

# Likelihood Evaluation of Models with Occasionally Binding Constraints\*

Pablo Cuba-Borda    Luca Guerrieri    Matteo Iacoviello    Molin Zhong<sup>†</sup>

December 31, 2017

## Abstract

Applied researchers often need to estimate key parameters of DSGE models. Except in a handful of special cases, both the solution and the estimation step will require the use of numerical approximation techniques that introduce additional sources of error between the “true” value of the parameter and its actual estimate. In this paper, we focus on likelihood evaluation of models with occasionally binding constraints. We highlight how solution approximation errors and errors in specifying the likelihood function interact in ways that can compound each other.

**KEYWORDS:** Dynamic Models, Occasionally Binding Constraints, Likelihood-Based Estimation, Particle Filter.

**JEL CLASSIFICATION:** C32, C53, C63.

Latest version at <https://www2.bc.edu/matteo-iacoviello/CGIZ.pdf>

---

\*The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of anyone else associated with the Federal Reserve System.

<sup>†</sup>The authors are economists at the Federal Reserve Board.

They can be contacted at: [PABLO.A.CUBABORDA@FRB.GOV](mailto:PABLO.A.CUBABORDA@FRB.GOV), [LUCA.GUERRIERI@FRB.GOV](mailto:LUCA.GUERRIERI@FRB.GOV), [MATTEO.IACOVIELLO@FRB.GOV](mailto:MATTEO.IACOVIELLO@FRB.GOV), [MOLIN.ZHONG@FRB.GOV](mailto:MOLIN.ZHONG@FRB.GOV).

# 1 Introduction

Applied researchers often need to estimate key parameters of DSGE models. Consider the example of a researcher who wishes to interpret data on consumption through the lens of a DSGE model in order to estimate the coefficient of relative risk aversion. Except in a handful of special cases, both the solution and the estimation steps will require the use of numerical approximation techniques that introduce sources of error between the “true” value of the parameter and its estimate.

In this paper, we explore how solution approximation and estimation errors affect inference regarding the parameters of a DSGE model. We focus on the likelihood evaluation of models with occasionally binding constraints. Several authors—see the Handbook of Macroeconomics chapter by [Fernández-Villaverde, Rubio-Ramírez, and Schorfheide \(2016\)](#), as well as the textbook treatment of [Herbst and Schorfheide \(2016\)](#)—have already analyzed estimation and inference issues for nonlinear DSGE models. However, the previous analyses have tended to focus on nonlinearities either triggered by non-recursive utility functions—such as Epstein-Zin preferences—or by time-varying volatility of shocks, while occasionally binding constraints (or, equivalently, models with endogenous regime shifts) have received relatively less attention, at least until the Global Financial Crisis and the advent of the zero-lower-bound era.<sup>1</sup> Our results highlight how solution approximation errors and errors in specifying the likelihood function interact in ways that can compound each other.

We consider three solution methods and three paths to specifying the likelihood function for a DSGE model. Throughout our analysis, we assume that the model’s most accurate solution is the data generating process (DGP) for the observables and that the model only includes primitive shocks internalized by the agents. We use this setup to highlight how solution approximation errors and likelihood specification errors affect inference about structural parameters, and how their interaction is magnified in models with occasionally binding constraints.

The solution methods that we consider fall on different points of the trade-off between speed and accuracy. They include a highly accurate global method that relies on value function iteration, a piecewise linear method—based on the OccBin toolkit of [Guerrieri and Iacoviello \(2015\)](#)—and

---

<sup>1</sup>For models with nonstandard preferences, see for instance [van Binsbergen, Fernández-Villaverde, Koijen, and Rubio-Ramírez \(2012\)](#). For models with time-varying shocks, see for instance [Justiniano and Primiceri \(2008\)](#), [Amisano and Tristani \(2011\)](#), [Fernandez-Villaverde, Guerron-Quintana, and Rubio-Ramirez \(2015\)](#). For models with the zero lower bound, see for instance [Guerrieri and Iacoviello \(2017\)](#), [Gust, Herbst, López-Salido, and Smith \(2017\)](#), and [Aruoba, Cuba-Borda, and Schorfheide \(2018\)](#).

a first-order perturbation method that disregards the occasionally binding constraint. We show that the less accurate a solution method is, the harder it gets to retrieve the parameter values that govern agents’ decisions, and as the quality of the solution deteriorates some parameters stop being identified. In our example application, the coefficient of relative risk aversion is not identified when the solution method relies on a first-order approximation around a point where the borrowing constraint is assumed to always bind.

The alternative approaches to forming the likelihood that we consider offer different degrees of generality and interact in different ways with the approximation errors for the alternative solution methods. We showcase an inversion filter that relies on characterizing the likelihood function analytically by inverting the decision rule for the model. We also consider a particle filter, a standard approach to forming the likelihood for nonlinear models based on a sequential Monte Carlo approach (Fernández-Villaverde and Rubio-Ramírez, 2007).<sup>2</sup> Like in the case of solution error, we show that as we increase measurement error, the likelihood misspecification also increases, making it harder to retrieve the parameter values that govern the DGP.

A common assumption for the practical implementation of the particle filter is to posit that the data generating process includes measurement error, and to fix the variance of this error to some constant value.<sup>3</sup> This assumption may seem to be an innocuous way to get around degeneracy issues when choosing a computationally manageable number of particles. Indeed, if measurement error is part of the data generating process and the variance of the measurement error is estimated alongside other parameters of interest, the particle filter delivers an unbiased estimate of the model’s likelihood.<sup>4</sup> However, in our setup—in which the true DGP does not contain measurement error—the approximation error involved in the particle filter grows with the size of the assumed measurement error.<sup>5</sup>

In particular, we show that estimation measurement error—blaming the data for discrepancy between the model and the data—can amplify model approximation error—blaming the model for discrepancy between the model and the data—and that inaccurate data are just as pernicious

---

<sup>2</sup>Whenever applicable, we also consider a Kalman filter. Of course, the choice of procedure for computing the likelihood is also related to the choice for solution method, so we do not cover all possible pairings.

<sup>3</sup>See, for instance, Bocola (2016), van Binsbergen, Fernández-Villaverde, Koijen, and Rubio-Ramírez (2012), Gust, Herbst, López-Salido, and Smith (2017), and Cuba-Borda (2014).

<sup>4</sup>When the particle filter is embedded in a Markov Chain Monte Carlo sampler, Andrieu, Doucet, and Holenstein (2010) show that one can sample from the correct posterior distribution of the parameters.

<sup>5</sup>Canova (2009) stresses that using measurement error for estimation can distort inference on otherwise properly identified structural parameters.

cious as an inaccurate model solution. Intuitively, measurement error makes it more difficult for the econometrician to distinguish between regimes based on limited observations in models with occasionally binding constraints. In turn, this difficulty in correctly identifying the regime may lead to a substantial deterioration in the inference about model parameters. In this sense, our paper complements the results of two strands of the literature. In the context of linearized DSGE models [Canova, Ferroni, and Matthes \(2014\)](#) show that incorrectly assuming measurement error may distort parameter inference. In the context of nonlinear regression models, measurement error—regardless of whether it is introduced on the right or left-hand side variable—can lead to inconsistent parameter estimates. For instance, [Hausman \(2001\)](#) discusses how a mismeasured left-hand side variable can lead to biased and inconsistent estimators in a large class of nonlinear models, such as binary choice, quantile regression, or duration and hazard models. In our example application, we show how such intuition applies to models in which occasionally binding constraints lead to regime changes that have to be inferred from the data.

There are many more approaches to forming the likelihood of a model than the ones considered here. Without attempting to offer a complete list, some additional alternatives include the extended Kalman filter, the unscented Kalman filter, and the central difference Kalman filter ([Andreasen, 2013](#)). Furthermore, it is certainly possible to deploy estimation methods that do not rely on the likelihood of the model.<sup>6</sup> We have a simple justification for the parsimonious choice of estimation methods considered here. We find the inversion filter appealing because it allows us to consider models in which all stochastic innovations, including measurement error if desired, are taken into account by the agents in the model. This characteristic is valuable, but we do not see how to apply this method to over-identified models. The particle filter is certainly more flexible in this last respect, and is a focal example that has received widespread attention for the estimation of nonlinear DSGE models. Finally, the ease of implementation of the Kalman filter and its resilience to the curse of dimensionality probably justify why this method is often applied in practice as a shortcut, even if it cannot encompass the occasionally binding constraints in the model of interest. Some pitfalls of linearization are already well documented.<sup>7</sup> We highlight how this shortcut may come at nontrivial costs for inference purposes.

The rest of the paper proceeds as follows. Section 2 presents the conceptual framework for

---

<sup>6</sup>[Ruge-Murcia \(2012\)](#), for example, uses simulated method of moments. See [Fernández-Villaverde, Rubio-Ramírez, and Schorfheide \(2016\)](#) for an overview of alternative solution and estimation methods.

<sup>7</sup>For instance, see [Kim and Kim \(2007\)](#).

the analysis. Section 3 discusses the model that we take as data-generating process for our Monte Carlo experiments, Section 4 gives model solution and likelihood evaluation details, and Section 5 describes the results of those experiments. Section 6 offers compares our results to those of related papers and Section 7 concludes.

## 2 Conceptual Framework

We consider dynamic stochastic models and their relationship to observed data through the lens of a nonlinear state-space representation. To fix notation, this general representation takes the form:

$$s_t = h(s_{t-1}, \eta_t; \theta), \tag{1}$$

$$y_t = g(s_t; \theta) + \zeta_t, \tag{2}$$

$$\eta_t \sim N(0, \Sigma). \tag{3}$$

$$\zeta_t \sim N(0, \Omega). \tag{4}$$

Equation (1) determines the evolution of the endogenous variables summarized in the vector  $s_t$ ,  $\eta_t$  is a vector of exogenous stochastic innovations that are normally distributed with mean 0 and covariance matrix given by  $\Sigma$ ; the vector  $\theta$  includes all other parameters. Equation (2) relates the observations summarized in the vector  $y_t$  to the endogenous variables in  $s_t$ , subject to white noise measurement error  $\zeta_t$  with variance  $\Omega$ . We are interested in characterizing the likelihood function of the model conditional on the matrix of observations through time  $T$ .

$$\mathcal{L} = \ell(\theta; y_{1:T}). \tag{5}$$

### 2.1 Model Approximation

The function  $h(\cdot)$  is determined by the economic model of interest.<sup>8</sup> An important class of models, including the simple example considered in this paper, does not support a closed-form solution. Accordingly, in practice, the function  $h(\cdot)$  is also dependent on the numerical method chosen to

---

<sup>8</sup>Without loss of generality, the function  $g(\cdot)$  can be reduced to the role of selecting particular elements of the vector  $s_t$ .

approximate the solution of the model (and its approximation error). We consider three alternative solution methods:

1. Value function iteration, as described in [Judd \(1998\)](#) or in [Ljungqvist and Sargent \(2004\)](#), which introduces bounds to and discretizes the support of the state variables; we denote the related solution function  $h_{vfi}(\cdot)$ .
2. A piecewise linear solution, as described and implemented in the OccBin toolkit of [Guerrieri and Iacoviello \(2015\)](#); we denote the related solution function  $h_o(\cdot)$ .
3. A first-order perturbation solution, with the simplifying assumption that all of the constraints of the model always bind.

## 2.2 Likelihood Approximation

We consider three approaches to computing the likelihood function.

**Inversion Filter.** Under certain conditions that include the same number of observed variables in the vector  $y_t$  as innovations in the vector  $\eta_t$  plus the vector  $\zeta_t$ , knowledge of the distributions of  $\eta_t$  and the measurement error  $\zeta_t$  can be used to characterize the likelihood of  $y_t$  by substituting Equation (1) into Equation (2) and inverting the resulting combination function to back out the innovations. [Fair and Taylor \(1983\)](#) outline this approach to estimating nonlinear DSGE models and [Guerrieri and Iacoviello \(2017\)](#) implement it for the case of a medium-scale model with occasionally binding constraints and no measurement error.<sup>9</sup> While we cannot generalize this approach to over-identified models, it has the advantage of producing an exact value for the likelihood. Accordingly, we take this approach as the benchmark against which we compare the other alternatives that we consider.

**Particle Filter.** The particle filter applies more generally than the inversion filter. In particular, it can handle not just models that are exactly identified but also models that are over-identified, in the sense that the number of stochastic innovations in the vector  $\eta_t$  plus the number of errors in the vector  $\zeta_t$  exceeds the number of variables in the the vector  $y_t$ . The typical configuration for the

---

<sup>9</sup>Section 4.3 starting on page 36 of [Guerrieri and Iacoviello \(2017\)](#) describes this method in detail and spells out conditions for its application in conjunction with a piecewise linear solution method. [Amisano and Tristani \(2011\)](#) applied the same approach to form the likelihood of a model subject to exogenous regime switches.

particle filter posits that the data generating process includes measurement error for each variable observed. Prosaically, this choice avoids issues of degeneracy associated with a finite number of particles. Nonetheless, while this choice is expedient for statistical purposes, it divorces the information set of the econometrician from the information set of the agents in the model. While such separation of the information set is not problematic for engineering applications, for which the particle filter was originally developed, it seems more difficult to justify in economic applications. This is not to say that economic series are observed without measurement error, but simply that measurement error in Equation (2) is, by construction, disregarded by the agents of the model – a curious omission, especially given the extra care usually needed in solving forward-looking models or models with rational expectations.

Given these considerations, we focus on a data generating process that excludes measurement error from the observation equation. In this respect, any small amount of measurement error needed to avoid degeneracy of the particle filter introduces a misspecification error into the likelihood function. A typical approach, in practice, is to assume that measurement error covers a fixed fraction of the variance of the observed variables. We consider alternative values for this fraction, starting from small levels of 1%, nonetheless sufficient to avoid degeneracy for our example, up to a level of 20%, a value commonly chosen for other empirical applications.

**Kalman Filter.** In conjunction with the first-order perturbation solution, we form the likelihood function using a Kalman filter that, given our data generating process, correctly excludes measurement error. In this case, the misspecification error is avoided, and our analysis will focus on the effect of model approximation error on the shape of the resulting likelihood function.

### 3 An Application: A Consumption Model with an Occasionally Binding Borrowing Constraint

We base our analysis on a simple model for the choice of consumption and saving subject to a constraint that limits maximum borrowing to a fraction of current income. The reasons are twofold. First, the economic intuition for how this model works is remarkably simple, and versions of this model—with its emphasis on consumption smoothing—are the backbone of a large class

of richer models in modern macroeconomics (see for instance the treatments in [Deaton 1992](#) and [Ljungqvist and Sargent 2012](#)). Second, the model structure allows for a precautionary saving motive and nonlinear, kinked decision rules that can be fully captured only by a global numerical solution. By contrast, the piecewise linear solution can capture the nonlinearity of the consumption function but introduces a small solution error by ignoring precautionary saving motives, whereas a linearized decision rule that assumes that the constraint is always binding introduces even larger solution errors.

### 3.1 The Model

A consumer maximizes

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma} - 1}{1-\gamma},$$

where  $\gamma$  is the coefficient of relative risk aversion, subject to the budget constraint and to an occasionally binding constraint stating that borrowing  $B_t$  cannot exceed a fraction  $m$  of income  $Y_t$ :

$$C_t + RB_{t-1} = Y_t + B_t, \tag{6}$$

$$B_t \leq mY_t. \tag{7}$$

Above,  $R$  denotes the gross interest rate. The discount factor  $\beta$  is assumed to satisfy the restriction that  $\beta R < 1$ , so that in the deterministic steady state that ignores shock uncertainty the borrowing constraint is binding. Given initial conditions, the impatient household prefers a consumption path that is falling over time, and attains this path by borrowing today. If income is constant, the household will eventually be borrowing constrained and will roll its debt over forever, and consumption will settle at a level given by income less the steady state debt service.

The log of income follows an AR(1) stochastic process of the form

$$\ln Y_t = \rho \ln Y_{t-1} + \sigma \epsilon_t \tag{8}$$

where  $\epsilon_t$  is an exogenous innovation distributed as standard normal, and  $\sigma$  its standard deviation.

Denoting with  $\lambda_t$  the Lagrange multiplier on the borrowing constraint given by Equation (7), the set of equations describing the system of necessary conditions for an equilibrium is given by

a system of four equations in the four unknowns  $\{C_t, B_t, \lambda_t, Y_t\}$ , which includes equation (6), equation (8), together with the consumption Euler equation and the Kuhn-Tucker conditions, given respectively by

$$C_t^{-\gamma} = \beta RE_t (C_{t+1}^{-\gamma}) + \lambda_t \tag{9}$$

$$\lambda_t (B_t - mY_t) = 0. \tag{10}$$

The transitional dynamics of this model depend on the gap between the discount rate and the interest rate, which can be measured as  $g = 1/\beta - R$ . When the gap is small, the economy can be characterized as switching between two regimes. In the first regime, more likely to apply when income and assets are relatively low, the borrowing constraint binds. In that case, borrowing moves in lockstep with income, and consumption is more volatile than income. In the second regime, more likely to apply when income and assets are relatively high, the borrowing constraint is slack, and current consumption can be high relative to future consumption even if borrowing is below the maximum amount allowed.

### 3.2 Calibration.

We set  $\gamma = 1$ , so that utility is logarithmic in consumption. We set the maximum borrowing at one year of income, so that  $m = 1$ . For the income process, we set  $\rho = 0.90$  and  $\sigma = 0.01$ , so that the standard deviation of  $\ln Y$  is 2.5 percent. Finally, we set  $R = 1.05$  and  $\beta = 0.945$ . Under this calibration, the borrowing constraint, which binds in the reference regime, is slack about 40 percent of the time using the full nonlinear solution.

## 4 Model Solution and Likelihood Evaluation

**Value Function Iteration.** We use dynamic programming to characterize a high-quality fully-nonlinear solution. The debt level  $B_t$  is the only state variable in the model. We seek a rule that will map the current state variable  $B_{t-1}$  and the realization of the stochastic process  $Y_t$  into a choice of  $B_t$ . We discretize and put boundaries on the support of the decision rule that we seek. We discretize the support of both  $B_{t-1}$  and  $Y_t$ . We consider a uniformly spaced set of points for  $B_{t-1}$  and  $B_t$ . The lower boundary for  $B_{t-1}$  is 25 percent below the non-stochastic steady state for

borrowing. The upper boundary is 8 percent above the non-stochastic steady state for borrowing. We constrain  $Y$  to lie within three standard deviations of its process, i.e.  $|\ln Y_t| \leq 3\sqrt{\frac{\sigma^2}{1-\rho^2}}$ . We follow [Tauchen \(1986\)](#) to discretize the process  $\ln Y_t$ . Overall, the grid we consider involves 200 points for debt and 15 points for the income process.

The value-function-iteration algorithm that we use follows closely Chapter 12 of [Judd \(1998\)](#) and Chapter 3 of [Ljungqvist and Sargent \(2012\)](#). To accelerate the convergence of the dynamic programming algorithm, we use the Howard improvement algorithm.

**Piecewise Linear Solution.** The economy features two regimes: a regime in which the collateral constraint binds and a regime in which it does not (but is expected to bind in the future). The piecewise linear method resolves the problem of computing decision rules that approximate the equilibrium adequately under both regimes. Essential this method linearizes the model at the constraint and away from the constraint and then joins the two systems of equations using a shooting algorithm that reduces the problem to only solving for the expected duration of each regime, rather than solving for paths of all endogenous variables. The implementation of this algorithm is the same as in [Guerrieri and Iacoviello \(2015\)](#), which gives further details of the numerical implementation.

**First-order Perturbation.** For this solution, we disregard the possibility that the constraint could ever be slack and linearize the model around the nonstochastic steady state. We use the solution computed by Dynare, a popular and convenient set of tools for solving and estimating DSGE models. See [Judd \(1998\)](#) for details of this standard solution algorithm.

## 4.1 The Solution: Policy Functions and Impulse Response Functions

We display the policy functions in terms of the optimal borrowing and optimal consumption chosen by the agent as a function of income, holding the level of debt at its nonstochastic steady-state value of 1.

Figure 1 shows contours of the model's policy functions for the different solution algorithms. The figure shows that for lower-than-average realizations of income the agent hits the borrowing constraint. In that case, the consumption function is relatively steep, the multiplier on the borrowing constraint is positive, and consumption is very sensitive to changes in income. For higher-than-average income, consumption is sufficiently high today relative to the future that it pays off to save for the future. In that case, the borrowing constraint becomes temporarily

slack, the multiplier on the borrowing constraint is zero, and consumption becomes less sensitive to changes in income.

Following [Judd \(1992\)](#), we use the Euler equation residuals in units of consumption to quantify the error in the intertemporal allocation. The policy functions using value function iteration are minimally affected by approximation error with Euler errors in the order of \$1 per \$100,000 of consumption.<sup>10</sup> Accordingly, we take this solution method as the benchmark against which we compare the errors of the other methods. The errors for the piecewise linear solution are typically in the order of \$1 per \$1000 of consumption, a modest if nontrivial amount. Finally, in the class of models with occasionally binding constraints that is the focus of this paper, errors for the linear solution method can rise to the substantial amount of \$1 per \$10 of consumption.

These different errors also transmit to the likelihood function and may result, by themselves, in substantive problems for inference. For instance, [Fernandez-Villaverde, Rubio-Ramirez, and Santos \(2006\)](#) point out that second-order differences in alternative solution methods may lead to first-order differences in the related likelihood functions. This is because any differences in the solution method may be compounded by the sample size when forming the likelihood.

[Figure 2](#) shows the responses to two shocks, starting from a nonstochastic steady state where income is 1 and the ratio of debt to income is at the maximum limit. The first shock, in period 2, brings up income by 2 percent. The second shock, in period 21, pushes down income by 2 percent. The red and blue lines denote the piecewise linear solution and the value function iteration solution, respectively. The dash-dotted lines denote the first-order perturbation solution, which incorrectly assumes that the borrowing constraint always binds. As the figure shows, like the value function iteration solution, the piecewise linear algorithm captures the asymmetric responses of consumption and debt well. A positive income shock makes the borrowing constraint slack and the Lagrange multiplier hits 0; borrowing rises less than income, and consumption rises less than it would were the constraint binding in all states of the world. Conversely, when income drops, the borrowing constraint binds, borrowing falls in proportion with income, and consumption reacts more than under a positive shock.

---

<sup>10</sup>See [Appendix A.1](#) and [Figure A.1](#) for further details.

## 4.2 Evaluating the Likelihood of the Model

**Inversion Filter** In the absence of measurement error, we can combine Equations (1) and (2) to obtain:

$$y_t = f(s_{t-1}, \eta_t; \theta) \quad (11)$$

Where  $f \equiv (g \circ h)$  maps  $\mathbb{R}^{n_s} \times \mathbb{R}^{n_\eta}$  into  $\mathbb{R}^{n_y}$ , with  $n_s$ ,  $n_y$ ,  $n_\eta$ , denoting the number of endogenous state variables, observables, and structural shocks, respectively. When the structural innovations  $\eta_t$  are drawn from a multivariate normal distribution with covariance matrix  $\Sigma$  we can write the log-likelihood of the model as:

$$\log(p(y_{1:T})) = -\frac{Tn_y}{2} \log(2\pi) - \frac{T}{2} \log(\det(\Sigma)) - \frac{1}{2} \sum_{t=1}^T \eta_t' \Sigma^{-1} \eta_t + \sum_{t=1}^T \log \left( \left| \det \frac{\partial \eta_t}{\partial y_t} \right| \right) \quad (12)$$

In practical applications computing the log-likelihood function via the inversion filter (Equation 12) poses two challenges. First, the true structural shocks  $\eta_t$  are unobserved and inverting the function  $f(\cdot)$  to recover them might be cumbersome. Second, and as we discussed in section 2, the function  $h(\cdot)$  has to be approximated numerically with different degrees of accuracy and hence the Jacobian matrix  $\frac{\partial \eta_t}{\partial y_t}$  – which is defined implicitly by Equation (11) – can be computed in closed form only in very special cases.

With respect to the first issue, Figure 1 shows that the policy function for consumption is monotonic despite the occasionally binding constraint of the model. Hence, in our application, we can invert the policy functions  $h(\cdot)_o$  and  $h(\cdot)_{vfi}$  using a standard nonlinear root-finding algorithm.<sup>11</sup> For the second issue, in our benchmark specification which uses the most accurate solution approximation given by  $h_{vfi}(\cdot)$ , we construct the Jacobian matrix using a multi-point finite difference method. When using a piecewise linear solution, the Jacobian matrix can be constructed analytically. As shown in Guerrieri and Iacoviello (2017), this is possible due to the local linearity in  $\eta_t$  of the matrices that define the approximated piecewise linear solution  $h_o(\cdot)$ .

**Particle Filter** Our implementation follows the bootstrap particle filter laid out in Fernández-Villaverde and Rubio-Ramírez (2007). The bootstrap particle filter follows several basic steps.<sup>12</sup> To

<sup>11</sup>In other applications, the nonlinearities in the model might be more severe and multiple sets of shocks may be consistent with the same observation. Standard results can be invoked to construct the likelihood function even for this general case of a correspondence.

<sup>12</sup>A detailed explanation of the particle filter can be found in the Appendix.

obtain an accurate approximation of the likelihood, we set the number of particles in our simulation  $N = 2,000,000$ . The particle filter as specified requires a certain amount of measurement error in the data to avoid issues of degeneracy.

0. **Initialization:** Simulate  $N$  particles from an initial distribution  $p(s_0)$  and set the weight of each particle  $i$ :  $w_0^{(i)} = \frac{1}{N}$ . This distribution comes from the steady state distribution of the states. We approximate it by taking a draw for each particle  $i$  following a simulation of the model for 50 periods.

Entering into time  $t$ , we have a swarm of  $N$  particles  $\{s_{t-1}^{(i)}\}_{i=1}^N$  and associated weights  $\{w_{t-1}^{(i)}\}_{i=1}^N$  that are distributed according to  $p(s_{t-1}|y_{1:t-1})$ .

1. **State forecast:** Take each particle  $s_{t-1}^{(i)}$  simulate it forward via the state transition equation  $h(s_{t-1}, \eta_t; \theta)$ . This then gives us a new swarm of particles  $\{s_t^{(i)}\}_{i=1}^N$  and associated weights  $\{w_{t-1}^{(i)}\}_{i=1}^N$ .
2. **Observable forecast:** For each particle  $s_t^{(i)}$ , we can calculate  $p(y_t|s_t^{(i)}, y_{1:t-1}) = p(y_t|s_t^{(i)})$ . The weighted average of these particles is the discrete approximation to the integral that defines  $p(y_t|y_{1:t-1}, \theta)$ . Call this approximation  $\hat{p}(y_t|y_{1:t-1}, \theta)$ .

$$\hat{p}(y_t|y_{1:t-1}, \theta) = \frac{1}{N} \sum_{i=1}^N p(y_t|s_t^{(i)})w_{t-1}^{(i)} \quad (13)$$

3. **State update:** We now incorporate time  $t$  information  $y_t$  to have the particles be distributed according to  $p(s_t|y_{1:t})$ . We calculate the new importance weights as follows:

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p(y_t|s_t^{(i)})}{\frac{1}{N} \sum_{i=1}^N p(y_t|s_t^{(i)})w_{t-1}^{(i)}} \quad (14)$$

4. **(Optional) Resampling step:** Every period, there is an optional step to replenish low weight particles with high weight ones. We draw  $N$  new particles from the set  $\{s_t^{(i)}\}_{i=1}^N$  with replacement according to their weights  $\{w_t^{(i)}\}_{i=1}^N$ . We enter this step when our effective sample size falls below  $1/3$  of  $N$ .

We iterate steps 1 – 4 until time  $T$ . This completes the recursion.

## 5 Findings

In our Monte Carlo experiment, the DGP is the consumption/saving model solved with the value-function-iteration method. We simulate 100 observations of consumption data ( $T = 100$ ).<sup>13</sup> Our data-generating process does not contain measurement error – or, equivalently, in Equation 4, the variance of measurement error  $\Omega$  is 0. We are focused on inference regarding the parameter  $\gamma$ , the coefficient of relative risk aversion in the model. In the DGP,  $\gamma = 1$ . All other parameters are fixed at their pseudo-true values as described in the calibration section above.

### 5.1 Overview

Figure 3 depicts the likelihood function for a single simulated sample of  $T = 100$  observations. The top left panel focuses on the value function iteration solution and forms the likelihood through the inversion filter. Of note, the likelihood peaks at 1, the pseudo-true value in the DGP. Because we use enough nodes to render the Euler equation residuals negligible and because the inversion filter avoids the misspecification error in the measurement equation, we take this case as the benchmark against which we assess the alternative combinations of solution methods and filters shown in the figure—the solid blue line is replicated in every other panel.

The panels in Figure 3 are arranged so that the left column only considers the likelihood functions that rely on a value function iteration at the solution step, while the right column showcases likelihood contours based on the OccBin solution method that introduces some solution error. The top row focuses on the inversion filter that avoids misspecification error. Moving down the figure, the rows below the first introduce increasing misspecification error – summarized by the variance of measurement error, expressed as percent of the variance of consumption. Accordingly, the differences in the likelihood contours in the left column highlight the effect of misspecification error. The difference in the top row highlight approximation error, and rows from the second on from the right column show the interaction between the two sources of error.

Each panel allows the computation of some headline comparison metrics. The distance between the peak of each contour quantifies a bias in the point estimates. The vertical distance between the likelihood contours in each panel can be interpreted as a percentage point difference in the

---

<sup>13</sup>Specifically, starting from the non-stochastic steady state, we simulate data of length  $T = 500$  and treat the first 400 observations as burn-in.

probability of any given value of  $\gamma$  (given the observations in the sample).

## 5.2 Solution Approximation

Focusing on the top right panel, the OccBin solution biases the estimate of  $\gamma$  upwards. The solution method ignores precautionary motives and results in a consumption function that, over some regions, is more sensitive to variation in income relative to the consumption function from the accurate value function iteration method. Accordingly, one way to match the observations is through a higher level of risk aversion relative to the data generating process, which results in the upward bias shown, but not in a substantial flattening of the likelihood contour.

## 5.3 Likelihood Misspecification

Moving to the first column of the second row, the panel labeled “Minimal Solution Error, 1% Measurement Error” isolates the effects of measurement error. As expected, measurement error leads to a flattening of the likelihood contour, but in this case, it also leads to an upward bias in the point estimate of  $\gamma$ , not the typical effect of measurement error. The bias is related to the kinked model decision rules used in the underlying data-generating process.

Figure 4 can be used to illustrate how this bias is related to the occasionally binding constraint in the model. The policy function underlying the DGP is represented by the blue, solid line. Notice that for realizations of the income lower than the average value of 1, the borrowing constraint binds and consumption is approximately linear in income. In turn, for higher than average realizations of income, the borrowing constraint is slack and consumption responds in a more muted way to high realizations of the income process. Accordingly, there is a kink in the consumption function right at the point where the borrowing constraint becomes slack. Notice also that the point where the constraint becomes slack is related to the underlying value of  $\gamma$ . Lower values of  $\gamma$  shift this point to the right. Values of  $\gamma$  lower than its assumed value of 1 (coinciding with less risk aversion) would cause the consumption function to be too steep and could be reconciled with the observations for consumption only via skewed and, thus, less likely estimates of the income shocks. Values of  $\gamma$  higher than its assumed value of 1 would cause the consumption function to be too flat, and again call for less likely estimates of the income shocks. In sum, the position of the kink and subsequent shape of the consumption function inferred from observations on consumption can

influence the estimates of  $\gamma$ .

For the sake of argument, let's consider a case slightly different from the one we posited, in which the DGP includes measurement error, but the econometrician does not realize it. Adding (normally distributed) measurement error changes both features of the consumption function: it makes consumption more volatile, and it reduces/washes away any skewness. Accordingly, the econometrician would think that consumption function is consistent with a lower value of  $\gamma$ , such as the one represented by the dashed, red line. Figure 5 illustrates this case, showing how “too little” assumed measurement error relative to what is embedded in the DGP biases the estimate of  $\gamma$  *downwards*, regardless of the solution error.

Our case is the mirror image of the one described above. The data-generating process does not include measurement error, but the econometrician assumes that it is part of the data-generating process. That is, the econometrician sees skewed and asymmetric consumption even after accounting for normally distributed, additive measurement error. Accordingly, the econometrician's estimates of  $\gamma$  are biased *upwards*, and this bias is greater, the greater the fraction of the observed variation incorrectly attributed to the measurement error. (As measurement error rises, so does the econometrician's estimate of  $\gamma$ , because higher values of  $\gamma$  can undo any reduction in skewness that measurement error introduces onto observed consumption.)

Figure 4 also highlights that equal size increases in  $\gamma$  induce progressively smaller shifts to the left for the kink of the policy function. Accordingly, the bias in  $\gamma$  can be predicted to increase nonlinearly in the size of the measurement error.

Going back to Figure 3, and moving down along the left-hand-side column, one can readily evince that the peak of the likelihood function shifts further to the right as the relative size of the measurement error increases. Strikingly, this increase is highly nonlinear in the size of the measurement error, in line with the discussion of the policy function above.

An additional effect of measurement error misspecification is to flatten out the likelihood function. As the last row of this column shows, with measurement error at 20% of the variance of consumption, the misspecification bias is so severe as to hinder the identification of  $\gamma$ .

## 5.4 Interaction of Solution Approximation and Likelihood Misspecification

Figure 3 also showcases the interaction between measurement error and misspecification error. The left-hand side panel in the second row, labeled “Some Solution Error, 1% Measurement Error,” can be compared against the panel to its left, which highlights misspecification error, and against the panel just above, which showcases approximation error. When the two sources of error interact, it is readily apparent that the bias is greater than the sum of the biases for each error in isolation.

Another way to understand the bias from the approximation error of the OccBin solution is that the inability to capture precautionary motives moves the kink in the consumption function to the right of its true position. So, again, one way to match the observations would be to realign the consumption function produced by OccBin inferring a value of  $\gamma$  greater than the pseudo-true value of 1 used in the experiment. But as we have seen above, additional shifts to the left for the kink in the consumption function require disproportionately larger increases in  $\gamma$ , which is why the cumulation of the two sources of error, which happen to go in the same direction, results in a sizable magnification of the bias.

Finally, Figure 6 shows results analogous to those presented in Figure 3. The solid, blue line denotes again the likelihood contour based on the global solution with value function iteration in conjunction with the inversion filter. The dashed, red line denotes the likelihood contour based on a first-order perturbation solution in conjunction with a Kalman filter that excludes measurement error, in line with the data-generating process. We focus here on approximation error. Since the model is linearized at the constraint and the consumption function is unaffected by  $\gamma$  at the constraint, as the figure makes apparent,  $\gamma$  is not identified in this case.

## 6 Discussion and Related Literature

While we are not aware of other work that has analyzed the the interaction between multiple sources of error on the ability to infer the parameters of a dynamic model, ours is certainly not the first paper to consider the effects of approximation error.

Fernandez-Villaverde, Rubio-Ramirez, and Santos (2006) show that approximation error is compounded by the length of the sample  $T$ , so second-order errors in the policy function may

rise to the level of first-order differences in the likelihood function. The analysis in [Fernandez-Villaverde, Rubio-Ramirez, and Santos \(2006\)](#) abstracts from measurement error misspecification. In particular, it assumes some measurement error in the true data generating process, and assumes a correct specification of the measurement error properties upon estimation. In practice, when estimating nonlinear models with the particle filter, the importance of measurement error is usually calibrated so that the variance of the error is a fixed fraction of the variance of the observed variables.

[Fernandez-Villaverde and Rubio-Ramirez \(2005\)](#) simulate data with measurement error from a nonlinear real business cycle model and compares the estimation performance of using a Kalman filter with a linearized model versus a particle filter with a nonlinear model. However, [Fernandez-Villaverde and Rubio-Ramirez \(2005\)](#) abstract from possible misspecification in the measurement equation. Additionally, we focus on a specific form of nonlinearity – occasionally binding constraints – and show that, in the presence of this nonlinearity, a linearization strategy with a Kalman filter leads to especially poor performance for parameter estimation purposes.

## 7 Conclusion

Occasionally binding constraints create challenges for standard numerical solution algorithm as well as for likelihood-based inference methods. We showed that model misspecification related to measurement error can, in a simple model of consumption, flatten the likelihood function and lead to biased estimates of the model parameters. We also showed that this misspecification error can interact with approximation error and that the bias resulting from this interaction is greater than the bias associated with each error in isolation.

Each model presents specific solution and estimation challenges, but some of the results for our simple example are bound to extend to other models. A challenge unique to models with occasionally binding constraints is that if the observations are subject to measurement error, this error will muddle the inference on whether the constraints bind or not. Imposing that the variance of measurement error is a fixed fraction of the variance of each observed variable is likely to result in biased estimates of the frequency at which the various occasionally binding constraints bind, and through that channel, bias the estimates of other parameters that influence that frequency.

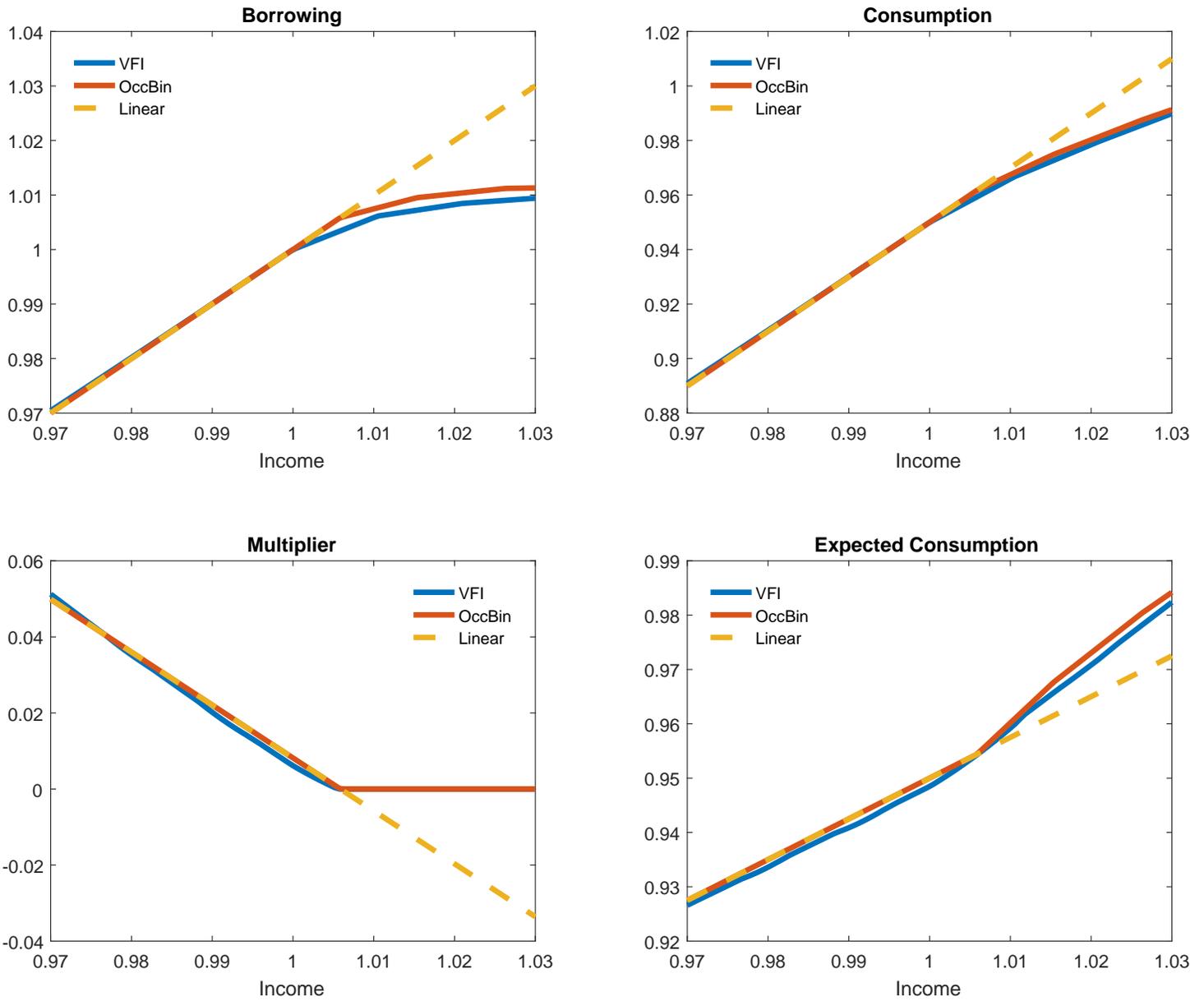
## References

- Amisano, G. and O. Tristani (2011). Exact likelihood computation for nonlinear DSGE models with heteroskedastic innovations. *Journal of Economic Dynamics and Control* 35(12), 2167–2185.
- Andreasen, M. M. (2013). Non-linear dsge models and the central difference kalman filter. *Journal of Applied Econometrics* 28(6), 929–955.
- Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(3), 269–342.
- Aruoba, B., P. Cuba-Borda, and F. Schorfheide (2018). Macroeconomic dynamics near the zlb: A tale of two countries. *The Review of Economic Studies* 85(1), 87–118.
- Bocola, L. (2016). The pass-through of sovereign risk. *Journal of Political Economy* 124(4), 879–926.
- Canova, F. (2009). How much structure in empirical models? In *Palgrave handbook of econometrics*, pp. 68–97. Springer.
- Canova, F., F. Ferroni, and C. Matthes (2014). Choosing the variables to estimate singular dsge models. *Journal of Applied Econometrics* 29(7), 1099–1117.
- Cuba-Borda, P. (2014). What explains the great recession and the slow recovery? *Manuscript, University of Maryland*.
- Deaton, A. (1992). *Understanding consumption*. Oxford University Press.
- Fair, R. C. and J. B. Taylor (1983). Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models. *Econometrica* 51(4), 1169–1185.
- Fernandez-Villaverde, J., P. Guerron-Quintana, and J. F. Rubio-Ramirez (2015). Estimating dynamic equilibrium models with stochastic volatility. *Journal of Econometrics* 185(1), 216–229.

- Fernandez-Villaverde, J. and J. F. Rubio-Ramirez (2005). Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics* 20(7), 891–910.
- Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007). Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies* 74(4), 1059–1087.
- Fernandez-Villaverde, J., J. F. Rubio-Ramirez, and M. S. Santos (2006, January). Convergence Properties of the Likelihood of Computed Dynamic Models. *Econometrica* 74(1), 93–119.
- Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). Chapter 9 - solution and estimation methods for dsge models. Volume 2 of *Handbook of Macroeconomics*, pp. 527 – 724. Elsevier.
- Guerrieri, L. and M. Iacoviello (2015). OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics* 70(C), 22–38.
- Guerrieri, L. and M. Iacoviello (2017). Collateral constraints and macroeconomic asymmetries. *Journal of Monetary Economics* 90(C), 28–49.
- Gust, C., E. Herbst, D. López-Salido, and M. E. Smith (2017). The empirical implications of the interest-rate lower bound. *American Economic Review* 107(7).
- Hausman, J. (2001, December). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives* 15(4), 57–67.
- Herbst, E. P. and F. Schorfheide (2016). *Bayesian Estimation of DSGE Models*. Princeton University Press.
- Judd, K. L. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory* 58(2), 410–452.
- Judd, K. L. (1998). *Numerical Methods in Economics*. The MIT Press.
- Justiniano, A. and G. E. Primiceri (2008). The Time-Varying Volatility of Macroeconomic Fluctuations. *American Economic Review* 98(3), 604–641.

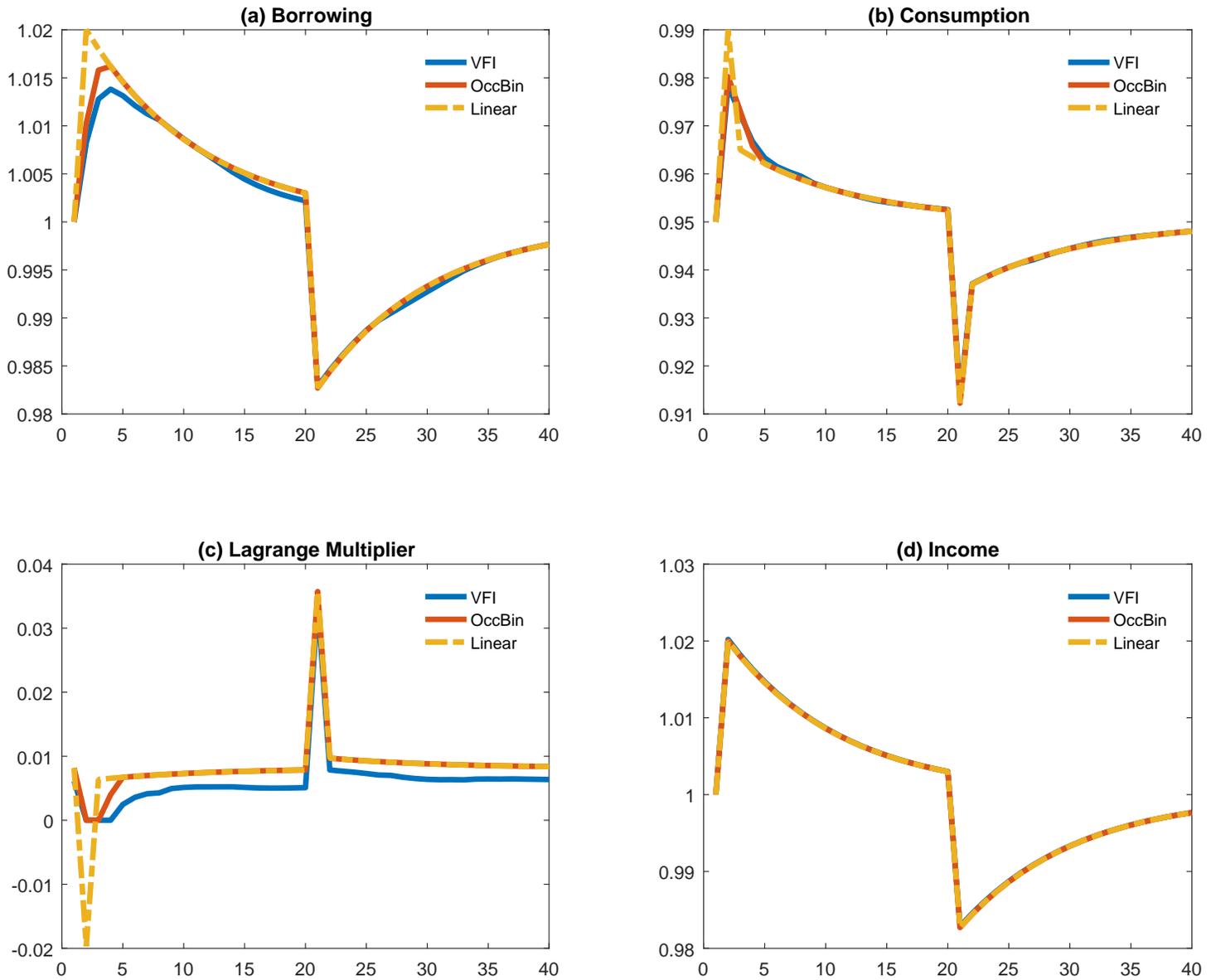
- Kim, J. and S. H. Kim (2007). Two Pitfalls of Linearization Methods. *Journal of Money, Credit and Banking* 39(4), 995–1001.
- Ljungqvist, L. and T. J. Sargent (2004). *Recursive Macroeconomic Theory, 2nd Edition*, Volume 1 of *MIT Press Books*. The MIT Press.
- Ljungqvist, L. and T. J. Sargent (2012). *Recursive macroeconomic theory*. MIT press.
- Ruge-Murcia, F. (2012). Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles. *Journal of Economic Dynamics and Control* 36(6), 914–938.
- Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters* 20(2), 177–181.
- van Binsbergen, J. H., J. Fernández-Villaverde, R. S. Koijen, and J. Rubio-Ramírez (2012). The term structure of interest rates in a dsge model with recursive preferences. *Journal of Monetary Economics* 59(7), 634 – 648.

Figure 1: POLICY FUNCTIONS OF THE CONSUMPTION-SAVINGS MODEL WITH OCCASIONALLY BINDING CONSTRAINTS



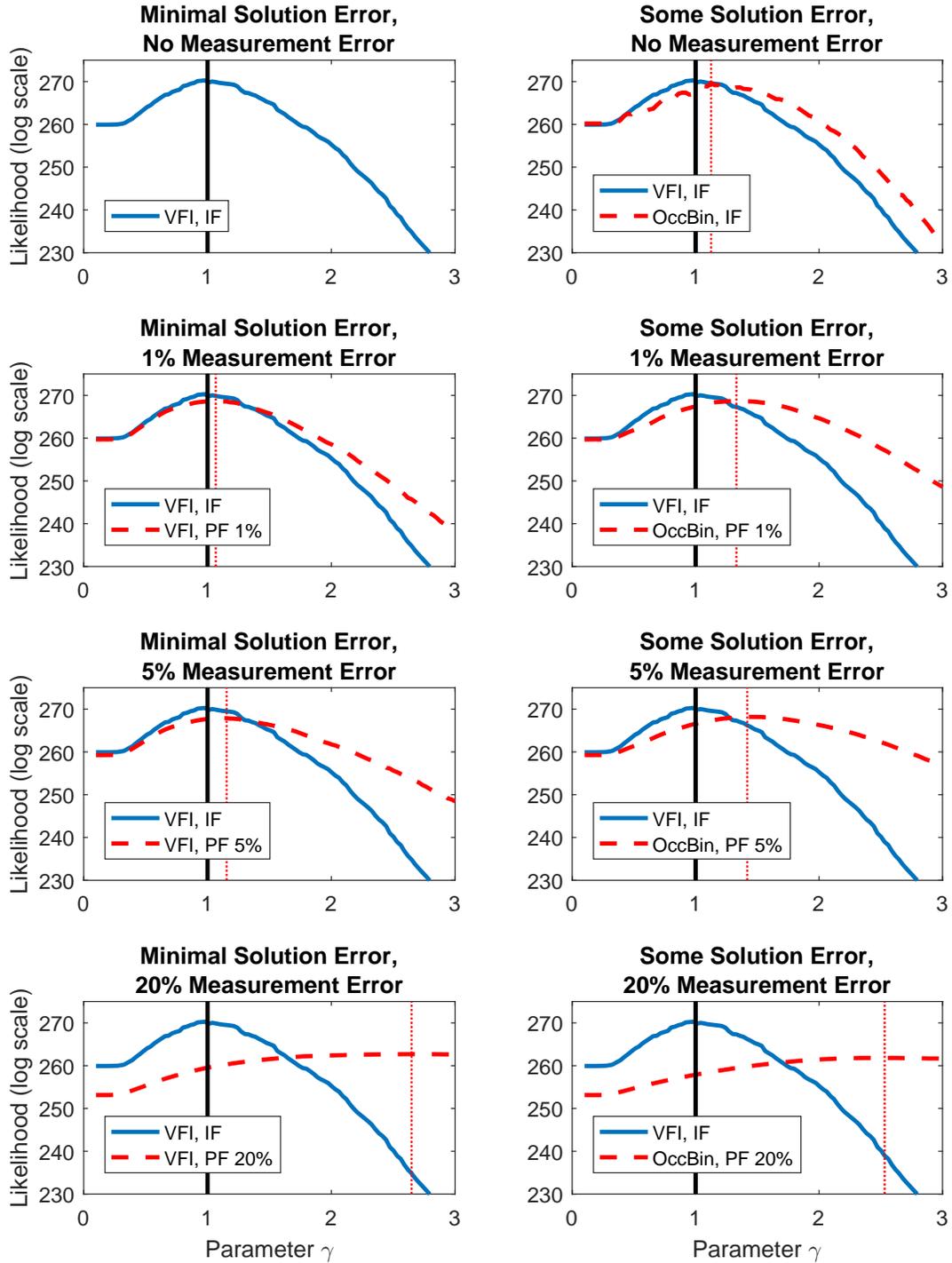
**Note:** “VFI” refers to the global solution method with value function iteration. “OccBin” refers to the piecewise linear solution in Guerrieri and Iacoviello (2015). “Linear” refers to a first-order perturbation solution. The policy functions shown are contours for an initial level of debt at its steady state value equal to 1.

Figure 2: IMPULSE RESPONSE FUNCTIONS OF THE CONSUMPTION-SAVINGS MODEL WITH OCCASIONALLY BINDING CONSTRAINTS



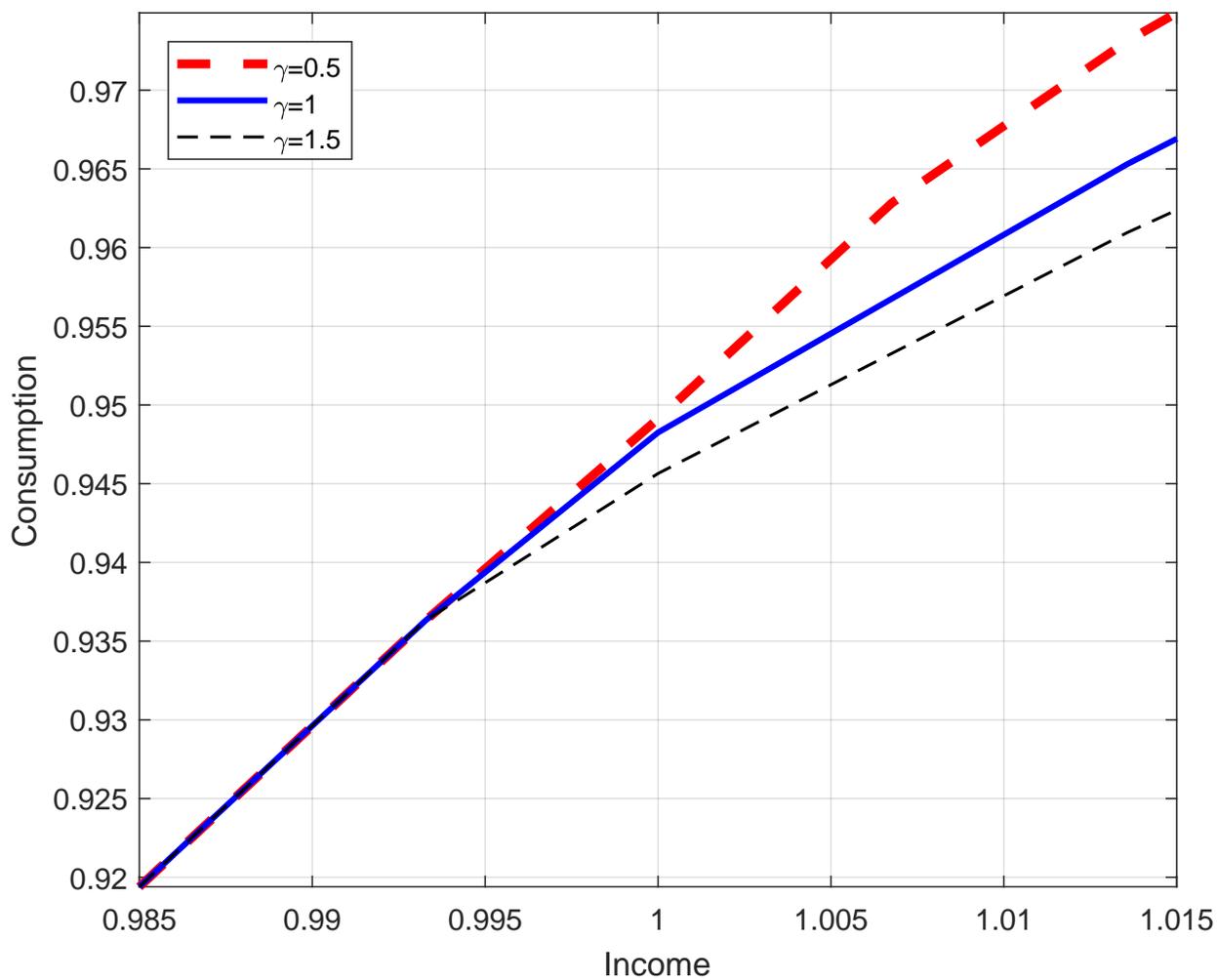
**Note:** “VFI” refers to the global solution method with value function iteration. “OccBin” refers to the piecewise linear solution in Guerrieri and Iacoviello (2015). “Linear” refers to a first-order perturbation solution. The figure shows two shocks to income. The first shock in period two brings up income by 2 percent. The second shocks in period 21 brings down income by 2 percent.

Figure 3: LIKELIHOOD CONTOURS FOR ALTERNATIVE SOLUTION METHODS AND FILTERS. THE CASE OF NO MEASUREMENT ERROR IN THE DGP



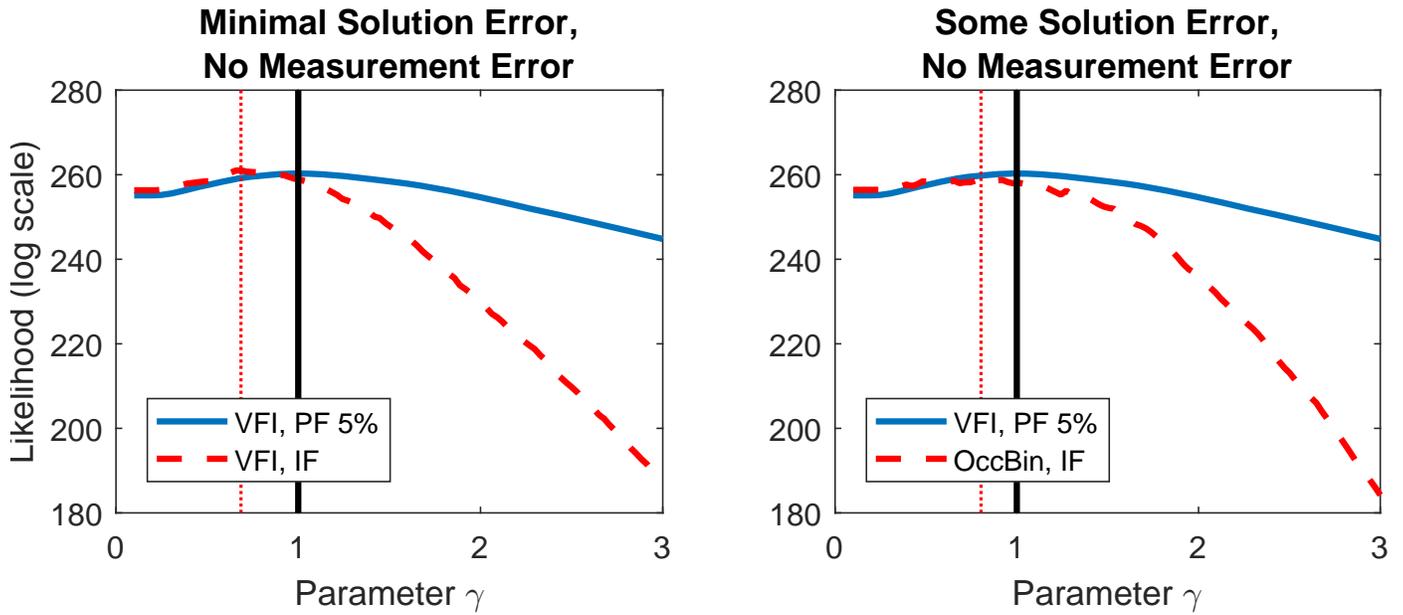
**Note:** VFI, OccBin and Linear refer, respectively, to the global solution with value function iteration, the OccBin solution in [Guerrieri and Iacoviello \(2015\)](#), and first-order perturbation solution. IF and PF refer, respectively, to the inversion filter and to the particle filter for varying levels of the measurement error. The benchmark “VFI, IF” combination excludes misspecification error and is least affected by approximation error. The vertical lines in each panel denote the peaks of the likelihood contours shown. Under the pseudo-true data generating process, the value coefficient of relative risk aversion,  $\gamma$  is 1.

Figure 4: CONSUMPTION FUNCTIONS FOR ALTERNATIVE VALUES OF THE COEFFICIENT OF RELATIVE RISK AVERSION,  $\gamma$



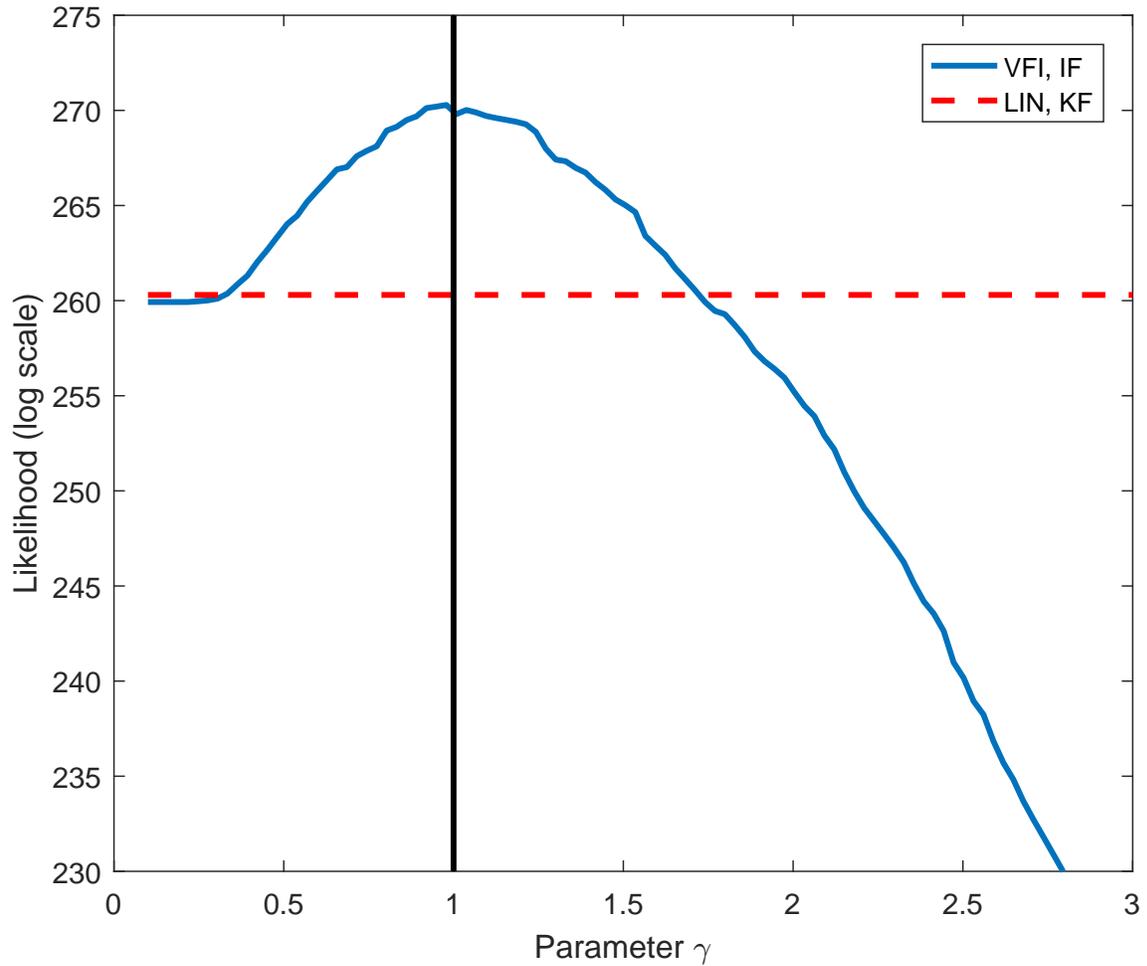
**Note:** The policy functions shown are contours, computed using the value function iteration solution, for an initial value of debt at its non-stochastic steady state level equal to 1.

Figure 5: CONTOURS OF THE LIKELIHOOD WHEN THE DGP INCORPORATES MEASUREMENT ERROR



**Note:** In this case, the DGP is the model solved with the value-function-iteration method and it includes 5 percent measurement error in the observation equation for consumption. The particle filter is calibrated to recognize the pseudo-true importance of measurement error in the DGP. The inversion filter incorrectly excludes measurement error.

Figure 6: COMPARISON OF LIKELIHOOD FUNCTIONS FROM ALTERNATIVE APPROXIMATIONS



**Note:** “VFI” refers to the global solution method with value function iteration. “LIN” refers to a first-order perturbation solution. “IF” refers to the inversion filter and “KF” to the kalman filter. The solid vertical line denotes the peak of the likelihood contour for the benchmark VFI, IF combination.

# Appendix

## A.1 Euler Errors

To gauge the quality of the approximation of each solution method that we consider, we follow the standard bounded rationality approach of [Judd \(1998\)](#). Accordingly, starting from the Euler equation for consumption of Equation (10) and reproduced here,

$$C_t^{-\gamma} = \beta RE_t (C_{t+1}^{-\gamma}) + \lambda_t,$$

we take each side to the power of  $-\frac{1}{\gamma}$  to obtain

$$C_t = [\beta RE_t (C_{t+1}^{-\gamma}) + \lambda_t]^{\frac{1}{\gamma}}.$$

This transformation allows us to express the Euler equation in units of consumption. We use each numerical solution method to evaluate the left-hand side and the right-hand side of the equation above and call the difference between the two sides an Euler residual function. [Figure A.1](#) plots contours of the residual function for each of the solution method considered. The residuals are expressed in log 10 scale. Accordingly, in the figure, a level of -1 can be interpreted as an error of \$1 per \$10 of consumption and a level of -5 as an error of \$1 per \$100,000 of consumption. The figure confirms the high accuracy of the value function iteration solution, modest errors for the OccBin solution and large errors for the linear solution that disregards the occasional binding constraint on borrowing.

## A.2 Particle filter details

This section gives additional details about the particle filter. Since its introduction to economics in [Fernández-Villaverde and Rubio-Ramírez \(2007\)](#), the particle filter has been a benchmark method used to estimate nonlinear DSGE models. It relies on simulation methods to approximate the nonstandard distributions of objects needed to calculate the likelihood.

Given structural parameters  $\theta$ , the likelihood of the model can be written in its prediction error decomposition form, as in Equation (15). In turn, the prediction error decomposition is related to the distribution of the hidden states  $s_t$  of the model in Equation (16) and the estimate of  $s_t$  given time  $t - 1$  data is related to the estimate of  $s_{t-1}$  given time  $t - 1$  data via Equation 17.

$$p(y_{1:T}|\theta) = \prod_{t=1}^T p(y_t|y_{1:t-1}, \theta) \quad (15)$$

$$p(y_t|y_{1:t-1}, \theta) = \int p(y_t|s_t, y_{1:t-1})p(s_t|y_{1:t-1})ds_t \quad (16)$$

$$p(s_t|y_{1:t-1}) = \int p(s_t|s_{t-1}, y_{1:t-1})p(s_{t-1}|y_{1:t-1})ds_{t-1} \quad (17)$$

In the linear Gaussian case, the Kalman filter delivers exact formulas for  $p(s_t|y_{1:t-1})$ ,  $p(s_{t-1}|y_{1:t-1})$ , and hence  $p(y_t|y_{1:t-1})$ . These exact formulas no longer hold in nonlinear models. The idea of the particle filter is to simulate many particles representing the hidden states. By appropriately propagating these particles through time, they will then approximate the distributions of interest necessary to calculate the likelihood.

Our paper relies on the most elementary implementation of the particle filter, known as the bootstrap particle filter (Fernandez-Villaverde and Rubio-Ramirez (2005), Fernandez-Villaverde, Rubio-Ramirez, and Santos (2006), Fernández-Villaverde and Rubio-Ramírez (2007)). The algorithm of the bootstrap particle filter is as follows:

0. **Initialization:** Simulate  $N$  particles from an initial distribution  $p(s_0)$ . Set the weight of each particle  $i$ :  $w_0^{(i)} = \frac{1}{N}$ .

Entering into time  $t$ , we have a swarm of  $N$  particles  $\{s_{t-1}^{(i)}\}_{i=1}^N$  and associated weights  $\{w_{t-1}^{(i)}\}_{i=1}^N$  that are distributed according to  $p(s_{t-1}|y_{1:t-1})$ .

1. **State forecast:** Take each particle  $s_{t-1}^{(i)}$  simulate it forward via the state transition equation  $h(s_{t-1}, \eta_t; \theta)$ . This then gives us a new swarm of particles  $\{s_t^{(i)}\}_{i=1}^N$  and associated weights  $\{w_{t-1}^{(i)}\}_{i=1}^N$  that are distributed according to  $p(s_t|y_{1:t-1})$ .

2. **Observable forecast:** For each particle  $s_t^{(i)}$ , we can calculate  $p(y_t|s_t^{(i)}, y_{1:t-1}) = p(y_t|s_t^{(i)})$ .

The weighted average of these particles is the discrete approximation to the integral that

defines  $p(y_t|y_{1:t-1}, \theta)$ . Call this approximation  $\hat{p}(y_t|y_{1:t-1}, \theta)$ .

$$p(y_t|y_{1:t-1}, \theta) = \int p(y_t|s_t, y_{1:t-1})p(s_t|y_{1:t-1})ds_t \approx \frac{1}{N} \sum_{i=1}^N p(y_t|s_t^{(i)})w_{t-1}^{(i)} = \hat{p}(y_t|y_{1:t-1}, \theta) \quad (18)$$

3. **State update:** We now incorporate time  $t$  information  $y_t$  to have the particles be distributed according to  $p(s_t|y_{1:t})$ . We know the following identity holds:

$$p(s_t|y_{1:t}) = \frac{p(y_t|s_t, y_{1:t-1})p(s_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \quad (19)$$

We already have particles distributed according to  $p(s_t|y_{1:t-1})$ . We can think about this as an importance sampling problem where we are using  $p(s_t|y_{1:t-1})$  as the proposal density. We can use the expression  $\frac{p(y_t|s_t, y_{1:t-1})}{p(y_t|y_{1:t-1})}$  to update each particle's importance weight. Therefore, the new importance weights are

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p(y_t|s_t^{(i)})}{\frac{1}{N} \sum_{i=1}^N p(y_t|s_t^{(i)})w_{t-1}^{(i)}} \quad (20)$$

4. **(Optional) Resampling step:** Every period, there is an optional step to replenish low weight particles with high weight ones. We draw  $N$  new particles from the set  $\{s_t^{(i)}\}_{i=1}^N$  with replacement according to their weights  $\{w_t^{(i)}\}_{i=1}^N$ . Usually this step is done when certain measures of particle degeneracy (such as effective sample size) reach critical thresholds.

We iterate steps 1 – 4 until time  $T$ . This completes the recursion.

#### *Importance of measurement error*

The particle filter relies on the existence of a greater or equal number of elements of stochasticity than observables:

$$\dim(\eta_t) + \dim(\zeta_t) \geq \dim(y_t) \quad (21)$$

With a lower dimension of randomness than the number of observables, the researcher runs into well known problems of stochastic singularity. In fact, the bootstrap particle filter as implemented in the algorithm actually relies on a more elements of randomness than observables. If the dimension of the shocks is exactly equal to the number of observables, only a finite number of

states  $s_t$  can exactly match the observables  $y_t$ , which makes the implementation of the state and observable forecast steps of the filter as currently discussed inoperable. In practice, researchers assume a certain amount of measurement error in each observable  $y_t$ .

Suppose that  $\zeta_t \sim N(0, \sigma_\zeta^2)$  and take a univariate  $y_t$  case for simplicity. The density  $p(y_t | s_t^{(i)})$  is then given by:

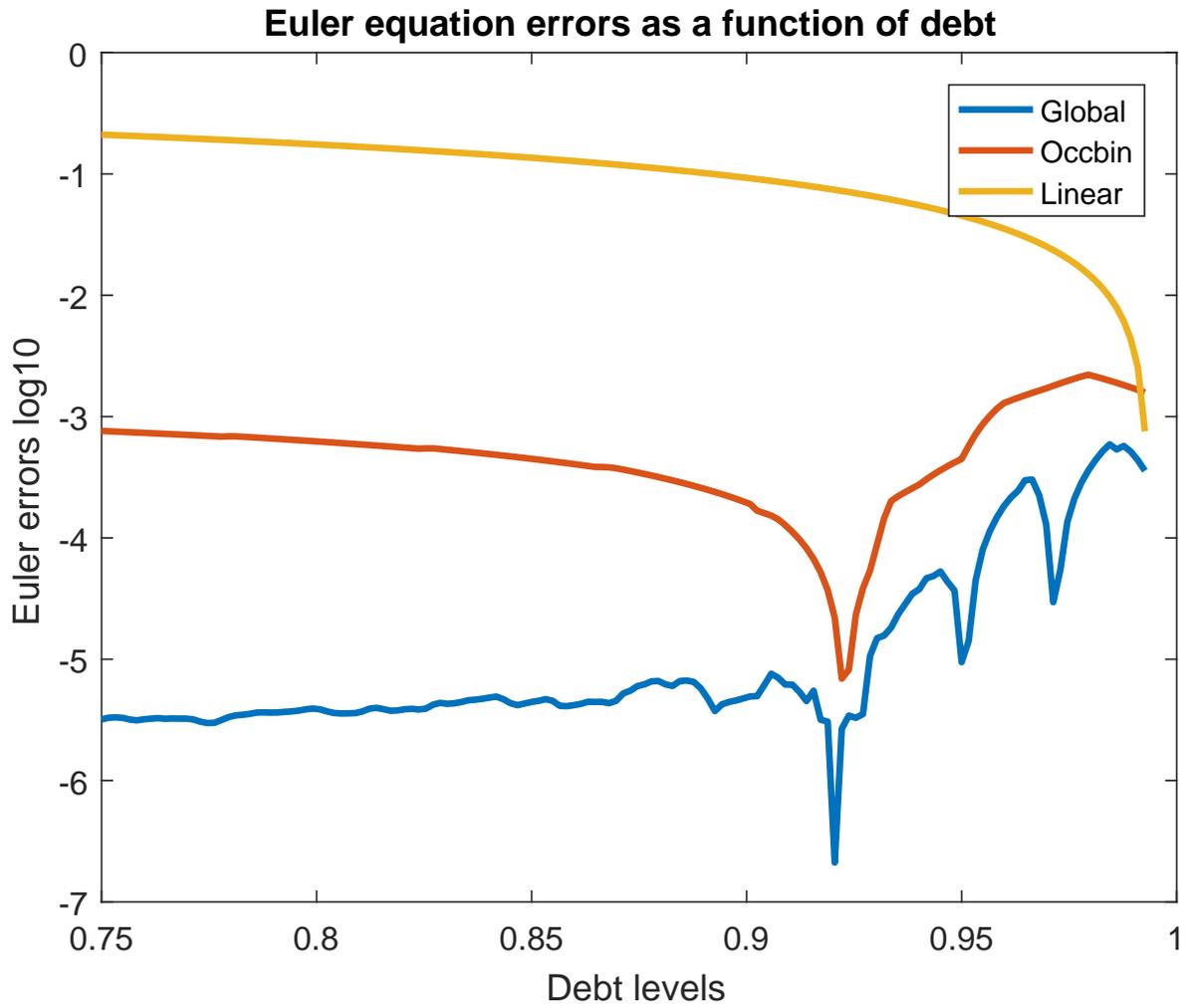
$$p(y_t | s_t^{(i)}) \propto \frac{1}{\sqrt{\sigma_e^2}} e^{-\frac{(y_t - g(s_t^{(i)}; \theta))^2}{2\sigma_e^2}} \quad (22)$$

As one can see from Equation (22), a smaller assumed standard deviation of the measurement error penalizes poor proposals  $s_t^{(i)}$  more. This increases the variance of  $p(y_t | s_t^{(i)})$  across particles and there also that of the weights  $w_t^{(i)}$ .

*Importance of the proposal density for  $s_t^{(i)}$*

Another important factor in determining the accuracy of the particle filter approximation to the likelihood is the proposal density used to generate the time  $t$  particles  $s_t^{(i)}$ . Given a particle swarm  $\{s_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^N$ , we are proposing the time  $t$  particles by simulating from the transition density. Therefore, our proposal distribution is  $p(s_t | s_{t-1})$ . Note however, that we are targeting  $p(y_t | s_t, y_{1:t-1})p(s_t | s_{t-1})$ . One can see this by examining Equation (19). Therefore, if we could bring in some information about  $p(y_t | s_t, y_{1:t-1})$  in our proposal density, this could improve the performance of the particle filter by tending to propose "high likelihood" particles. [Herbst and Schorfheide \(2016\)](#) discuss alternative strategies in forecasting the states. Another brute force option is to increase the number of particles in the particle filter.

Figure A.1: EULER ERRORS



**Note:**  $\log_{10}$  of absolute value of Euler equation errors.