country Model," manuscript, University of Chicago Business School, 1991b.

KYDLAND, F. AND E. PRESCOTT, "Time to Build and Aggregate Fluctuations," *Econometrica* 50 (1982), 1345–1370.

——— AND ———, "The Econometrics of the General Equilibrium Approach to Business Cycles," *Scandinavian Journal of Economics* 93 (1991), 161–178.

LEE, B. S. AND B. INGRAM, "Simulation Estimators of Time Series Models," *Journal of Econometrics* 47 (1991), 197–206.

MARCET, A., "Solving Nonlinear Rational Expectations Models by Parametrizing Expectations," Working Paper No. 5, Universitat Pompeu Fabra, 1992.

MCGRATTEN, E., "The Macroeconomic Effects of Distortionary Taxation," Discussion Paper No. 37, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, 1990.

MERHA, R. AND E. PRESCOTT, "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15 (1985), 145–162.

NIEDERREITER, H., "Quasi Monte Carlo Methods for Multidimensional Numerical Integration," *International Series of Numerical Mathematics* 85 (1988), 157–171.

NOVALES, A., "Solving Nonlinear Rational Expectations Models: A Stochastic Equilibrium Model of Interest Rates," *Econometrica* 58 (1990), 93–111.

PAGAN, A. AND SHANNON, "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, 1985).

——— AND A. ULLAH, "The Econometric Analysis of Models with Risk Terms," *Journal of Applied Econometrics* 3 (1988), 87–105.

——— AND ———, "Non-Parametric Estimation," manuscript, University of Rochester, 1991.

PHILLIPS, P.C.B., "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics* 6 (1991), 333–354.

SARGENT, T., *Dynamic Macroeconomic Theory* (Cambridge: Harvard Press, 1987).

SHOVEN, J. AND J. WHALLEY, "Applied General Equilibrium Models of Taxation and International Trade: An Introduction and Survey," *Journal of Economic Literature* 22 (1984), 1007–1051.

SIMS, C., "Solving Nonlinear Stochastic Equilibrium Models Backward," manuscript, University of Minnesota, 1984.

———, "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamic and Control* 12 (1988), 463–474.

———, "Models and Their Users," *American Journal of Agricultural Economics* 71 (1989), 489–494.

SMITH, T., "Estimating Non-Linear Time Series Models Using Simulated VAR: Two Approaches," *Journal of Applied Econometrics* 8 (1993), S63–S84.

STOCK, J. AND M. WATSON, "Interpreting the Evidence of Money-Income Causality," *Journal of Econometrics* 40 (1989), 161–182.

TANNER, M. AND W. WONG, "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association* 87 (1987), 528–550.

TAUCHEN, G. AND R. HUSSEY, "Quadrature Based Methods for Obtaining Approximate Solutions to Integral Equations of Nonlinear Asset Pricing Models," *Econometrica* 59 (1991), 371–397.

WATSON, M., "Measures of Fit for Calibrated Models," *Journal of Political Economy* 101 (1993), 1011–1041.

WOLF, F., *Meta-Analysis: Quantitative Methods for Research Synthesis* (Beverly Hills: Sage, 1986).