Estimating Nonlinear DSGE Models by the Simulated Method of Moments*

Francisco J. Ruge-Murcia†

First draft: March 2010
Revised: November 2010

Abstract

This paper studies the application of the simulated method of moments (SMM) for the estimation of nonlinear dynamic stochastic general equilibrium (DSGE) models. Monte Carlo analysis is employed to examine the small-sample properties of SMM in specifications with different curvature. Results show that SMM is computationally efficient and delivers accurate estimates, even when the simulated series are relatively short. However, asymptotic standard errors tend to overstate the actual variability of the estimates and, consequently, statistical inference is conservative. A simple strategy to incorporate priors in a method of moments context is proposed. An empirical application to the macroeconomic effects of rare events indicates that negatively skewed productivity shocks induce agents to accumulate additional capital and can endogenously generate asymmetric business cycles.

JEL Classification: C15, C11, E2
Key Words: Monte-Carlo analysis; priors; perturbation methods, rare events, skewness.

*I received helpful comments and suggestions from Silvia Gonçalves, Jinill Kim, Bill McCausland, and Serena Ng. The financial support of the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

†Département de sciences économiques and CIREQ, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada. E-mail: francisco.ruge-murcia@umontreal.ca. Codes and future versions of this paper are available at my Web page (*www.cireq. umontreal.ca/personnel/ruge.html*).

1 Introduction

The econometric analysis of nonlinear dynamic stochastic general equilibrium (DSGE) models opens exciting possibilities in macroeconomics and finance because they relax the certainty equivalence and linear propagation that characterize linear models (that is, models solved using first-order approximations alone). Allowing uncertainty to effect economic choices is essential to study, among others, asset prices and optimal policy, and nonlinear dynamics are necessary to capture asymmetries, time-varying volatility, and other nonlinear features of the data.[1] As in the case of linear models, estimation means that parameters are obtained by imposing on the data the restrictions of the model of interest,[2] that parameter uncertainty may be explicitly incorporated in impulse-response analysis, and that statistical inference may be used for hypothesis testing and model selection.

This paper studies the application of the simulated method of moments (SMM) to the estimation of nonlinear DSGE models. Two theoretical results suggest that SMM is an attractive strategy to follow. First, Duffie and Singleton (1993) show that under general regularity conditions the SMM estimator is consistent and asymptotically normal. Of course, other estimators, for example maximum likelihood (ML), have these desirable properties and the difference between them is, therefore, one of statistical efficiency and computational ease.[3] Second, Santos and Peralta-Alva (2006) show that the moments computed using simulated observations from an approximate solution to the model converge to those of the invariant distribution of the model as the approximation error tends to zero. This continuity property is important because finding the exact solution of DGSE models is usually not possible and so the best a researcher can do is to derive and estimate an approximate solution.

More specifically, this paper is concerned with the small-sample properties and computational efficiency of SMM. Both issues are of practical importance. Small-sample properties are important because the time series available to estimate DSGE models are relatively short and there may well be discrepancies between the asymptotic and finite-sample distributions of the estimates. In turn, these discrepancies may have implications for statistical inference. In order to study this issue, I follow the usual Monte-Carlo approach standard in econometrics and consider various configurations of a simple dynamic economy based on Brock and Mirman (1972). Computational efficiency is also important because a quick evaluation of the

[1]For example, Kim and Ruge-Murcia (2009) estimate a nonlinear DSGE model where nominal wages are downwardly rigid and find that productivity shocks induce asymmetric effects on the rates of price and wage inflation, as was conjectured by Tobin (1972).

[2]This addresses the concern that the assumptions of the micro studies that generate calibration parameters might be inconsistent with those of the model.

[3]For a discussion of the issues concerning the ML estimation of nonlinear DSGE models, see Fernández-Villaverde and Rubio-Ramirez (2007).

statistical objective function permits the use of genetic algorithms for its optimization and of numerically-intensive methods, like the block bootstrap, for the construction of accurate confidence intervals.

This paper also proposes a simple strategy to incorporate prior information in a method of moments framework. The strategy is inspired by the mixed estimation approach due to Theil and Goldberger (1961), where priors are treated as additional observations and combined with the data to deliver a penalized statistical objective function. In the same spirit, I treat priors as additional moments and write a penalized objective function where the penalty increases as the parameters deviate from the priors. The small-sample properties of this quasi-Bayesian SMM estimate are also studied using Monte-Carlo analysis.

Monte-Carlo results show that SMM delivers accurate parameter estimates, even when the simulated series are relatively short. However, asymptotic standard errors tend to overstate the true variability of the estimates and, consequently, statistical inference is conservative. The computational cost of SMM increases approximately linearly with the length of the simulated series used to calculate the moments implied by the model but, overall, the procedure is computationally efficient because the evaluation of the statistical objective function is cheap. For example, the average time required to estimate the growth model varies between 30 and 95 seconds, depending on whether the simulated series are five or twenty times larger than the sample size.

In addition, this paper contributes to the literature on the solution to dynamic general equilibrium models by showing how to derive third-order polynomial approximations to the policy functions, studying the economic implications of the third-order terms, and making available MATLAB codes to implement this approximation. In particular, it is shown that third-order approximate solutions allow the skewness of the shocks to affect agents' choices. This idea is explored empirically by estimating the model of an economy subject to asymmetric productivity shocks. Rietz (1988) and Barro (2006, 2009) argue that extreme, low probability events—presumably arising from an asymmetric distribution—have substantial implications for asset pricing. The application here focuses on the consequences of rare events for consumption, capital accumulation, and labor supply. This application also illustrates the use of SMM for the estimation of a nonlinear dynamic model using actual data and shows that it can easily accommodate non-Normal distributions. Results show that in an economy where productivity shocks are negatively skewed, agents accumulate more capital than those in an economy where shocks are symmetric. In turn, the larger capital stock finances more leisure and consumption in the former compared with the latter. In addition, higher-order moments are closer to those in the data, and impulse-responses to equally-likely positive and negative productivity shocks are asymmetric.

The rest of the paper is organized as follows. Section 2 presents the data generating process (DGP) and discusses the implications of third-order approximate solutions to dynamic general equilibrium models. Section 3 describes the simulated method of moments, proposes a simple strategy to incorporate prior information, and compares SMM with the generalized method of moments. Section 4 outlines the Monte-Carlo design and reports the results of various experiments involving different model curvature, weighting matrices, and degrees of approximation. Section 5 uses actual U.S. data to estimate a simple nonlinear DSGE model with asymmetrically-distributed productivity innovations. Finally, Section 6 concludes. Codes and replication material for this paper are made separately available in the author's Web page.

2 The Data Generating Process

This section describes the data generating process (DGP) used in the analysis. The DGP consists of a simple dynamic stochastic general equilibrium model and its numerical solution. The model is a version of the stochastic growth model (Brock and Mirman, 1972) augmented to incorporate inelastic labor supply and habit formation in consumption. The model solution is obtained by means of a perturbation method.

2.1 Economic Model

Consider a benevolent central planner that maximizes

$$E_s \sum_{t=s}^{\infty} \beta^{t-s} \left(\frac{(c_t - ac_{t-1})^{1-\gamma}}{1-\gamma} + b(1-n_t) \right), \tag{1}$$

where E_s denotes the expectation conditional on information available at time s, $\beta \in (0,1)$ is the discount factor, c_t is consumption, n_t is hours worked, γ and b are strictly positive preference parameters, $a \in [0,1)$ represents the importance of habit formation, and the time endowment has been normalized to one. In the case where $a = 0$ utility is time separable in consumption, while in the case where $a > 0$ consumptions in two consecutive periods are complements. The linear representation of the disutility of labor is based on the indivisible-labor model due to Hansen (1985). The population size is constant and normalized to one.

The only good in this economy is produced using the technology $z_t k_t^{\alpha} n_t^{1-\alpha}$, where $\alpha \in (0,1)$ is a constant parameter, k_t is the capital stock, and z_t is an exogenous productivity shock. The central planner is subject to the resource constraint

$$c_t + k_{t+1} = z_t k_t^{\alpha} n_t^{1-\alpha} + (1-\delta)k_t, \tag{2}$$

where $\delta \in (0, 1]$ is the rate of depreciation. The productivity shock follows the process

$$\ln(z_t) = \rho \ln(z_{t-1}) + \epsilon_t, \tag{3}$$

where $\rho \in (-1, 1)$, ϵ_t is an innovation assumed to be identically and independently distributed (i.i.d.) with mean zero, standard deviation equal to σ, and skewness equal to s.

In addition to the transversality condition, the first-order necessary conditions associated with the optimal choice of c_t and n_t are

$$\lambda_t = \beta E_t \left(\lambda_{t+1} (1 + \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha} - \delta) \right), \tag{4}$$

$$b/\lambda_t = (1-\alpha) z_t k_t^{\alpha} n_t^{-\alpha}, \tag{5}$$

where $\lambda_t = (c_t - ac_{t-1})^{-\gamma} - a\beta E_t (c_{t+1} - ac_t)^{-\gamma}$ is the marginal utility of consumption. Condition (4) is the Euler equation of consumption that equates the marginal benefit of consuming an extra unit of good with the marginal benefit of saving it in the form of capital. Condition (5) equates the marginal rate of substitution of labor and consumption with the marginal productivity of labor.

2.2 Solution Method

As it is well known, this model admits an exact solution only under stringent assumptions.[4] More generally, the solution of dynamic general equilibrium models must be found numerically and necessarily involves some degree of approximation. In this paper, I employ a perturbation method that approximates the planner's decision rules by means of a polynomial in the state variables and characterizes the local dynamics around the deterministic steady state. Higher-order polynomial approximations capture the nonlinear relation between choice and state variables, and relax the certainty equivalence implicit in first-order solution methods. In particular, I use here second- and third-order polynomial approximations to the policy rules. For a general explanation of this approach and solvability conditions, see Jin and Judd (2002).

Second-order polynomial approximations are studied in detail by Schmitt-Grohé and Uribe (2004), Kim, Kim, Schaumburg, and Sims (2008), and Lombardo (2010), who also produce codes for the implementation of their respective approaches.[5] Compared with a linear approximation, a second-order approximation includes quadratic terms in the state

[4] Assumptions include, for example, complete depreciation and logarithmic consumption preferences.

[5] In this project, I use MATLAB codes adapted from those originally written by Stephanie Schmitt-Grohé and Martin Uribe for a second-order approximation to the policy rules. Their codes are also used to produce the first- and second-order terms used as input for my third-order approximation.

variables, including all possible cross products, and a risk-adjustment term that is proportional to the variance of the shock innovations.

For the third-order polynomial approximation, I follow the recursive approach in Jin and Judd (2002). That is, I take as input the previously-computed coefficients of the first- and second-order terms of the policy rule and a construct linear system of equations that can be solved to deliver the coefficients of the third-order terms. Provided that the non-singularity condition in Theorem 3 of Jin and Judd (2002, p. 19) is satisfied, the solution is unique. Appendix A shows the analytical steps necessary to derive the third-order coefficients and describes MATLAB codes that can be used to compute them.

A third-order approximation to the policy rules includes, in addition to linear and quadratic terms, four sets of terms (see Appendix A). First, cubic terms in the state variables, including all possible cross products. Second, cubic terms that involve cross products between standard deviations and squares of the state variables. Appendix A shows, however, that the coefficients of all these terms are zero. A comparable result for second-order approximations is reported by Schmitt-Grohé and Uribe (2004, p. 763) for the coefficients of cross products between standard deviations and levels of the state variables. Third, cubic terms that involve cross products between variances and levels of the state variables. Finally, terms proportional to the third moments of the shock innovations. Appendix A shows that in the special case where the distribution of the innovations is symmetric—and, hence, the skewness is zero—these terms are zero. However, in the more general case where the distribution is asymmetric, these terms may be positive or negative depending on the skewness of the innovations and the values of other structural parameters.

In order to develop the reader's intuition regarding the nonlinearity of the model, Figure 1 plots the policy functions of next-period capital, consumption and hours worked as a function of the current capital stock and productivity shock. The policy functions are obtained using first-, second- and third-order polynomial approximations. The parameters used to construct this figure are $\alpha = 0.36$, $\beta = 0.95$, $\delta = 0.025$, $\gamma = 2$, $a = 0.9$, $\rho = 0.85$, $\sigma = 0.1$, and $s = -2$. (In order to put the skewness in perspective, recall that a chi-squared distribution with 2 degrees of freedom has a skewness of +2.) The weight of leisure in the utility function, b, is set so that the proportion of time spent working in the deterministic steady state is one-third. In the figure, vertical axes, and horizontal axes in the upper panels, are percentage deviations from the deterministic steady state, while horizontal axes in the lower panels are the standard deviation of the productivity shock.

Notice that under the higher-order approximations, consumption is generally lower, and hours and next-period capital generally higher, than under the first-order approximation. The reason is that higher-order approximations allow uncertainty to affect economic choices and so a prudent agent consumes less, saves more, and works more than an agent in a certainty-equivalent world. In the case of the third-order approximation, the skewness of the shock also has a level effect on the policy rules. There is no appreciable difference between the second- and third-order approximations to the decision rules of next-period capital, but in the case of consumption and hours the difference can be large.

Finally, notice that (by construction) the relation between the endogenous variables and the productivity shock is nonlinear under the second- and third-order approximations. This has important implications for the impulse-response analysis of the DSGE model. For example, it is clear in Figure 1 that a positive productivity shock of size $+2$ standard deviations would induce a larger change (in absolute value) in the next-period capital stock than a shock of size -2. Hence, in contrast to first-order approximate DSGE models, a normalization does not summarize the dynamic responses to a shock. Instead, responses will typically depend on the sign and size of the shock, and the state of the system when the shock occurs.[6]

3 The Simulated Method of Moments

The simulated method of moments (SMM) was originally developed by McFadden (1989) and Pakes and Pollard (1989) to estimate discrete-choice models in i.i.d. environments, and extended by Lee and Ingram (1991) and Duffie and Singleton (1993) to time-series models with serially correlated shocks. Duffie and Singleton (1993) show the consistency and asymptotic normality of the SMM estimators under fairly general conditions.

Consider a fully-specified model with unknown parameters $\theta \in \Theta$, where θ is a $q \times 1$ vector and $\Theta \subset \Re^q$ is a compact set. In our case, the model is a nonlinear DSGE model and θ may contain, for example, the structural parameters $\{\alpha, \beta, \rho, \sigma, \delta, \gamma, a\}$. A sample of T observations of economic data, $\{x_t\}$, is available to estimate the model, with x_t stationary and ergodic. The stationarity and ergodicity of x_t may have been induced by a prior transformation of the raw data, for example by means of a detrending procedure. Denote by

$$(1/T)\sum_{t=1}^{T}\mathbf{m}(x_t) \tag{6}$$

the $p \times 1$ vector of statistics or moments computed based on the time average of some function of the data. A necessary, but not sufficient, condition for identification is $p \geqslant q$. Under the

[6] For a complete treatment of impulse-response analysis in nonlinear systems, see Gallant, Rossi and Tauchen (1993), and Koop, Pesaran, and Potter (1996).

assumption of ergodicity and by the Law of large numbers

$$(1/T)\sum_{t=1}^{T}\mathbf{m}(x_t) \rightarrow E(\mathbf{m}(x_t)) \text{ almost surely, as } T \rightarrow \infty.$$

Provided that the model can be expressed in terms of trend-free processes, and under the maintained hypothesis that the model is a correct description of (some aspect of) the economy when $\theta = \theta_0$, then there exists a synthetic counterpart of the observed data $\{x_t\}$, namely $\{x_\iota(\theta_0)\}$, obtained by simulating the model given a draw of random shocks. Loosely speaking, the maintained assumption is that the simulated and observed data are drawn from the same statistical distribution when the former is generated using the true parameter values. More generally, consider the synthetic series $\{x_\iota(\theta)\}$ simulated using the parameter values θ. The length of the simulated series is τT, where $\tau \geqslant 1$ is an integer. Using these data, it is possible to compute a $p \times 1$ vector of moments analog to (6)

$$(1/\tau T)\sum_{\iota=1}^{\tau T}\mathbf{m}(x_\iota(\theta)).$$

Under Assumption 2 in Duffie and Singleton (1993, p. 939), which states that for all $\theta \in \Theta$, the process $\{x_\iota(\theta)\}$ is geometrically ergodic, and by Lemma 2 (p. 938), which is a Uniform Weak Law of large numbers,

$$(1/\tau T)\sum_{\iota=1}^{\tau T}\mathbf{m}(x_\iota(\theta)) \rightarrow E(\mathbf{m}(x_\iota(\theta)) \text{ almost surely, as } \tau T \rightarrow \infty,$$

for any $\theta \in \Theta$.[7] Finally, under the assumption that the model is correctly specified

$$E(\mathbf{m}(x_\iota(\theta_0)) = E(\mathbf{m}(x_t)).$$

Then, the SMM estimator is defined as

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \mathbf{M}(\theta)'\mathbf{W}\mathbf{M}(\theta), \tag{7}$$

where

$$\mathbf{M}(\theta) = \left((1/T)\sum_{t=1}^{T}\mathbf{m}(x_t) - (1/\tau T)\sum_{\iota=1}^{\tau T}\mathbf{m}(x_\iota(\theta))\right), \tag{8}$$

and $\mathbf{W}$ is a positive-definite weighting matrix of dimension $p \times p$. Intuitively, the SMM estimator is the value of θ that minimizes the (weighted) distance between the moments

[7] Additional restrictions are imposed on the simulated series $\{x_\iota(\theta)\}$ to account for the fact that the initial draw may not necessarily come from its ergodic distribution and that the transition law of state variables may be affected by the simulation. See Duffie and Singleton (1993) for details.

implied by the model and those computed from the observed data, where the former are obtained using artificial data simulated from the model.

Under the regularity conditions spelled out in Duffie and Singleton (1993), $\widehat{\theta}$ is a consistent estimator of θ_0, and its asymptotic distribution is

$$\sqrt{T}(\widehat{\theta} - \theta_0) \to N(0,(1 + 1/\tau)(\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}\mathbf{J}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{J}(\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}), \tag{9}$$

where

$$\mathbf{S} = \lim_{T\to\infty} Var\left((1/\sqrt{T})\sum_{t=1}^{T}\mathbf{m}(x_t)\right), \tag{10}$$

and $\mathbf{J} = E(\partial\mathbf{m}(x_\iota(\theta))/\partial\theta)$ is a finite matrix of full column rank and dimension $p \times q$. Although $\widehat{\theta}$ is consistent for any positive-definite weighting matrix, the smallest asymptotic variance in (9) is obtained when the weighting matrix is the inverse of the long-run variance of the moments, that is, $\mathbf{W} = \mathbf{S}^{-1}$. In this case the asymptotic distribution of the SMM estimator simplifies to

$$\sqrt{T}(\widehat{\theta} - \theta_0) \to N(0,(1 + 1/\tau)(\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}). \tag{11}$$

When p is strictly larger than q, that is when the model is over-identified, it is possible to construct a general specification test using the chi-square statistic proposed in Lee and Ingram (1991, p. 204) and based on Hansen (1982). The test statistic is easiest to compute in the case where $\mathbf{W} = \mathbf{S}^{-1}$. Then,

$$T(1 + 1/\tau)\left(\mathbf{M}(\widehat{\theta})'\mathbf{W}\mathbf{M}(\widehat{\theta})\right) \to \chi^2(p - q), \tag{12}$$

where $\mathbf{M}(\widehat{\theta})'\mathbf{W}\mathbf{M}(\widehat{\theta})$ is the value of the objective function at the optimum. Hayashi (2000) shows how to derive Wald, Lagrange and an analog to the Likelihood Ratio tests for method of moments estimators.

3.1 Incorporating Priors

In many applications, the researcher may be interested in incorporating additional information obtained from micro data or in ruling out parts of the parameter space that are considered economically uninteresting. This section proposes a simple strategy to do so in the context of method of moments estimation. The strategy is inspired by the mixed estimation approach in Theil and Goldberger (1961), where priors are treated as additional observations and combined with the data to deliver a penalized statistical objective function.

Theil and Goldberger's approach was originally developed for the linear regression model and leads to a Generalized Least Squares (GLS) estimator that optimally incorporates the prior information. Stone (1954) gives a maximum likelihood interpretation to this GLS estimator and Hamilton (1994, p. 359) shows that its mean and variance are exactly those of the Bayesian posterior distribution. In the maximum likelihood framework, the mixed estimation strategy yields a likelihood function that consists of the likelihood of the data and a penalty function—that is, the likelihood of the priors—which increases as parameters deviate from the priors. Hamilton (1991) and Ruge-Murcia (2007) study the application of this strategy for the estimation of mixtures of normal distributions and linearized DSGE models, respectively.

In the same spirit, the statistical objective function of the method of moments estimator may be augmented with an extra term that explicitly incorporates prior information about parameter values. In order to keep the notation simple, assume that the researcher has priors about all structural parameters and write the penalized SMM objective function as

$$\mathbf{M}(\theta)'\mathbf{W}\mathbf{M}(\theta) + (\theta - \theta^*)'\mathbf{\Omega}(\theta - \theta^*), \tag{13}$$

where θ^* is a $q \times 1$ vector of priors about θ, and $\mathbf{\Omega}$ is a $q \times q$ positive-definite weighting matrix that represents the researcher's confidence in the prior information.[8] Thus, while mixed estimation treats priors as additional observations, the function (13) treats priors as additional moments. Notice that the penalty $(\theta - \theta^*)'\mathbf{\Omega}(\theta - \theta^*)$ is monotonically increasing in the distance between θ and θ^*. Then, a quasi-Bayesian SMM estimate of the parameters of the DSGE model is

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \ [\mathbf{M}(\theta) \ (\theta - \theta^*)]'\mathbf{V}[\mathbf{M}(\theta) \ (\theta - \theta^*)], \tag{14}$$

where $[\mathbf{M}(\theta) \ (\theta - \theta^*)]$ is a $(p+q) \times 1$ vector and

$$\mathbf{V} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} \end{bmatrix}.$$

The upper and lower off-diagonal matrices of $\mathbf{V}$ are $p \times q$ and $q \times p$, respectively, with all elements equal to zero. For the Monte Carlo experiments in Section 4, the asymptotic variance-covariance matrix of this quasi-Bayesian estimate is computed as

$$(1 + 1/\tau)(\mathbf{B}'\mathbf{V}\mathbf{B})^{-1}\mathbf{B}'\mathbf{V}\mathbf{Q}\mathbf{V}\mathbf{B}(\mathbf{B}'\mathbf{V}\mathbf{B})^{-1},$$

where $\mathbf{B} = [\mathbf{J} \ \mathbf{I}]'$, $\mathbf{I}$ is a $q \times q$ identity matrix, and

$$\mathbf{Q} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} \end{bmatrix}.$$

[8] Although the discussion and Monte-Carlo experiments in this paper are for nonlinear DSGE models, it is clear that this idea may also be applied to the estimation of linearized DSGE models and to other minimum distance estimators, like the generalized method of moments, the extended method of simulated moments (see Smith, 1993), and the matching of theoretical and empirical impulse responses.

3.2 Comparison with GMM

Let us consider the case where it is possible to derive analytical expressions for the unconditional moments as a function of the parameters. If one collects these expressions in the $p \times 1$ vector $E(\mathbf{m}(\theta))$, and uses $E(\mathbf{m}(\theta))$ instead of the simulation-based estimate $(1/\tau T) \sum_{i=1}^{\tau T} \mathbf{m}_i(\theta)$ in the statistical objective function, then the resulting generalized method of moments (GMM) estimator is

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \ \mathbf{M}(\theta)' \mathbf{W} \mathbf{M}(\theta), \tag{15}$$

where

$$\mathbf{M}(\theta) = \left((1/T) \sum_{t=1}^{T} \mathbf{m}(x_t) - E(\mathbf{m}(\theta)) \right),$$

and $\mathbf{W}$ is a $p \times p$ positive-definite weighting matrix. Under the regularity conditions in Hansen (1982), this estimator is consistent and has asymptotic distribution

$$\sqrt{T}(\widehat{\theta} - \theta_0) \to N(0, (\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}\mathbf{J}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{J}(\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}), \tag{16}$$

where $\mathbf{S}$ was defined in (10) and $\mathbf{J} = \partial E(\mathbf{m}(\theta))/\partial\theta$ is a $p \times q$ matrix of full column rank.

In the case of linearized models, it is possible to compute $E(\mathbf{m}(\theta))$ from the (linear) decision rules that solve the model, and so the GMM estimator is feasible. In general, GMM delivers more statistically efficient estimates than SMM because there is no simulation uncertainty (see below). For small-scale models, GMM is also more computationally efficient than SMM (see Ruge-Murcia, 2007). However, for large-scale model the matrix inversions required to go from the decision rules to the moments imply that simulation may be a faster way to compute the moments and, thus, SMM may be preferable to GMM for that reason.

In the case of nonlinear models, it is not always possible to derive analytical expressions for the unconditional moments. Simulation is then an attractive alternative because the simulation-based estimate $(1/\tau T) \sum_{i=1}^{\tau T} \mathbf{m}_i(\theta)$ is consistent for $E(\mathbf{m}(\theta))$ and simulation uncertainty can be controlled by the econometrician through a judicious choice of τ. To see the latter point, compare the asymptotic variance-covariance matrices under SMM and GMM in (9) and (16), respectively. Note that, since $\mathbf{W}$ and $\mathbf{S}$ depend only on the data and the simulated moments converge to the analytical ones as the size of the artificial sample increases, the difference in the standard errors of both estimates is primarily due to the term $(1 + 1/\tau)$ in the distribution (9). This term captures the increase in sample uncertainty due to the use of simulation to compute population moments. Note, however, that $(1 + 1/\tau)$ decreases quickly towards 1 as τ increases: For example, when $\tau = 5$, 10 and 20, the asymptotic SMM standard errors are $1.10, 1.05$ and 1.025 times larger than those implied by GMM.

One instance where it is possible to derive analytical expressions for the second moments predicted by a nonlinear model is when the solution is obtained using the pruning algorithm proposed by Kim, Kim, Schaumburg and Sims (2008). Hence, in this instance the GMM estimator is also feasible. However, in the case of second-order approximate solutions, the second moments implied by the nonlinear solution are (by construction) the same as those implied by the linear solution. Thus, the GMM estimates of the linear and nonlinear models are identical. This equivalence does not arise in the case of third- and higher-order approximate solutions because their second moments depend on terms of order higher than two and are, therefore, different from those implied by the linear solution.

4 Monte-Carlo Experiments

4.1 Design

The basic model with no habit formation has seven structural parameters but, for reasons to be made clear below, I concentrate on four of them in the Monte Carlo experiments. These four parameters are the subjective discount factor (β), the autocorrelation coefficient of the technology shock (ρ), the standard deviation of the technology innovation (σ), and the curvature parameter of consumption (γ) . Thus, $\theta = (\beta, \rho, \sigma, \gamma)'$ is a 4×1 vector. In all experiments, the capital share (α) is fixed to 0.36, the depreciation rate (δ) is fixed to 0.025, and the weight of leisure in the utility function (b) is set so that the time spent working in steady state is one third of the time endowment. Fixing the value of some parameters replicates actual practice by researchers who estimate DSGE models, sidesteps the weak identification of α and δ intrinsic to this model,[9] and reduces the computational burden in the Monte Carlo. The data from the model are generated using $\beta = 0.95$, $\rho = 0.85$, $\sigma = 0.04$, and three possible values for γ (that is, $\gamma = 1$, 2 and 5). The simulations of the nonlinear model, both for data generating and for SMM estimation, are based on the pruned version of the model, as suggested by Kim, Kim, Schaumburg and Sims (2008). The data series are consumption and hours worked in deviations from their deterministic steady state values and the moments are the variances, the covariance, and the first- and second-order autocovariances of these series.

For the version of the model with habit formation $\theta = (\beta, \rho, \sigma, \gamma, a)'$ is a 5×1 vector and the value of the habit parameter used to generate the data is $a = 0.8$. For these experiments I use two possible values for the standard deviation of the productivity innovation,

[9] Canova and Sala (2009) show that the rational-expectations solution of a real business cycle model with inelastic labor supply implies that the dynamics of the capital stock are only weakly influenced by α and ρ, and insensitive to proportional changes in δ and β.

namely $\sigma = 0.04$ and $\sigma = 0.08$, and the three possible values for γ listed above. These different configurations allow me to study the properties of SMM in DSGE models with different curvature and departure from certainty equivalence. All experiments are based on 200 replications using a sample of 200 observations. The sample size is comparable to that employed in practice to estimate DSGE models (for example, it is equivalent to, say, quarterly observations for a period of fifty years).

For all parameter configurations, I use three different values for τ, that is $\tau = 5$, 10, 20, meaning that the simulated series are, respectively, five, ten and twenty times larger than the sample size. Exploring the effects of using different values of τ is important for two reasons. First, the asymptotic distribution of the estimates depends on τ because simulation uncertainty depends on the length of the simulated series relative to the sample size. Second, since the computational cost is increasing in τ, it is useful to know whether it is worthwhile to use long artificial series.

Finally, for the weighting matrix $\mathbf{W}$, I use the inverse of the matrix with the long-run variance of the moments, that is $\mathbf{W} = \mathbf{S}^{-1}$, where $\mathbf{S}$ was defined in (10) and is computed using the Newey-West estimator with a Barlett kernel and bandwidth given by the integer of $4(T/100)^{2/9}$. This weighting matrix is optimal in that it delivers the smallest possible asymptotic variance among the class of positive-definite matrices. The asymptotic distribution of the SMM estimator in the case where $\mathbf{W} = \mathbf{S}^{-1}$ is given in (11). I also carry out experiments using two other (sub-optimal) weighting matrices. They are 1) the identity matrix and 2) the inverse of a matrix with diagonal elements equal to those of $\mathbf{S}$ and off-diagonal elements equal to zero. These experiments allow me to evaluate the efficiency loss associated with weighting matrices that are not optimal but which have practical advantages in actual applications.[10]

4.2 Results

Results of Monte Carlos experiments for basic model are reported in Table 1 and those for the models with habit formation are reported in Tables 2 and 3. In the tables, Mean and Median are, respectively, the mean and median of the estimated parameters, A.S.E. is the median of the asymptotic standard errors, and S.D. is the standard deviation of the estimates. These statistics were computed using the 200 replications for each experiment. Size is the proportion of times that the null hypothesis that the parameter takes its true value is rejected using a t test with nominal size of five percent. Or, put differently, Size

[10]See, for example, Cochrane (2001, p. 215), who argues that in many instances it may be desirable to use a weighting matrix that pays attention to economically, rather than only statistically, interesting moments.

is the empirical size of the t test. S.E. is the standard error of this empirical size and is computed as the standard deviation of a Bernoulli variable. Finally, OI is the empirical size of the chi-square test of the overidentification restrictions.

These tables support the following conclusions. First, SMM estimates are quantitatively close to the true values used to generate the data in all cases. To the see this, note that in all tables the mean and median of the estimated parameters are very similar to the true values. This result, of course, is driven by the consistency of the SMM estimator, but it is important to know that SMM delivers accurate parameter estimates for relatively small samples and for versions of the model with different curvature. Second, asymptotic standard errors computed from (11) often overstate the actual variability of the parameter estimates. To see this, note that in most cases the A.S.E. is larger than the standard deviation of the estimates. This suggests a discrepancy between the small-sample and the asymptotic distributions. A similar results is reported by Ruge-Murcia (2007) in the context of linear DSGE models. Third, the empirical size of the t test of the null hypothesis that the parameter takes its true value usually differs from the nominal size of five percent. Finally, note that in all cases the empirical size of the chi-square test of the over-identification restrictions is well below its nominal size of five percent. The result that the chi-square test easily fails to detect a misspecified model is well known in the literature (see, for example, Newey, 1985), but the results reported here indicate that this result also holds in the case of fully-specified nonlinear DSGE models.

Asymptotic standard errors and the empirical standard deviations of the estimates tend to decline with size of the simulated sample (that is, with the value of τ) as predicted by the theory. This result is visually confirmed by the empirical distributions of the estimates reported Figures 2 through 4. (These figures correspond to the experiments for $\gamma = 2$ but the distributions for experiments for $\gamma = 1, 5$ support the same conclusions.) However, this increase in statistical efficiency needs to be balanced with the increased computational cost associated with using a larger artificial sample (see Section 4.6).

Figures 2 through 4 also illustrate the earlier result that the asymptotic distribution may not always be a good approximation to the small-sample distribution. For example, the discount factor (β) is estimated much more precisely in practice than the asymptotic approximation would have us believe. Also, in some cases the small-sample distributions are skewed, rather than symmetric. For example, the small-sample distributions of the autoregressive coefficient (ρ) are mildly skewed to the left, and so the mean is usually smaller than the median in Tables 1 through 3, and both are somewhat smaller than the true value of 0.85 used to generate the data.

From Tables 1 through 3 it appears that small-sample distortions diminish somewhat as the consumption curvature increases. A priori one would have expected the asymptotic approximation to worsen as the model becomes more nonlinear. However, the nonlinearity also seems to sharpen identification leading to smaller standard errors. To see this, consider Figure 5, which plots the objective function of along the dimension of the discount factor (β), the autoregressive coefficient (ρ) and the standard deviation of the innovation (σ) for each of the three possible values of the curvature parameter (γ) and holding the other parameters fixed to their true value. From this Figure, it is apparent that as the consumption curvature increases, the convexity of the statistical objective function increases as well.

4.3 Other Weighting Matrices

In this section I study the small-sample properties of SMM in the case where the weighting matrix is 1) the identity matrix and 2) the inverse of the matrix with diagonal elements equal to those of $\mathbf{S}$ and off-diagonal elements equal to zero. In what follows, I denote the latter matrix by $\mathbf{D}^{-1}$.

The identity matrix gives the same weight to all moments and the objective function is, therefore, proportional to the (squared) Euclidean distance between the empirical and theoretical moments. A potential problem with the identity matrix is that the objective function is not scale independent. Thus, the algorithm used to numerically minimize the objective function may converge either too quickly, before the true minimum is found, or not converge at all, if the converge criterion is too tight given the units of the moments. An easy way to deal with this problem is to appropriately scale the objective function or the weighting matrix. For example, rather than using $\mathbf{W} = \mathbf{I}$, the researcher may use $\mathbf{W} = \zeta\mathbf{I}$ where ζ is a scaling constant. The second weighting matrix (that is, $\mathbf{D}^{-1}$) is attractive because the objective function is scale free and it gives a larger weight to moments that are more precisely estimated.

Results of experiments using these two matrices are reported in Tables 4 and 5 and Figures 6 and 7. As in the previous experiments (where $\mathbf{W} = \mathbf{S}^{-1}$), SMM estimates are close to the true values used to generate the data, asymptotic standard errors tend to overstate the actual variability of the estimates, and the empirical size of the t test is often different from the nominal size. Comparing standard errors across different weighting matrices, notice that standard errors obtained when $\mathbf{W} = \mathbf{I}$ are larger than when $\mathbf{W} = \mathbf{D}^{-1}$, which in turn are larger than when $\mathbf{W} = \mathbf{S}^{-1}$. This was the expected ranking since $\mathbf{D}$ shares (by construction) the same diagonal as $\mathbf{S}$ and $\mathbf{S}^{-1}$ is the optimal weighting matrix. The difference between standard errors is relatively small when the curvature parameter $\gamma = 1$ but becomes substantial when $\gamma = 5$. This suggests that the efficiency gains associated with using the the optimal weighting matrix increase with the nonlinearity of the model. On the other hand, at least for the examples considered here, the difference between standard errors is not so large as to overcome economic considerations that may motivate the use of alternative weighting matrices (see, Cochrane, 2001 p. 215).

4.4 Third-Order Approximation

Since perturbation methods are only locally accurate, it is important to study the econometric implications of estimating a model approximated to an order higher than two. These results are reported in Table 6 and are similar to those previously reported in that estimates are quantitatively close to the true values, there is a discrepancy between asymptotic and small-sample standard errors, and the t test of the null hypothesis that the parameter takes its true value is subject to size distortions.[11]

There are, however, two important differences. First, the standard deviations of the empirical estimates are generally smaller than those obtained using a second-order approximation. This may be seen by comparing the empirical distributions in Figures 8 and 2 and the S.D.'s in Tables 1 and 4. This suggests that parameters are more precisely estimated when using a third- rather than a second-order approximation. The most striking example of this observation is the estimate of the discount factor, β. On the other hand, since asymptotic standard errors are similar in both cases, this means that the asymptotic distribution is a poorer approximation to the small-sample distribution in this case than in the ones reported in the previous sections. Second, although the chi-square test of the over-identification restrictions tends to over-reject, the discrepancy between actual and nominal size are much smaller than for the second-order approximation.

Overall, these results indicate some statistical advantages in estimating a third- rather than a second-order approximation to the model. A third-order solution is also attractive because in addition to higher accuracy, it allows the explicit effect of skewness on economic choices and permits the use of third-order moments for the estimation of the model. However, there are two caveats. First, as will see below, the computational cost of estimating a third-order approximation of the model is higher than that of estimating a second-order. Second, while the pruning algorithm proposed by Kim, Kim, Schaumburg and Sims (2008) insures the stability of the second-order system (provided first-order dynamics are stable), this is no longer the case when this algorithm is applied to a third-order solution.

[11] Note that since the analysis is carried out under the null hypothesis, the DGP in these experiments is not the same as in the previous ones: The data in the previous section is generated using a second-order approximate solution, while the data here is generated using a third-order.

4.5 Incorporating Priors

For the Monte-Carlo experiments concerning the small sample performance of the quasi-bayesian SMM, I use as priors for β, ρ and γ their true values and assume a flat prior for σ. Results are reported in Table 7 and the empirical distributions (for the case where $\gamma = 2$) are plotted in Figure 9. Since the priors are exactly the same values used to generate the data, the estimates are much closer to the true values than in the previous experiments without priors. As before, I find that small-sample and asymptotic distributions differ. In particular, the asymptotic standard errors greatly overstate (understate) the actual variability of the parameter estimates for which priors are (are not) imposed.

4.6 Computing Time

While there are no large differences in statistical efficiency for the values of τ studied here, there does appear to be some differences in computational efficiency. (Recall that τ measures the size of the artificial series compared with that of the data.) The left-hand side panel of Figure 10 plots the average time required to estimate one replication of second-order approximate models as a function of τ. Notice that computing time increases approximately linearly with the length of the simulated sample, and that, as one would expect, models with additional state variables (for example, the model with habit formation) require additional time. For example, the second-order approximate basic and habit models with $\tau = 5$ take about 30 and 68 seconds, respectively.

The right-hand side panel of the same Figure includes the average time required to estimate third-order approximate models and shows that it is one order of magnitude larger than that of second-order models. For example, the third-order basic model with $\tau = 5$ takes 362 seconds, compared with 30 seconds for the second-order. As before, computing time increases approximately linearly with the length of the simulated sample.

Overall, these results suggest that computing time is not a constraint for the estimation of nonlinear DSGE models by the simulated method of moments.

5 The Macroeconomic Effects of Rare Events

A standard assumption in macroeconomics is that shock innovations are drawn from a symmetric distribution. In symmetric distributions the shape of the left side of the central maximum is a mirror image of the right side, and so (loosely speaking) negative realizations of a given magnitude are as likely as positive realizations. Among symmetric distributions, the Normal distribution is by far the most widely used in empirical research.

On the other hand, quantitative results by Rietz (1988) and Barro (2006, 2009) suggest that extreme, low probability events—presumably arising from an asymmetric distribution—may have large economic implications in a setup where agents are risk averse. In particular, Rietz and Barro are concerned with catastrophic events (e.g., wars) as a possible explanation for the large return on equities compared with the riskless asset.

The goal of this Section is to study the implications of rare, though not necessarily catastrophic, events for consumption, capital accumulation, and labor supply in the simple economy described in Section 2.1. To that effect, I estimate versions of the model where productivity innovations are drawn from asymmetric distributions that allow very large negative realizations from time to time. I consider two asymmetric distributions, namely the Rayleigh and the Skew normal distributions.[12] The Rayleigh distribution is a one parameter distribution with support $[0, \infty)$ and positive skewness equal to $2\sqrt{\pi}(\pi - 3)/(4 - \pi)^{3/2} = 0.6311$. I work with the negative of the Rayleigh distribution (therefore the skewness is -0.6311) and adjust its location so that the distribution has a zero mean. The Skew normal distribution due to O'Hagan and Leonhard (1976) is a three parameter distribution (a location, a scale, and a shape parameters) with support $(-\infty, \infty)$. The skewness of the Skew normal distribution depends only on the scale and shape parameters and can take values between -1 and 1. I also adjust the location of this distribution so that it has a zero mean. As a benchmark, I estimate a version of the model with normally distributed productivity innovations.

As was discussed in Section 2.2, neither a first- nor a second-order approximate solution to the model can capture the effects of skewness on the agents' economic choices. Hence, I solve and estimate here a third-order approximate solution to the model. In addition to studying the economic implications of rare events, this Section illustrates the application of simulated method of moments for the estimation of a nonlinear dynamic model using actual data, and shows that it can easily accommodate non-Normal distributions.

5.1 Data and SMM Estimation

The model is estimated using 205 quarterly U.S. observations of the growth rate of consumption, the growth rate of investment, and hours worked for the period 1959:Q2 to 2010:Q2. The raw data were taken from the FRED database available at the Federal Reserve Bank of St. Louis Web site (*www.stls.frb.org*). Hours worked is measured by the quarterly average of the weekly hours of production and non-supervisory employees in manufacturing. Consumption is measured by personal consumption expenditures on nondurable goods and services, and investment is measured by private nonresidential fixed investment plus per-

[12] The use of the Skew normal distribution was suggested to me by Bill McCausland.

sonal consumption expenditures on durable goods. The latter two measures are closest to the definitions of consumption and investment in the model. Since the raw data are nominal, I converted them into real *per capita* terms by dividing the series by the quarterly average of the consumer price index (CPI) for all urban consumers and the mid-month U.S. population estimate produced by the Bureau of Economic Analysis (BEA). All data are seasonally adjusted at the source.

In this application, I use the identity matrix as the weighting matrix in the statistical objective function, and compute the long-run variance of the moments using the Newey-West estimator with a Barlett kernel and bandwidth given by the integer of $4(T/100)^{2/9}$ where T is the sample size. Since $T = 205$ the bandwidth is, therefore, four. The number of simulated observations is five times larger than the sample size, that is $\tau = 5$. As seen in the Monte-Carlo results, $\tau = 5$ is much more computationally efficient than larger values of τ and only mildly less statistically efficient. The simulations of the nonlinear model are based on the pruned version of the third-order approximate solution, as suggested by Kim, Kim, Schaumburg and Sims (2008). The moments used to estimate the model are the variances, covariances, first- and second-order autocovariances, and the third-order moments of the three data series—that is, fifteen moments.

I estimate the discount factor (β), the consumption curvature of the utility function (γ), the autoregressive coefficient of the productivity shock (ρ), and the parameters of the innovation distribution. Since the mean of the innovations is fixed to zero, I only estimate one parameter in the case of the Rayleigh and Normal distributions, and two (the scale and shape parameters) in the case of the Skew Normal distribution. During the estimation procedure, the production function parameter (α) was fixed to 0.36, which is in line with share of capital in total income according to the National Income and Product Accounts (NIPA), and the depreciation rate (δ) was set to 0.025. Finally, the weight of leisure in the utility function (b) was set so that the proportion of time spent working in the deterministic steady state is one-third.

The theoretical results concerning the statistical properties of SMM were derived under the assumption that the series are stationary (see Section 3 above). In order to verify whether this assumption is appropriate for the data used to estimate the model, I performed Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit-root tests. In all cases, the estimated alternative was an autoregression with a constant intercept but no time trend. The level of augmentation of the ADF test (that is, the number of lagged first differences in the regression) was selected using recursive t tests and the Modified Information Criterion (MIC) in Ng and Perron (2001). The truncation lag of the PP test was set to four in all cases, but conclusions are robust to using other number of lags.

Unit-root test results are reported in Table 8. The null hypothesis of a unit root in consumption growth and in investment growth can be safely rejected at the five percent significance level using either test. Regarding hours worked, the null hypothesis can be rejected at the five percent level using the PP test and at the ten percent level using the ADF test with augmentation selected by the MIC. The null hypothesis cannot be rejected at the ten percent level using the ADF test with augmentation selected using recursive t tests, but this conclusion is marginal because the test statistic is -2.53 while the critical value is -2.58. Overall, these results suggest that the data may be well represented by stationary processes around a constant mean.

5.2 Parameter Estimates

Parameter estimates are reported in Table 9. Estimates of the discount factor are very similar across distributions, ranging only from 0.9942 (Rayleigh) to 0.9972 (Skew normal). On the other hand, there are some differences in the estimates of the consumption curvature parameter and the autoregressive coefficient of the productivity shock. In particular, the consumption parameter is smaller when the innovation distributions are asymmetric than when they are normally distributed. (Still, it is important to note that standard errors are large enough that the statistical difference among estimates is probably nil.)

The most obvious difference is, of course, in the estimates of the innovation distributions. The standard deviations are 0.0061, 0.0059, and 0.0041 for the Normal, Rayleigh, and Skew normal distributions, respectively, while the skewness are 0, -0.6311, and -0.8539. In the two former cases, the skewness is fixed (by construction), while in the latter case it depends on the estimated shape and slope parameters. Histograms of the estimated distributions, computed on the basis of a simulation of 5000 observations, are plotted in Figure 11. The negative skewness of the Rayleigh and Skew normal distributions are apparent from this Figure: Extreme negative realizations of the productivity innovation may occasionally happen but extreme positive realizations are relatively uncommon (or not possible in the case of the Rayleigh distribution, which is bounded from above). In contrast, the symmetric Normal distribution allows both possible and negative realizations with the same frequency. Notice that in order to accommodate the large negative realizations of the innovations, which seem to be a feature of the data (see below), the Normal distribution may overstate the actual standard deviation of the productivity innovations.

5.3 Medians

Table 10 reports the median percentage deviation from the deterministic steady state for some key variables. Recall that since certainty equivalence does not hold in this model, the variance and skewness of the shocks affect the agents' decisions. Table 10 shows that when productivity shocks are normally distributed, median consumption is about 1 percent below what it would be in a certainty-equivalent world. The median investment and capital stock are also below the deterministic steady state by about around 1.3 percent, while hours worked are above by 0.4 percent. Since the shock distribution is symmetric (and, thus, the skewness is zero), these effects are primarily due to the variance of the shocks. Results for the mildly skewed Rayleigh distribution are similar.

In contrast, when shocks follow a Skew normal distribution, agents protect themselves against volatility and possible large decreases in productivity by accumulating more capital and working more hours compared with the certainty-equivalent case. The increase in wealth actually allows a small increase in consumption (0.15 percent) above the deterministic steady state.

Since the estimates of the preference parameters are not drastically different across the three versions of the model, the differences reported in Table 10 are attributable to the differences in the shock distributions and the agents' reaction to their different variance and skewness.

5.4 Higher-Order Moments

I now derive and evaluate the model predictions concerning the higher-order moments of consumption growth, investment growth, and hours worked. Table 11 compares the moments predicted by the model (computed from a simulation of 5000 observations) and those of the U.S. data. In the data, consumption growth is considerably less volatile than investment growth and hours. All three models capture this feature of the data, but the Skew normal distribution under-predicts the standard deviation of consumption growth. Also, in the data, all three series are negatively skewed. Although the three models imply negative skewness, the Normal distribution implies lower skewness than the data in all cases.[13] In contrast, the Rayleigh and Skew normal distributions imply skewness closer to the data. Notice, however, that all three distributions tend to under-predict the skewness of hours worked and that the Skew normal distribution over-predicts the skewness of investment growth. Finally, in

[13]Notice that in contrast to linear DSGE models that inherit their higher-order properties directly from the shock innovations, the nonlinear propagation mechanism in this model means that economic variables may be non-Normal, even if the productivity innovations are Normal.

the data, the kurtosis is larger than the value of 3 implied by the Normal distribution. All three models tend to under-predict kurtosis but the asymmetric distributions generally imply kurtosis larger than 3 while the Normal distribution implies kurtosis smaller than 3 for all variables.

Table 11 also reports the Jarque-Bera test statistic for the null hypothesis that a series follows a Normal distribution (with unspecified mean and variance). For all U.S. data series, the hypothesis can be rejected at the five percent significance level. In contrast, for artificial series with 5000 observations from the model with normally distributed productivity innovations, the hypothesis cannot be rejected for consumption and investment growth, although it can be rejected for hours worked. When the artificial observations are generated using either a Rayleigh or a Skew normal distribution, the hypothesis can be rejected in all cases.

In summary, these results suggest that asymmetric productivity innovations deliver higher-order moments that are in better agreement with the U.S. data compared with a symmetric Normal distribution.

5.5 Impulse-Response Analysis

In this Section, I study the economy responses to a productivity shock. Since in nonlinear systems the effect of shocks generally depend on their sign, size, and timing (see Gallant, Rossi and Tauchen, 1993, and Koop, Pesaran, and Potter, 1996), I consider shocks of different sign and sizes and assume that they take occur when the system is at the stochastic steady state (i.e., when all variables are equal to their unconditional mean). Since the distributions are not the same in each of the versions of the model, I fixed the percentile (rather than actual size) of the shock. In particular, I consider innovations in the 25th and 75th percentiles, along with rare events in the form of innovation in the 1st and 99th percentiles.

Results are reported in Figure 12 through 14 for the Normal, Rayleigh, and Skew normal distributions respectively. In these figures, the horizontal axes are periods and the vertical axes are percentage deviations from the stochastic steady state. Notice that in all cases a positive (negative) productivity shock induces an increase (a decrease) in consumption, investment and hours worked. However, in the case of the Normal distribution responses are close to symmetric despite the fact the model is nonlinear. That is, the rare bonanza due to the large productivity innovation in the 99th percentile induces increases in the variables which are (almost) mirror images of the decreases due to the rare loss when the innovation is in the 1st percentile.

On the other hand, in the case of the asymmetric distributions, negative innovations deliver larger decreases in consumption, investment, and hours worked than an equally-likely positive realization. Of course, the size (in absolute value) of innovation in the 1st and 99th (and the 25th and 75th) percentile are not same because the distributions are asymmetric. The point is, however, that the likelihood of these two realizations is the same. For example, in the case of the Skew normal distribution, the rare bonanza due to a large productivity innovation in the 99th percentile induces increases of 0.3, 3.5, and 1 percent in consumption, investment, and hours worked, respectively, while the rare loss due an innovation in the 1st percentile induces decreases of -0.5, 6.5 and 1.9 percent, respectively. There is also asymmetry for the (smaller) innovations in the 25th and 75th percentiles, but in this case, the positive innovations induce quantitatively larger responses than the equally-likely negative innovation.

6 Conclusions

This paper describes in a pedagogical manner the application of the simulated method of moments for the estimation of nonlinear DSGE models, studies its small-sample properties in models with different curvature, provides evidence about its computational efficiency, proposes a simple strategy to incorporate prior information based on the mixed-estimation approach due to Theil and Goldberger (1961), and shows how to compute third-order approximate solutions to the decision rules of a DSGE model. Monte-Carlo results that SMM is delivers accurate and computationally efficient parameter estimates, even when the simulated series are relatively short. However, asymptotic standard errors tend to overstate the true variability of the estimates and, consequently, statistical inference is conservative.

An application to effects of rare events shows that relaxing the standard assumption of symmetry in shock distributions may be important for economic and statistical reasons: Skewed distributions allow agents to contemplate the possibility of extreme shock realizations and make their economic choices accordingly, and deliver higher-order moments that are generally closer to those in U.S. data.

A Solving DSGE Models Using a Third-Order Approximation to the Policy Function

A.1 Analytics

As in Schmitt-Grohé and Uribe (2004), the model solution is written as[14]

$$\begin{aligned} y_t &= g(x_t, \sigma), \\ x_{t+1} &= h(x_t, \sigma) + \eta\sigma\varepsilon_{t+1} \end{aligned}$$

where y_t is a $n_y \times 1$ vector of non-predetermined variables, x_t is a $n_x \times 1$ vector of predetermined variables, ε_t is a $n_\varepsilon \times 1$ vector of innovations to the exogenous state variables in x_t, η is $n_x \times n_\varepsilon$ vector of (scaling) parameters, and σ is a perturbation parameter. The innovations are assumed to be i.i.d. with mean zero, variance-covariance matrix equal to the identity matrix, I, and a cube of third moments, S, with possible non-zero values along the main diagonal. The goal is to approximate the policy functions $g(x_t, \sigma)$ and $h(x_t, \sigma)$ with a polynomial obtained by means of a Taylor series expansion around the deterministic steady state where $x_t = \overline{x}$ and $\sigma = 0$. Exploiting the recursive nature of the problem, define

$$F(x, \sigma) = E_t f(g(h(x, \sigma) + \eta\sigma\varepsilon', \sigma), g(x, \sigma), h(x, \sigma) + \eta\sigma\varepsilon', x) = 0$$

where $E_t f(\cdot)$ denote the set of equilibrium conditions of the model, E_t is the expectation conditional on information known in period t, and the prime indicates variables at time $t+1$.

Using tensor notation, the third-order approximation of $g(x_t, \sigma)$ and $h(x_t, \sigma)$ around $(\overline{x}, 0)$ are of the form

$$\begin{aligned} [g(x,\sigma)]^i &= [g(\overline{x},0)]^i + [g_x(\overline{x},0)]^i_a[(x-\overline{x})]^a + [g_\sigma(\overline{x},0)]^i[\sigma] \\ &+(1/2)[g_{xx}(\overline{x},0)]^i_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b \\ &+(1/2)[g_{x\sigma}(\overline{x},0)]^i_a[(x-\overline{x})]^a[\sigma] \\ &+(1/2)[g_{\sigma x}(\overline{x},0)]^i_a[(x-\overline{x})]^a[\sigma] \\ &+(1/2)[g_{\sigma\sigma}(\overline{x},0)]^i[\sigma][\sigma] \\ &+(1/6)[g_{xxx}(\overline{x},0)]^i_{abc}[(x-\overline{x})]^a[(x-\overline{x})]^b[(x-\overline{x})]^c \\ &+(1/6)[g_{xx\sigma}(\overline{x},0)]^i_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b[\sigma] \\ &+(1/6)[g_{x\sigma x}(\overline{x},0)]^i_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b[\sigma] \\ &+(1/6)[g_{\sigma xx}(\overline{x},0)]^i_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b[\sigma] \end{aligned}$$

[14]In order facilitate the comparison with Schmitt-Grohé and Uribe (2004), I adopt their notation in this Appendix.

$$
\begin{aligned}
&+(1/6)[g_{\sigma\sigma x}(\overline{x},0)]^i_a[(x-\overline{x})]^a[\sigma][\sigma]\\
&+(1/6)[g_{\sigma x\sigma}(\overline{x},0)]^i_a[(x-\overline{x})]^a[\sigma][\sigma]\\
&+(1/6)[g_{x\sigma\sigma}(\overline{x},0)]^i_a[(x-\overline{x})]^a[\sigma][\sigma]\\
&+(1/6)[g_{\sigma\sigma\sigma}(\overline{x},0)]^i[\sigma][\sigma][\sigma],
\end{aligned}
$$

and

$$
\begin{aligned}
[h(x,\sigma)]^j \;=\; & [h(\overline{x},0)]^j+[h_x(\overline{x},0)]^j_a[(x-\overline{x})]^a+[h_\sigma(\overline{x},0)]^j[\sigma]\\
&+(1/2)[h_{xx}(\overline{x},0)]^j_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b\\
&+(1/2)[h_{x\sigma}(\overline{x},0)]^j_a[(x-\overline{x})]^a[\sigma]\\
&+(1/2)[h_{\sigma x}(\overline{x},0)]^j_a[(x-\overline{x})]^a[\sigma]\\
&+(1/2)[h_{\sigma\sigma}(\overline{x},0)]^j[\sigma][\sigma]\\
&+(1/6)[h_{xxx}(\overline{x},0)]^j_{abc}[(x-\overline{x})]^a[(x-\overline{x})]^b[(x-\overline{x})]^c\\
&+(1/6)[h_{xx\sigma}(\overline{x},0)]^j_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b[\sigma]\\
&+(1/6)[h_{x\sigma x}(\overline{x},0)]^j_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b[\sigma]\\
&+(1/6)[h_{\sigma xx}(\overline{x},0)]^j_{ab}[(x-\overline{x})]^a[(x-\overline{x})]^b[\sigma]\\
&+(1/6)[h_{\sigma\sigma x}(\overline{x},0)]^j_a[(x-\overline{x})]^a[\sigma][\sigma]\\
&+(1/6)[h_{\sigma x\sigma}(\overline{x},0)]^j_a[(x-\overline{x})]^a[\sigma][\sigma]\\
&+(1/6)[h_{x\sigma\sigma}(\overline{x},0)]^j_a[(x-\overline{x})]^a[\sigma][\sigma]\\
&+(1/6)[h_{\sigma\sigma\sigma}(\overline{x},0)]^j[\sigma][\sigma][\sigma],
\end{aligned}
$$

where $i=1,\ldots,n_y, a,b,c=1,\ldots,n_x$, and $j=1,\ldots,n_x$. Schmitt-Grohé and Uribe (2004) explain in detail how to compute the terms $[g(\overline{x},0)]$, $[g_x(\overline{x},0)]^i_a$, $[g_\sigma(\overline{x},0)]^i$, $[g_{xx}(\overline{x},0)]^i_{ab}$, $[g_{x\sigma}(\overline{x},0)]^i_a$, $[g_{\sigma x}(\overline{x},0)]^i_a$, $[g_{\sigma\sigma}(\overline{x},0)]^i$, $[h(\overline{x},0)]$, $[h_x(\overline{x},0)]^j_a$, $[h_\sigma(\overline{x},0)]^j$, $[h_{xx}(\overline{x},0)]^j_{ab}$, $[h_{x\sigma}(\overline{x},0)]^j_a$, $[h_{\sigma x}(\overline{x},0)]^j_a$, and $[h_{\sigma\sigma}(\overline{x},0)]^j$ required for the second-order approximation to the policy rules. Such terms are inputs to the computation of the third-order terms which I now derive.

Consider first $g_{xxx}(\overline{x},0)$ and $h_{xxx}(\overline{x},0)$. Since all order derivatives of $F(x,\sigma)$ should equal zero (see Jin and Judd, 2002), the terms $g_{xxx}(\overline{x},0)$ and $h_{xxx}(\overline{x},0)$ can be obtained from the solution to the system

$$
\begin{aligned}
[F_{xxx}(\overline{x},0)]^i_{jkm} \;=\; & [f_{y'y'y'}]^i_{\alpha\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m\left([g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j\right)\\
&+[f_{y'y'y}]^i_{\alpha\gamma\mu}[g_x]^\mu_m\left([g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j\right)\\
&+[f_{y'y'x'}]^i_{\alpha\gamma\nu}[h_x]^\nu_m\left([g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j\right)\\
&+[f_{y'y'x}]^i_{\alpha\gamma m}\left([g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j\right)
\end{aligned}
$$

$$+[f_{y'y'}]^i_{\alpha\gamma}\left(\left([g_{xx}]^\gamma_{\delta\nu}[h_x]^\nu_m[h_x]^\delta_k+[g_x]^\gamma_\delta[h_{xx}]^\delta_{km}\right)[g_x]^\alpha_\beta\right)[h_x]^\beta_j$$

$$+[f_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_x]^\delta_k[g_{xx}]^\alpha_{\beta\nu}[h_x]^\nu_m[h_x]^\beta_j$$

$$+[f_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_\beta[h_{xx}]^\beta_{jm}$$

$$+[f_{y'yy'}]^i_{\alpha\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_k[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'yy}]^i_{\alpha\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_k[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'yx'}]^i_{\alpha\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_k[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'yx}]^i_{\alpha\gamma m}[g_x]^\gamma_k[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'y}]^i_{\alpha\gamma}\left(\left([g_{xx}]^\gamma_{km}[g_x]^\alpha_\beta+[g_x]^\gamma_k[g_{xx}]^\alpha_{\beta\nu}[h_x]^\nu_m\right)[h_x]^\beta_j+[g_x]^\gamma_k[g_x]^\alpha_\beta[h_{xx}]^\beta_{jm}\right)$$

$$+[f_{y'x'y'}]^i_{\alpha\delta\mu}[g_x]^\mu_\nu[h_x]^\nu_m[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j+[f_{y'x'y}]^i_{\alpha\delta\mu}[g_x]^\mu_m[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'x'x'}]^i_{\alpha\delta\nu}[h_x]^\nu_m[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j+[f_{y'x'x}]^i_{\alpha\delta m}[h_x]^\delta_k[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'x'}]^i_{\alpha\delta}\left(\left([h_{xx}]^\delta_{km}[g_x]^\alpha_\beta+[h_x]^\delta_k[g_{xx}]^\alpha_{\beta\nu}[h_x]^\nu_m\right)[h_x]^\beta_j+[h_x]^\delta_k[g_x]^\alpha_\beta[h_{xx}]^\beta_{jm}\right)$$

$$+[f_{y'xy'}]^i_{\alpha k\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\alpha_\beta[h_x]^\beta_j+[f_{y'xy}]^i_{\alpha k\mu}[g_x]^\mu_m[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'xx'}]^i_{\alpha k\nu}[h_x]^\nu_m[g_x]^\alpha_\beta[h_x]^\beta_j+[f_{y'xx}]^i_{\alpha km}[g_x]^\alpha_\beta[h_x]^\beta_j$$

$$+[f_{y'x}]^i_{\alpha k}\left([g_{xx}]^\alpha_{\beta\nu}[h_x]^\nu_m[h_x]^\beta_j+[g_x]^\alpha_\beta[h_{xx}]^\beta_{jm}\right)$$

$$+[f_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_x]^\delta_m[g_{xx}]^\alpha_{\beta\delta}[h_x]^\delta_k[h_x]^\beta_j+[g_x]^\alpha_\beta[h_{xx}]^\beta_{jk}$$

$$+[f_{y'y}]^i_{\alpha\gamma}[g_x]^\gamma_m[g_{xx}]^\alpha_{\beta\delta}[h_x]^\delta_k[h_x]^\beta_j+[g_x]^\alpha_\beta[h_{xx}]^\beta_{jk}$$

$$+[f_{y'x'}]^i_{\alpha\delta}[h_x]^\delta_m[g_{xx}]^\alpha_{\beta\delta}[h_x]^\delta_k[h_x]^\beta_j+[g_x]^\alpha_\beta[h_{xx}]^\beta_{jk}$$

$$+[f_{y'x}]^i_{\alpha m}[g_{xx}]^\alpha_{\beta\delta}[h_x]^\delta_k[h_x]^\beta_j+[g_x]^\alpha_\beta[h_{xx}]^\beta_{jk}$$

$$+[f_{y'}]^i_\alpha\left([g_{xxx}]^\alpha_{\beta\delta\nu}[h_x]^\nu_m[h_x]^\delta_k+[g_{xx}]^\alpha_{\beta\delta}[h_{xx}]^\delta_{km}\right)[h_x]^\beta_j$$

$$+[f_{y'}]^i_\alpha\left([g_{xx}]^\alpha_{\beta\delta}[h_x]^\delta_k[h_{xx}]^\beta_{jm}+[g_{xx}]^\alpha_{\beta\nu}[h_x]^\nu_m[h_{xx}]^\beta_{jk}+[g_x]^\alpha_\beta[h_{xxx}]^\beta_{jkm}\right)$$

$$+[f_{yy'y'}]^i_{\alpha\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_j+[f_{yy'y}]^i_{\alpha\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_j$$

$$+[f_{yy'x'}]^i_{\alpha\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_j+[f_{yy'x}]^i_{\alpha\gamma m}[g_x]^\gamma_\delta[h_x]^\delta_k[g_x]^\alpha_j$$

$$+[f_{yy'}]^i_{\alpha\gamma}\left(\left([g_{xx}]^\gamma_{\delta\nu}[h_x]^\nu_m[h_x]^\delta_k+[g_x]^\gamma_\delta[h_{xx}]^\delta_{km}\right)[g_x]^\alpha_j+[g_x]^\gamma_\delta[h_x]^\delta_k[g_{xx}]^\alpha_{jm}\right)$$

$$+[f_{yyy'}]^i_{\alpha\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_k[g_x]^\alpha_j+[f_{yyy}]^i_{\alpha\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_k[g_x]^\alpha_j$$

$$+[f_{yyx'}]^i_{\alpha\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_k[g_x]^\alpha_j+[f_{yyx}]^i_{\alpha\gamma m}[g_x]^\gamma_k[g_x]^\alpha_j$$

$$+[f_{yy}]^i_{\alpha\gamma}\left(g_{xx}]^\gamma_{km}[g_x]^\alpha_j+[g_x]^\gamma_k[g_{xx}]^\alpha_{jm}\right)$$

$$+[f_{yx'y'}]^i_{\alpha\delta\mu}[g_x]^\mu_\nu[h_x]^\nu_m[h_x]^\delta_k[g_x]^\alpha_j+[f_{yx'y}]^i_{\alpha\delta\mu}[g_x]^\mu_m[h_x]^\delta_k[g_x]^\alpha_j$$

$$+[f_{yx'x'}]^i_{\alpha\delta\nu}[h_x]^\nu_m[h_x]^\delta_k[g_x]^\alpha_j+[f_{yx'x}]^i_{\alpha\delta m}[h_x]^\delta_k[g_x]^\alpha_j$$

$$+[f_{yx'y'}]^i_{\alpha\delta}\left([h_{xx}]^\delta_{km}[g_x]^\alpha_j+[h_x]^\delta_k[g_{xx}]^\alpha_{jm}\right)$$

$$+[f_{yxy'}]^i_{\alpha k\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^a_j+[f_{yxy}]^i_{\alpha k\mu}[g_x]^\mu_m[g_x]^a_j$$

$$
\begin{aligned}
&+[f_{yxx'}]^i_{\alpha k\nu}[h_x]^\nu_m[g_x]^a_j + [f_{yxx}]^i_{\alpha km}[g_x]^a_j \\
&+[f_{yx}]^i_{\alpha k}[g_{xx}]^\alpha_{jm} + [f_{yy'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_x]^\delta_m[g_{xx}]^a_{jk} \\
&+[f_{yy}]^i_{\alpha\gamma}[g_x]^\gamma_m[g_{xx}]^a_{jk} + [f_{yx'}]^i_{\alpha\delta}[h_x]^\delta_m[g_{xx}]^a_{jk} \\
&+[f_{yx}]^i_{\alpha m}[g_{xx}]^a_{jk} + [f_y]^i_\alpha[g_{xxx}]^\alpha_{jkm} \\
&+[f_{x'y'y'}]^i_{\beta\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_\delta[h_x]^\delta_k[h_x]^\beta_j + [f_{x'y'y}]^i_{\beta\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_\delta[h_x]^\delta_k[h_x]^\beta_j \\
&+[f_{x'y'x'}]^i_{\beta\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_\delta[h_x]^\delta_k[h_x]^\beta_j + [f_{x'y'x}]^i_{\beta\gamma m}[g_x]^\gamma_\delta[h_x]^\delta_k[h_x]^\beta_j \\
&+[f_{x'y'}]^i_{\beta\gamma}\left(\left([g_{xx}]^\gamma_{\delta\nu}[h_x]^\nu_m[h_x]^\delta_k + [g_x]^\gamma_\delta[h_{xx}]^\delta_{km}\right)[h_x]^\beta_j + [g_x]^\gamma_\delta[h_x]^\delta_k[h_{xx}]^\beta_{jm}\right) \\
&+[f_{x'yy'}]^i_{\beta\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_k[h_x]^\beta_j + [f_{x'yy}]^i_{\beta\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_k[h_x]^\beta_j \\
&+[f_{x'yx'}]^i_{\beta\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_k[h_x]^\beta_j + [f_{x'yx}]^i_{\beta\gamma m}[g_x]^\gamma_k[h_x]^\beta_j \\
&+[f_{x'y}]^i_{\beta\gamma}\left([g_{xx}]^\gamma_{km}[h_x]^\beta_j + [g_x]^\gamma_k[h_{xx}]^\beta_{jm}\right) \\
&+[f_{x'x'y'}]^i_{\beta\delta\mu}[g_x]^\mu_\nu[h_x]^\nu_m[h_x]^\delta_k[h_x]^\beta_j + [f_{x'x'y}]^i_{\beta\delta\mu}[g_x]^\mu_m[h_x]^\delta_k[h_x]^\beta_j \\
&+[f_{x'x'x'}]^i_{\beta\delta\nu}[h_x]^\nu_m[h_x]^\delta_k[h_x]^\beta_j + [f_{x'x'x}]^i_{\beta\delta m}[h_x]^\delta_k[h_x]^\beta_j \\
&+[f_{x'x'}]^i_{\beta\delta}\left([h_{xx}]^\delta_{km}[h_x]^\beta_j + [h_x]^\delta_k[h_{xx}]^\beta_{jm}\right) \\
&+[f_{x'xy'}]^i_{\beta k\mu}[g_x]^\mu_\nu[h_x]^\nu_m[h_x]^\beta_j + [f_{x'xy}]^i_{\beta k\mu}[g_x]^\mu_m[h_x]^\beta_j \\
&+[f_{x'xx'}]^i_{\beta k\nu}[h_x]^\nu_m[h_x]^\beta_j + [f_{x'xx}]^i_{\beta km}[h_x]^\beta_j + [f_{x'x}]^i_{\beta k}[h_{xx}]^\beta_{jm} \\
&+[f_{x'y'}]^i_{\beta\gamma}[g_x]^\gamma_\delta[h_x]^\delta_m[h_{xx}]^\beta_{jk} + [f_{x'y}]^i_{\beta\gamma}[g_x]^\gamma_m[h_{xx}]^\beta_{jk} \\
&+[f_{x'x'}]^i_{\beta\delta}[h_x]^\delta_m[h_{xx}]^\beta_{jk} + [f_{x'x}]^i_{\beta m}[h_{xx}]^\beta_{jk} + [f_{x'}]^i_\beta[h_{xxx}]^\beta_{jkm} \\
&+[f_{xy'y'}]^i_{j\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_\delta[h_x]^\delta_k + [f_{xy'y}]^i_{j\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_\delta[h_x]^\delta_k \\
&+[f_{xy'x'}]^i_{j\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_\delta[h_x]^\delta_k + [f_{xy'x}]^i_{j\gamma m}[g_x]^\gamma_\delta[h_x]^\delta_k \\
&+[f_{xy'}]^i_{j\gamma}\left([g_{xx}]^\gamma_{\delta\nu}[h_x]^\nu_m[h_x]^\delta_k + [g_x]^\gamma_\delta[h_{xx}]^\delta_{km}\right) \\
&+[f_{xyy'}]^i_{j\gamma\mu}[g_x]^\mu_\nu[h_x]^\nu_m[g_x]^\gamma_k + [f_{xyy}]^i_{j\gamma\mu}[g_x]^\mu_m[g_x]^\gamma_k \\
&+[f_{xyx'}]^i_{j\gamma\nu}[h_x]^\nu_m[g_x]^\gamma_k + [f_{xyx}]^i_{j\gamma m}[g_x]^\gamma_k + [f_{xy}]^i_{j\gamma}[g_{xx}]^\gamma_{km} \\
&+[f_{xx'y'}]^i_{j\delta\mu}[g_x]^\mu_\nu[h_x]^\nu_m[h_x]^\delta_k + [f_{xx'y}]^i_{j\delta\mu}[g_x]^\mu_m[h_x]^\delta_k \\
&+[f_{xx'x'}]^i_{j\delta\nu}[h_x]^\nu_m[h_x]^\delta_k + [f_{xx'x}]^i_{j\delta m}[h_x]^\delta_k + [f_{xx'}]^i_{j\delta}[h_{xx}]^\delta_{km} + \\
&+[f_{xxy'}]^i_{jk\mu}[g_x]^\mu_\nu[h_x]^\nu_m + [f_{xxy}]^i_{jk\mu}[g_x]^\mu_m + [f_{xxx'}]^i_{jk\nu}[h_x]^\nu_m + [f_{xxx}]^i_{jkm} \\
= \quad & 0,
\end{aligned}
$$

where $i = 1, \ldots, n$, $j, k, \beta, \delta, \nu, m = 1, \ldots, n_x$, and $\alpha, \gamma, \mu = 1, \ldots, n_y$. Note that this is a linear system of equations in $g_{xxx}(\overline{x}, 0)$ and $h_{xxx}(\overline{x}, 0)$ and, provided the solvability conditions in Jin and Judd (2002) are met, it has a unique solution.

Similarly, $g_{\sigma xx}(\overline{x},0)$ and $h_{\sigma xx}(\overline{x},0)$ can be obtained from the solution to the system

$$\begin{aligned}[F_{\sigma xx}(\overline{x},0)]^i_{jm} &= [f_{y'}]^i_\alpha[g_x]^\alpha_\beta[h_{\sigma xx}]^\beta_{jm} + [f_{y'}]^i_\alpha[g_{\sigma xx}]^\alpha_{\gamma\nu}[h_x]^\nu_m[h_x]^\gamma_j + [f_y]^i_\alpha[g_{\sigma xx}]^\alpha_{jm} + [f_{x'}]^i_\beta[h_{\sigma xx}]^\alpha_{jm} \\ &= 0,\end{aligned}$$

where $i = 1,\dots,n,\ j,\beta,\nu,m = 1,\dots,n_x,\ \alpha,\gamma = 1,\dots,n_y$, and terms whose conditional expectation is zero have been dropped to save space. Note that this is a homogeneous system whose solution, if it exists, is $g_{\sigma xx} = h_{\sigma xx} = 0$. It follows that $g_{x\sigma x} = h_{x\sigma x} = g_{xx\sigma} = h_{xx\sigma} = 0$, as well.

Consider now $g_{x\sigma\sigma}(\overline{x},0)$ and $h_{x\sigma\sigma}(\overline{x},0)$. Again, these terms can be obtained from the solution to the system

$$\begin{aligned}[F_{x\sigma\sigma}(\overline{x},0)]^i_{jm} &= [f_{y'y'y'}]^i_{\alpha\gamma\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_\beta[H_x]^\beta_j \\
&+[f_{y'y'x'}]^i_{\alpha\delta\nu}[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_\beta[H_x]^\beta_j \\
&+[f_{y'y'}]^i_{\alpha\gamma}\left([g_x]^\gamma_\delta[h_{\sigma\sigma}]^\delta[g_x]^\alpha_\beta[H_x]^\beta_j + [g_{xx}]^\gamma_{\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_\beta[H_x]^\beta_j\right) \\
&+[f_{y'y'}]^i_{\alpha\gamma}\left([g_x]^\gamma_\delta[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\nu}[\eta]^\nu_\varsigma[I]^\xi_\varsigma[H_x]^\beta_j + [g_{\sigma\sigma}]^\mu[g_x]^\alpha_\beta\right) \\
&+[f_{y'y}]^i_{\alpha\gamma}[g_{\sigma\sigma}]^\gamma[g_x]^\alpha_\beta[H_x]^\beta_j \\
&+[f_{y'x'}]^i_{\alpha\delta}[h_{\sigma\sigma}]^\delta[g_x]^\alpha_\beta[H_x]^\beta_j \\
&+[f_{y'x'y'}]^i_{\alpha\delta\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_\beta[H_x]^\beta_j \\
&+[f_{y'x'x'}]^i_{\alpha\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_\beta[H_x]^\beta_j \\
&+[f_{y'x'}]^i_{\alpha\delta}[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\nu}[\eta]^\nu_\varsigma[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\nu}[\eta]^\nu_\varsigma[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{y'x'}]^i_{\alpha\delta}[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\nu}[\eta]^\nu_\varsigma[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{y'}]^i_\alpha\left([g_{xx}]^\alpha_{\beta\nu}[h_{\sigma\sigma}]^\nu[H_x]^\beta_j + [g_{xxx}]^\alpha_{\beta\delta\nu}[\eta]^\nu_\varsigma[\eta]^\nu_\varsigma[I]^\xi_\varsigma[H_x]^\beta_j\right) \\
&+[f_{y'}]^i_\alpha\left([g_{x\sigma\sigma}]^\alpha_\beta[H_x]^\beta_j + [g_x]^\alpha_\beta[H_{x\sigma\sigma}]^\beta_j\right) \\
&+[f_{yy'y'}]^i_{\alpha\gamma\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_j \\
&+[f_{yy'x'}]^i_{\alpha\delta\nu}[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_j \\
&+[f_{yy'}]^i_{\alpha\gamma}\left([g_x]^\gamma_\delta[h_{\sigma\sigma}]^\delta[g_x]^\alpha_j + [g_{xx}]^\gamma_{\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_j + [g_{\sigma\sigma}]^\mu[g_x]^\alpha_j\right) \\
&+[f_{yy}]^i_{\alpha\gamma}[g_{\sigma\sigma}]^\gamma[g_x]^\alpha_j \\
&+[f_{yx'y'}]^i_{\alpha\delta\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_j \\
&+[f_{yx'x'}]^i_{\alpha\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[g_x]^\alpha_j \\
&+[f_{yx'}]^i_{\alpha\delta}[h_{\sigma\sigma}]^\delta[g_x]^\alpha_j \\
&+[f_y]^i_\alpha[g_{x\sigma\sigma}]^\alpha_j\end{aligned}$$

$$
\begin{aligned}
&+[f_{x'y'y'}]^i_{\beta\gamma\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{x'y'x'}]^i_{\beta\gamma\nu}[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{x'y'}]^i_{\beta\gamma}\left([g_x]^\gamma_\delta[h_{\sigma\sigma}]^\delta[H_x]^\beta_j+[g_{xx}]^\gamma_{\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[I]^\xi_\varsigma[H_x]^\beta_j+[g_{\sigma\sigma}]^\mu[H_x]^\beta_j\right) \\
&+[f_{x'y}]^i_{\beta\gamma}[g_{\sigma\sigma}]^\gamma[H_x]^\beta_j \\
&+[f_{x'x'y'}]^i_{\beta\delta\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma\eta]^\delta_\xi[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{x'x'x'}]^i_{\beta\delta\nu}[\eta]^\nu_\varsigma\eta]^\delta_\xi[I]^\xi_\varsigma[H_x]^\beta_j \\
&+[f_{x'x'}]^i_{\beta\delta}[h_{\sigma\sigma}]^\delta[H_x]^\beta_j \\
&+[f_{x'}]^i_\beta[H_{x\sigma\sigma}]^\beta_j \\
=\ & 0,
\end{aligned}
$$

where $i=1,\ldots,n$, $j,\beta,\nu=1,\ldots,n_x$, $\alpha,\gamma=1,\ldots,n_y$, and terms whose conditional expectation is zero have been dropped to save space.

Finally, $g_{\sigma\sigma\sigma}(\overline{x},0)$ and $h_{\sigma\sigma\sigma}(\overline{x},0)$ can be obtained from the solution to the system

$$
\begin{aligned}
[F_{\sigma\sigma\sigma}(\overline{x},0)]^i_j =\ & [f_{y'y'y'}]^i_{\alpha\gamma\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[g_x]^\alpha_\beta[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{y'y'x'}]^i_{\alpha\delta\nu}[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[g_x]^\alpha_\beta[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{y'y'}]^i_{\alpha\gamma}\left([g_{xx}]^\gamma_{\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[g_x]^\alpha_\beta[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi}+[g_x]^\gamma_\delta[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\delta}[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi}\right) \\
&+[f_{y'x'y'}]^i_{\alpha\delta\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[\eta]^\delta_\xi[g_x]^\alpha_\beta[\eta]^\beta_\phi[\eta][S]^\varsigma_{\xi\phi} \\
&+[f_{y'x'x'}]^i_{\alpha\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[g_x]^\alpha_\beta[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{y'x'}]^i_{\alpha\delta}[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\nu}[\eta]^\nu_\varsigma[\eta]^\beta_\phi[S]^\xi_{\varsigma\phi} \\
&+[f_{y'}]^i_\alpha[g_x]^\alpha_\beta[h_{\sigma\sigma\sigma}]^\beta \\
&+[f_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\delta}[\eta]^\delta_\varsigma[\eta]^\beta_\phi[S]^\xi_{\varsigma\phi} \\
&+[f_{y'x'}]^i_{\alpha\delta}[\eta]^\delta_\xi[g_{xx}]^\alpha_{\beta\delta}[\eta]^\delta_\varsigma[\eta]^\beta_\phi[S]^\xi_{\varsigma\phi} \\
&+[f_{y'}]^i_\alpha[g_{xxx}]^\alpha_{\beta\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{y'}]^i_\alpha[g_{\sigma\sigma\sigma}]^\alpha \\
&+[f_y]^i_{\alpha\mu}[g_{\sigma\sigma\sigma}]^\mu \\
&+[f_{x'y'y'}]^i_{\beta\gamma\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{x'y'x'}]^i_{\beta\gamma\nu}[\eta]^\nu_\varsigma[g_x]^\gamma_\delta[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{x'y'}]^i_{\beta\gamma}[g_{xx}]^\gamma_{\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{x'x'y'}]^i_{\beta\delta\vartheta}[g_x]^\vartheta_\nu[\eta]^\nu_\varsigma[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\xi_{\varsigma\phi} \\
&+[f_{x'x'x'}]^i_{\beta\delta\nu}[\eta]^\nu_\varsigma[\eta]^\delta_\xi[\eta]^\beta_\phi[S]^\varsigma_{\xi\phi} \\
&+[f_{x'}]^i_\beta[h_{\sigma\sigma\sigma}]^\beta
\end{aligned}
$$

$$= 0$$

where $i = 1, \ldots, n$, $j, k, \beta, \delta, \nu, m = 1, \ldots, n_x$, and $\alpha, \gamma, \mu = 1, \ldots, n_y$, and terms whose conditional expectation is zero have been dropped to save space. In the especial case where all elements along the of S are zero, meaning that the distributions of all structural shocks are symmetric, then the system above is homogeneous and the solution if it exists, is $g_{\sigma\sigma\sigma} = h_{\sigma\sigma\sigma} = 0$. More generally, however, when the distribution of shock is skewed, then solution to the system will deliver non-zero coefficients.

A.2 MATLAB Codes

MATLAB codes to compute the coefficients of the third-order polynomial approximation to the policy rules are available from my Web site (*www.cireq.umontreal.ca/personnel/ruge.html*). These codes are general in that they may be adapted to any DSGE model and are compatible with those written by Stephanie Schmitt-Grohé and Martin Uribe for a second-order approximation. Note that their codes are used to produced the first- and second-order terms used as input for my third-order approximation.

B Notes to Tables

Notes to Table 1: Mean is the average of the estimated parameter values; A.S.E. is the median asymptotic standard error; Median and S.D. are, respectively, the median and standard deviation of the empirical parameter distribution; Size is the empirical size of the t test, O.I. is the empirical size of the chi-square test of the overidentification restrictions, and S.E. is the standard error of the empirical test size.

Notes to Table 2: See notes to Table 1.

Notes to Table 3: See notes to Table 1.

Notes to Table 4: See notes to Table 1.

Notes to Table 5: See notes to Table 1.

Notes to Table 6: See notes to Table 1.

Notes to Table 7: See notes to Table 1.

Notes to Table 8: The superscripts $*$ and $\dagger$ denote the rejection of the null hypothesis of a unit root at the five and ten percent significance levels respectively. The alternative is a stationary autoregression with a constant term.

Notes to Table 9: The superscripts $*$ and $\dagger$ denote statistical significance at the five and ten percent significance levels respectively. The location parameter of the Skew normal distribution was fixed to zero.

Notes to Table 10: The medians predicted by the model were computed using simulated samples of 5000 observations.

Notes to Table 11: The moments of the U.S. data were computed using quarterly observation for the period 1959:Q2 to 2010:Q2. The moments predicted by the model were computed using simulated samples of 5000 observations. The superscript $*$ denotes the rejection of the null hypothesis of Normality at the five percent significance level.

Table 1. Basic Model

$\beta = 0.95$		$\rho = 0.85$		$\sigma = 0.04$		$\gamma = 1, 2, 5$		
Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	O.I. S.E.
				$\gamma = 1, \tau = 5$				
0.9498	0.9502	0.8307	0.8355	0.0400	0.0399	1.0103	1.0080	
0.0103	0.0084	0.0252	0.0369	0.0027	0.0031	0.0553	0.0400	0.0050
0.0250	0.0110	0.1900	0.0277	0.0950	0.0207	0.0100	0.0070	0.0050
				$\gamma = 1, \tau = 10$				
0.9477	0.9499	0.8315	0.8358	0.0386	0.0386	0.9892	0.9888	
0.0112	0.0086	0.0253	0.0402	0.0025	0.0026	0.0544	0.0389	0.0150
0.0250	0.0110	0.2000	0.0283	0.0900	0.0202	0.0150	0.0086	0.0086
				$\gamma = 1, \tau = 20$				
0.9474	0.9499	0.8284	0.8324	0.0395	0.0385	0.9835	0.9849	
0.0104	0.0084	0.0252	0.0399	0.0024	0.0029	0.0528	0.0384	0.0050
0.0200	0.0099	0.2100	0.0288	0.1750	0.0269	0.0550	0.0161	0.0050
				$\gamma = 2, \tau = 5$				
0.9490	0.9501	0.8339	0.8405	0.0396	0.0394	2.0028	2.0023	
0.0134	0.0079	0.0302	0.0317	0.0032	0.0040	0.1414	0.0619	0.0150
0.0100	0.0070	0.0900	0.0202	0.1100	0.0221	0.0000	0.0000	0.0086
				$\gamma = 2, \tau = 10$				
0.9494	0.9501	0.8377	0.8399	0.0381	0.0383	1.9942	1.9939	
0.0143	0.0073	0.0305	0.0339	0.0031	0.0032	0.1448	0.0614	0.0200
0.0000	0.0000	0.0700	0.0180	0.1350	0.0242	0.0000	0.0000	0.0099
				$\gamma = 2, \tau = 20$				
0.9473	0.9500	0.8288	0.8343	0.0379	0.0379	1.9802	1.9857	
0.0140	0.0078	0.0303	0.0374	0.0030	0.0034	0.1404	0.0647	0.0150
0.0100	0.0070	0.1350	0.0242	0.2000	0.0283	0.0000	0.0000	0.0086
				$\gamma = 5, \tau = 5$				
0.9489	0.9502	0.8330	0.8386	0.0382	0.0378	4.9945	4.9901	
0.0183	0.0101	0.0493	0.0401	0.0044	0.0041	0.4309	0.2405	0.0200
0.0150	0.0086	0.0700	0.0180	0.0950	0.0207	0.0100	0.0070	0.0099
				$\gamma = 5, \tau = 10$				
0.9511	0.9503	0.8395	0.8394	0.0368	0.0367	5.0175	4.9825	
0.0160	0.0115	0.0428	0.0423	0.0037	0.0041	0.3789	0.2785	0.0050
0.0100	0.0070	0.0850	0.0197	0.2050	0.0285	0.0000	0.0000	0.0050
				$\gamma = 5, \tau = 20$				
0.9516	0.9503	0.8425	0.8422	0.0366	0.0365	5.0452	5.0098	
0.0161	0.0110	0.0420	0.0419	0.0038	0.0037	0.3890	0.2665	0.0050
0.0100	0.0070	0.0750	0.0186	0.1700	0.0266	0.0050	0.0050	0.0050

Table 2. Habit Formation

$\beta = 0.95$		$\rho = 0.85$		$\sigma = 0.04$		$\gamma = 1, 2, 5$		$a = 0.8$		
Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	
A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.	O.I.
Size	S.E.	Size	S.E.	Size	S.E.	Size	S.E.	Size	S.E.	S.E.
				$\gamma = 1$, $\tau = 5$						
0.9484	0.9500	0.8325	0.8353	0.0393	0.0390	1.0474	1.0463	0.7360	0.7667	
0.0140	0.0059	0.0261	0.0304	0.0039	0.0041	0.1647	0.1723	0.1070	0.1365	0.0050
0.0000	0.0000	0.1150	0.0226	0.0600	0.0168	0.0300	0.0121	0.0150	0.0086	0.0050
				$\gamma = 1$, $\tau = 10$						
0.9475	0.9499	0.8304	0.8373	0.0384	0.0375	1.0119	1.0124	0.7642	0.7886	
0.0118	0.0060	0.0257	0.0368	0.0038	0.0043	0.1321	0.1554	0.0791	0.1072	0.0050
0.0000	0.0000	0.1400	0.0245	0.1100	0.0221	0.0300	0.0121	0.0250	0.0110	0.0050
				$\gamma = 1$, $\tau = 20$						
0.9470	0.9498	0.8302	0.8378	0.0392	0.0390	0.9905	0.9880	0.7696	0.7990	
0.0115	0.0053	0.0262	0.0334	0.0038	0.0044	0.1370	0.1757	0.0746	0.1219	0.0000
0.0000	0.0000	0.1450	0.0249	0.1200	0.0230	0.0300	0.0121	0.0300	0.0121	0.0000
				$\gamma = 2$, $\tau = 5$						
0.9463	0.9501	0.8310	0.8419	0.0390	0.0387	2.0466	2.0472	0.7609	0.7869	
0.0182	0.0102	0.0470	0.0358	0.0062	0.0056	0.1744	0.2342	0.0646	0.0964	0.0050
0.0250	0.0110	0.0650	0.0174	0.0650	0.0174	0.1000	0.0212	0.1150	0.0226	0.0050
				$\gamma = 2$, $\tau = 10$						
0.9475	0.9501	0.8319	0.8395	0.0373	0.0365	1.9748	1.9742	0.7886	0.8005	
0.0160	0.0076	0.0437	0.0319	0.0058	0.0049	0.1761	0.1958	0.0479	0.0753	0.0050
0.0000	0.0000	0.0200	0.0099	0.1000	0.0212	0.0350	0.0130	0.0750	0.0186	0.0050
				$\gamma = 2, \tau = 20$						
0.9473	0.9500	0.8336	0.8416	0.0379	0.0378	1.9938	2.0065	0.7833	0.7921	
0.0151	0.0082	0.0415	0.0336	0.0058	0.0051	0.1736	0.2099	0.0482	0.0717	0.0100
0.0300	0.0121	0.0700	0.0180	0.1000	0.0212	0.0400	0.0139	0.1050	0.0217	0.0070
				$\gamma = 5, \tau = 5$						
0.9475	0.9501	0.8384	0.8454	0.0383	0.0385	5.0675	5.1183	0.7800	0.7859	
0.0208	0.0119	0.0601	0.0365	0.0073	0.0058	0.4770	0.3804	0.0453	0.0644	0.0000
0.0350	0.0130	0.0350	0.0130	0.0500	0.0154	0.0200	0.0099	0.1000	0.0212	0.0000
				$\gamma = 5, \tau = 10$						
0.9480	0.9500	0.8388	0.8468	0.0368	0.0363	5.0049	5.0214	0.7892	0.7962	
0.0172	0.0114	0.0499	0.0395	0.0062	0.0059	0.5000	0.3296	0.0344	0.0599	0.0050
0.0050	0.0050	0.0100	0.0070	0.1150	0.0226	0.0250	0.0110	0.1650	0.0262	0.0050
				$\gamma = 5, \tau = 20$						
0.9484	0.9501	0.8403	0.8470	0.0368	0.0364	4.9848	4.9847	0.7928	0.7986	
0.0171	0.0108	0.0500	0.0355	0.0063	0.0052	0.5062	0.3335	0.0355	0.0559	0.0050
0.0100	0.0070	0.0150	0.0086	0.0850	0.0197	0.0200	0.0099	0.1600	0.0259	0.0050

Table 3. Habit Formation and High Variance

$\beta = 0.95$		$\rho = .85$		$\sigma = 0.08$		$\gamma = 1, 2, 5$		$a = 0.8$		
Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	O.I. S.E.
				$\gamma = 1, \tau = 5$						
0.9484	0.9500	0.8340	0.8370	0.0789	0.0786	1.0414	1.0474	0.7320	0.7795	
0.0147	0.0063	0.0269	0.0331	0.0080	0.0092	0.1678	0.1977	0.1051	0.1628	0.0100
0.0000	0.0000	0.1250	0.0234	0.0900	0.0202	0.0450	0.0147	0.0300	0.0121	0.0070
				$\gamma = 1, \tau = 10$						
0.9475	0.9499	0.8331	0.8387	0.0776	0.0768	1.0077	1.0078	0.7635	0.7898	
0.0119	0.0049	0.0258	0.0315	0.0078	0.0090	0.1288	0.1633	0.0776	0.1211	0.0150
0.0000	0.0000	0.1050	0.0217	0.1100	0.0221	0.0450	0.0147	0.0450	0.0147	0.0086
				$\gamma = 1, \tau = 20$						
0.9478	0.9499	0.8284	0.8332	0.0767	0.0762	0.9958	0.9919	0.7746	0.7920	
0.0115	0.0047	0.0254	0.0332	0.0075	0.0085	0.1280	0.1387	0.0719	0.0916	0.0000
0.0000	0.0000	0.1500	0.0252	0.1250	0.0234	0.0150	0.0086	0.0200	0.0099	0.0000
				$\gamma = 2, \tau = 5$						
0.9470	0.9500	0.8334	0.8426	0.0784	0.0769	2.0420	2.0561	0.7702	0.7877	
0.0184	0.0097	0.0473	0.0352	0.0124	0.0109	0.1708	0.2290	0.0640	0.0879	0.0000
0.0150	0.0086	0.0400	0.0139	0.0450	0.0147	0.0800	0.0192	0.1050	0.0217	0.0000
				$\gamma = 2, \tau = 10$						
0.9480	0.9501	0.8371	0.8428	0.0756	0.0745	1.9698	1.9781	0.7926	0.8023	
0.0157	0.0085	0.0425	0.0313	0.0116	0.0094	0.1779	0.2102	0.0468	0.0749	0.0100
0.0100	0.0070	0.0350	0.0130	0.0750	0.0186	0.0600	0.0168	0.1050	0.0217	0.0070
				$\gamma = 2, \tau = 20$						
0.9498	0.9502	0.8404	0.8446	0.0741	0.0731	1.9902	2.0035	0.7906	0.7995	
0.0161	0.0029	0.0430	0.0180	0.0113	0.0113	0.1758	0.2206	0.0467	0.0648	0.0000
0.0100	0.0070	0.0100	0.0070	0.1450	0.0249	0.0550	0.0161	0.0450	0.0147	0.0000
				$\gamma = 5, \tau = 5$						
0.9507	0.9501	0.8491	0.8484	0.0753	0.0751	5.1258	5.1496	0.7828	0.7871	
0.0224	0.0046	0.0629	0.0149	0.0153	0.0132	0.5427	0.5297	0.0480	0.0675	0.0250
0.0000	0.0000	0.0050	0.0050	0.0500	0.0154	0.0150	0.0086	0.0000	0.0000	0.0110
				$\gamma = 5, \tau = 10$						
0.9507	0.9501	0.8493	0.8484	0.0708	0.0700	5.0637	5.0721	0.7892	0.7919	
0.0177	0.0054	0.0520	0.0179	0.0129	0.0109	0.5538	0.4670	0.0375	0.0579	0.0100
0.0100	0.0070	0.0150	0.0086	0.1050	0.0217	0.0350	0.0130	0.0450	0.0147	0.0070
				$\gamma = 5, \tau = 20$						
0.9505	0.9501	0.8482	0.8490	0.0711	0.0711	5.0593	5.0410	0.7890	0.7913	
0.0177	0.0047	0.0507	0.0176	0.0126	0.0137	0.5593	0.4913	0.0379	0.0612	0.0150
0.0050	0.0050	0.0050	0.0050	0.1550	0.0256	0.0150	0.0086	0.0350	0.0130	0.0086

Table 4. Diagonal Weighting Matrix

$\beta = 0.95$		$\rho = .85$		$\sigma = 0.04$		$\gamma = 1, 2, 5$	
Mean	Median	Mean	Median	Mean	Median	Mean	Median
A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.
Size	S.E.	Size	S.E.	Size	S.E.	Size	S.E.
			$\gamma = 1, \tau = 5$				
0.9507	0.9506	0.8247	0.8325	0.0394	0.0390	0.9699	0.9912
0.0145	0.0105	0.0328	0.0753	0.0054	0.0050	0.1394	0.1340
0.0500	0.0154	0.1800	0.0272	0.0350	0.0130	0.0150	0.0086
			$\gamma = 1, \tau = 10$				
0.9498	0.9501	0.8299	0.8444	0.0373	0.0373	0.9624	0.9863
0.0166	0.088	0.0362	0.0662	0.0055	0.0050	0.1408	0.1269
0.0200	0.0099	0.1650	0.0262	0.0800	0.0192	0.0100	0.0070
			$\gamma = 1, \tau = 20$				
0.9486	0.9501	0.8293	0.8378	0.0380	0.0381	0.9634	0.9820
0.0164	0.0120	0.0341	0.0574	0.0054	0.0046	0.1385	0.1143
0.0550	0.0161	0.1550	0.0256	0.01550	0.0256	0.0050	0.0050
			$\gamma = 2, \tau = 5$				
0.9510	0.9505	0.8319	0.8343	0.0394	0.0394	1.9706	1.9897
0.0197	0.0044	0.0377	0.0343	0.0062	0.0048	0.1787	0.0963
0.0200	0.0099	0.0550	0.0161	0.0550	0.0161	0.0050	0.0050
			$\gamma = 2, \tau = 10$				
0.9504	0.9502	0.8398	0.8437	0.0374	0.0372	1.9655	1.9804
0.0228	0.0036	0.0405	0.0347	0.0063	0.0044	0.1720	0.0883
0.0100	0.0070	0.0350	0.0130	0.0350	0.0130	0.0050	0.0050
			$\gamma = 2, \tau = 20$				
0.9503	0.9502	0.8387	0.8430	0.0376	0.0374	1.9613	1.9735
0.0219	0.0025	0.0394	0.0312	0.0061	0.0048	0.1660	0.0881
0.0050	0.0050	0.0200	0.0099	0.0800	0.0192	0.0050	0.0050
			$\gamma = 5, \tau = 5$				
0.9503	0.9500	0.8364	0.8498	0.0391	0.0386	4.9465	4.9974
0.0230	0.0008	0.0552	0.0291	0.0060	0.0061	0.4684	0.1770
0.0000	0.0000	0.0000	0.0000	0.0650	0.0174	0.0050	0.0050
			$\gamma = 5, \tau = 10$				
0.9502	0.9500	0.8410	0.8499	0.0374	0.0371	4.9314	4.9619
0.0224	0.0007	0.0500	0.0253	0.0052	0.0053	0.4419	0.1719
0.0000	0.0000	0.0000	0.0000	0.1200	0.0230	0.0000	0.0000
			$\gamma = 5, \tau = 20$				
0.9502	0.9500	0.8392	0.8500	0.0375	0.0369	4.9386	4.9643
0.0218	0.0006	0.0496	0.0231	0.0051	0.0056	0.4363	0.1557
0.0000	0.0000	0.0000	0.0000	0.1050	0.0217	0.0000	0.0000

Table 5. Identity Weighting Matrix

$\beta = 0.95$		$\rho = .85$		$\sigma = 0.04$		$\gamma = 1, 2, 5$	
Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.
			$\gamma = 1, \tau = 5$				
0.9519	0.9510	0.8188	0.8284	0.0387	0.0385	0.9550	0.9822
0.0138	0.0125	0.0332	0.0711	0.0050	0.0055	0.1425	0.1578
0.0550	0.0161	0.1400	0.0245	0.0950	0.0207	0.0250	0.0110
			$\gamma = 1, \tau = 10$				
0.9514	0.9505	0.8287	0.8389	0.0361	0.0364	0.9513	0.9801
0.0156	0.0102	0.0336	0.0675	0.0049	0.0046	0.1366	0.1369
0.0300	0.0121	0.1650	0.0262	0.1250	0.0234	0.0150	0.0086
			$\gamma = 1, \tau = 20$				
0.9521	0.9503	0.8348	0.8472	0.0367	0.0368	0.9405	0.9782
0.0150	0.0120	0.0329	0.0623	0.0048	0.0047	0.1361	0.1338
0.0750	0.0186	0.1600	0.0259	0.1100	0.0221	0.0100	0.0070
			$\gamma = 2, \tau = 5$				
0.9503	0.9504	0.8336	0.8363	0.0396	0.0393	1.9937	2.0044
0.0207	0.0008	0.0396	0.0302	0.0065	0.0047	0.1759	0.0772
0.0000	0.0000	0.0100	0.0070	0.0400	0.0139	0.0000	0.0000
			$\gamma = 2, \tau = 10$				
0.9500	0.9501	0.8413	0.8455	0.0379	0.0374	1.9774	1.9901
0.0242	0.0009	0.0434	0.0341	0.0068	0.0040	0.1691	0.0806
0.0000	0.0000	0.0100	0.0070	0.0400	0.0139	0.0000	0.0000
			$\gamma = 2, \tau = 20$				
0.9501	0.9502	0.8367	0.8401	0.0373	0.0373	1.9636	1.9849
0.0232	0.0008	0.0427	0.0341	0.0065	0.0043	0.1678	0.0980
0.0000	0.0000	0.0250	0.0110	0.0650	0.0174	0.0050	0.0050
			$\gamma = 5, \tau = 5$				
0.9500	0.9500	0.8483	0.8500	0.0370	0.0370	5.1125	5.0029
0.0343	0.0003	0.0792	0.0159	0.0092	0.0066	0.5526	0.2706
0.0000	0.0000	0.0000	0.0000	0.0350	0.0130	0.0000	0.0000
			$\gamma = 5, \tau = 10$				
0.9499	0.9500	0.8514	0.8500	0.0366	0.0368	5.0366	5.0000
0.0353	0.0003	0.0755	0.0145	0.0090	0.0066	0.5003	0.2513
0.0000	0.0000	0.0000	0.0000	0.0600	0.0168	0.0100	0.0070
			$\gamma = 5, \tau = 20$				
0.9499	0.9500	0.8516	0.8500	0.0368	0.0372	5.0333	5.0000
0.0340	0.0003	0.0734	0.0112	0.0088	0.0056	0.4869	0.2151
0.0000	0.0000	0.0000	0.0000	0.0200	0.0099	0.0000	0.0000

Table 6. Third-Order Approximation

$\beta = 0.95$		$\rho = 0.85$		$\sigma = 0.04$		$\gamma = 1, 2, 5$		
Mean	Median	Mean	Median	Mean	Median	Mean	Median	
A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.	A.S.E.	S.D.	O.I.
Size	S.E.	Size	S.E.	Size	S.E.	Size	S.E.	S.E.
				$\gamma = 1, \tau = 5$				
0.9495	0.9503	0.8249	0.8256	0.0399	0.0396	1.0014	0.9987	
0.0108	0.0072	0.0263	0.0385	0.0027	0.0032	0.0568	0.0408	0.0250
0.0150	0.0086	0.2250	0.0295	0.0900	0.0202	0.0350	0.0130	0.0110
				$\gamma = 1, \tau = 10$				
0.9487	0.9500	0.8290	0.8352	0.0381	0.0380	0.9922	0.9950	
0.0111	0.0049	0.0260	0.0392	0.0025	0.0029	0.0547	0.0309	0.0150
0.0150	0.0086	0.2000	0.0283	0.1950	0.0280	0.0250	0.0110	0.0086
				$\gamma = 1, \tau = 20$				
0.9484	0.9501	0.0860	0.8310	0.0379	0.0380	0.9903	0.9937	
0.0109	0.0058	0.0254	0.0416	0.0025	0.0031	0.0561	0.0332	0.0250
0.0150	0.0086	0.2300	0.0298	0.2300	0.0298	0.0200	0.0099	0.0110
				$\gamma = 2, \tau = 5$				
0.9500	0.9503	0.8325	0.8336	0.0392	0.0394	2.0054	2.0015	
0.0142	0.0021	0.0303	0.0287	0.0032	0.0039	0.1468	0.0477	0.0400
0.0000	0.0000	0.0500	0.0154	0.1500	0.0252	0.0000	0.0000	0.0139
				$\gamma = 2, \tau = 10$				
0.9498	0.9501	0.8358	0.8399	0.0379	0.0374	1.9928	1.9920	
0.0143	0.0031	0.0307	0.0279	0.0031	0.0035	0.1451	0.0472	0.0200
0.0000	0.0000	0.0650	0.0174	0.1550	0.0256	0.0000	0.0000	0.0099
				$\gamma = 2, \tau = 20$				
0.9497	0.9501	0.8351	0.8412	0.0373	0.0373	1.9970	1.9990	
0.0138	0.0030	0.0302	0.0315	0.0030	0.0038	0.1431	0.0484	0.0100
0.0100	0.0070	0.0650	0.0174	0.2400	0.0302	0.0000	0.0000	0.0070
				$\gamma = 5, \tau = 5$				
0.9499	0.9502	0.8368	0.8408	0.0393	0.0392	4.9797	4.9838	
0.0184	0.0061	0.0491	0.0272	0.0045	0.0042	0.4409	0.1635	0.0250
0.0100	0.0070	0.0150	0.0086	0.0800	0.0192	0.0000	0.0000	0.0110
				$\gamma = 5, \tau = 10$				
0.9503	0.9502	0.8426	0.8449	0.0374	0.0372	5.0083	4.9976	
0.0164	0.0026	0.0423	0.0210	0.0038	0.0042	0.3097	0.1175	0.0150
0.0050	0.0050	0.0050	0.0050	0.1450	0.0249	0.0000	0.0000	0.0086
				$\gamma = 5, \tau = 20$				
0.9507	0.9502	0.8430	0.8439	0.0368	0.0365	5.0350	5.0128	
0.0161	0.0045	0.0429	0.0244	0.0037	0.0040	0.3869	0.1833	0.0250
0.0050	0.0050	0.0150	0.0086	0.1700	0.0266	0.0000	0.0000	0.0110

Table 7. Incorporationg Priors

$\beta = 0.95$		$\rho = 0.85$		$\sigma = 0.04$		$\gamma = 1, 2, 5$		
Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	Mean A.S.E. Size	Median S.D. S.E.	O.I. S.E.
				$\gamma = 1, \tau = 5$				
0.9504	0.9500	0.8497	0.8497	0.0381	0.0383	1.000	1.000	
0.0021	0.0017	0.0024	0.0005	0.0021	0.0047	0.0024	0.0003	0.2450
0.0300	0.0121	0.0000	0.0000	0.3900	0.0345	0.0000	0.0000	0.0304
				$\gamma = 1, \tau = 10$				
0.9505	0.9504	0.8497	0.8498	0.0367	0.0370	0.9999	1.0000	
0.0020	0.0016	0.0023	0.0005	0.0019	0.0042	0.0024	0.0002	0.2150
0.0250	0.0110	0.0000	0.0000	0.4350	0.0351	0.0000	0.0000	0.0290
				$\gamma = 1, \tau = 20$				
0.9507	0.9506	0.8497	0.8497	0.0362	0.0368	0.9999	0.9999	
0.0020	0.0018	0.0023	0.0005	0.0019	0.0048	0.0023	0.0003	0.2300
0.0450	0.0147	0.0000	0.0000	0.4500	0.0352	0.0000	0.0000	0.0298
				$\gamma = 2, \tau = 5$				
0.9507	0.9505	0.8496	0.8496	0.0383	0.0387	2.0000	2.0000	
0.0021	0.0018	0.0024	0.0007	0.0021	0.0052	0.0024	0.0001	0.2600
0.0500	0.0154	0.0000	0.0000	0.3250	0.0331	0.0000	0.0000	0.0310
				$\gamma = 2, \tau = 10$				
0.9508	0.9505	0.8497	0.8498	0.0370	0.0375	2.0000	2.0000	
0.0020	0.0018	0.0023	0.0007	0.0019	0.0041	0.0023	0.0001	0.1950
0.0500	0.0154	0.0000	0.0000	0.4350	0.0351	0.0000	0.0000	0.0280
				$\gamma = 2, \tau = 20$				
0.9509	0.9507	0.8495	0.8496	0.0362	0.0365	2.0000	2.0000	
0.0020	0.0016	0.0023	0.0006	0.0018	0.0046	0.0023	0.0001	0.1850
0.0400	0.0139	0.0000	0.0000	0.5100	0.0353	0.0000	0.0000	0.0275
				$\gamma = 5, \tau = 5$				
0.9507	0.9506	0.8494	0.8494	0.0376	0.0378	5.0000	5.0000	
0.0019	0.0020	0.0024	0.0008	0.0021	0.0045	0.0024	0.0001	0.2150
0.0550	0.0161	0.0000	0.0000	0.3800	0.0343	0.0000	0.0000	0.0290
				$\gamma = 5, \tau = 10$				
0.9509	0.9508	0.8494	0.8495	0.0365	0.0368	5.0000	5.0000	
0.0018	0.0015	0.0023	0.0008	0.0019	0.0040	0.0023	0.0001	0.1900
0.0600	0.0168	0.0000	0.0000	0.4600	0.0352	0.0000	0.0000	0.0277
				$\gamma = 5, \tau = 20$				
0.9506	0.9506	0.8496	0.8497	0.0366	0.0371	5.0000	5.0000	
0.0018	0.0016	0.002	0.0008	0.0019	0.0043	0.0023	0.0001	0.1400
0.0450	0.0147	0.0000	0.0000	0.4450	0.0351	0.0000	0.0000	0.0245

Table 8. Unit-Root Tests

	Test statistic		
	ADF		
Variable	t tests	MIC	PP
Growth rate of consumption	-4.886^*	-4.904^*	-9.354^*
Growth rate of investment	-5.731^*	-5.104^*	-5.304^*
Hours worked	-2.532	$-2.637^\dagger$	-3.402^*

Table 9. SMM Estimates

Description	Distribution		
	Normal	Rayleigh	Skew Normal
Discount factor	0.9955*	0.9942*	0.9972*
	(0.0227)	(0.0073)	(0.0061)
Consumption curvature	1.4225†	1.0353*	0.9951
	(0.7780)	(0.2167)	(1.2680)
Autoregressive coefficient	0.9842*	0.9952*	0.9510*
	(0.0117)	(0.0082)	(0.1024)
Standard deviation	0.0062*	–	–
	(0.0015)		
Rayleigh parameter	–	0.0089*	–
		(0.0022)	
Shape parameter	–	–	0.0066
			(0.0053)
Slope parameter	–	–	−0.9810*
			(0.0607)
Skewness	0	−0.6311	−0.8539
Value of function at the minimum	49.19	4.63	4.06

Table 10. Median Deviation from Deterministic Steady State

Variable	Distribution		
	Normal	Rayleigh	Skew Normal
Consumption	−1.069	−1.095	0.168
Investment	−1.297	−1.197	0.291
Capital stock	−1.393	−1.034	0.278
Hours worked	0.409	0.081	0.063

Table 11. Implications for Higher-Order Moments

Variable	U.S. Data	Distribution		
		Normal	Rayleigh	Skew Normal
			Standard Deviation	
Growth rate of consumption	0.565	0.321	0.406	0.176
Growth rate of investment	2.243	2.183	2.097	2.239
Hours worked	1.541	1.570	1.387	1.511
			Skewness	
Growth rate of consumption	−0.574	−0.267	−0.557	−0.632
Growth rate of investment	−0.572	−0.357	−0.584	−0.726
Hours worked	−0.475	−0.218	−0.245	−0.277
			Kurtosis	
Growth rate of consumption	5.644	2.845	3.045	3.309
Growth rate of investment	4.550	2.891	3.103	3.510
Hours worked	3.178	2.678	2.724	3.061
			Jarque–Bera Test Statistic	
Growth rate of consumption	68.33*	5.65	258.8*	352.6*
Growth rate of investment	5.99*	3.59	285.9*	493.0*
Hours worked	7.80*	61.44*	65.88*	64.67*

References

[1] Barro, R. J., 2006. Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics* 121, 823-866.

[2] Barro, R. J., 2006. Rare disasters, asset prices, and welfare costs. *American Economic Review* 99, 243-264.

[3] Brock, W., Mirman, L., 1972. Optimal economic growth and uncertainty: The discounted case. *Journal of Economic Theory* 4, 479-513.

[4] Canova, F., Sala, L., 2009. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* 56, 431-449.

[5] Cochrane, J. H., 2001. *Asset Pricing*. Princeton University Press: Princeton.

[6] Duffie, D., Singleton, K. J., 1993. Simulated moments estimation of markov models of asset prices. *Econometrica* 61, 929-952.

[7] Fernández-Villaverde, J., Rubio-Ramírez, J., 2007. Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies*, 74, 1059-1087.

[8] Gallant, A. R., Rossi, P. E., Tauchen, G., 1993. Nonlinear Dynamic Structures. *Econometrica* 61, 871-908.

[9] Hamilton, J. D., 1991. A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions. *Journal of Business and Economic Statistics* 9, 27-39.

[10] Hamilton, J. D., 1994. *Time Series Analysis*. Princeton University Press: Princeton.

[11] Hansen, G. D., 1985. Indivisible labor and the business cycle. *Journal of Monetary Economics* 16, 309-327.

[12] Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1929-1954.

[13] Hayashi, F., (2000). *Econometrics.* Princeton University Press: Princeton.

[14] Jin, H., Judd, K. L., 2002. Perturbation methods for general dynamic stochastic models. Hoover Institution, Mimeo.

[15] Kim. J., Kim, S., Schaumburg E., Sims, C., 2008. Calculating and Using Second-Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models. *Journal of Economic Dynamics and Control* 32, 3397-3414.

[16] Kim, J., Ruge-Murcia, F. J., 2009, How much inflation is necessary to grease the wheels? *Journal of Monetary Economics* 56, 365-377.

[17] Koop, G., Pesaran, M. H., Potter, S., 1996. Impulse Response Analysis in Nonlinear Multivariate Models. *Journal of Econometrics* 74, 119-147.

[18] Lee, B.-S., Ingram, B. F., 1991. Simulation estimation of time-series models. *Journal of Econometrics* 47, 195-205.

[19] Lombardo, G., 2010. On approximating DSGE models by series expansions. ECB Working Paper No. 1264.

[20] McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57, 995-1026.

[21] Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519-1554.

[22] Newey, W. K., 1985. Generalized method of moments specification testing. *Journal of Econometrics* 29, 229-256.

[23] O'Hagan, A., Leonhard, T., 1976. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201-202.

[24] Pakes, A., Pollard, D., 1989. The asymptotic distribution of simulation experiments. *Econometrica* 57, 1027-1057.

[25] Rietz, T., 1988. The equity risk premium: A solution. *Journal of Monetary Economics* 22, 117-131.

[26] Ruge-Murcia, F. J., 2007. Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control* 31, 1599-2636.

[27] Santos, M., Peralta-Alva, A., 2005. Accuracy of Simulations for Stochastic Dynamic Models. *Econometrica* 73, p. 1939-1976.

[28] Schmitt-Grohé, S., Uribe, M., 2004. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control* 28, 755-775.

[29] Smith, A. A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8, 63-84.

[30] Stone, J. R. N., 1954. *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom.* Cambridge University Press: Cambridge.

[31] Theil, H., Goldberger, A. S., 1961. On pure and mixed statistical estimation in economics. *International Economic Review* 2, 65-78.

[32] Tobin, J., 1972. Inflation and Unemployment. *American Economic Review* 62, 1-18.

Figure 1: Policy Rules Computed Using Different Polynomial Approximations

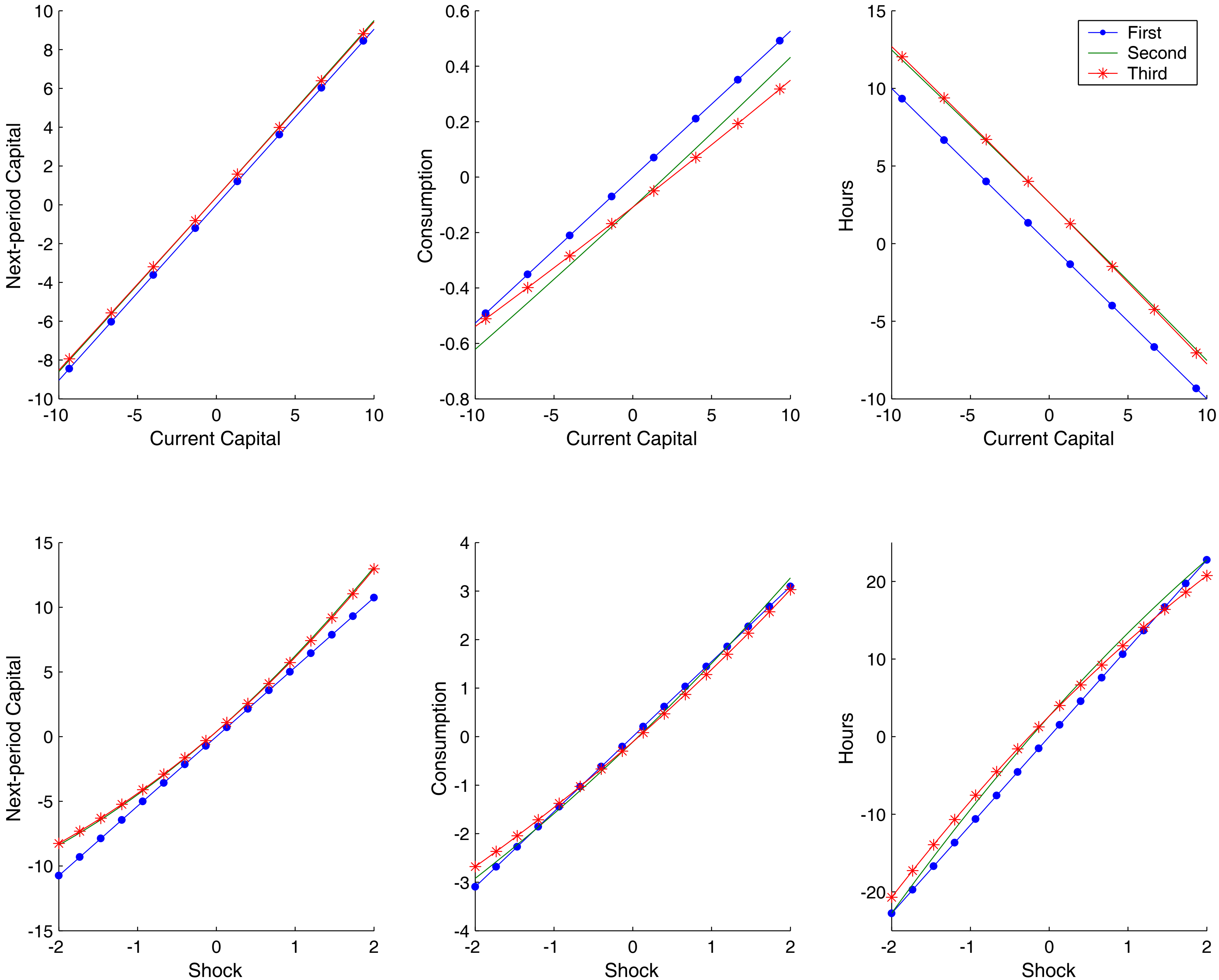

Figure 2: Empirical Distributions (Basic Model)

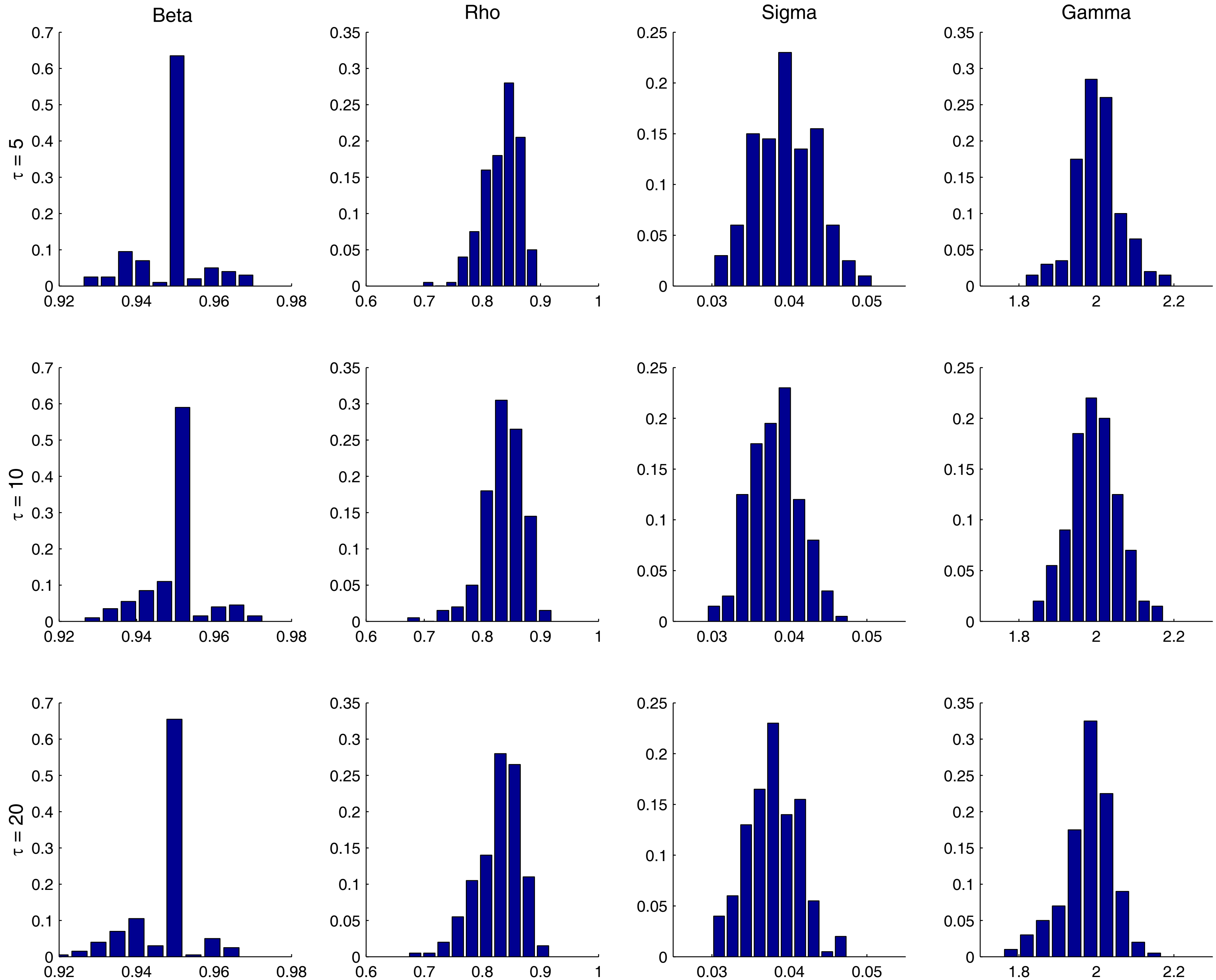

Figure 3: Empirical Distributions (Model with Habit Formation)

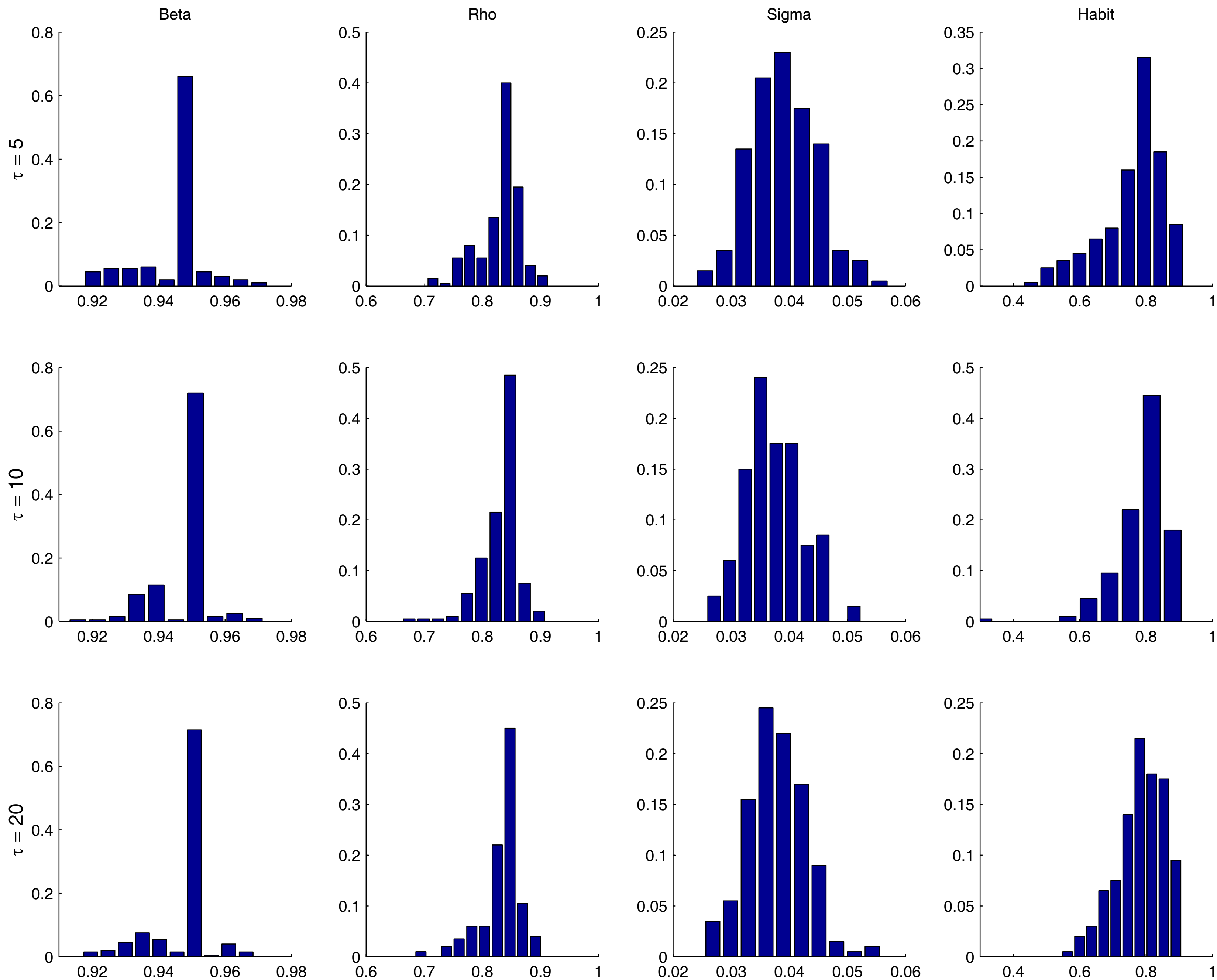

Figure 4: Empirical Distributions (Model with Habit Formation and High Variance)

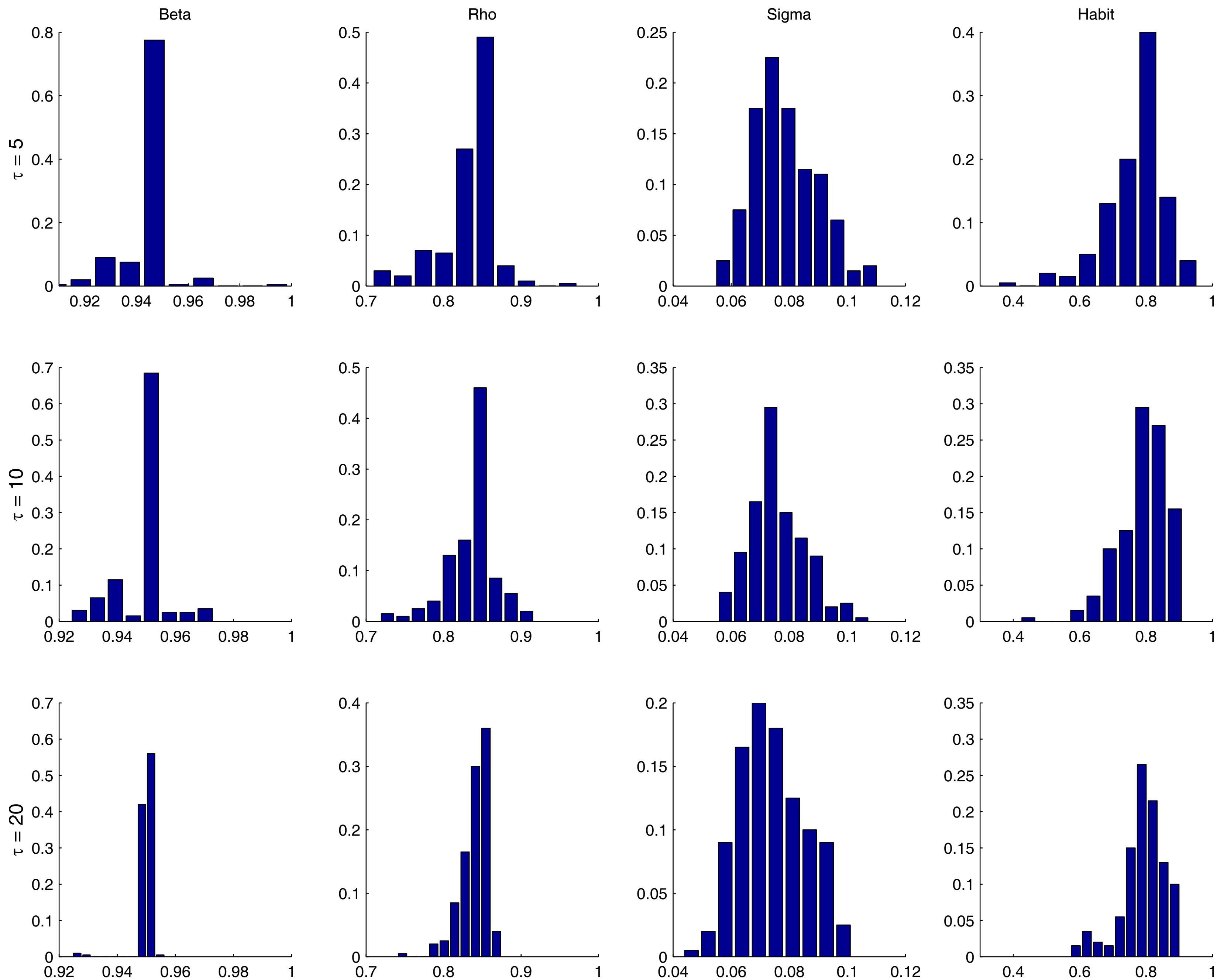

Figure 5: Objective Functions

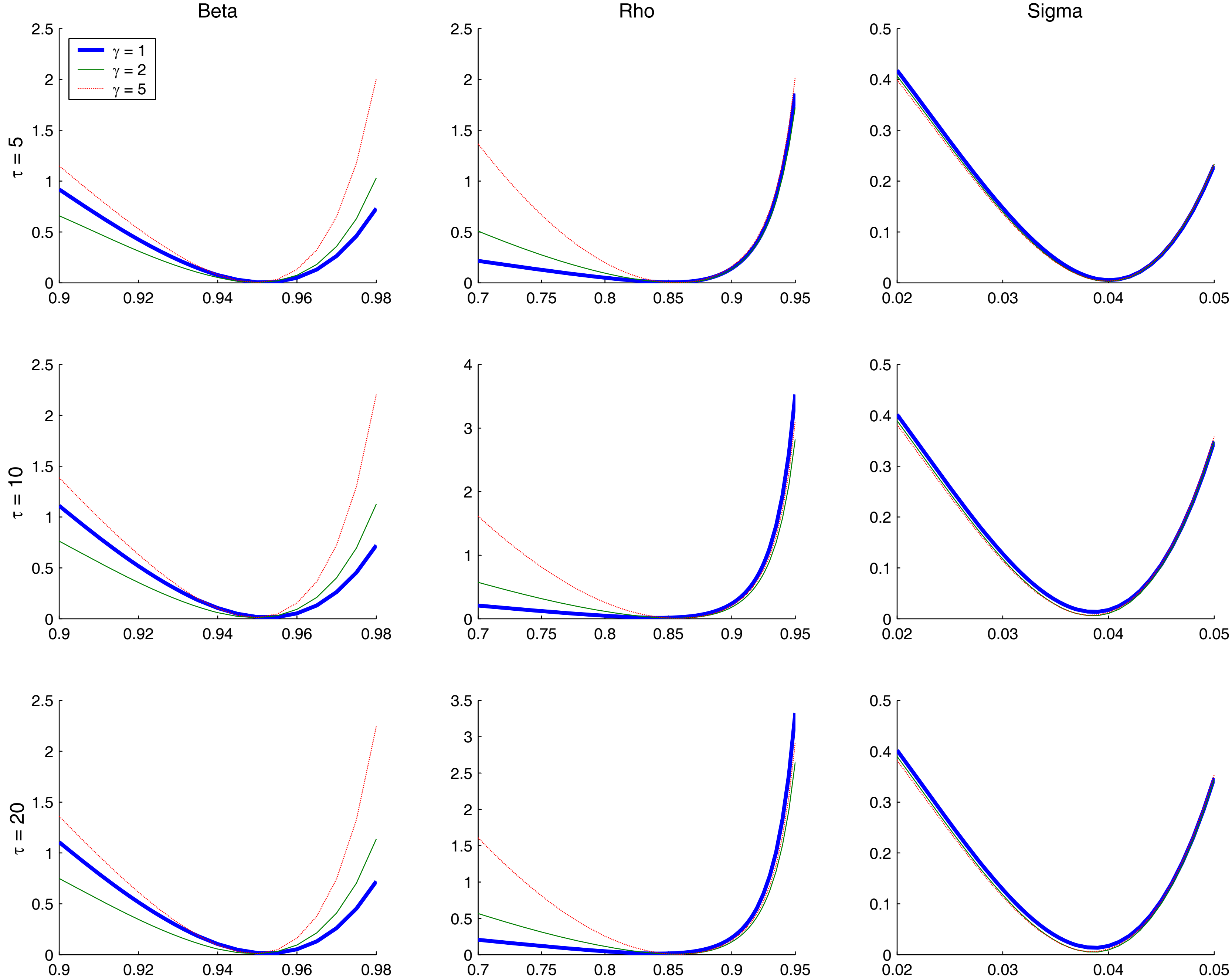

Figure 6: Empirical Distributions (Diagonal Weighting Matrix)

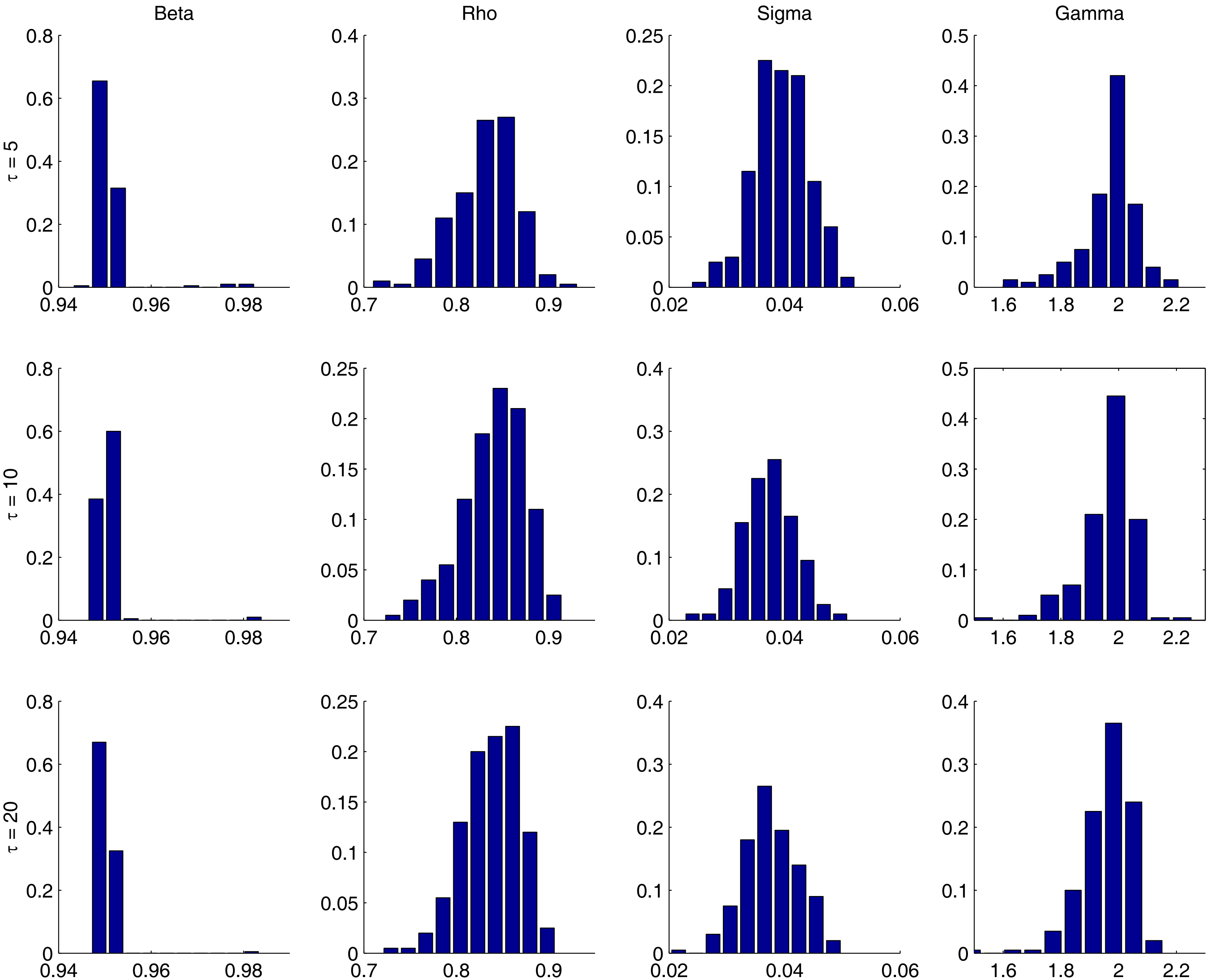

Figure 7: Empirical Distributions (Identity Weighting Matrix)

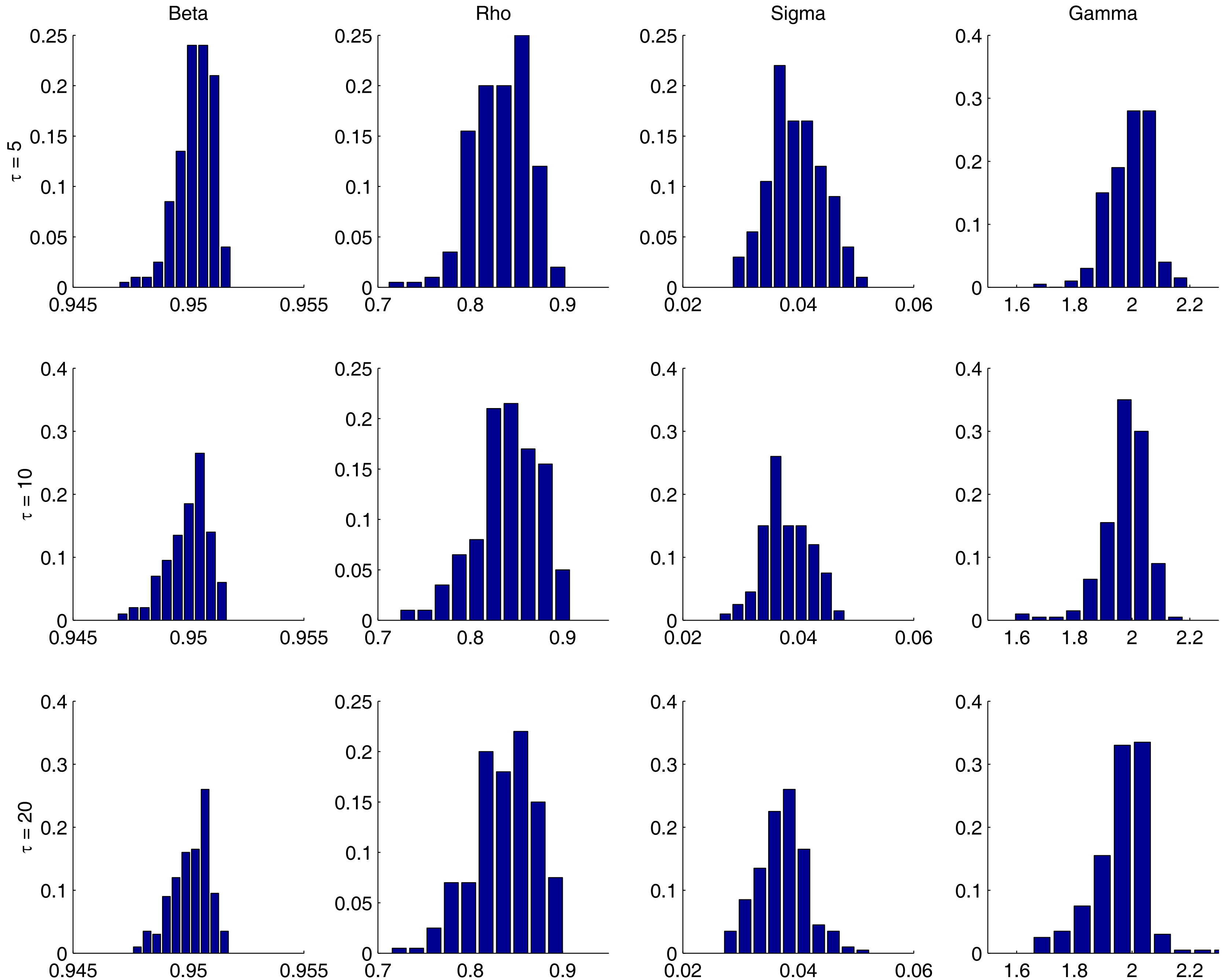

Figure 8: Empirical Distributions (Third-Order Expansion)

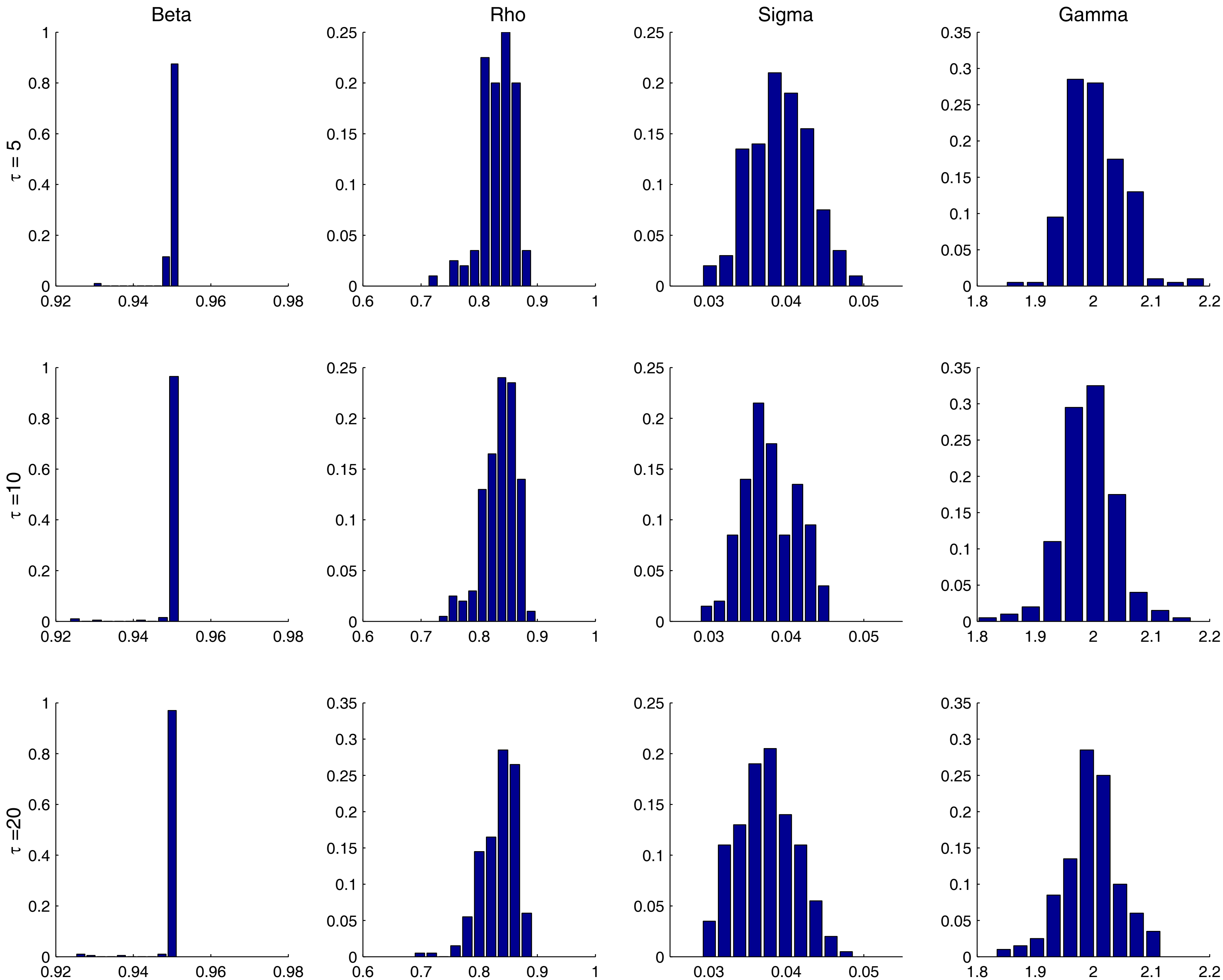

Figure 9: Empirical Distributions (Incorporating Priors)

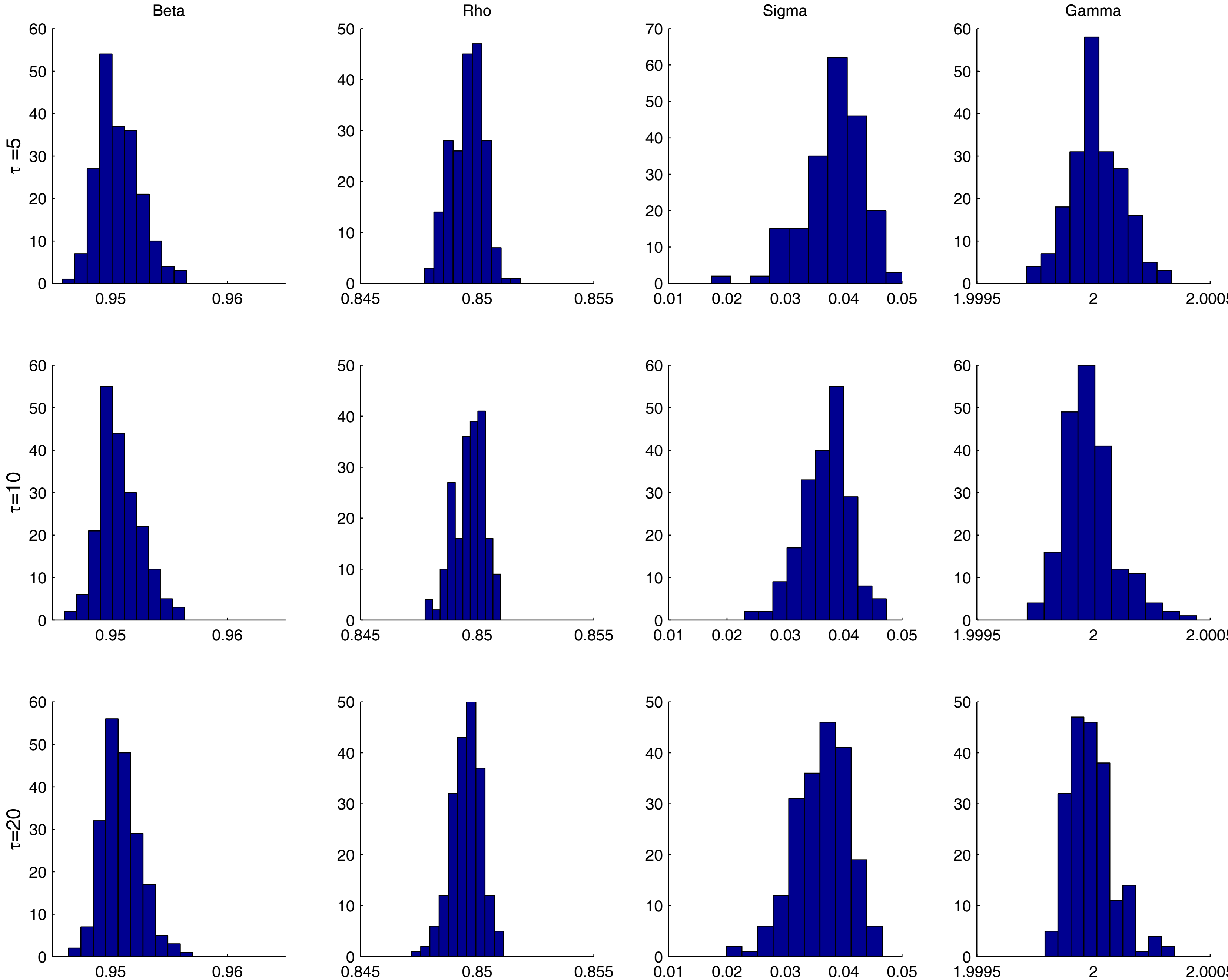

Figure 10: Computing Time

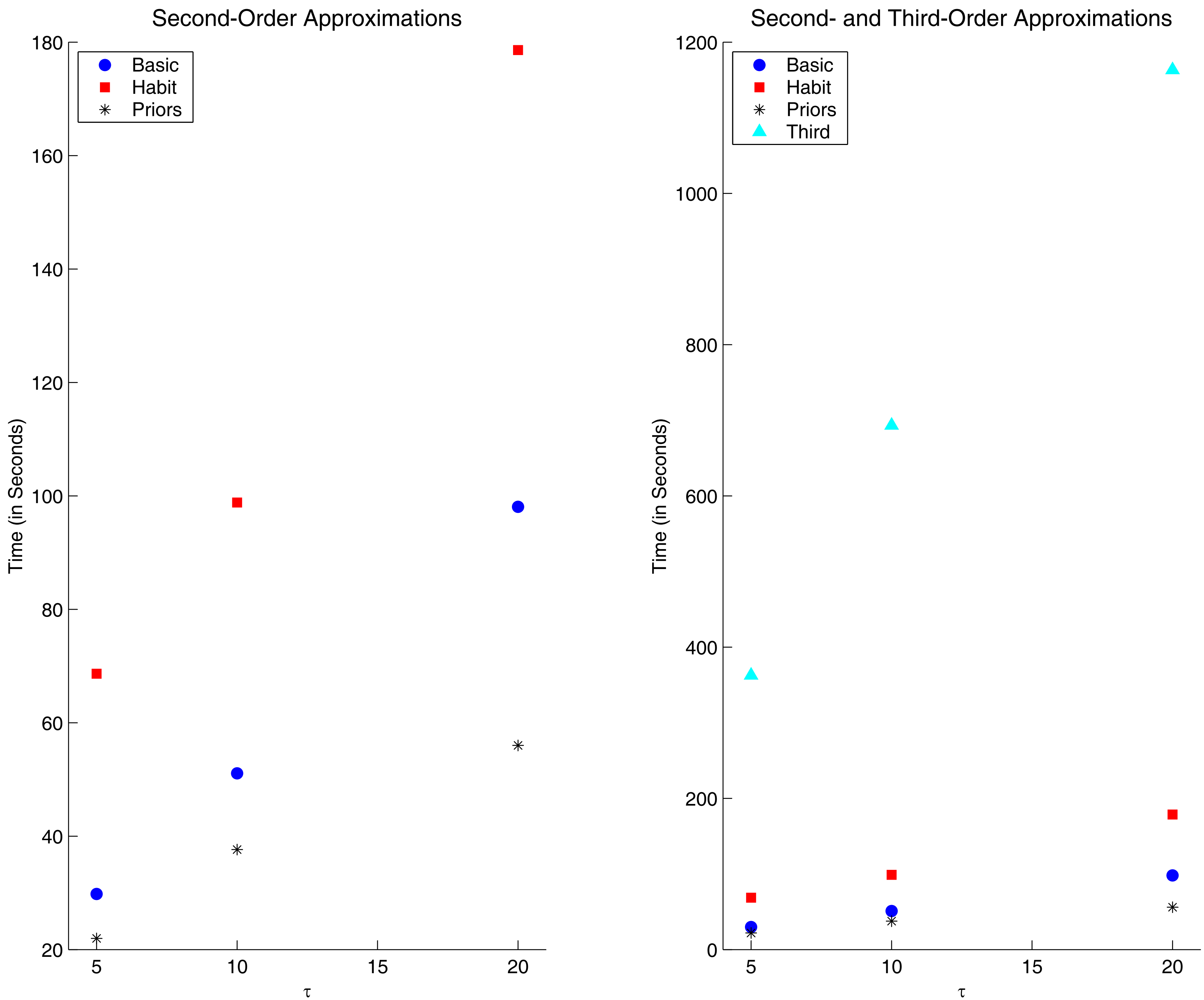

Figure 11: Estimated Distributions

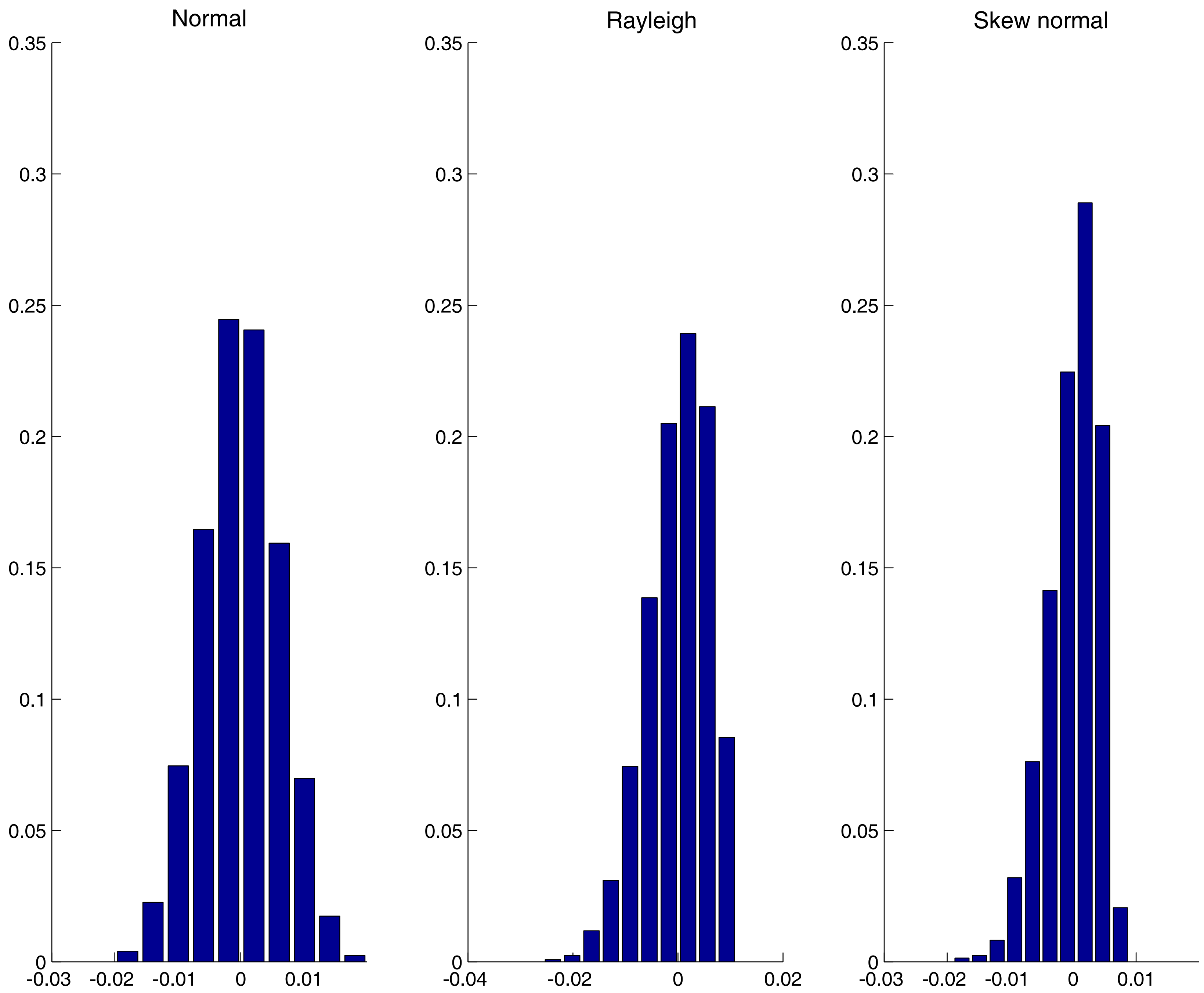

Figure 12: Impulse Responses (Normal Innovations)

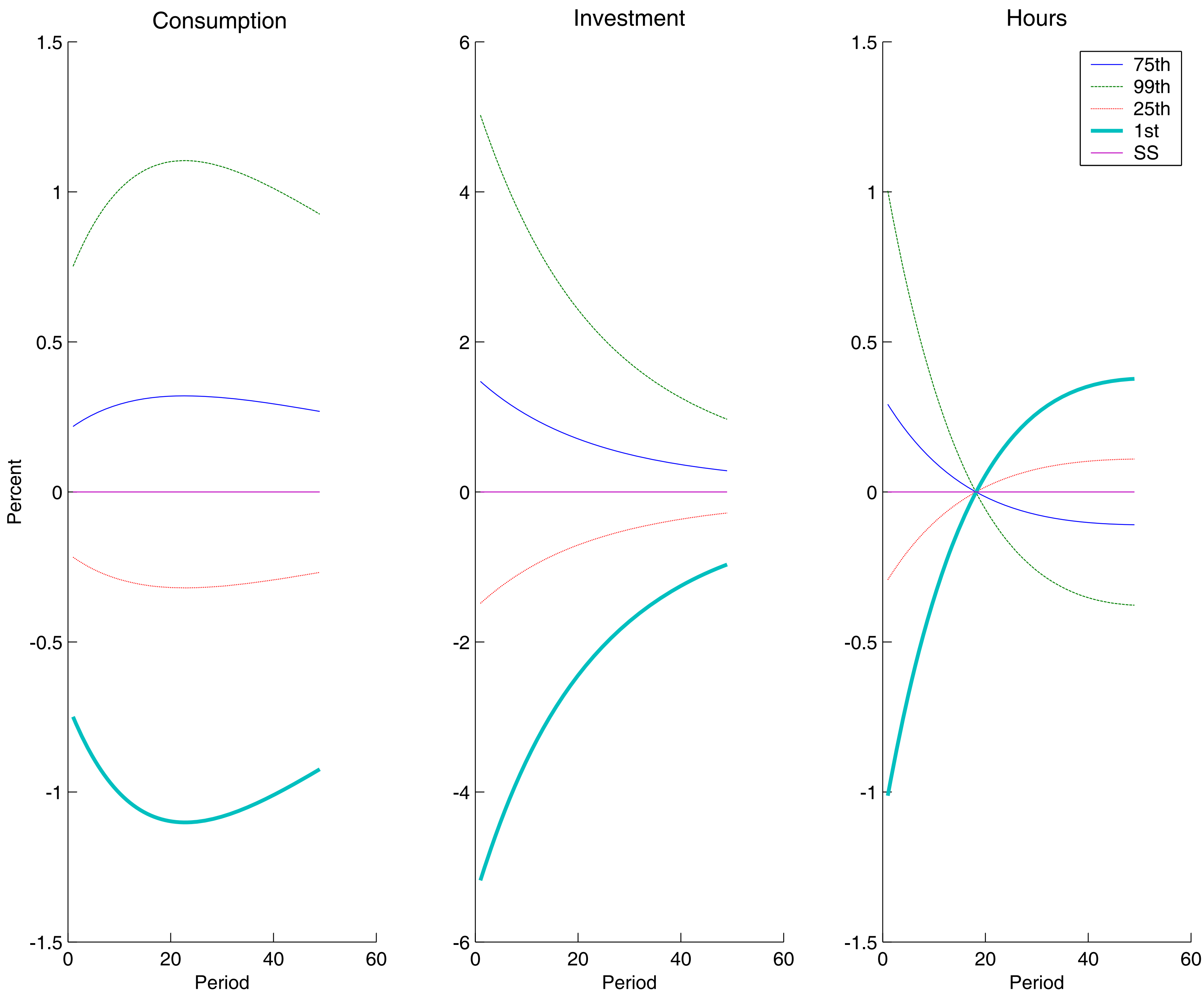

Figure 13: Impulse Responses (Rayleigh Innovations)

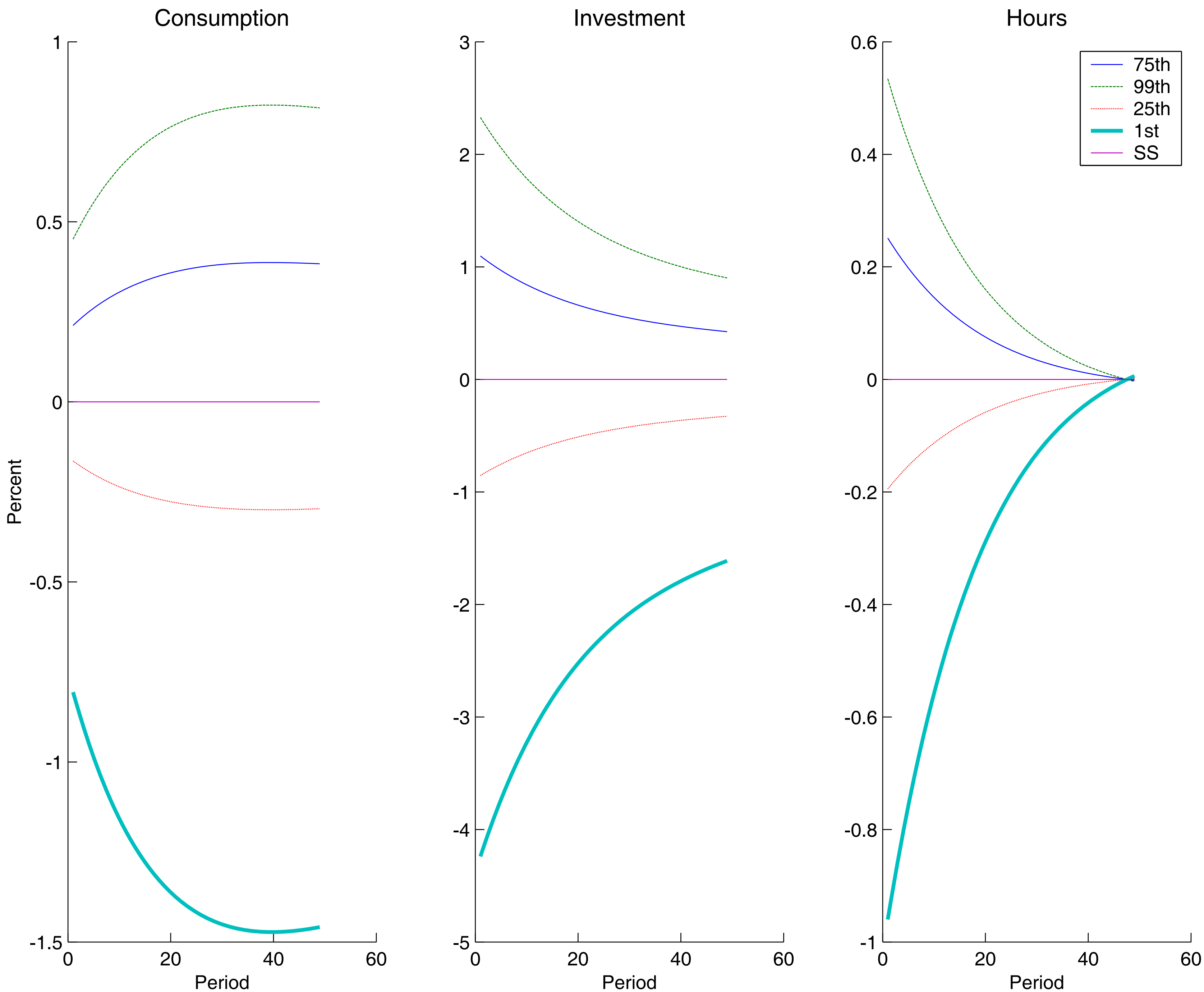

Figure 14: Impulse Responses (Skew Normal Innovations)