Computational Experiments and Reality

John Geweke
University of Minnesota and Federal Reserve Bank of Minneapolis

First version: February 12, 1999
This version: June 9, 1999

Abstract

This study explores three alternative econometric interpretations of dynamic, stochastic general equilibrium (DSGE) models. Under a strong econometric interpretation, these models provide likelihood functions for observed sequences of prices and quantities. Given this interpretation, most DSGE models are rejected using classical econometrics and assigned zero probability in a Bayesian approach. Under a weak econometric interpretation, commonly made in the calibration literature (Kydland and Prescott, 1996), a DSGE model mimics the world only along a carefully specified set of dimensions. Computational experiments provide predictive distributions in these dimensions, which are then compared with the corresponding observed values. The weak econometric interpretation shares the same assumptions as the strong econometric interpretation, however, and it therefore leads again to the conclusion that most DSGE models are not credible. Under a minimal econometric interpretation, introduced by DeJong, Ingram and Whiteman (1996) and further developed in this study, DSGE models provide only population moments for certain functions of prices and quantities. They have empirical content only in the presence of auxiliary models which link population moments with observables. Such models may be entirely atheoretical. This study shows how DSGE models and an atheoretical model may be integrated into a common probability model, and used together to draw conclusions about different DSGE models. The methods and conclusions of the study are all illustrated using models of the equity premium.

Research assistance was provided by Jack Stecher. The National Science Foundation provided partial financial support through grant SBR-9514865. Comments from seminar participants at Johns Hopkins University, New York University, the University of Iowa, and Stanford University were helpful in writing and revising the paper. Responsibility for errors rests with the author, and the views expressed are not necessarily those of the Federal Reserve Bank of Minneapolis or the University of Minnesota Department of Economics.

1. Introduction

The dynamic, stochastic, general equilibrium (DSGE) model has become a central analytical tool in studying aspects of economic behavior in which aggregate uncertainty is important. These models abstract sufficiently from measured economic behavior that clarification of the dimensions of reality they are intended to mimic is essential if they are to deepen our understanding of real economies. If the relation between DSGE models and measured economic behavior can be made formal, explicit, and simple, then the analytic power of this approach and our understanding of economic behavior will be enhanced. If this relation is arbitrary, vague, and convoluted, then the usefulness of DSGE models will be neither tested nor widely appreciated. The objective of this study is to provide such a formal, explicit and simple characterization of the relation between DSGE models and measured economic behavior.

The approach taken here is to examine three alternative interpretations of the relationship. The first, called the strong econometric interpretation, leads to conventional, likelihood based, econometric methods. It is widely understood that DSGE models fare badly under this interpretation, and the DSGE literature consistently denies its appropriateness given the level of abstraction in the models. The second, called the weak econometric interpretation, greatly reduces the dimensions of observed behavior a DSGE model is designed to explain. It is the interpretation advanced by Kydland and Prescott (1996). Its assumptions are in fact no weaker than those that lead to likelihood based econometrics, and so DSGE models fare badly under this interpretation as well, although the failure is not so immediately evident. This study develops a third, minimal econometric interpretation of DSGE models introduced by DeJong, Ingram and Whiteman (1996). The assumptions underlying this interpretation are much weaker, and it is immune to the difficulties encountered in likelihood based econometrics. To be capable of explaining measured aggregate economic behavior, however, DSGE models under this interpretation must be married to econometric models that provide empirically plausible descriptions of measured behavior. This study shows how to do this in a way that is formal, explicit, simple, and easy to implement.

The three econometric interpretations are all presented with reference to a particular substantive application – DSGE models designed to explain the equity premium. The paper begins, in Section 2, by setting forth four such models. The alternative econometric interpretations are taken up in Sections 3, 4, and 5. In each case numerical and graphical methods are used to illustrate the application to equity premium models. The weak econometric interpretation (Section 4) corroborates the findings of the DSGE literature regarding the equity premium puzzle – as it must, for this is the interpretation used there. The minimal econometric interpretation (Section 5) reverses some of the findings widely regarded as established by DSGE models. These findings illustrate some of the returns to a formal, explicit and simple approach to inference in DSGE models.

2. The essential elements of DSGE models

Dynamic, stochastic general equilibrium (DSGE) models have several common elements. They specify preferences of economic agents over alternative paths of consumption, a technology of production, and perhaps a government sector. They assume that all economic agents choose their most preferred path of consumption. They allow stochastic perturbations to the production technology. They use the principle of competition to determine equilibrium paths of quantities and prices, as functions of tastes, technology, and stochastic shocks. Tastes, technology, and the assumption of competition transform the technology shock distribution to a distribution of quantities and prices.

To isolate the econometrically relevant implications of these models, let "A" denote the assumptions of a particular model. For example, this could include the assumptions that preferences are time separable with constant relative risk aversion in each period, production is Cobb-Douglas, shocks to technology are log-normal and first order autoregressive, and equilibrium is competitive. Let "θ_A" denote the parameters that provide quantitative content for the model – for example, the specification that labor's share is .70, the coefficient of relative risk aversion is 2.0, and so on. Finally, let "$\mathbf{y}$" denote an observable, finite sequence of quantities and prices whose equilibrium values the model describes, for example 90 years of annual asset returns and output growth.

If the model has a unique equilibrium then it implies a distribution of **y**, given the values of the parameters. A generic expression for this distribution is $\mathrm{p}(\mathbf{y}|\theta_A, A)$. In most DSGE models, $\mathrm{p}(\mathbf{y}|\theta_A, A)$ cannot be derived in closed form. However, it is typically not difficult to learn about $\mathrm{p}(\mathbf{y}|\theta_A, A)$ by means of forward simulations: given a value of θ_A, pseudo-random vectors $\tilde{\mathbf{y}}$ can be drawn independently and repeatedly from $\mathrm{p}(\mathbf{y}|\theta_A, A)$. In many cases, this ability to simulate is sufficient to draw formal conclusions about the model and use it to study the substantive questions it was designed to address.

2.1. An example: General equilibrium models of the equity premium

Average annual real returns on relatively riskless short-term securities in the U.S. have been about one percent during the past one hundred years. Average annual real returns on equities over the same period have averaged above six percent. The equity premium—the difference between the return to equities and the return to relatively riskless short-term securities—has exceeded five percent during the past century in the U.S. Many simple general equilibrium models predict average returns on riskless assets that are much higher than the observed average value, and average equity premia that are much lower, given parameter values generally regarded as reasonable. This predictive failure has become known as the equity premium puzzle. Kocherlakota (1996) provides a recent review of the literature.

In the simplest general equilibrium model of the equity premium there is a single perishable good produced and consumed each period. Let period *t* production of the good be y_t, and denote the period-to-period gross growth rate of output by $x_t = y_t / y_{t-1}$. The representative agent orders preferences over random paths of consumption $\{y_t\}$ by

(2.1.1) $\mathrm{E}_t\left[\sum_{s=0}^{\infty} \delta^s \,\mathrm{U}(y_{t+s})\right]$.

In this expression $\delta \in (0,1)$ is the subjective discount factor, and E_t denotes expectation conditional on time *t* information. The instantaneous utility function is the constant relative risk aversion (CRRA) utility function

(2.1.2) $$\mathrm{U}(c_t) = c_t^{1-\alpha} / (1-\alpha),$$

it being understood that $\mathrm{U}(c_t) = \log(c_t)$ when $\alpha = 1$. The parameter α is the coefficient of relative risk aversion in the instantaneous utility function (2.1.2), and is also proportional to the marginal rate of intertemporal substitution in the preference ordering (2.1.1).

Define a riskless asset to be a claim to one unit of consumption in the next period. If such an asset is held in this economy, its period *t* price must be

(2.1.3) $p_t = \delta\,\mathrm{E}_t\left[\mathrm{U}'(y_{t+1}) / \mathrm{U}'(y_t)\right] = \delta\,\mathrm{E}_t\left(x_{t+1}^{-\alpha}\right).$

Define one share of equity to be a claim to the fraction *f* of output in all future periods. If this asset is held in this economy, its period *t* price must be

(2.1.4)
$$\begin{aligned} q_t &= f \cdot \mathrm{E}_t\left[\sum_{s=1}^{\infty} \delta^s\, \mathrm{U}'(y_{t+s}) y_{t+s} \Big/ \mathrm{U}'(y_t)\right] \\ &= f y_t\, \mathrm{E}_t\left[\sum_{s=1}^{\infty} \delta^s\, \mathrm{U}'(y_{t+s}) y_{t+s} \Big/ \mathrm{U}'(y_t) y_t\right] = f y_t\, \mathrm{E}_t\left[\sum_{s=1}^{\infty} \delta^s \prod_{j=1}^{s} x_{t+j}^{1-\alpha}\right], \end{aligned}$$

from which

(2.1.5) $$q_t = \delta\,\mathrm{E}_t\left[\mathrm{U}'(y_{t+1})(f y_{t+1} + q_{t+1}) / \mathrm{U}'(y_t)\right] = \delta\,\mathrm{E}_t\left[x_{t+1}^{-\alpha}(f y_{t+1} + q_{t+1})\right].$$

From (2.1.4), the share price is proportional to output. If the growth rate x_t is stationary, then q_t / y_t is also stationary even though output y_t is not. This is a consequence of the assumption that instantaneous utility is of the CRRA form (2.1.2) – in fact, (2.1.2) is the unique instantaneous utility function with this property in (2.1.1) (King, Plosser and Rebelo, 1990).

2.2. The Mehra-Prescott and Rietz models

Mehra and Prescott (1985) assume that the growth rate x_t is a first order Markov chain with *n* discrete states. The growth rate is λ_j in state *j*. Assume that the time *t* information set includes the history of growth rates, and let P_t denote probability conditional on time *t* information. Then the Mehra-Prescott assumption can be expressed

(2.2.1) $$\mathrm{P}_t\left(x_{t+1} = \lambda_j \middle| x_t = \lambda_i\right) = \phi_{ij}.$$

Suppose that this economy is in state *i* at time *t*. Then from (2.1.3) and (2.2.1), the price of the riskless asset is

$$p_t = p^{(i)} \equiv \delta \sum_{j=1}^{n} \phi_{ij} \lambda_j^{-\alpha},$$

and the return to the riskless asset held from period t to period $t+1$ is

$$r_{t+1} = r^{(i)} \equiv 1/p^{(i)} - 1.$$

From (2.1.4) the share price q_t is proportional to output y_t. Hence for $x_t = \lambda_i$, denote the share price $q_t = w_i y_t$. Substituting in (2.1.5),

$$w_i = \delta \sum_{j=1}^{n} \phi_{ij} \lambda_j^{-\alpha} \left(f\lambda_j + w_j \lambda_j \right) = \delta \sum_{j=1}^{n} \phi_{ij} \lambda_j^{(1-\alpha)} \left(f + w_j \right) \quad (i = 1, \ldots, n).$$

Solving this system of n linear equations for $(w_1, \ldots, w_n)$ yields the share prices $w_i y_t$. If $x_{t-1} = \lambda_i$ and $x_t = \lambda_j$, then the net return to equity holding from period $t-1$ to period t is

$$s_t = s^{(i,j)} \equiv \frac{q_t + fy_t - q_{t-1}}{q_{t-1}} = \frac{w_j y_t + fy_{t-1}\lambda_j - w_i y_{t-1}}{w_i y_{t-1}} = \frac{\lambda_j \left(w_j + f \right)}{w_i} - 1.$$

Mehra and Prescott (1985) take up the case $n = 2$, and restrict $\phi_{11} = \phi_{22} = \phi$. They choose $\lambda_1 = 1.054$, $\lambda_2 = 0.982$, and $\phi = 0.43$ to match the mean, standard deviation, and first order autocorrelation in the annual growth rate of per capita U.S. real consumption between 1889 and 1978. They then examine whether there are values of α less than 10 and any values of $\delta \in (0, 1)$ consistent with the observed average annual real returns of 0.0080 for short-term relatively riskless assets, and 0.0698 for the Standard and Poor's Composite Stock Price Index, over the same period. Their conclusion is negative.

Rietz (1988) uses the same model but adds a third state for output growth ($n = 3$). The third state occurs with low probability, the growth rate in this state is quite negative, and return to one of the two normal growth states occurs with certainty in the next period. Rietz concludes that this model is consistent with the observed average returns to riskless assets and the Stock Price Index, for some combinations of the parameter values – for example, α in the range of 5 to 7, δ above .98, and a probability of about 0.1% of a growth rate in which half of output is lost.

2.3. The Labadie and mixture models

Labadie (1989) takes

$$\log x_t = \beta_0 + \beta_1 \log x_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{IID}{\sim} \mathrm{N}\left(0, \sigma^2\right).$$

Tsionas (1994) generalizes this to

$$\log x_t = \beta_0 + \beta_1 \log x_{t-1} + \omega_t^{1/2} \varepsilon_t, \quad \varepsilon_t \overset{IID}{\sim} \mathrm{N}(0, 1), \quad \omega_t \overset{IID}{\sim} \mathrm{p}_{\omega}(\cdot),$$

with $\{\omega_t\}$ and $\{\varepsilon_t\}$ mutually independent. The riskless asset price follows from (2.1.3):

$$p_t = \delta\,\mathrm{E}_t\left(x_{t+1}^{-\alpha}\right) = \delta \exp\left[-\alpha\left(\beta_0 + \beta_1 \log x_t\right)\right]\mathrm{E}_t\left[\exp\left(-\alpha\omega_{t+1}^{1/2}\varepsilon_{t+1}\right)\right]$$
$$= \delta \exp\left(-\alpha\beta_0\right)x_t^{-\alpha\beta_1}\,\mathrm{E}_t\left(\alpha^2\omega_{t+1}/2\right) = \delta\exp\left(-\alpha\beta_0\right)x_t^{-\alpha\beta_1}\,\mathrm{M}\left(\alpha^2/2\right),$$

where $\mathrm{M}(\cdot)$ denotes the moment generating function of ω, $\mathrm{M}(t) = \mathrm{E}\left[\exp(\omega t)\right]$. The net return on holding the riskless asset is thus

(2.3.1) $$r_t = \exp\left(\alpha\beta_0\right)x_t^{\alpha\beta_1}\Big/\delta\,\mathrm{M}\left(\alpha^2/2\right) - 1.$$

Direct substitution in (2.1.5) leads to

(2.3.2) $$q_t / f y_t = \sum_{s=1}^{\infty} c_s m_s x_t^{a_s} ,$$

with

$$a_s = \rho\beta_1\left(1-\beta_1^s\right)\Big/\left(1-\beta_1\right),$$
$$c_s = \exp\left[\frac{\rho s\beta_0}{1-\beta_1} - \frac{\rho\beta_0\beta_1\left(1-\beta_1^s\right)}{\left(1-\beta_1\right)^2}\right],$$
$$m_s = \delta^s \prod_{j=1}^{s}\mathrm{M}\left[\rho^2\left(1-\beta_1^j\right)^2\Big/\rho^2\left(1-\beta_1\right)^2\right],$$

where $\rho = 1-\alpha$. The expression (2.3.2) converges, and equilibrium with finite equity prices exists, if and only if $\mathrm{M}\left[\rho^2\big/2\left(1-\beta_1^2\right)\right] < \delta^{-1}$. Defining the left side of (2.3.2) to be h_t, the return to equity is then $s_t = x_t\left(h_t + 1\right)\big/h_{t-1} - 1$.

Tsionas (1994) thus extends Labadie (1989) by permitting the growth shock to be a scale mixture of normals. The best known scale mixture of normals is the Student-*t* distribution, corresponding to an inverted gamma mixing distribution $\mathrm{p}_\omega(\cdot)$. However, the inverted gamma distribution has no moment generating function: the implicit integral on the left side of (2.3.1) diverges, and there is no equilibrium with finite asset prices. An attractive flexible family of symmetric distributions is the finite scale mixture of normals, for which the moment generating function is trivial and always exists. A distribution in this family has *n* components, with component *i* assigned probability p_i. Conditional on component *i*, $\omega_t = \omega_{(i)}$ $(i = 1,\ldots,n)$. Thus, the p.d.f. of $u_t = \omega_t^{1/2}\varepsilon_t$ is

$$\mathrm{p}\left(u_t\right) = (2\pi)^{-1/2}\sum_{i=1}^{n} p_i\omega_{(i)}^{-1/2}\exp\left(-u_t^2\big/2\omega_{(i)}\right).$$

To extend Labadie's model in much the same way that Rietz extended Mehra and Prescott, let $n = 2$, let $i = 1$ denote the "normal" state, and let $i = 2$ denote the "high variance" state. We refer to this subsequently as the mixture model.

The Labadie and mixture models can be calibrated in the same way as the Mehra-Prescott and Rietz models. In the Labadie model choose β_0, β_1, and σ^2 to match the same three moments used by Mehra and Prescott: the mean, standard deviation, and first-order autocorrelation coefficient of U.S. consumption growth for 1889 through 1978. For the mixture model do the same thing, except to substitute $\omega_{(1)}$ for σ^2. To parallel the treatment of Rietz, let $p_1 = .99$ and $p_2 = .01$ in the mixture model.

3. Strong econometric interpretation of DSGE models

The strong econometric interpretation of a dynamic, stochastic general equilibrium (DSGE) model is that the model provides a predictive distribution for an observable sequence of quantities and/or prices $\mathbf{y}$. Given the parameter values, $\mathrm{p}(\mathbf{y}|\theta_A, A)$ is the *ex ante*, predictive distribution for the observables $\mathbf{y}$. Then, letting $\mathbf{y}^o$ denote the observed value of $\mathbf{y}$, $\mathrm{L}(\theta_A;\mathbf{y}^o, A) \equiv \mathrm{p}(\mathbf{y}^o|\theta_A, A)$ is the likelihood function *ex post*. To provide such a predictive distribution the model must specify values of the parameters θ_A, or indicate a reasonable range for parameter values and a distribution over that range, not just a distribution conditional on an unknown parameter vector θ_A. Many studies that construct and calibrate DSGE models provide this sort of information about parameter values, at least informally, by means of reference to values in the literature and through their choice of calibrated values used in simulations.

Given a strong econometric interpretation, there is potentially a rich agenda of positive and normative economics. On the positive side, predictive distributions of observables can be used to compare alternative models, discard some, and improve others. On the normative side, DSGE models provide predictive distributions of the welfare implications of the policy interventions they are designed to accommodate. Unfortunately, under a strong econometric interpretation the normative promise of DSGE models cannot be realized, because they fail completely in positive comparisons with alternative models. This result emerges in both Bayesian and classical approaches.

3.1. Classical econometrics

Formal, classical approaches to model evaluation and comparison are squarely in the tradition of Neyman and Pearson, and have become known as specification testing. Let $z = \mathrm{t}(\mathbf{y})$ be a scalar function of the observables, and let $\mathrm{p}(z|\theta_A, A)$ be the density of z implied by $\mathrm{p}(\mathbf{y}|\theta_A, A)$. If z is pivotal—i.e., the distribution of z is the same for all $\theta_A \in \Theta_A$—then one may unambiguously write the probability density function of z as $\mathrm{p}(z|A)$ and the cumulative distribution function as $\mathrm{P}(z|A)$. In this case, the model A implies that $1 - \mathrm{P}(z|A)$ is uniformly distributed on the unit interval. Then $z^o = \mathrm{t}(\mathbf{y}^o)$ is, formally, a test statistic, and $1 - \mathrm{P}(z^o|A)$ is the marginal significance level of z^o. It is then a short step to state that the model A is rejected at marginal significance level $1 - \mathrm{P}(z^o|A)$. The specification A is rejected if z^o is sufficiently large, or equivalently $1 - \mathrm{P}(z^o|A)$ is too small. More realistically z^o or $1 - \mathrm{P}(z^o|A)$ may simply be presented as a useful summary of evidence against model specification A.

Underlying this familiar use of a test statistic $z^o = \mathrm{t}(\mathbf{y}^o)$ is the belief, perhaps unarticulated, that there is some plausible alternative specification $\mathrm{p}(\mathbf{y}|\theta_B, B)$, $\theta_B \in \Theta_B$, and that for at least some $\theta_B \in \Theta_B$, $\mathrm{P}(z|\theta_B, B) < \mathrm{P}(z|A)$. It is this belief that makes it reasonable to reject A in a test of, say, size 5% if $1 - \mathrm{P}(z^o|A) < .05$, as opposed to rejecting A if $.376 < 1 - \mathrm{P}(z^o|A) < .426$. In formal terms, the power as well as the size of a test must be considered. But the concept of power requires an alternative to the model under consideration. Model evaluation cannot be divorced from model comparison: underlying any specification test is a hypothesis test against a class of alternative models.

The choice of the test statistic $z = \mathrm{t}(\mathbf{y})$ typically involves both science and art. Finding a pivotal test statistic is possible only in special simple circumstances: Student's t for a single linear restriction in a linear model is the leading example. In most cases the best a classical econometrician can do is to establish that $z = \mathrm{t}(\mathbf{y})$ is pivotal asymptotically. This technical requirement substantially limits the range of test statistics that can be considered. Determination of $\mathrm{P}(z|A)$ in finite sample for pivotal z or in large sample for asymptotically pivotal z is then straightforward, for simulation can be used to create the necessary tables if an analytic derivation is not at hand. Especially in complex models, showing that $z = \mathrm{t}(\mathbf{y})$ is asymptotically pivotal is the dominant technical concern. The choice of z within this class is more art: whether there are other tests more powerful than a given $z = \mathrm{t}(\mathbf{y})$ is generally unknown.

3.2. Bayesian econometrics

The formal Bayesian approach to model comparison is grounded in complete specification of the joint distribution of models, parameters, and observables. If two models, A and B, are being considered, this complete specification can be achieved by stipulating a density $\mathrm{p}(\theta_A|A)$ with support Θ_A, a density $\mathrm{p}(\theta_B|B)$ with support Θ_B, and model probabilities $\mathrm{p}(A)$ and $\mathrm{p}(B)$ with $\mathrm{p}(A)+\mathrm{p}(B)=1$. Then

(3.2.1) $$\mathrm{p}(A,B,\theta_A,\theta_B,\mathbf{y}) = \mathrm{p}(A)\,\mathrm{p}(\theta_A|A)\,\mathrm{p}(\mathbf{y}|\theta_A,A) + \mathrm{p}(B)\,\mathrm{p}(\theta_B|B)\,\mathrm{p}(\mathbf{y}|\theta_B,B).$$

Given (3.2.1), the evidence in the data about models and parameters follows from the laws of conditional probability. For example, the marginal likelihood of model A is

$$\mathrm{p}(\mathbf{y}^o|A) = \int_{\Theta_A} \mathrm{p}(\theta_A|A)\,\mathrm{p}(\mathbf{y}^o|\theta_A,A)\,d\nu(\theta_A)$$

the posterior distribution of θ_A is

$$\mathrm{p}(\theta_A|\mathbf{y}^o,A) = \mathrm{p}(\theta_A|A)\,\mathrm{p}(\mathbf{y}^o|\theta_A,A)\big/\mathrm{p}(\mathbf{y}^o|A),$$

and the posterior odds ratio in favor of model A is

(3.2.2) $$\frac{\mathrm{p}(A|\mathbf{y}^o)}{\mathrm{p}(B|\mathbf{y}^o)} = \frac{\mathrm{p}(A)\,\mathrm{p}(\mathbf{y}^o|A)}{\mathrm{p}(B)\,\mathrm{p}(\mathbf{y}^o|B)}.$$

This approach generalizes immediately to any finite number of models.

The posterior odds ratio directly addresses the question so frequently asked in applied economics, "having looked at the data, what do we now think about alternative theories A and B?" The clarity of (3.2.2) in answering this question reflect the demands of Bayesian econometrics for an explicit representation of beliefs about models ($\mathrm{p}(A)$ and $\mathrm{p}(B)$, and $\mathrm{p}(\theta_A|A)$ and $\mathrm{p}(\theta_B|B)$). Technical work in Bayesian econometrics focuses on two tasks. One is the tractable representation of prior beliefs in densities like $\mathrm{p}(\theta_A|A)$, and the related problem of communicating information about posterior odds ratios to audiences in which prior beliefs may be heterogeneous. The second task is the computation of (3.2.2) and the evaluation of posterior moments of generic form

$$\text{(3.2.3)} \qquad \mathrm{E}\left[\mathrm{g}\left(\mathbf{y}^o, \theta_A, A\right)\middle|\mathbf{y}^o, A\right] = \int_{\Theta_A} \mathrm{g}\left(\mathbf{y}^o, \theta_A\right)\mathrm{p}\left(\theta_A\middle|\mathbf{y}^o, A\right)d\nu\left(\theta_A\right),$$

which generally entails constructing a posterior simulator – an algorithm that produces pseudo-random vectors $\theta_A^{(m)}$ with common density $\mathrm{p}\left(\theta_A\middle|\mathbf{y}^o, A\right)$.

3.3. The failure of DSGE models under the strong econometric interpretation

To appreciate the failure of DSGE models in either classical specification testing or Bayesian model comparison, consider the equity premium models described in Section 2. In the Mehra-Prescott and Rietz variants, the number of states is finite. Given n states, there can be at most n^2 different observable combinations of consumption growth and asset returns. Obviously this property does not characterize the data in any literal way. A trivial classical specification test would reject the model, and the marginal likelihood in any Bayesian formulation would be zero. Less obviously, and more substantively, the observed combinations of the risk free rate and equity premium have a very dispersed support, as illustrated in Figure 3.3.1.

Similar problems with respect to the support of the distribution of observables arise in the Labadie and Tsionas variants of the model. Corresponding to any growth rate there is exactly one riskless return (2.3.1), and to any two successive growth rates, one risky return based on (2.3.2). Thus, for example, these models imply $\min_{(b_0, b_1, b_2)} \sum_{t=1}^{T}\left(r_t - b_0 - b_1 x_t^{b_2}\right)^2 = 0$. This condition is violated in the data. In any classical specification test the model is rejected in a test of any positive size, and in a Bayesian model comparison the marginal likelihood of the model is zero. The observed combinations of growth rates and risk free rates, displayed in Figure 3.3.2, do not even suggest such a relationship.

The restriction of observables to a degenerate space of lower dimension is a well documented failure of most DSGE models. Watson (1993), for example, has illustrated that the reduction in dimension is not even approximately true as a characterization of the data in the one-sector neoclassical model of King, Plosser, and Rebelo (1988). The problem derives from the small number of shocks—often just one, as is the case here—and the larger number of observables. Smith (1993) presents a simple real business cycle model with two shocks and two observables, and employs a formal, likelihood-based approach to make inferences about parameter values. Since observables are not restricted to a space of lower dimension, his model is not trivially rejected under the strong econometric interpretation. The difficulty lies not in the economics of dynamic general equilibrium, but in the fact that the technology of building this kind of model is not generally developed to the point of accommodating a sufficiently large number of shocks in a credible way. A strong econometric interpretation of DSGE models requires an explicit accounting for the dimensions of variation observed in the data, that are not accounted for in the model.

4. Weak econometric interpretation of DSGE models

Most macroeconomists who work with DSGE models eschew the strong econometric interpretation. For example, Mehra and Prescott (1985) in constructing their model of consumption growth, the riskless return, and the equity premium, plainly state that the model is intended to explain the first moments in returns, but not the second moments. That is, the model purports to account for sample average values of the riskless return and the equity premium, but not for the volatility in returns (Mehra and Prescott, 1985, p. 146). Kydland and Prescott (1996, p. 69) also emphasize that the model economy is intended to "mimic the world along a carefully specified set of dimensions."

To begin the process of formalizing this interpretation of DSGE models, let $\mathbf{z} = \mathrm{f}(\mathbf{y})$ denote the dimensions of the model that are intended to mimic the real world. The weak econometric interpretation of a dynamic general equilibrium model is that the model provides a predictive distribution for the functions $\mathbf{z} = \mathrm{f}(\mathbf{y})$ of the observable, finite sequences of quantities and/or prices $\mathbf{y}$. This section argues that this is the interpretation most frequently given to DSGE models by macroeconomists, including Kydland and

Prescott (1996). While many applications of DSGE models resort to *ad hoc* comparison of predictive distributions with observed behavior, careful investigators, including Kydland and Prescott (1996), use the weak econometric interpretation of DSGE models presented here. This section illustrates this interpretation in the context of the equity premium model introduced in Section 2. Finally, this section shows that this implementation of the weak econometric interpretation in fact makes the same assumptions as the strong econometric interpretation. A corollary is that to deny the strong econometric interpretation of a DSGE model, while examining the implications of the weak econometric interpretation in the manner of most of the applied DSGE literature, is internally inconsistent.

4.1. Formalizing the weak econometric interpretation

Let $\mathbf{z} = \mathrm{f}(\mathbf{y})$ denote the dimensions of the model intended to mimic the real world. In the DSGE literature such dimensions are typically sample moments – means, variances, autocorrelations, and the like. Given a complete, probabilistic specification of the model along the lines outlined in Section 3.2, $\mathrm{p}(\mathbf{z}|\theta_A, A)$ is implied by $\mathrm{p}(\mathbf{y}|\theta_A, A)$ and $\mathbf{z} = \mathrm{f}(\mathbf{y})$. Hence there is a predictive density

$$\mathrm{p}(\mathbf{z}|A) = \int_{\Theta_A} \mathrm{p}(\theta_A|A)\,\mathrm{p}(\mathbf{z}|\theta_A, A)\,d\nu(\theta_A)$$

for **z**.

In the DSGE literature, this predictive density is typically investigated by means of simulation, or computational experiments. Often, θ_A is fixed, or a few different values of θ_A are considered to allow for uncertainty about θ_A. These are simply particular forms of the prior density $\mathrm{p}(\theta_A|A)$. Formally, a computational experiment is $\tilde{\theta}_A^{(m)} \sim \mathrm{p}(\theta_A|A)$, $\tilde{\mathbf{y}}^{(m)} \sim \mathrm{p}\left(\mathbf{y}\middle|\tilde{\theta}_A^{(m)}, A\right)$, $\tilde{\mathbf{z}}^{(m)} = \mathrm{f}\left(\tilde{\mathbf{y}}^{(m)}\right)$ $(m = 1, \ldots, M)$. The pseudo-random vectors $\tilde{\mathbf{z}}^{(m)}$ characterize the predictive distribution of the model, and can be compared with the observed value, $\mathbf{z}^o$.

Kydland and Prescott (1996, p. 70) are quite clear about this process:

> If the model has aggregate uncertainty…then the model will imply a process governing the random evolution of the economy. In the case of uncertainty, the computer can generate any number of independent

> realizations of the equilibrium stochastic process, and these relations, along with statistical estimation theory, are then used to measure the sampling distribution of any desired set of statistics of the model economy.

And, again (Kydland and Prescott, 1996, pp. 75-76):

> If the model economy has aggregate uncertainty, first a set of statistics that summarize relevant aspects of the behavior of the actual economy is selected. Then the computational experiment is used to generate many independent realizations of the equilibrium process for the model economy. In this way, the sampling distribution of this set of statistics can be determined to any degree of accuracy for the model economy and compared with the values of the set of statistics for the actual economy. In comparing the sampling distribution of a statistic for the model economy to the value of that statistic for the actual data, it is crucial that the same statistic be computed for the model and the real world. If, for example, the statistic for the real world is for a 50-year period, then the statistic for the model economy must also be for a 50-year period.

What aspects of the predictive density $\mathrm{p}(\mathbf{z}|A)$, as represented by the $\tilde{\mathbf{z}}^{(m)}$, should be compared with $\mathbf{z}^o$? In a classical approach, if $\mathbf{z}$ were pivotal then $\mathrm{p}(\mathbf{z}|A) = \mathrm{p}(\mathbf{z}|\theta_A^*)$ for any $\theta_A^* \in \Theta_A$; the prior distribution for θ_A would be irrelevant, and $\mathbf{z}^o$ would be compared with the relevant tail of $\mathrm{p}(\mathbf{z}|\theta_A^*)$. In fact, however, the dimensions of the real world, $\mathbf{z}$, that the model is intended to describe are typically anything but pivotal. For example, in the equity premium model set forth in Section 2, the sample means of the riskless rate and equity premium that the model is intended to describe are sensitive to the risk aversion parameter α and discount parameter δ.

A formal Bayesian approach conditions on $\mathbf{z}^o$, the observed dimensions of the real world the model is intended to address. Given two competing models, A and B, the posterior odds ratio is then

$$\frac{\mathrm{p}(A|\mathbf{z}^o)}{\mathrm{p}(B|\mathbf{z}^o)} = \frac{\mathrm{p}(A)\,\mathrm{p}(\mathbf{z}^o|A)}{\mathrm{p}(B)\,\mathrm{p}(\mathbf{z}^o|B)}.$$

For purposes of model evaluation and comparison, therefore, it is the predictive density of $\mathbf{z}$ at the observed value $\mathbf{z}^o$ that matters. The key technical problem is the approximation of $\mathrm{p}(\mathbf{z}^o|A)$ on the basis of the computational experiments that generated

$\tilde{\mathbf{z}}^{(m)}$ $(m = 1, \ldots, M)$. Predictive means, $\mathrm{E}(\mathbf{z}|A) \approx M^{-1}\sum_{m=1}^{M}\tilde{\mathbf{z}}^{(m)}$, which are so often presented and compared with $\mathbf{z}^{o}$ in the DSGE macroeconomics literature, are irrelevant, under either a classical or a Bayesian approach.

If the order of the vector $\mathbf{z}$ is small—say, three or less—then numerical approximation of $\mathrm{p}(\mathbf{z}^{o}|A)$ is straightforward, and much simpler than the numerical approximation of the full marginal likelihoods in (3.2.2) under the strong econometric interpretation of the DSGE model. The latter requires backward simulation, for example by means of a Markov chain Monte Carlo algorithm. The former only requires the $\tilde{\mathbf{z}}^{(m)}$ produced through the forward simulation in the familiar computational experiments of the DSGE literature. Conventional smoothing procedures like kernel density methods will provide a numerical approximation to $\mathrm{p}(\mathbf{z}^{o}|A)$:

$$\mathrm{p}(\mathbf{z}^{o}|A) \approx M^{-1}\sum_{m=1}^{M}\mathrm{K}\left(\mathbf{z}^{(m)};\mathbf{z}^{o}\right),$$

where the kernel smoother $\mathrm{K}(\mathbf{z};\mathbf{z}^{o})$ is a nonnegative function of $\mathbf{z}$ that is concentrated near $\mathbf{z}^{o}$ and integrates to one. This computational procedure also provides the numerical approximation

$$\mathrm{E}\left[\mathrm{g}(\theta_A,\mathbf{y}^{o})|\mathbf{z}^{o},A\right] \approx \sum_{m=1}^{M}\mathrm{K}\left(\mathbf{z}^{(m)};\mathbf{z}^{o}\right)\mathrm{g}\left(\theta^{(m)},\mathbf{y}^{o}\right)\Big/\sum_{m=1}^{M}\mathrm{K}\left(\mathbf{z}^{(m)};\mathbf{z}^{o}\right)$$

to any posterior moment. This is also much less demanding that the posterior simulation required to approximate (3.2.3).

As a byproduct, the forward simulation exercise produces the full predictive distribution of $\mathbf{z}$, including points far from the observed value. Careful examination of these points can lead to further insights into the model. We turn now to some examples.

4.2. Weak econometric interpretation of the equity premium model

The Mehra-Prescott and Labadie models completely specify the distribution of growth. The Rietz and mixture models each specify the distribution of growth up to a single, unknown parameter: λ_3, growth in the event of a crash, and $\omega_{(2)}$, the high variance, respectively. For the Rietz model, adopt the prior distribution

$$\log\left[\lambda_3/(1-\lambda_3)\right] \sim \mathrm{N}\left(.036, 1.185^2\right),$$

and for the mixture model use

$$1.33/\omega_{(2)} \sim \chi^2(1.66).$$

Deciles for both distributions are given in Table 4.2.1. The distribution of λ_3 is centered at $\lambda_3 = 0.509$, halving of expected output, which is the intermediate of the three examples taken up in Rietz (1988). The distribution of $\omega_{(2)}$ centers the standard deviation in the high variance state about 1.26, implying that in this state output is about as likely to be between one-third and triple its normal value as it is to be outside this range.

All other parameters in the models pertain to the consumption growth process. In the exercises reported here these parameters were held fixed at their calibrated values, which are chosen to reproduce the mean, standard deviation, and first order autocorrelation of the consumption growth rate. Modifying the analysis by introducing prior distributions for these parameters increases the technical complexity of the exercise, because the dimension of the predictive distribution is increased from two to five, but has little effect on the final results.

None of the models fix the relative risk aversion parameter α or the subjective rate of discount δ. This analysis employs priors that should provide substantial probabilities to the ranges most economists would regard as plausible, and permit some unreasonable values as well. For α, take

$$\log\alpha \sim \mathrm{N}\left(.4055, 1.3077^2\right),$$

and for δ,

$$\log\left[\delta/(1-\delta)\right] \sim \mathrm{N}\left(3.476, 1.418^2\right).$$

Deciles for these prior distributions are also shown in Table 4.2.1. The prior distribution for α is centered at $\alpha = 1.5$, and a centered 80% prior credible interval for α is (0.281, 8.0). The prior distribution for δ is centered at 0.97, and a centered 80% prior credible interval is (0.84, 0.995).

These prior distributions, together with the data densities described in Section 2, provide predictive densities for all four models: Mehra-Prescott, Rietz, Labadie, and mixtures. For each model, draws from predictive densities for output growth rate and asset returns can be made by (1) drawing from the prior distribution for the unknown parameter; (2) conditional on the drawn parameters, generating a sample of 90 successive years of growth rates from the probability density for $\{x_t\}$; (3) solving for the riskless and risky returns in each year as indicated in Section 2. Draws from the predictive density for any function of output growth rate and asset returns, are then just the corresponding functions of this generated, synthetic sample. For example, to draw from the predictive density for 90-year means of the risk free rate and the equity premium, following step (3), just construct these functions and record them. Notice that the predictive density for the mean risk free rate and mean equity premium accounts for both uncertainty about parameter values (by means of the draws from the prior) and sampling variation due to 90-year averaging (be means of the 90-year simulation).

Results of these exercises can be portrayed graphically. Figures 4.2.1–4.2.4 show the predictive distributions for 90-year averages of the risk free return and equity premium, as represented by 2,500 points $\tilde{\mathbf{z}}^{(m)}$ drawn from $\mathrm{p}(\mathbf{z}|A)$ for each model. In each figure, the vertical line indicates the observed value of .008 for the risk free rate and the horizontal line indicates the observed value of .0618 for the equity premium. The supports of the Rietz and mixture model predictive densities include the observed values, but those of the Mehra-Prescott and Labadie model predictive densities do not. This is qualitative corroboration of the failure of the latter two models to explain the equity premium puzzle in the weak econometric DSGE literature, and the ability of the Rietz model to account for the observed means. It extends this corroboration to the mixture model.

Figures 4.2.1–4.2.4 can be used to explore the workings of all four models in some detail, by examining the parameter values or the history of consumption growth underlying the points plotted. This can be done by color coding the points for values of one of the parameters, like the relative risk aversion parameter α, or by software that provides a full display of parameter values and the simulated time series after clicking on a point. For example, high values of the risk free rate correspond to low values of δ. Negative values of the risk free rate and high equity premia in the Rietz and mixture models typically reflect high risk aversion in conjunction with a low probability of very negative growth rates. The values of the risk free rate and equity premium in the Rietz and mixture models close to the historical averages typically correspond to situations in which very negative growth rates were possible but did not occur during the simulated 90-year history.

Table 4.2.2 provides approximations of the log marginal likelihood, $\log\left[p\left(\mathbf{z}^o|A\right)\right]$, of each of the four models under the weak econometric interpretation. A bivariate Gaussian density kernel was centered at the observed sample mean for the riskless return and equity premium, and 25,000 points $\tilde{\mathbf{z}}^{(m)}$ were drawn from the predictive density $p\left(\mathbf{z}|A\right)$. The density kernel was symmetric. Various standard deviations were used as indicated in the left column of Table 4.2.2. The most concentrated kernel (top line) puts almost all of its weight on returns within 0.001 of the observed value. The least concentrated kernel (bottom line) extends weight to returns within 0.02 of the observed value. As one moves down the rows, approximations show greater bias (because they include values of $\tilde{\mathbf{z}}^{(m)}$ farther from the data point) but less variance (because more points are given weight). Asymptotic standard errors are indicated parenthetically. Table 4.2.2 shows that the marginal likelihoods of the Mehra-Prescott and Labadie models are zero. For the more concentrated kernels the Rietz and mixture models have indistinguishable marginal likelihoods. For the less concentrated kernels the mixture model is favored, but the Bayes factor is never more than about 2:1.

4.3. Logical problems with the weak econometric interpretation

As the DSGE literature emphasizes, all models are approximations of reality, and it is important to clarify which aspects of reality a model is intended to mimic. In the strong econometric interpretation of a model, this limited scope is recognized in the choice of the random vector $\mathbf{y}$. If, subsequently, attention is shifted to only a subset of the original variables, there are no conceptual difficulties: one simply works with the marginal distributions of the included variables.

The dimensions of reality addressed by DSGE models entail a reduction of a different kind. For example, the equity premium models are intended to explain sample means of the riskless return and equity premium, but no other aspects of these returns. This is not possible. If the model accounts for $(T+1)$-year averages as well as T-year averages, then the model also has implications for the year-to-year returns. One cannot choose to believe the former but not the latter.

More significantly, the DSGE literature takes the short-run dynamics of these models literally, in establishing the sampling distribution of the set of statistics **z** that summarize the relevant aspects of the behavior of the actual economy. This fact is emphasized in Kydland and Prescott (1996). It is made quite clear in careful calibration studies, for example Gregory and Smith (1991, p. 298) and Christiano and Eichenbaum (1992, pp. 436 and 439). It is inherent in the formal Bayesian implementation of the weak econometric interpretation set forth here. On the other hand, it is ignored in that part of the DSGE literature that reduces econometrics to an informal comparison of two numbers, the observed value and the predictive mean. However, the sampling distribution of the statistic **z** is often a function of profoundly unrealistic aspects of these models, aspects that lie outside the dimensions of reality the models were intended to mimic. For example, in the equity premium models the sampling distribution of average asset returns over the 90-year period are closely related to the variances of these returns, through the usual arithmetic for the standard deviation of a sample mean. In establishing the sampling distribution of these means through repeated simulation of the model, one is taking literally the second moments of returns inherent in the model. These are precisely the dimensions the original model was not intended to capture (Mehra and Prescott, 1985, p. 146), and the models are unrealistic in these dimensions. For example, the sample standard deviation of the equity premium is .164 over the period 1889-1979, whereas at prior median values the standard deviation is .055 in the Mehra-Prescott model and .258 in the Rietz model. The weak econometric interpretation of DSGE models leads to formal methods for model comparison that are easy to implement and have an unambiguous interpretation. As a byproduct, there are some interesting and useful visual displays. But the assumptions that underlie the weak econometric interpretation are in fact the same as those made in the strong econometric interpretation: the model is assumed to account for all aspects of the observed sequence of quantities and/or prices.

5. Minimal econometric interpretation of DSGE models

The logical problems encountered in the claim that DSGE models account for only a few sample moments of observed sequences of quantities and prices prevents the development of this notion into coherent methods of inference about these models. To broaden the claim to assert that DSGE models in fact provide likelihood functions leads to outright dismissal of most of these models (Section 3). This section considers a more modest claim for DSGE models, also studied by DeJong, Ingram and Whiteman (1996): only that they account for population moments of specified, observable functions of sequences of prices and/or quantities. A DSGE model, *A*, with a given parameter vector θ_A, implies population moments $\mathbf{m} = \mathrm{E}\left[\mathbf{z} | \theta_A, A\right]$, where $\mathbf{z} = \mathrm{f}(\mathbf{y})$ is the same vector of sample moments considered under the weak econometric interpretation. If *A* is endowed with a prior distribution $\mathrm{p}(\theta_A | A)$, then *A* provides a distribution for **m** as well. If the mapping from θ_A to **m** is one-to-one, then the DSGE model rationalizes any **m** in terms of behavioral parameters of the model *A*.

By not claiming to predict sample moments, the minimal econometric interpretation avoids the logical pitfall that predicting sample moments and carrying out inference based on the model's predictive distribution leads inevitably back to a conventional likelihood function. The cost entailed in this retreat is that the DSGE model, by itself, now has no implications for anything that might be observed. To endow such a model with empirical content it is necessary to posit, separately, a link between the population moment **m** and the observable sequence of prices and/or quantities **y**. DeJong, Ingram and Whiteman (1996) also noted the need for such a link. This section shows how to do this formally, and provides some examples of the procedure. The result is an integration of atheoretical, data-based macroeconometric models with DSGE models.

5.1. Formal development

Let *A* and *B* denote two alternative DSGE models, each describing the same vector of population moments **m** by means of the respective densities $\mathrm{p}(\mathbf{m} | A)$ and $\mathrm{p}(\mathbf{m} | B)$. The densities could be degenerate at a point but in general are not because of subjective uncertainty about parameter values in both models.

Introduce a third, econometric model *E*, that provides a posterior distribution for the moment vector **m**, given the data $\mathbf{y}^o$: $\mathrm{p}(\mathbf{m} | \mathbf{y}^o, E)$. The moment vector **m** is the same vector of population moments described by *A* and *B*. For example, *A* and *B* might refer to two of the equity premium models introduced in Section 2, $\mathbf{m}$ could be the 2×1 vector of unconditional means for the riskless rate and the equity premium, and E might be a vector autoregression for the riskless rate and equity premium. The posterior distribution $\mathrm{p}\left(\mathbf{m}|\mathbf{y}^{o},E\right)$ could be obtained using conventional posterior simulation methods.

Assumption 1. For any E, $\mathrm{p}\left(\mathbf{y}|\mathbf{m},A,E\right)=\mathrm{p}\left(\mathbf{y}|\mathbf{m},B,E\right)\equiv\mathrm{p}\left(\mathbf{y}|\mathbf{m},E\right)$.

Assumption 1 acknowledges that since A and B claim only to describe $\mathbf{m}$, then if $\mathbf{m}$ is known in the context of E, A and B can have nothing further to say about $\mathbf{y}$. As a consequence:

Result 1. $\mathrm{p}\left(A|\mathbf{m},\mathbf{y},E\right)=\dfrac{\mathrm{p}\left(\mathbf{y}|\mathbf{m},A,E\right)\mathrm{p}\left(A|\mathbf{m},E\right)}{\mathrm{p}\left(\mathbf{y}|\mathbf{m},E\right)}=\mathrm{p}\left(A|\mathbf{m},E\right)$.

Of course, $\mathrm{p}\left(B|\mathbf{m},\mathbf{y},E\right)=\mathrm{p}\left(B|\mathbf{m},E\right)$ as well. Since the models A and B have no implications for $\mathbf{y}$ beyond $\mathbf{m}$, then if we knew $\mathbf{m}$ we could draw conclusions about A and B without even collecting data $\mathbf{y}^{o}$.

In fact, we don't know $\mathbf{m}$. We will use E to link the unknown $\mathbf{m}$ to the observed $\mathbf{y}^{o}$, but it will be convenient if E itself does not take a stance on $\mathbf{m}$.

Assumption 2. $\mathrm{p}\left(\mathbf{m}|E\right)\propto const$, $\mathrm{p}\left(\mathbf{m}|A,E\right)=\mathrm{p}\left(\mathbf{m}|A\right)$, and $\mathrm{p}\left(\mathbf{m}|B,E\right)=\mathrm{p}\left(\mathbf{m}|B\right)$.

Heuristically, E says nothing about $\mathbf{m}$ either absolutely or relative to A and B.

As a consequence of Assumptions 1 and 2, one can draw conclusions about A and B conditional on the data $\mathbf{y}^{o}$, and the model E used to link population moments and data.

Result 2.

$$\mathrm{p}\left(A|\mathbf{y}^{o},E\right)=\int\mathrm{p}\left(A|\mathbf{m},\mathbf{y}^{o},E\right)\mathrm{p}\left(\mathbf{m}|\mathbf{y}^{o},E\right)d\mathbf{m}=\int\mathrm{p}\left(A|\mathbf{m},E\right)\mathrm{p}\left(\mathbf{m}|\mathbf{y}^{o},E\right)d\mathbf{m}$$

$$= \int \frac{p(\mathbf{m}|A,E)\,p(A|E)}{p(\mathbf{m}|E)} \cdot p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m} \propto p(A|E) \int p(\mathbf{m}|A)\, p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m} \ .$$

As a consequence of Result 2,

(5.1.1) $$\frac{p\left(A\middle|\mathbf{y}^o, E\right)}{p\left(B\middle|\mathbf{y}^o, E\right)} = \frac{p(A|E)}{p(B|E)} \cdot \frac{\int p(\mathbf{m}|A)\, p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m}}{\int p(\mathbf{m}|B)\, p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m}}.$$

The posterior odds ratio, on the left, is the product of a prior odds ratio and the Bayes factor

$$\frac{p\left(\mathbf{y}^o|A,E\right)}{p\left(\mathbf{y}^o|B,E\right)} = \frac{\int p(\mathbf{m}|A)\, p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m}}{\int p(\mathbf{m}|B)\, p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m}}.$$

The expression $\int p(\mathbf{m}|A)\, p\left(\mathbf{m}\middle|\mathbf{y}^o, E\right) d\mathbf{m}$ is the convolution of two densities: the density for **m** implied by the DSGE model, *A*, and the posterior density for **m** implied by the econometric model *E* given the data $\mathbf{y}^o$. Loosely speaking, the greater the overlap between these two densities, the greater the Bayes factor in favor of Model *A*. This looser interpretation underlies the confidence interval criterion proposed in DeJong, Ingram and Whiteman (1996), for univariate **m**. The odds ratio (5.1.1) provides an exact interpretation for multivariate **m**. The development here also emphasizes the importance of a flat prior for the moments (Assumption 2), as opposed to simply a convenient diffuse prior for **m**.

Models *A* and *B* can be compared on the basis of three simulations of the moment vector **m**: one drawn from *A*, one from *B*, and one from the posterior distribution in *E*. Informal comparison can be based on a visual inspection of the clouds of points from these three models. A more formal comparison can be made by means of the kernel density approximation,

(5.1.2) $$p\left(A\middle|\mathbf{y}^o, E\right) \dot{\propto} p(A|E)(MN)^{-2} \sum_{r=1}^{M} \sum_{s=1}^{N} \mathrm{K}\left(\mathbf{m}_A^{(r)}, \mathbf{m}_E^{(s)}\right).$$

It is important to keep in mind that the comparison between DSGE models under the minimal econometric interpretation is always contingent on the choice of the common econometric model *E* that provides the link to reality. This is inescapable. It is therefore prudent to be explicit about the chosen model E, and to explore the sensitivity of results to the choice of E.

5.2. Minimal econometric interpretation of the equity premium model

In the equity premium example the vector **m** consists of the population means for the risk free rate and the equity premium. Here we use perhaps the simplest econometric model E with implications for **m**: a first-order Gaussian bivariate autoregression for the risk free rate at equity premium, with stationarity imposed:

$$\mathbf{y}_t - \mathbf{m} = \mathbf{F}(\mathbf{y}_{t-1} - \mathbf{m}) + \varepsilon_t \quad (t = 1879, \ldots .1978)$$

where the 2×1 vector $\mathbf{y}_t$ consists of the observed risk free rate and equity premium in the indicated year, and $\varepsilon_t \overset{IID}{\sim} \mathrm{N}(\mathbf{0}, \Sigma)$.

Draws of **m** from the posterior distribution were obtained using a Metropolis within Gibbs posterior simulation algorithm. An improper prior for $(\beta, \mathbf{F}, \Sigma)$, flat subject to the stationarity condition on **F**, was employed. This prior satisfies Assumption 2 above. Some posterior moments for the parameters are indicated in Table 5.2.1. There is modest autocorrelation in the riskless rate (about .4), less in the equity premium (about .2), and very little cross correlation between the two time series. The innovation variance in the equity premium exceeds that of the risk free rate by a factor of more than 10. The implied standard deviation for the equity premium is over .16, and that for the riskless rate is over .05.

The posterior distribution of **m** is presented graphically in Figure 5.2.1. The range of values well within the support of the posterior distribution extends far beyond the observed sample means. A centered 90% posterior credible interval for the mean of the risk free rate extends from -0.9% to 2.6%. For the equity premium the range is much larger: from 2.2% to 9.7%. Even with 90 years of data, there is great uncertainty about the population mean of the equity premium. This uncertainty is due to the great variance in the equity premium from year to year. It is not due to drift: there has been no tendency for the equity premium to rise or fall secularly (Mehra and Prescott, 1985, Table 1), and the autocorrelation in the simple model used here is only .4.

Figure 5.2.1 provides $\mathrm{p}\left(\mathbf{m}\middle|\mathbf{y}^{o},E\right)$. Figures 5.2.2–5.2.5 provide $\mathrm{p}(\mathbf{m}|A)$ for each of the four models A. For each model, parameters were drawn from the same prior distributions used in the weak econometric interpretation. (These priors are summarized in Table 4.2.1.) Then, the corresponding population moments were computed. For the Mehra-Prescott and Rietz models there are closed form expressions for these moments. For the Labadie and mixture models, a simulation of 1,000 periods was made corresponding to each set of parameter values drawn from the prior. Then a second antithetic simulation (i.e., shocks with signs reversed) was made. Then the mean of the risk free rate and equity premium averaged over the two 1,000-period simulations was used in lieu of the population mean.

Comparisons of Figures 5.2.2–5.2.5, with their respective counterparts in Figures 4.2.1–4.2.4, reveal similar patterns. But the distributions in the earlier figures are more diffuse, relative to those in the latter figures which are more neatly demarcated and somewhat more compact. The difference reflects the sampling variation in 90-year averages, which is present in Figures 4.2.1–4.2.4 but not Figures 5.2.2–5.2.5.

In the minimal econometric interpretation, a model receives support to the extent that the posterior density $\mathrm{p}\left(\mathbf{m}\middle|\mathbf{y}^{o},E\right)$, presented in Figure 5.2.1, overlaps with the model predictive density $\mathrm{p}(\mathbf{m}|A)$, presented in one of Figures 5.2.2–5.2.5, in a manner that is made explicit in equation (5.1.1). Because the posterior density $\mathrm{p}\left(\mathbf{m}\middle|\mathbf{y}^{o},E\right)$ is so diffuse, there is overlap between this density and each of the model densities, as careful inspection of Figures 5.2.1–5.2.5 will indicate. This is true even of the Mehra-Prescott and Labadie models, which received no support under the weak econometric interpretation. The reason is that the weak econometric interpretation takes literally the very small variance in the riskless return and equity premium implied by these two models. Given this small sampling variation the observed averages cannot be accounted for in these models. Given the much greater sampling variation that emerges in the bivariate autoregression used here, it becomes clear that the evidence in the historical record to be explained is much weaker. This finding underscores the point made forcefully by Eichenbaum (1991, p. 611) that assuming the population moment is equal to the sample moment can be treacherous. “Data” and “facts” are not the same.

Formal approximation using (5.1.2) underscores these informal findings. Using 1,000 draws of the posterior mean (every tenth draw from the Gibbs sampler, after discarding the first 1,000 draws) and 5,000 independent draws of population moments from each of the models, the approximations presented in Table 5.2.2 were obtained. The alternative Gaussian kernels employed are the same as those for Table 4.2.2. The marginal likelihood for the mixture model is higher than that for the Labadie model (Bayes factor about 5:1), and that for the Labadie model is higher than for either the Rietz or Mehra-Prescott models. For the Labadie and mixture models log marginal likelihoods are stable across kernels of different bandwidths. For the Mehra-Prescott and Rietz models, log marginal likelihoods increase with bandwidth, reflecting the more tangential relation between $\mathrm{p}(\mathbf{m}|A)$ and $\mathrm{p}(\mathbf{m}|\mathbf{y}^o, E)$ in these models.

6. Summary and conclusion

This study has examined three ways in which computational experiments can be used to see how well dynamic, stochastic general equilibrium (DSGE) models explain observed behavior. Since these models imply distributions for the paths of prices and quantities, a straightforward, likelihood based approach—termed the strong econometric interpretation in this study—is perhaps the most obvious. It is widely recognized that most DSGE models fail under this interpretation because they predict exact relations that are not found in the data.

The widespread interpretation of DSGE models in the macroeconomics literature is that they are intended only to mimic the world along a carefully specified set of dimensions. This interpretation is sometimes reduced to a list of sample moments, on the one hand, and a list of corresponding moments of the model's predictive distribution, on the other. Careful investigators recognize that some basis for comparison of these two sets of moments is needed. Kydland and Prescott (1996) clearly indicate that what is at stake is whether the sample moments are consistent with the predictive distribution of the model for those moments. This inherently Bayesian approach—termed the weak econometric interpretation in this study—takes the period-to-period dynamics of the models literally. While it confines itself to just a few dimensions of the data, in accounting for sampling variation it makes the same assumptions as does the strong econometric interpretation. It is therefore subject to the same criticism, that those assumptions are inconsistent with what is observed.

To isolate the idea that DSGE models explain only certain dimensions, in a way that does not run afoul of the literal incredibility of these models, this study examined the implications of the claim that DSGE models predict only certain specified population moments of observable data. Since population moments are never observed, an auxiliary econometric model must also be assumed if the DSGE model is to have any refutable implications. The study showed that atheoretical econometric models with a Bayesian interpretation can perform this function. Under this set of assumptions—termed the minimal econometric interpretation in this study—formal Bayesian model comparison is possible, and is free of the logical problems associated with the weak econometric interpretation.

These ideas were illustrated using the "equity premium puzzle'" models of Mehra and Prescott (1985), Rietz (1988), Labadie (1989), and Tsionas (1994). The weak econometric interpretation reiterated both the inability of the Mehra-Prescott and Labadie models to account for the sample average risk free rate and equity premium in the U.S., and the ability of the Rietz model and the mixture models developed by Tsionas to do so. This reflects the fact that it is the weak econometric interpretation that is dominant in the DSGE literature of macroeconomics. This application provided a rich graphical interpretation of these models as well as Bayes factors for the comparison of models.

The minimal econometric interpretation of the same models greatly changed the nature of the findings, and underscores that point that the methodological issues raised in this study have substantive implications for macroeconomics. The most important finding was that we in fact have limited information about the population mean of the equity premium, because year-to-year fluctuations have been so great. The posterior distribution for the mean of the risk free rate and the equity premium supports values consistent with the original Mehra-Prescott model, the other models considered in this study, and quite likely with all other DSGE models designed to address this question. In short, there is no evidence of an equity premium puzzle.

These substantive findings illustrate the dangers of informal or undisciplined readings of the evidence in macroeconomic time series. To push ahead with methods based on assumptions known to be incompatible with what is observed, or to ignore uncertainty that is manifest in the record, can lead large and well-intentioned groups of investigators astray. The benefits of an analytically rigorous economic theory will be realized only when harnessed to the same high standards for measurement.

References

Christiano, L.J., and M. Eichenbaum, 1992, "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review* **82**, 430-450.

DeJong, D.N., B.F. Ingram, and C.H. Whiteman, 1996, "A Bayesian Approach to Calibration," *Journal of Business and Economic Statistics* **14**: 1-10.

Eichenbaum, M., 1991, "Real Business Cycle Theory: Wisdom or Whimsy?" *Journal of Economic Dynamics and Control* **15**: 607-626.

Gregory, A.W. and G.W. Smith, 1991, "Calibration as Testing: Inference in Simulated Macroeconomic Models," *Journal of Business and Economic Statistics* **9**: 297-304.

King, R. G., C.I. Plosser, and S. T. Rebelo, 1988, "Production, Growth and Business Cycles: II. New Directions," *Journal of Monetary Economics* **21**: 309-341.

King, R. G., C.I. Plosser, and S. T. Rebelo, 1990, "Production, Growth and Business Cycles: Technical Appendix," University of Rochester mimeo.

Kocherlakota, N., 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature* **34**: 42-71.

Kydland, F.E., and E.C. Prescott, 1996, "The Computational Experiment: An Econometric Tool," *Journal of Economic Perspectives* **10**:1, 69-86.

Labadie, P., 1989, "Stochastic Inflation and the Equity Premium," *Journal of Monetary Economics* **24**: 277-298.

Mehra, R., and E.C. Prescott, 1985, "The Equity Premium: A Puzzle," *Journal Monetary Economics* **15**: 145–161.

Rietz, T.A., 1988, "The Equity Premium: A Solution," *Journal of Monetary Economics* **22**: 17–33.

Smith, A.A., 1993, "Estimating Nonlinear Time-series Models using Simulated Vector Autoregressions," *Journal of Applied Econometrics* **8**: S63-S84.

Tsionas, E.G., 1994, "Asset Returns in General Equilibrium with Scale-Mixture-of-Normals Endowment Processes," University of Minnesota unpublished Ph.D. dissertation.

Watson, M.W., 1993, "Measures of Fit for Calibrated Models," *Journal of Political Economy* **101**: 1-11-1041.

Table 4.2.1

Deciles of prior distributions

	λ_3	$\omega_{(2)}$	α	δ
0.1	.185	.332	.281	.840
0.2	.277	.491	.499	.907
0.3	.358	.677	.756	.939
0.4	.434	.917	1.077	.958
0.5	.509	1.26	1.500	.970
0.6	.583	1.77	2.089	.979
0.7	.659	2.68	2.978	.986
0.8	.738	4.61	4.509	.991
0.9	.826	11.12	8.016	.995

Table 4.2.2

Weak econometric interpretation

Log marginal likelihoods

Gaussian kernel smoothing standard deviation	Mehra-Prescott model	Rietz model	Labadie model	Mixture model
.0005	$-\infty$	2.60 (.99)	$-\infty$	2.46 (.78)
.0010	-406	2.09 (.64)	-575	2.40 (.43)
.0020	-101	2.22 (.28)	-143	2.83 (.23)
.0050	-17.6 (1.0)	2.11 (.13)	-21.9 (.4)	2.88 (.09)
.0100	-3.01 (.03)	2.16 (.06)	-2.95 (.04)	3.04 (.04)

Table 5.2.1

Bivariate first order autoregression
Risk-free rate (r_t) and equity premium (e_t)

$$\begin{pmatrix} r_t - m_1 \\ e_t - m_2 \end{pmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{pmatrix} r_{t-1} - m_1 \\ e_{t-1} - m_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}; \quad \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right\}$$

Parameter	Posterior mean	Posterior standard deviation
m_1	.0088	.0106
m_2	.0591	.0227
f_{11}	.4362	.0907
f_{12}	-.0972	.0303
f_{21}	-.0065	.3143
f_{22}	.2003	.1077
σ_{11}	.0022	.0003
σ_{22}	.0268	.0041
σ_{12}	-.0009	.0008

Table 5.2.2

Minimal econometric interpretation

Log marginal likelihoods

Gaussian kernel smoothing standard deviation	Mehra-Prescott model	Rietz model	Labadie model	Mixture model
.0005	.53 (.14)	.93 (.16)	2.11 (.12)	3.62 (.06)
.0010	1.28 (.06)	.91 (.08)	2.14 (.06)	3.86 (.02)
.0020	1.61 (.03)	.99 (.04)	2.24 (.03)	4.05 (.01)
.0050	1.89 (.01)	2.066 (.008)	2.22 (.01)	4.141 (.004)
.0100	2.321 (.003)	1.975 (.005)	2.514 (.005)	4.205 (.002)

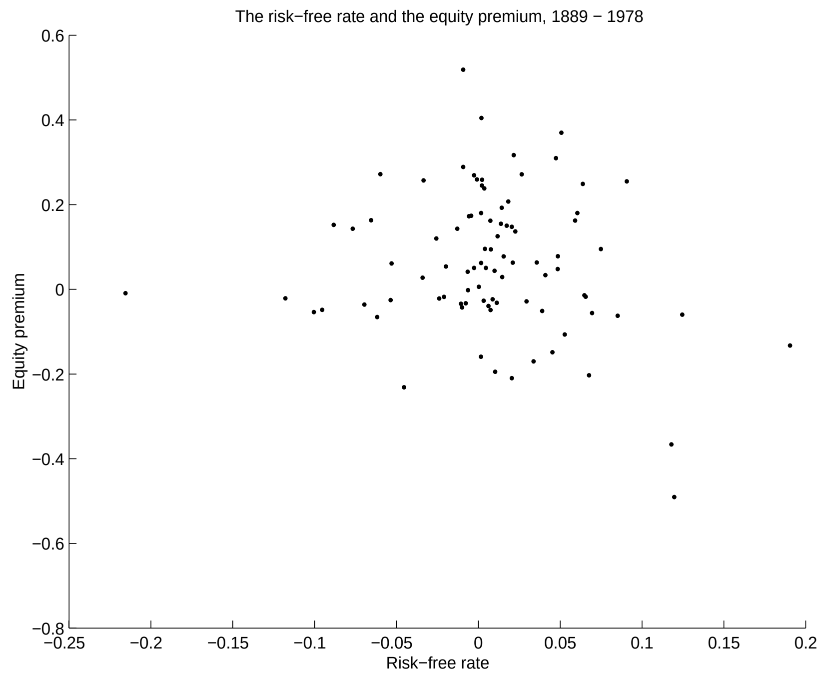

Figure 3.3.1 Annual combinations of the risk free rate and equity premium, 1889-1978.

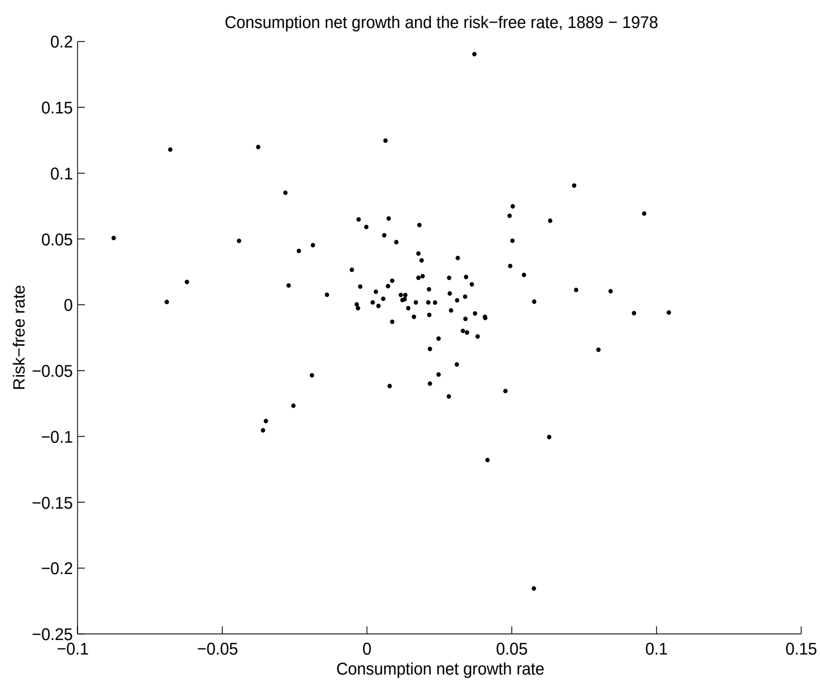

Figure 3.3.2 Annual combinations of the net growth rate in consumption and the risk free rate, 1889-1978.

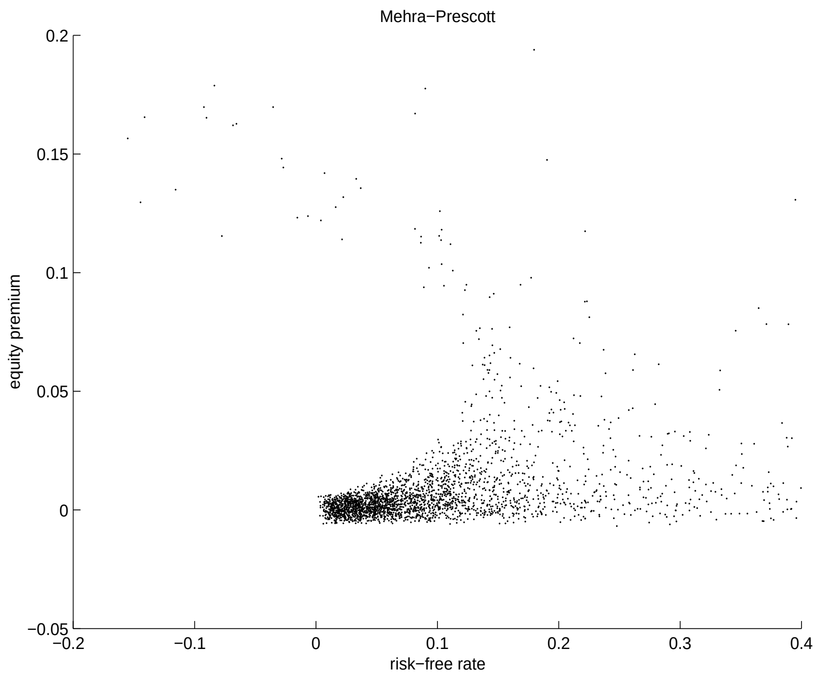

Figure 4.2.1 Predictive distribution for the risk free rate and the equity premium under the weak econometric interpretation in the Mehra-Prescott model.

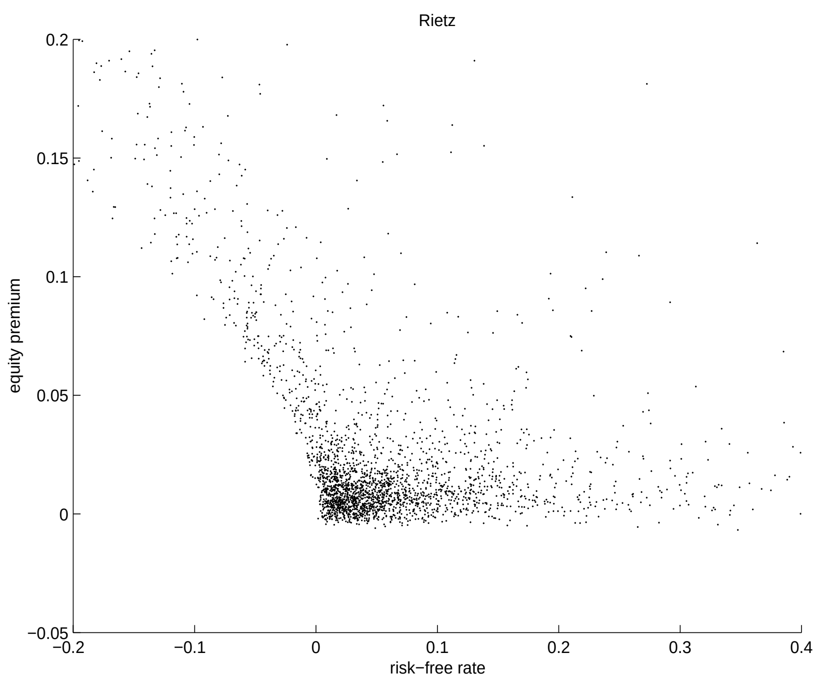

Figure 4.2.2 Predictive distribution for the risk free rate and the equity premium under the weak econometric interpretation in the Rietz model.

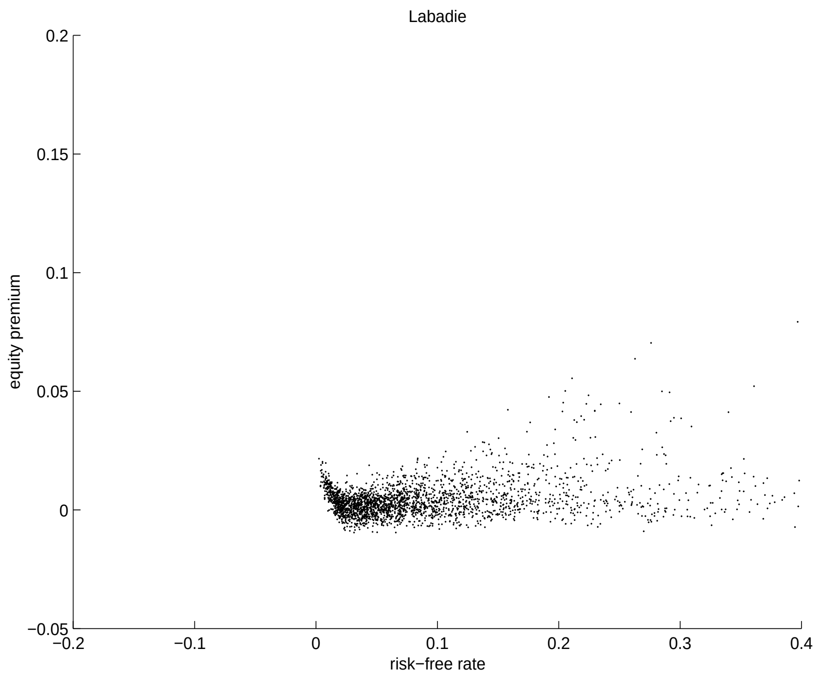

Figure 4.2.3 Predictive distribution for the risk free rate and the equity premium under the weak econometric interpretation in the Labadie model.

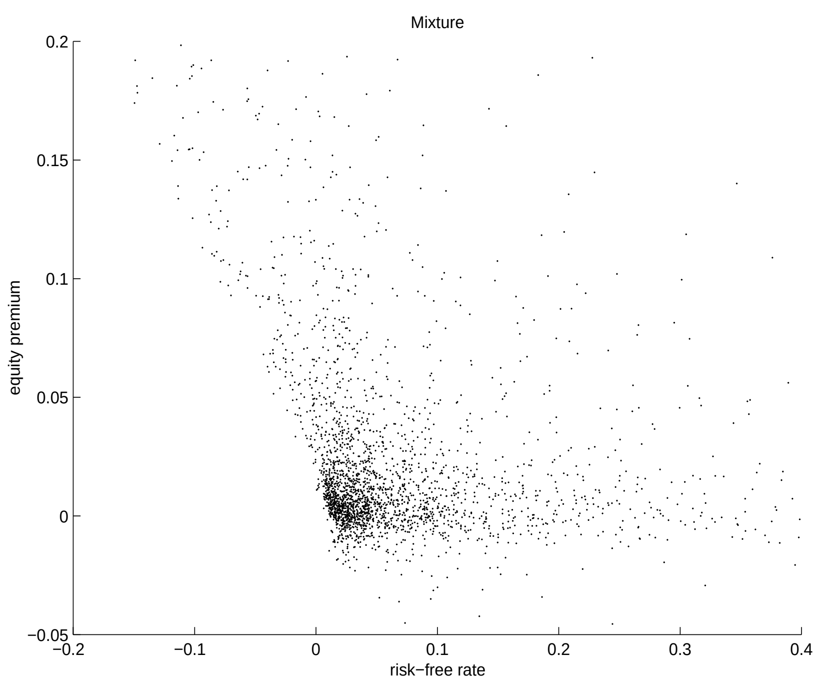

Figure 4.2.4 Predictive distribution for the risk free rate and the equity premium under the weak econometric interpretation in the Mixture model.

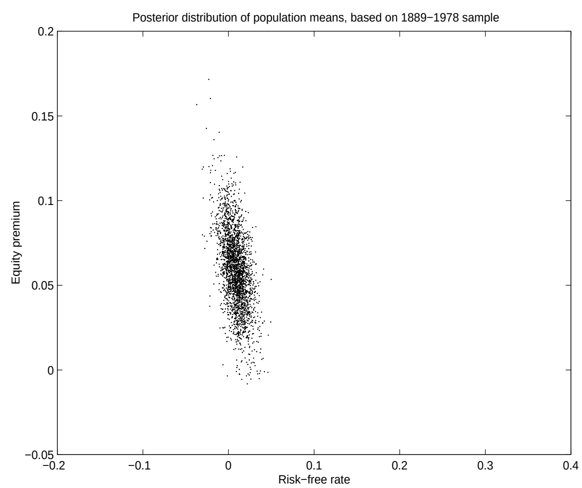

Figure 5.2.1 Posterior distribution of the population mean values of the risk free rate and equity premium, based on 1889-1978 annual data and a bivariate, first-order autoregression.

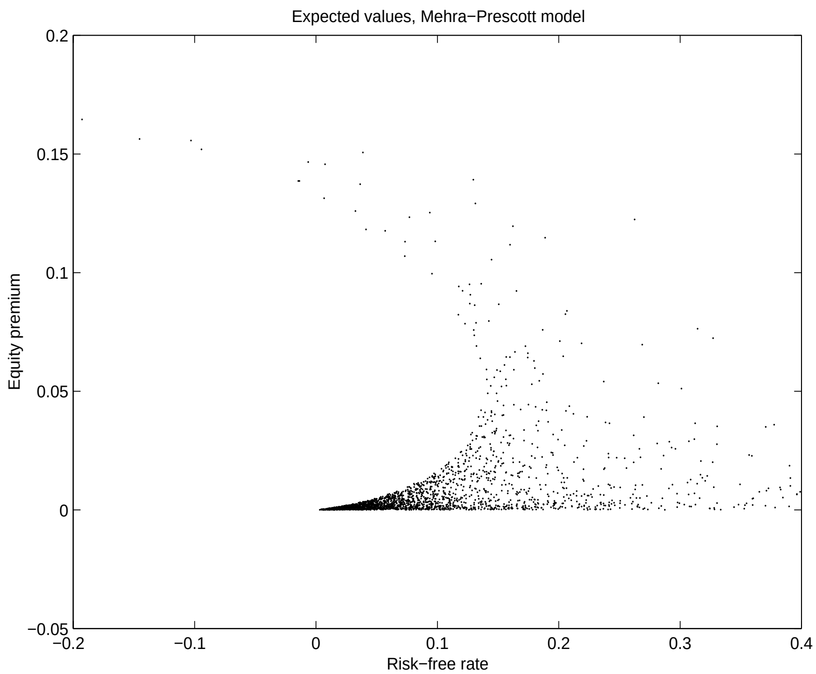

Figure 5.2.2 Predictive distribution for the expectation of the risk free rate and the equity premium under the minimal econometric interpretation of the Mehra-Prescott model.

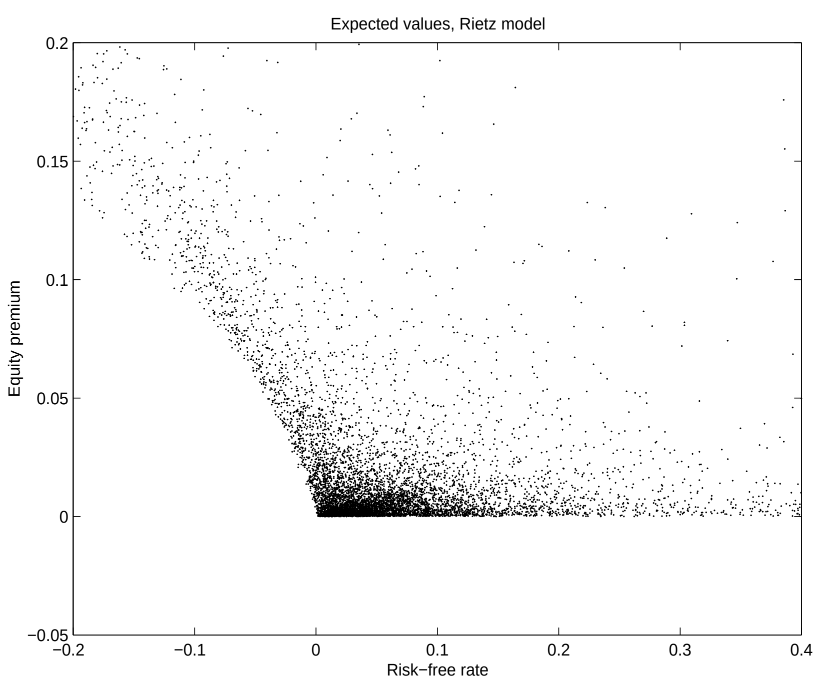

Figure 5.2.3 Predictive distribution for the expectation of the risk free rate and the equity premium under the minimal econometric interpretation of the Rietz model.

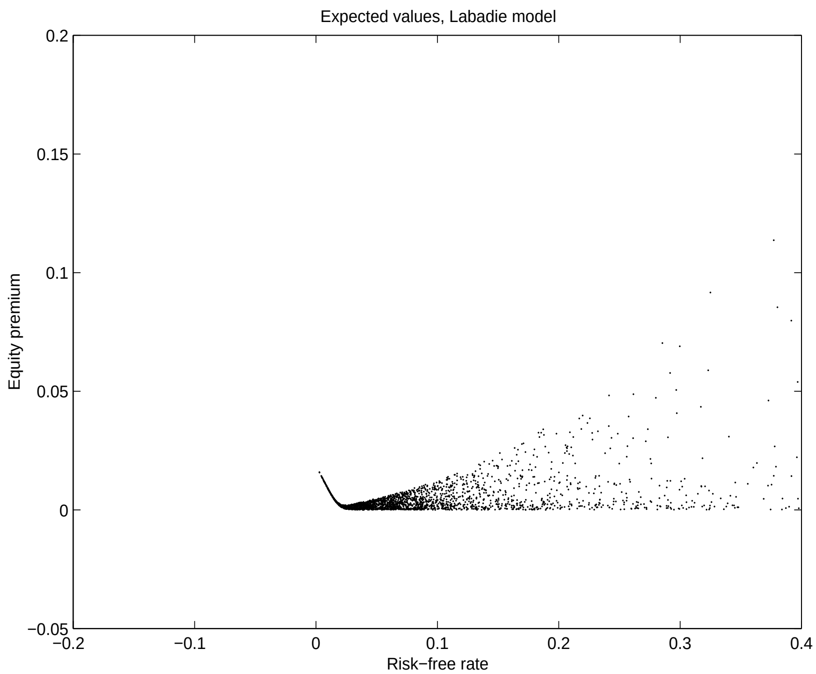

Figure 5.2.4 Predictive distribution for the expectation of the risk free rate and the equity premium under the minimal econometric interpretation of the Labadie model.

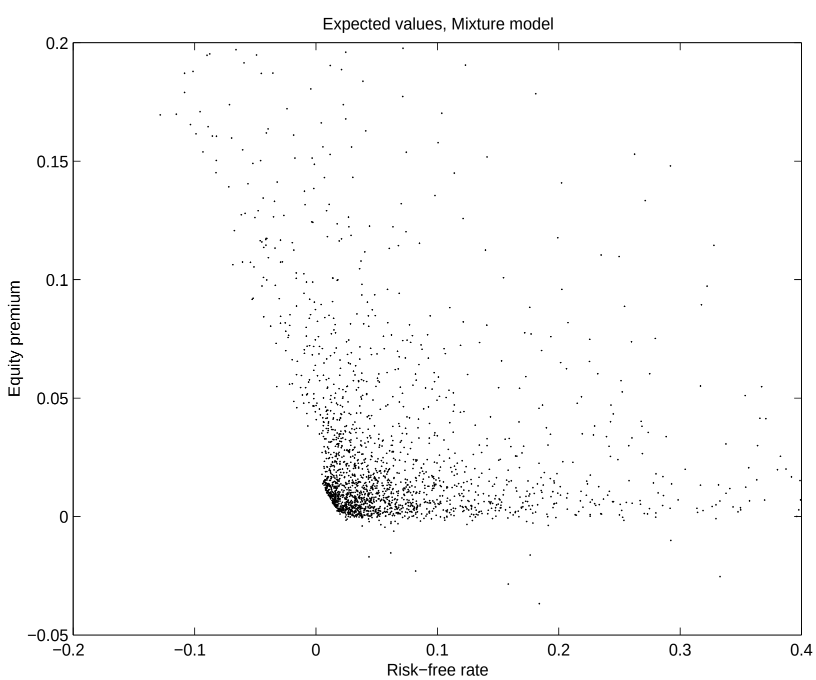

Figure 5.2.5 Predictive distribution for the expectation of the risk free rate and the equity premium under the minimal econometric interpretation of the Mixture model.