# Conditional Initialization for Enhanced Inference in Nonlinear State-Space Models

Daniel Fehrle[a]

[a]Kiel University, Department of Economics, Wilhelm-Seelig-Platz 1, 24118 Kiel, Germany, fehrle@economics.uni-kiel.de

January 9, 2026

JEL classification: C32, C52, C68, E37
Keywords: Inversion Filter, Nonlinear DSGE Models, Conditional Likelihood

**Abstract**

This paper advances inference in nonlinear state-space models by deriving the distribution of initial endogenous states conditional on initial observations, addressing the general unknown nature of the unconditional distribution. Together with the inversion filter, this allows the likelihood to be evaluated deterministically and conditional on the initial observations. Such conditional likelihoods are commonly used in practice, e.g., when estimating autoregressive processes via least squares. Monte Carlo studies demonstrate that employing the conditional distribution of endogenous states can significantly enhance both Frequentist and Bayesian inference concerning states and parameters. The paper illustrates the method's practical relevance with two applications: first, a Monetary Business Cycle Accounting analysis of the COVID-19 recession and subsequent inflation surge in the US and Eurozone using a global solution; second, an evaluation of uncertainty shocks on the US business cycle during the Great Moderation via a third-order perturbation method. Additionally, the paper discusses limitations of the inversion filter and proposes remedies, particularly for spurious many-to-one policy function mappings that arise with higher-order perturbation solutions.

Acknowledgment: I am grateful to Christopher Heiberger and Johannes Huber for our prior collaborations, from which I reused several code snippets. I thank Gregor Boehl and Nikolay Iskrev for valuable comments and discussions. This research was supported through high-performance computing resources available at the Kiel University Computing Centre which is partly founded by the German Research Foundation (DFG)—project number 440395346.

## 1 INTRODUCTION

Macroeconomics requires nonlinear solutions to capture complex or asymmetric responses to shocks, as well as global solutions to handle large disturbances such as the COVID-19 pandemic. Conducting inference and parameter estimation in these models, in turn, necessitates the use of nonlinear filters. The inversion filter (Fair and Taylor, 1983) offers a computationally attractive method for this purpose, recovering latent shocks by inverting the observation equation. The change-of-variables theorem allows the likelihood of the data to be expressed as the product of the conditional shock density and the Jacobian determinant. Yet, the filter faces a critical limitation: in applications featuring latent endogenous states, the unknown distribution of the initial endogenous states and the recursive generation of subsequent states forces the results, including the likelihood, to be strictly conditional on these unknown initial values.[1] Conditioning on unknown initial endogenous states biases the resulting likelihood and, consequently, likelihood-based inference (Boehl and Strobel, 2024).

This paper resolves this limitation by characterizing the distribution of initial endogenous states conditional on initial observations. Using this distribution, the likelihood function obtained from the inversion filter can be expressed conditional only on the initial observations, which is standard practice in applied macroeconometrics, for example, when estimating autoregressive processes via least squares. The derivation of the conditional distribution of the endogenous states follows the logic of the inversion filter. Under local invertibility, the initial endogenous states and observations uniquely determine the corresponding exogenous shocks. Applying the change-of-variables theorem then allows the conditional distribution of the endogenous states to be expressed in terms of the stationary distribution of the shocks and a volume-adjustment term associated with the mapping to the exogenous shocks.[2]

The conditional endogenous state distribution enables three distinct likelihood functions: i) integrating out the initial endogenous states to obtain the exact conditional likelihood, ii) treating the initial states as parameters in the likelihood, or iii) constructing a profile likelihood in which the likelihood is maximized with respect to the initial states

[1] In practice, the endogenous states are set to their steady-state values, and the first 10-20 per-period likelihood contributions are typically discarded to mitigate the impact of the initial condition (e.g., Guerrieri and Iacoviello, 2017; Kollmann, 2017).

[2] For mappings from $\mathbb{R}^n \to \mathbb{R}^m$ with $m > n$, this term generalizes the usual Jacobian determinant and is also referred to as the Gram determinant of the Jacobian.

while remaining conditional on the initial observations. Crucially, when the initial endogenous states are treated as parameters in linearized models, the Jacobian determinant is constant with respect to these states and is therefore usually ignored during maximization with respect to these states.

After deriving the conditional distribution of the endogenous states and the various forms of the likelihood, I conduct a Monte Carlo analysis using a Dynamic Stochastic General Equilibrium (DSGE) model with one endogenous state and four first-order autoregressive exogenous shocks—the exogenous states. Here, I use a linear-Gaussian approximation of the model, applied to both the data-generating process and the econometric model, which allows to calculate the likelihood analytically and thereby provides a tractable benchmark distribution of estimators and posteriors (Herbst and Schorfheide, 2015; Farmer, 2021; Fehrle et al., 2025).

The first exercise focuses on state estimation when the model parameters are known. Using the conditional distribution of the endogenous states reduces the Root Mean Squared Error (RMSE) of the initial states by up to 25% compared to a naive initialization at the model's stable fixed point. When combined with additional information from the likelihood, the RMSE decreases by nearly 40% relative to the naive initialization. For comparison, Kalman filtering reduces the RMSE of the initial states by slightly more than 40%, and smoothing achieves reductions of up to 50%.

Building on the state estimation results, which delivers the conditional shock density needed for the derivation of likelihood, the subsequent exercises shift focus to parameter estimation using Maximum Likelihood (ML) and Bayesian estimation exercises for two structural parameters and all parameters of the stochastic processes. The usage of the conditional distribution of the endogenous states decreases the RMSE to the exact maximum likelihood estimates up to 50% compared to a naive initialization. This improvement remains consistent even when incorporating a 10-to-20-period burn-in phase to the naive initialization. As one of the exogenous states can be inferred independent of the endogenous state, i.e., directly from the data, the corresponding estimates are equivalent to estimating an autoregressive process using least squares. Most of the remaining parameter estimates yield RMSE of a similar magnitude when using the conditional endogenous state distribution.

When evaluating the whole posterior with uninformative priors, the improvement is even greater, with the RMSE reduced by up to 75%. The benefits of using the conditional endogenous state distribution are also evident in state estimation with estimated parame-

ters and in forecasting based on estimated parameters.

I use these insights for two applications. First, I conduct a Monetary Business Cycle Accounting (MBCA) analysis (Šustek, 2011; Chari et al., 2007) of the COVID-19-induced recession and the subsequent inflation surge in the US and the European Monetary Union (EMU) using a global solution method. It turns out that the labor wedge dominates the recession and drives inflationary pressure at the pandemic peak, while the investment wedge acts countercyclical but also inflationary. Production efficiency contributes modestly. The bond and monetary wedge act inflationary after the pandemic but suppressed inflation during the pandemic itself.

As a second application, I measure the impact of uncertainty shocks on the US business cycle during the Great Moderation (1985 – 2019) using the model from Basu and Bundick (2017) with a third-order perturbation. This New-Keynesian model features uncertainty shocks that can generate co-moving responses in consumption, investment, and hours worked. By estimating the underlying shock process parameters and a single structural parameter, together with inferring the latent states, the analysis indicates that uncertainty shocks account for most of the fluctuations during the Great Moderation, whereas productivity shocks have little influence on the cycle.

Unfortunately, the inversion filter comes not without limitations. In general, due to the filters logic, it requires a one-to-one (bijective) policy function, which also implies that the number of observables must equal the number of exogenous states. The initialization method proposed here introduces a third condition: the number of endogenous states must not exceed the number of exogenous states. I discuss how these restrictions, while potentially limiting in theory, do not pose significant constraints for large classes of applications in macroeconomics and dynamic economics. Moreover, I outline possible remedies and extensions that can relax or work around these limitations in practice, including cases where errors from local approximation transform an otherwise one-to-one policy function into a spurious many-to-one mapping.

Kollmann (2017) addresses the problem of spurious many-to-one mappings that arise from higher-order Taylor expansions by inverting the pruned policy function of Kim et al. (2008) and setting the remaining higher-order terms that depend solely on exogenous states equal to their expected values. In this study, I propose instead using the Lagrange Inversion Theorem, which provides a formal and systematic method for locally inverting an analytic function via a series expansion.

The analyses of Amisano and Tristani (2011), Guerrieri and Iacoviello (2017), and Cuba-

Borda et al. (2019) do not encounter spurious many-to-one mappings, as they consider MIT regime-switching models with otherwise linear policy functions.[3] However, Boehl and Strobel (2024) report root-finding problems even in one-to-one policy functions due to non-convergence of numerical algorithms. In such cases, the Lagrange Inversion Theorem can also be beneficial. Likewise, although sufficiently accurate global solution methods should preclude spurious many-to-one mappings when the actual policy function is one-to-one, the Lagrange Inversion Theorem may provide a remedy when root-finding algorithms fail to converge.

The literature has already documented several advantages of the inversion filter. Fair and Taylor (1983), Guerrieri and Iacoviello (2017), Kollmann (2017), Cuba-Borda et al. (2019), Atkinson et al. (2020), Huber (2022), and Fehrle and Huber (2023) emphasize its numerical efficiency. Moreover, Fehrle and Huber (2023) show that, when the model is linearized, second moments of the innovations can be estimated analytically.[4] Finally, Cuba-Borda et al. (2019) highlight that the inversion filter does not require the introduction of measurement errors generally. In this paper, I further emphasize the advantages of the filter's likelihood differentiability. In particular, I show that this approach can substantially improve accuracy in ML estimation and in sampling with a Random-Walk Metropolis-Hastings algorithm. This enhancement is notable when compared to a benchmark bootstrap particle filter, primarily because the filter allows for derivative-based optimization.[5] That said, the goal of this study is not to compare different classes of nonlinear filters, but to enhance the statistical efficiency of the inversion filter and thereby to broaden the toolbox for nonlinear state-space estimation.

In early studies on inversion filters, endogenous states were not latent (Fair and Taylor, 1983) or non-existent (Amisano and Tristani, 2011). Studies with latent endogenous states set these endogenous states to the steady state (Guerrieri and Iacoviello, 2017; Kollmann, 2017; Atkinson et al., 2020; Huber, 2022; Fehrle and Huber, 2023; Boehl and Strobel, 2024). Kollmann (2017) and Guerrieri and Iacoviello (2017) discard the first 10 and 20 per-period likelihood contributions, respectively. Cuba-Borda et al. (2019) conduct a Monte Carlo study, where the initial states are set to their true values. Fehrle and

[3] Holden (2023) and Boehl and Strobel (2024) emphasize that many-to-one mappings may arise when the possibility of switching regimes is anticipated.

[4] Note that in the linear case and an initialization at the steady-state, the inversion filter is equivalent to a steady-state Kalman filter.

[5] Accuracy in the Bayesian sampler is enhanced as the proposal distribution's covariance matrix can be set to the inverse of the negative Hessian at the posterior mode (Herbst and Schorfheide, 2015, Chapter 4).

Huber (2023) use the results from the inversion filter as an initial guess for a more time-consuming estimation based on the exact likelihood. Huber (2022) shows how the exact likelihood in linear state-space models can be reconstructed from the recursion that is conditional on the initial states, thereby reducing computational time for the exact likelihood.

Lastly, it is worth noting that the Monte Carlo analyses of Cuba-Borda et al. (2019), Atkinson et al. (2020), and Boehl and Strobel (2024) compare the inversion filter either to the true parameter values or to other likelihood approximations. In contrast, as mentioned, in this paper I compare the inversion filter directly to the exact likelihood, evaluating how the distribution of the estimator derived from the inversion filter matches the true estimator distribution. In this way, the paper contributes to the broader understanding of the use of the inversion filter.

The remainder of the paper is organized as follows. The first part presents the methodological advancements, including the derivation of the conditional distribution of the endogenous states and the corresponding likelihoods. The analytic derivation is followed by a Monte Carlo study and a discussion that addresses, among other topics, how to handle spurious many-to-one mappings. The second part presents the two applications: MBCA and uncertainty shocks. Finally, the paper concludes. An appendix provides additional results and information, particularly confirming the robustness of the analysis presented here.

## 2 Inversion Filter

### 2.1 Analytical Framework

To make the use of the conditional distribution of the initial endogenous states for the likelihood function straightforward, the exposition begins with a model of purely exogenous latent states, as applied, for example, by Amisano and Tristani (2011). In this setting, both the exact likelihood and the likelihood conditional on the initial observations can be obtained using the inversion filter. For models that also include endogenous states, the presentation of the likelihood conditional on the initial endogenous states then becomes straightforward. This likelihood corresponds to the benchmark inversion filter's in the recent literature.

I then present the paper's main contribution. I first derive the distribution of the initial endogenous states conditional on the initial observations using the change-of-variables

theorem. I then show how this distribution can be used to marginalize over the initial endogenous states to obtain the likelihood conditional solely on the initial observations again. As an alternative, I construct a likelihood that explicitly includes the probability distribution of the initial endogenous states and the profile likelihood with respect to the initial endogenous states.

Consider now the first case: a state-space model $y_t = f(z_t)$, where $y_t$ is a set of observables and $z_t$ is a vector of purely exogenous latent states following a stationary Markov process. The conditional transition density is given by $p(z_{t+1} \mid z_t) = p_\epsilon(\epsilon_{t+1})$ and the stationary unconditional distribution by $p_z(z_t)$. Both distributions are closed form. If $\dim(z_t) = \dim(y_t)$ and $f$ is one-to-one and continuously differentiable (i.e., a diffeomorphism on the support of $z_t$), then $z_t$ is uniquely determined by $y_t = f(z_t)$ and the change-of-variables theorem yields the log-likelihood:

$$\mathscr{L} = \mathscr{L}(Y_{1:T} \mid \Theta) = \ln\Big[p_z(z_1)\Big|\det \tfrac{\partial z_1}{\partial Y_1}\Big|\Big] + \sum_{t=1}^{T-1} \ln\Big[p_\epsilon(\epsilon_{t+1})\Big|\det \tfrac{\partial \epsilon_{t+1}}{\partial y_{t+1}}\Big|\Big],$$

where $z_t = f^{-1}(y_t)$, $\epsilon_{t+1} = z_{t+1} - \mathbb{E}_t[z_{t+1}]$, and $Y_{1:T} = \{y_t\}_{t=1}^T$.

Conditioning on the first $T^b$ observations gives

$$\mathscr{L}^{T_b} = \mathscr{L}(Y_{T^b+1:T} \mid Y_{1:T^b}; \Theta) = \sum_{t=T^b}^{T-1} \ln\Big[p_\epsilon(\epsilon_{t+1})\Big|\det \tfrac{\partial \epsilon_{t+1}}{\partial y_{t+1}}\Big|\Big].$$

For illustration, a least squares estimation of an observable $q$-order autoregressive process is equivalent with a conditional ML estimation with $T^b = q$.

Now consider the standard case, where the model includes an additional vector of endogenous states $x_t$, with observations given by $y_t = f(x_t, z_t)$ and dynamics $x_{t+1} = g(x_t, z_t)$. Given $x_1$ and $y_1$, the invertibility of $f$ in $z_t$ implies $z_1$ is uniquely determined—for $\dim(y) = \dim(z)$. Then, $(x_t, z_t) \mapsto (x_{t+1}, z_{t+1})$ is recursively determined: from $(x_t, z_t)$, the function $g$ yields $x_{t+1}$; the observation $y_{t+1}$ then identifies $z_{t+1}$. Iterating this procedure determines $(x_t, z_t)$ for $t = 1, \ldots, T$, given $x_1$ and $Y_{1:T}$. While the exact likelihood function is unknown, as long as the unconditional, stationary distribution $p_x(x_t)$ is unknown, the likelihood function conditional on $x_1$ and the first $T^b$ observables read

$$\mathscr{L}^{T_b,x} = \mathscr{L}(Y_{T^b+1:T} | Y_{1:T^b}, x_1; \Theta) = \sum_{t=T^b}^{T-1} \ln\left(p_\epsilon(\epsilon_{t+1}(x_1))\left|\det \frac{\partial \epsilon_{t+1}}{\partial y_{t+1}}\right|\right). \tag{1}$$

In general, $x_1$ is unobserved, and any misspecification of this initial condition biases the likelihood. However, the bias diminishes gradually under certain conditions as $t$ increases. Thus, the literature on inverted filters typically discards the first few per-period likelihood contributions, setting $T^b >> 1$ (Kollmann, 2017; Guerrieri and Iacoviello, 2017).

While the problem of the unknown unconditional stationary distribution $p_x(x_t)$ stays unsolved, the conditional density of $x_1$ given $y_1$ ($p(x_1|y_1)$) can be obtained by a change of variables. Specifically, regarding $y_1$ as deterministic, inverting the system delivers $z_1 = f_z^{-1}(y_1, x_1)$. As long as the number of endogenous states does not exceed the number of exogenous states ($\dim(x) \leq \dim(z)$), the conditional density of the initial state is well-defined. Specifically, if we define the Jacobian

$$J(x_1) = \frac{\partial f_z^{-1}(y_1, x_1)}{\partial x_1} \in \mathbb{R}^{\dim(z)\times\dim(x)},$$

then the density can be expressed as

$$p(x_1 \mid y_1) = p_z\big(f_z^{-1}(y_1, x_1)\big)\,\sqrt{\det\big(J(x_1)^\top J(x_1)\big)}. \tag{2}$$

In the exactly identified case ($\dim(x) = \dim(z)$), this reduces to the usual determinant $|\det J|$. In overidentified cases ($\dim(x) < \dim(z)$), the factor $\sqrt{\det(J^\top J)}$ generalizes the Jacobian to a rectangular mapping, giving the appropriate volume scaling of $x_1$ in $z$-space. This is known as the Gram determinant.

From this conditional distribution of $x_1$, the likelihood, solely conditional on the first $T^b$ observations, is

$$\mathscr{L}^{T_b} = \mathscr{L}(Y_{T^b+1:T} \mid Y_{1:T^b}; \Theta) = \ln \int p(x_1 \mid Y_1) \prod_{t=T^b}^{T-1} p_\epsilon\big(\epsilon_{t+1}(x_1)\big) \left|\det \tfrac{\partial \epsilon_{t+1}}{\partial y_{t+1}}\right| dx_1. \tag{3}$$

Note that for standard DSGE models, where $z_t$ is solely first-order autoregressive, $T^b = 1$ is sufficient.

Alternatively, if $x_1$ is treated as a parameter the likelihood reads

$$\mathscr{L}_f^{T_b} = \mathscr{L}(Y_{T^b+1:T} \mid Y_{1:T^b}; \{x_1, \Theta\}) = \ln p(x_1 \mid Y_1) + \sum_{t=T^b}^{T-1} \ln\Big[p_\epsilon\big(\epsilon_{t+1}(x_1)\big) \left|\det \tfrac{\partial \epsilon_{t+1}}{\partial y_{t+1}}\right|\Big], \tag{4}$$

and the corresponding profile likelihood

$$\mathscr{L}_p^{T_b} = \mathscr{L}_p(Y_{T^b+1:T} \mid Y_{1:T^b}; \Theta) = \max_{x_1} \mathscr{L}(Y_{T^b+1:T} \mid Y_{1:T^b}; \{x_1, \Theta\}). \tag{5}$$

### 2.2 Monte Carlo analysis

In this subsection, I present the Monte Carlo analysis, which quantifies the improvements from using the conditional distribution of the endogenous states for both state and parameter inference in a practical setting.

In line with the Monte Carlo study from Kollmann (2017), the setup is a real business cycle model with distorted first-order conditions. The model reads in canonical form

$$\alpha \frac{y_t}{n_t} = \exp(z_{Nt}) \theta \frac{c_t}{1-n_t} \tag{6}$$

$$1 = \exp(z_{Bt}) \beta \mathbb{E}_t \left[ \left( \frac{c_{t+1}}{c_t} \right)^{\sigma} \left( \frac{1-n_{t+1}}{1-n_t} \right)^{\sigma(1-\theta)} \left( 1 - \delta + \alpha \frac{y_{t+1}}{k_{t+1}} \right) \right] \tag{7}$$

$$y_t = \exp(z_{At}) k_t^{\alpha} n_t^{1-\alpha}, \tag{8}$$

$$y_t = i_t + c_t + g_t, \tag{9}$$

$$k_{t+1} = i_t + (1-\delta) k_t, \tag{10}$$

$$g_t = \gamma y_t \exp(z_{Gt}), \tag{11}$$

$$z_{it+1} = \rho_i z_{it} + \omega_i \epsilon_{it+1}, \ \epsilon_{it+1} \sim \mathscr{N}(0,1), \ i \in \{A, B, G, N\}, \tag{12}$$

and the parameters $\{\beta, \sigma, n^*, \alpha, \delta, \rho_i\} = \{0.99, 2, 0.3, 0.37, 0.014, 0.95\}$ as well as $\omega_i = 0.01$ for $i \in \{A, G, N\}$ and $0.0025$ for $i = B$. As standard in the DSGE literature, instead of calibrating $\theta$, I pin down hours worked in the steady state $n^*$, which in turn determines $\theta$ from eq. (6), evaluated at the model's deterministic steady state. Note that the states $z_{it}$ follow univariate first order autoregressive processes and correspond to the exogenous states $z_t$ in the section above, while the state $k_t$ arises endogenously in line with $x_t$ in the section above.

Although the primary goal is to improve non-linear filtering, I conduct the analyses using the linearized version of the model—both as the data-generating process and as the econometric model. This is a common approach (e.g., Herbst and Schorfheide, 2015, Part III; Farmer 2021,Fehrle et al. 2025) as it entails no loss of generality while offering two key advantages. First, solving the linearized model is computationally efficient, which allows us to use a large number of Monte Carlo replications and thereby reduce sampling variabil-

ity. Second, and more importantly, the exact likelihood function be evaluated analytically only for the linear case by means of the Kalman filter. This provides a clean benchmark for assessing filtering performance.

The Monte Carlo setup is as follows. I simulate the model $N = 1024$ times, each over 1200 periods, with the first 1000 periods serving as a burn-in phase. I then use the controls $\{Y_t\}_{t=1001}^{1200} = \{y_t, c_t, i_t, n_t\}_{t=1001}^{1200}$ as the observable variables, in line with $y_t$ in the section above. This results in a total of $n_Y \times T \times N = 4 \times 200 \times 1024$ observations. The model is calibrated at a quarterly frequency, so the effective time span of the retained data corresponds to 50 years—comparable to the data availability in many developed and emerging economies.

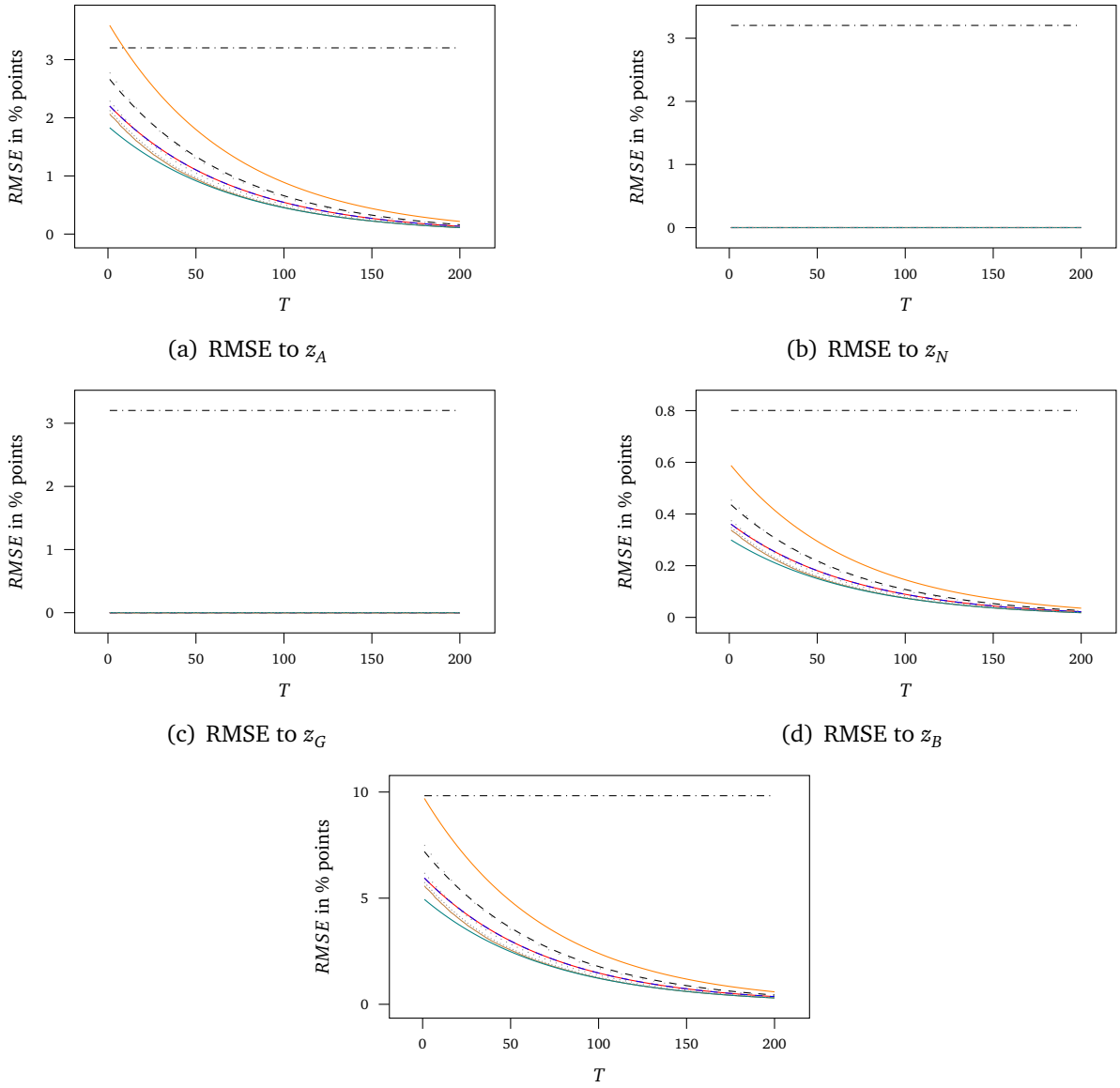
### 2.2.1 State estimation

In a first step, I investigate the impact of different initialization methods on the estimates of the latent states, assuming all parameters are known. Figure 1 plots the RMSE of these estimates. The yellow line corresponds to the inverted filter initialized at the unconditional mean of the endogenous state, $\mathbb{E}k_t$ ($\mathbb{E}x_1$). The black dashed line shows the inverted filter initialized at the point estimate of $k_1$ given $Y_1$ ($(\mathbb{E}_{y_1} x_1)$ eq. (2)), while the red line represents the inverted filter initialized with the full distribution of $k_1$ conditional on $Y_1$ ($\mathscr{L}^1$, eq. (3))—effectively implementing smoothing conditional on $Y_1$. The blue dashed line reflects the maximum likelihood estimate of $k_1$, obtained from the likelihood conditional on $Y_1$ ($\mathscr{L}_p^1$, eq. (4)). The brown line shows the estimates from the Kalman filter, and the green line those from the Kalman smoother. The gray lines—loosely, normally, and densely dotted—represent estimates using an approximation of the full (unconditional) likelihood from bootstrap particle filters with $10^2$, $10^3$, and $10^4$ particles, respectively. Finally, the dashed-dotted line indicates the RMSE when using the unconditional mean of the state as the estimate for all $t$.

It is worth noting that, since we observe $y_t, c_t, i_t$, we can infer $g_t$ exactly from the linearized version of equation (9), and consequently also recover $z_{Gt}$ from the linearized version of equation (11). In the same manner, using $y_t, c_t, n_t$ and equation (6), we can exactly infer $z_{Nt}$. As a result, all filters and smoothers estimate the states $z_{Gt}$ in Panel (c) and $z_{Nt}$ in Panel (d) up to numerical precision correctly.

Regarding the endogenous state $k_t$ in Panel (e), the filter that initializes $k_0$ at the unconditional mean $\mathbb{E}k_t$ results in an RMSE close to 10 percentage points—the unconditional

**Figure 1:** State estimation




**Notes:** RMSE of various state filter and smoother from 1024 samples in %. $\mathbb{E}(x_1)$ – initial endogenous state equal first moment; $\mathbb{E}_{y_1}(x_1)$ – initial endogenous state conditional on $Y_1$; $\mathscr{L}^1$ – smoothing with likelihood conditioned on $y_1$; $\mathscr{L}_p^1$ – likelihood maximizing initial endogenous state; $10^x$ Particles – filtering with exact likelihood, approximated using $10^x$, $x \in \{2,3,4\}$ particles; $\mathbb{E}([x,z]')$ – unconditional state expectation. All estimates based on time series with 200 periods.

standard deviation of $k_t$. Using the first observation to estimate the initial $k_t$ (eq. (2)) reduces the RMSE by 25%, as indicated by the dotted line. Both smoothing with the likelihood conditional on $Y_1$ ($\mathscr{L}^1$) and using the profile likelihood further reduce the RMSE, bringing it nearly 40% lower than the naive initialization at $\mathbb{E}k$. The conditional likelihood approach yields $k_t$ estimates nearly as accurate as those from the Kalman filter, which results in an RMSE slightly more than 40% lower than the naive initialization ($k_1 = \mathbb{E}k$). The Kalman smoother outperforms all other estimators, with an RMSE that is 50% lower than the naive initialization. Hence, using the conditional likelihood from the inverted filter brings the RMSE for $k_1$ down to 80% of the best estimator. The accuracy of the particle filter with a small number of particles ($10^2$) is similar to that obtained by using the first observation, $Y_1$, to estimate the endogenous state (eq. (1)). Increasing the number of particles by a factor of 10 improves accuracy, delivering an RMSE similar to that of smoothing with the conditional likelihood on $Y_1$ and maximizing $k_1$ given this likelihood. Increasing the amount of particles by another factor of 10 further closes the gap to the analytical filter by more than half.[6]

The patterns of RMSE of $z_{At}$ and $z_{Bt}$ in Panel (a) and (d) are similar to the estimates for $k_t$. Notably, initializing the endogenous state at its mean results in a larger RMSE for the estimation of $z_{A1}$ compared to a naive estimation using the unconditional mean $\mathbb{E}z_{At}$.

### 2.2.2 Full estimation

Next, I perform a Bayesian estimation with respect to the parameter vector

$$\Theta = \{\sigma, n^*, \rho_A, \rho_N, \rho_G, \rho_B, \omega_A, \omega_N, \omega_G, \omega_B\},$$

assuming flat, uninformative priors, as specified in Table 1. The remaining parameters are fixed at their true values.

Given the model, the observables, and the known parameters, this analysis examines how different sources of uncertainty affect the estimation of states and parameters. First, since $\gamma$ is known, we can still infer the state $z_{Gt}$ with certainty. Consequently, the differences between the exact, unconditioned likelihood estimates and the conditional likelihood estimates for $\rho_G$ and $\omega_A$ are analogous to the difference between estimating an first-order autoregressive process using the exact likelihood versus least squares, where

[6] Note that smoothing with particle filters is a non-trivial task. As a result, states are almost always estimated using filtering, not smoothing.

**Table 1:** Prior distribution

| Parameter | Lower Bound | Upper Bound | Prior Distribution |
|---|---|---|---|
| $n^*$ | 0.1 | 0.5 | $\mathcal{U}(0.1, 0.5)$ |
| $\sigma$ | 0.01 | 10 | $\mathcal{U}(0.01, 10)$ |
| $\rho_A$ | 0 | 0.999 | $\mathcal{U}(0, 0.999)$ |
| $\rho_N$ | 0 | 0.999 | $\mathcal{U}(0, 0.999)$ |
| $\rho_G$ | 0 | 0.999 | $\mathcal{U}(0, 0.999)$ |
| $\rho_B$ | 0 | 0.999 | $\mathcal{U}(0, 0.999)$ |
| $\omega_A$ | 0.001 | 0.1 | $\mathcal{U}(0.001, 0.1)$ |
| $\omega_N$ | 0.001 | 0.1 | $\mathcal{U}(0.001, 0.1)$ |
| $\omega_G$ | 0.001 | 0.1 | $\mathcal{U}(0.001, 0.1)$ |
| $\omega_B$ | 0.00025 | 0.025 | $\mathcal{U}(0.00025, 0.025)$ |

the latter is the standard in applied times series analysis. Second, because $\theta$ depends on the estimates of the steady state $n^*$, now the estimates of $z_N$ and the related parameters $\rho_N$ and $\omega_N$ are influenced by the uncertainty in $n^*$. Third, since $\alpha$ is known, the estimates of $z_A$ and the related parameters $\rho_A$ and $\omega_A$ are affected solely by the uncertainty regarding the initial endogenous state $k_0$ via equation (8). Lastly, the estimates of $z_B$ and the related parameters $\rho_B$ and $\omega_B$ as well as $\sigma$ and $n^*$ depend on the entire uncertainty.

**Posterior's mode—constraint maximum likelihood** First, I compare the posterior's mode resulting from the different conditional likelihoods with the unconditional one. Since the priors are uniformly distributed, the parameter values at the posterior's mode coincide with their (constrained) maximum likelihood estimates. To avoid potential issues with local maxima, I perform global maximization using 200 stage-one points and 1,000 trial points. Additionally, I conduct the analysis using a particle filter with $10^3$ Particles initialized with draws from the stationary distribution of the states. The maximization of the particle-filter likelihood uses a particle swarm filter using 100 particles in each generation.

Table 2 presents the RMSE of the ML estimates of the parameters and the resulting initial states to the exact ML estimates. The RMSE of the parameters is relative to the prior range. Again, $\mathscr{L}^1$ indicate the results from the likelihood conditioned on the first observations of $Y_1$ (eq. (3)) and $\mathscr{L}_p^1$ is the profile likelihood (5)). The likelihood conditioned on $k_1 = \mathbb{E}k$ is represented by $\mathscr{L}^{1,x=\mathbb{E}x}$, while $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ indicate estimates from the same initialization, however, the first 10 and 20 periods are burned (eq. 1, $T^b \in \{1, 10, 20\}$). Additionally, I report estimates using an approximation of the full (unconditional) likeli-

hood from a bootstrap particle filter with $10^3$particles. One, two, and three upper asterisks indicate $p$-values lower than 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$ and lower asterisks for the null hypothesis that the estimate equals that from $\mathscr{L}_p^1$.

As part of the main contribution of this study, the estimates solely conditional on the first observations of $Y_t$ ($\mathscr{L}^1$, $\mathscr{L}_p^1$) are significant superior than the estimates from $\mathscr{L}^{1,x=\mathbb{E}x}$ or statistically not different. The latter holds for $\rho_N$ and by construction for all estimations regarding $g$ ($\rho_g$, $\omega_G$, $z_{G1}$). As the RMSEs regarding $g$ are equivalent to RMSEs of the common approach to estimate autoregressive processes with least squares instead of the exact likelihood, we can consider the RMSEs regarding $g$ as an acceptable magnitude. From this point of view in turn, we can consider the improvements due to conditioning on $Y_1$ regarding the reaming shock parameters as high, as they bring down the RMSEs from up to more than twice as high as the $g$-estimates to be partly even lower than of the $g$-estimates.

Of additional interest is that, contrary to the intuition posed by the literature, the loss of information due to the burned periods is more decisive than the gain due to only using more exact states. Except of $\rho_A$, where the burn-in gives indeed slightly more accurate estimates, the RMSE is up to four times larger with a burn-in compared to the likelihood conditional on the endogenous state but without a burn-in. Considering the estimates obtained using the particle filter, I find that it outperforms the method based on $x_1 = \mathbb{E}x$ for most $z_A$- and $k$-related estimates, but performs worse for the remaining variables. More importantly, however, the particle filter's estimates have a higher RMSE than the likelihoods conditioning on the initial observation of $Y_t$—namely, $\mathscr{L}^1$ and $\mathscr{L}_p^1$. Lastly, the estimates from the likelihood conditioned solely on the first observations $\mathscr{L}^1$ are significantly superior than the estimates from profile likelihood $\mathscr{L}_p^1$ or statistically not different. However, the differences are minor.

Table 3 represents the one and four-period Root Mean Squared Forecast Error (RMSFE) resulting from forecasts of the estimates from conditional likelihoods to the forecasts from the estimates from the exact likelihood. Again $\mathscr{L}^1$ delivers significantly most accurate results, followed by $\mathscr{L}_p^1$. Further, $\mathscr{L}^{1,x=\mathbb{E}x}$ more precise forecasts than estimates from likelihoods with a burn-in phase ($\mathscr{L}^{10,x=\mathbb{E}x}$, $\mathscr{L}^{20,x=\mathbb{E}x}$) and all conditional filters achieve higher accuracy than the particle filter.

Given that the exercise analyzes the evaluation of both the posterior mode and the maximum likelihood estimator, I repeat these exercises for shorter ($T = 100$) and longer

**Table 2:** RMSEs of ML estimates from conditional likelihoods to true ML estimates (parameters in % of prior range)

| | $\mathscr{L}^1$ | $\mathscr{L}_p^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|
| $n^*$ | 2.30 | 2.12 | $2.68^{***}_{***}$ | $3.52^{***}_{***}$ | $4.08^{***}_{***}$ | $10.22^{***}_{***}$ |
| $\sigma$ | 10.95 | 10.29 | $13.78^{***}_{***}$ | $14.95^{***}_{***}$ | $16.00^{***}_{***}$ | $52.84^{***}_{***}$ |
| $\rho_A$ | 0.80 | $0.98^{***}_{***}$ | $2.01^{***}_{***}$ | $1.97^{***}_{***}$ | $1.92^{***}_{***}$ | $1.46^{***}_{***}$ |
| $\rho_N$ | 0.74 | 0.70 | 0.71 | $1.04^{***}_{***}$ | $1.31^{***}_{***}$ | $1.21^{***}_{***}$ |
| $\rho_G$ | 0.85 | 0.85 | 0.85 | $1.02^{***}_{***}$ | $1.18^{***}_{***}$ | $0.92^{**}_{**}$ |
| $\rho_B$ | 1.11 | 1.17 | $1.13^{***}_{***}$ | $1.26^{***}_{***}$ | $1.42^{***}_{***}$ | $3.74^{***}_{***}$ |
| $\omega_A$ | 0.02 | $0.03^{***}_{***}$ | $0.06^{***}_{***}$ | $0.13^{***}_{***}$ | $0.18^{***}_{***}$ | $0.15^{***}_{***}$ |
| $\omega_N$ | 0.17 | 0.16 | $0.19^{***}_{***}$ | $0.29^{***}_{***}$ | $0.33^{***}_{***}$ | $0.66^{***}_{***}$ |
| $\omega_G$ | 0.04 | 0.04 | 0.04 | $0.12^{***}_{***}$ | $0.17^{***}_{***}$ | $0.15^{***}_{***}$ |
| $\omega_B$ | 9.62 | 8.99 | $11.31^{**}_{**}$ | $12.40^{***}_{***}$ | $14.42^{***}_{***}$ | $55.51^{***}_{***}$ |
| $x_1$ | 4.20 | 4.17 | $9.29^{***}_{***}$ | $9.29^{***}_{***}$ | $9.29^{***}_{***}$ | $4.48^{***}_{***}$ |
| $z_{A1}$ | 1.55 | 1.54 | $3.44^{***}_{***}$ | $3.44^{***}_{***}$ | $3.44^{***}_{***}$ | $1.66^{***}_{***}$ |
| $z_{N1}$ | 0.10 | 0.09 | $0.13^{***}_{***}$ | $0.17^{***}_{***}$ | $0.18^{***}_{***}$ | $0.39^{***}_{***}$ |
| $z_{G1}$ | 0.00 | 0.00 | 0.00 | $0.00^{**}_{*}$ | $0.00^{*}$ | $0.00^{***}_{***}$ |
| $z_{B1}$ | 1.04 | 0.99 | $1.24^{***}_{***}$ | $1.29^{***}_{***}$ | $1.35^{***}_{***}$ | $3.92^{***}_{***}$ |

**Notes:** RMSE of the ML estimates from estimates to the true likelihood from 1032 samples, relative to the prior range in %. Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_p^1$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. Significance levels: $^*$, $^{**}$, $^{***}$ denote $p$-values of 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$; $_*$, $_{**}$, $_{***}$ indicate the same significance levels for the null that the estimate equals that from $\mathscr{L}_p^1$. Global maximization used 200 stage-one points and 1,000 trial points (default), particle filter estimates are done via swarm particles optimization (100 particles—default).

**Table 3:** RMSFE of ML estimates from conditional likelihoods to true ML estimates (in percentage points of the steady state)

| | $\mathscr{L}^1$ | $\mathscr{L}_p^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|
| $\hat{y}_{T+1}$ | $0.04^{***}_{***}$ | $0.05^{***}$ | $0.07^{***}_{***}$ | $0.07^{***}_{***}$ | $0.07^{***}_{***}$ | $0.10^{***}_{***}$ |
| $\hat{c}_{T+1}$ | $0.04^{**}_{**}$ | $0.04^{**}$ | $0.05^{***}_{***}$ | $0.05^{***}_{***}$ | $0.06^{***}_{***}$ | $0.12^{***}_{***}$ |
| $\hat{i}_{T+1}$ | $0.19^{***}_{***}$ | $0.22^{***}$ | $0.25^{***}_{***}$ | $0.27^{***}_{***}$ | $0.29^{***}_{***}$ | $0.62^{***}_{***}$ |
| $\hat{n}_{T+1}$ | $0.04^{**}_{**}$ | $0.05^{**}$ | $0.05^{***}_{***}$ | $0.06^{***}_{***}$ | $0.07^{***}_{***}$ | $0.14^{***}_{***}$ |
| $\hat{y}_{T+4}$ | $0.14^{***}_{***}$ | $0.16^{***}$ | $0.23^{***}_{***}$ | $0.23^{***}_{***}$ | $0.25^{***}_{***}$ | $0.33^{***}_{***}$ |
| $\hat{c}_{T+4}$ | $0.13^{*}_{*}$ | $0.14^{*}$ | $0.17^{***}_{***}$ | $0.18^{***}_{***}$ | $0.19^{***}_{***}$ | $0.38^{***}_{***}$ |
| $\hat{i}_{T+4}$ | $0.65^{***}_{***}$ | $0.71^{***}$ | $0.84^{***}_{***}$ | $0.92^{***}_{***}$ | $0.97^{***}_{***}$ | $2.00^{***}_{***}$ |
| $\hat{n}_{T+4}$ | $0.15^{*}_{*}$ | $0.16^{*}$ | $0.18^{***}_{***}$ | $0.20^{***}_{***}$ | $0.22^{***}_{***}$ | $0.45^{***}_{***}$ |

**Notes:** One and four periods RMSFE of the ML estimated models to estimates from the true likelihood estimated model from 1024 samples, in percentage points of the respective variable's steady state. Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_p^1$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. Significance levels: $^{*}$, $^{**}$, $^{***}$ denote $p$-values of 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$; $_{*}$, $_{**}$, $_{***}$ indicate the same significance levels for the null that the estimate equals that from $\mathscr{L}_p^1$.

($T = 400$) samples.[7] The results of the estimates are presented in Appendix Tables 11 and 12, and the forecasts in Appendix Tables 13 and 14. The findings largely replicate those discussed. Notably, in the short sample, the particle filter provides statistically significantly better estimates for $\rho_G$ and achieves numerically superior, albeit not statistically significant, results for $k_1$ and $z_A$. In all other cases—including the forecasts—estimates based on conditioning on the first observation statistically significantly outperform all other estimations.

Finally, Table 4 compare the time needed for one maximization. Except for the particle filter, it turns out that the Kalman filter is most time consuming. The Kalman filter estimates, the benchmark, need around 4 times as long as the second most time consuming evaluation— $\mathscr{L}^1$. The maximization using $\mathscr{L}^1$ needs around twice as long as $\mathscr{L}_p^1$ which needs around 30% longer than the likelihood specifications where the initial endogenous state set to the unconditional first moment. However, the computation is executed on single cores and the evaluation $\mathscr{L}^1$ is parallelizable, which can decrease the time needed significantly. The particle filter maximization takes to longest time, partly twice as long as the Kalman filter maximizations. While the particle filter is also parallelizable, the ini-

[7] The $T = 100$ sample consists of the first 100 periods of the $T = 200$ sample, which in turn corresponds to the first 200 periods of the $T = 400$ sample.

**Table 4:** Average maximization time per sample

| | Kalman | $\mathscr{L}^1$ | $\mathscr{L}^1_p$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|---|
| T=100 | 02:02 | 00:28 | 00:17 | 00:13 | 00:14 | 00:14 | 2:37 |
| T=200 | 03:13 | 00:41 | 00:22 | 00:17 | 00:18 | 00:18 | 5:30 |
| T=400 | 04:54 | 01:02 | 00:29 | 00:24 | 00:24 | 00:25 | 11:37 |

**Notes:** Time on one core of an AMD Epyc 7313 (Milan) (3.0GHz). Format: MM:SS. Global maximization used 200 stage-one points and 1,000 trial points (default) on 1024 samples, particle filter estimates are done via swarm particles optimization (100 particles—default). Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}^1_p$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution.

tialization of $k_0$ was directly drawn from the stationary Gaussian distribution, which in non-linear cases requires a time-consuming burn-in phase.

**Posterior distribution** I continue with the evaluation of the posterior's mean, quartiles, as well as the first and the last decile from the various likelihoods. I consider the differences between the posterior from the analytically evaluated exact likelihood and from the expected likelihood conditioned solely on $Y_1$ ($\mathscr{L}^1$, eq. (3)), from a specification which includes drawings from $k_1$ using the likelihood conditioned solely on $Y_1$ ($\mathscr{L}^1_f$, eq. (4)), from the likelihood conditioned on $k_1 = \mathbb{E}k$ ($\mathscr{L}^{1,x=\mathbb{E}x}$, eq. (1)), and an approximation of the exact likelihood using $10^3$ particles.[8]

To evaluate the posterior, I use the last 100,000 draws out of 150,000 from a benchmark Random Walk Metropolis Hastings sampler (Herbst and Schorfheide, 2015, Chapter 4) deploying a multivariate Gaussian proposal distribution. For differentiable kernels, the corresponding unscaled covariance equals the negative of the inverse Hessian at the kernel's mode, which also correspond to the initial draw. In the other case, for the particle filter, I use the covariance from the first 10,000 draws, which are sampled using the prior's variance and the initial draw equal the prior's mean.[9] I scale the proposal distribution by $0.25^2$ during the burn-in phase and re-scale after the burn-in once to approach an acceptance of rate of 33% as a rate between 20% and 40% is considered to be optimal (Herbst and Schorfheide, 2015, Chapter 4). Figure 2 display the distribution of the acceptance rates of the samplings from the benchmark kernel—the analytical exact likelihood—and shows a successful scaling given the goal. I present here only results from simulations

[8] The results from the likelihood conditioned on $k_1 = \mathbb{E}k$ and a burn-in phase of the maximum likelihood analysis indicates that burning brings no advancement. I evaluated this and it holds also for the whole posterior, which is why I omit to present the results here for reasons of clarity and comprehensibility.

[9] I got better results using the prior's mean than using the mode evaluated with a derivative free optimizer.

**Figure 2:** Acceptance rate, sampler with Kalman filter

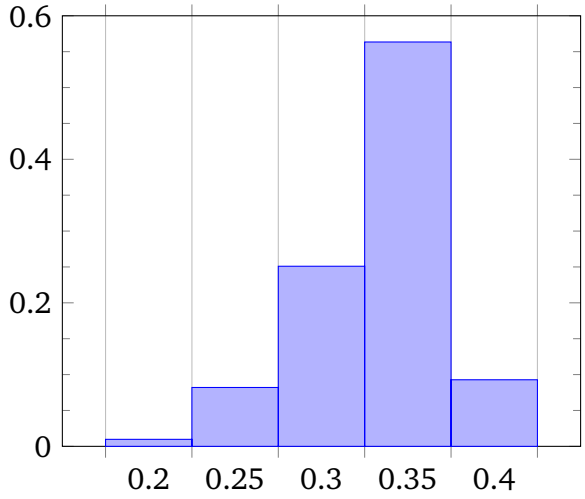



**Table 5:** Excluded posterior samples

| Kalman | $\mathscr{L}^1$ | $\mathscr{L}_f^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $10^3$ Particles | $N_{suc}/N$ |
|---|---|---|---|---|---|
| 1.46% | 2.34% | 10.55% | 2.64% | 2.05% | 864/1024 |

**Notes:**% of samples are discarded due to non-optimal acceptance rate. These samples are than excluded across all likelihood specifications. $N_{suc}$ number of overall successful posterior evaluations. Likelihood specifications: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $Y_1$; $\mathscr{L}_f^1$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution.

where the acceptance rate of the draws lie within the range of 20% and 40%. Discarded simulation are discarded in all likelihood specifications and in the Appendix A.1 the results from all draws.

Figure 5 shows the exclusion results. The highest rate of unsatisfactory acceptances occurs under the specification that draws includes $k_1$ using the likelihood conditioned only on $Y_1$ ($\mathscr{L}_f^1$), with a failure rate of nearly 11%. All other specifications fail less than 3% of the time, resulting in a total of 864 successful samples across all specifications, corresponding to an overall success rate of 84.3%.

Figure 3 presents the results for $n^*$ and $\sigma^*$. For $n^*$, the expected likelihood conditioned solely on $Y_1$ ($\mathscr{L}^1$) performs best, followed closely by the specification that includes drawings from $k_1$ using the likelihood conditioned on $Y_1$ ($\mathscr{L}_f^1$). The likelihood conditioned on $k_1 = \mathbb{E}k$ and $Y_1$ yields values approximately 10% higher than the best-performing specification. The approximation based on $10^3$ particles performs nearly twice as poorly as the best approach, although the difference is still smaller than in the maximum likelihood estima-

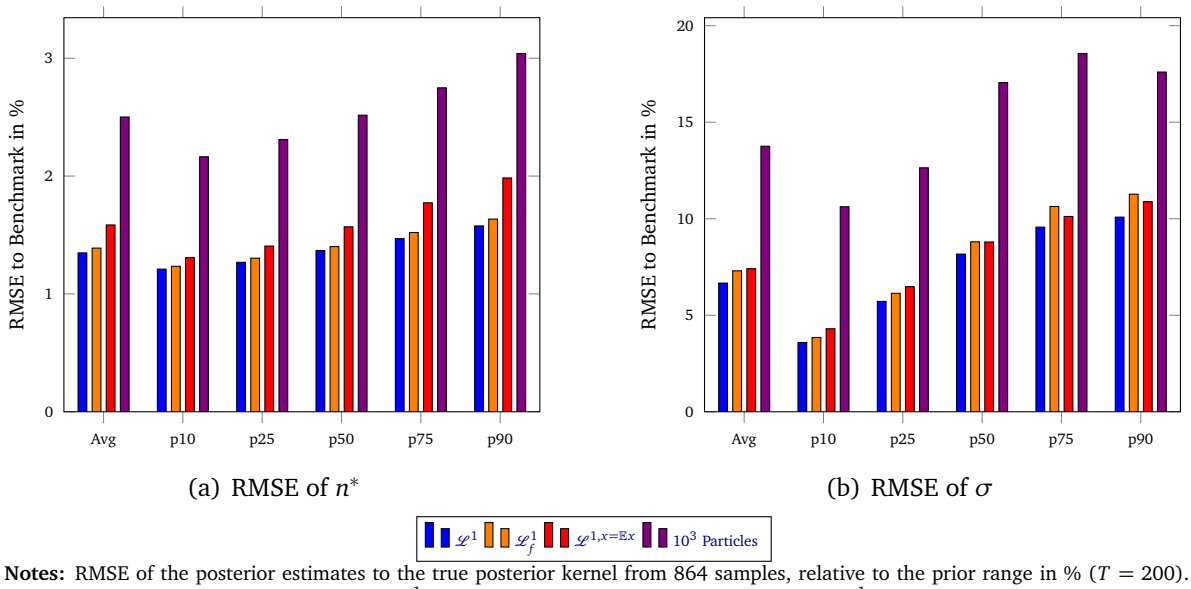
tion. For $\sigma^*$, the expected likelihood conditioned solely on $Y_1$ again delivers the most accurate results, while the specifications using drawings from $k_1$ and conditioning on $k_1 = \mathbb{E}k$ show similar performance across the evaluated measures. The particle-based approximation with $10^3$ particles performs noticeably worse than the best-performing method, though the magnitude of the difference is smaller than that observed for $n^*$.

Figure 4 presents the results for the autoregressive parameters. For $\rho_A$, the expected likelihood conditioned solely on $Y_1$ and the specification that includes drawings from $k_1$ perform best, with the particle-based approximation using $10^3$ particles trailing behind by more than 25%. There is a substantial gap to the specification conditioned on $k_1 = \mathbb{E}k$, which performs three to six times worse. For $\rho_N$ and $\rho_G$, the conditional filters yield similar RMSEs, with the particle-based approach achieving slightly lower errors—approximately one third smaller. However, all methods result in relatively small errors overall. For $\rho_B$, the expected likelihood conditioned solely on $Y_1$, the specification using drawings from $k_1$, and the particle-based approximation perform similarly and outperform the method based on conditioning on $k_1 = \mathbb{E}k$, whose RMSE is about 10% larger.

The RMSE of the posteriors' of the innovations' standard deviations is reported in Figure 5. The ordering of results for $\omega_A$ mirrors that of $\rho_A$, although the differences across specifications are smaller. Interestingly, for $\omega_N$, the particle-based approximation with $10^3$ particles performs significantly worse, with an RMSE roughly 50% higher than the best-performing specification—the specifications with likelihoods conditioned on $Y_1$, closely followed by the specification conditioned on $\mathbb{E}k$. For $\omega_G$, the results resemble those of $\rho_G$, but with a smaller differences across specifications; the approach using drawings from $k_1$ performs slightly worse for larger quantiles. For $\omega_B$, the particle-based approximation underperforms substantially, yielding RMSE values about 100% higher than those from the conditional specifications, which all perform similarly.

Figure 10–12 in the appendix reports results when no samples are discarded. In this setting, the particle-based approximation using $10^3$ particles performs worse in all cases except for $\rho_A$, where it still yields noticeably poorer results compared to the one presented here. Moreover, for both $\sigma$ and $\omega_B$, the specification drawing from $k_1$ also exhibits a decline in performance, becoming worse relative to the conditional likelihood alternatives. Other differences are not substantial enough to warrant discussion.

**Figure 3:** Posterior errors of estimated steady state and behavioral parameters (% of prior range)




(a) RMSE of $n^*$

(b) RMSE of $\sigma$

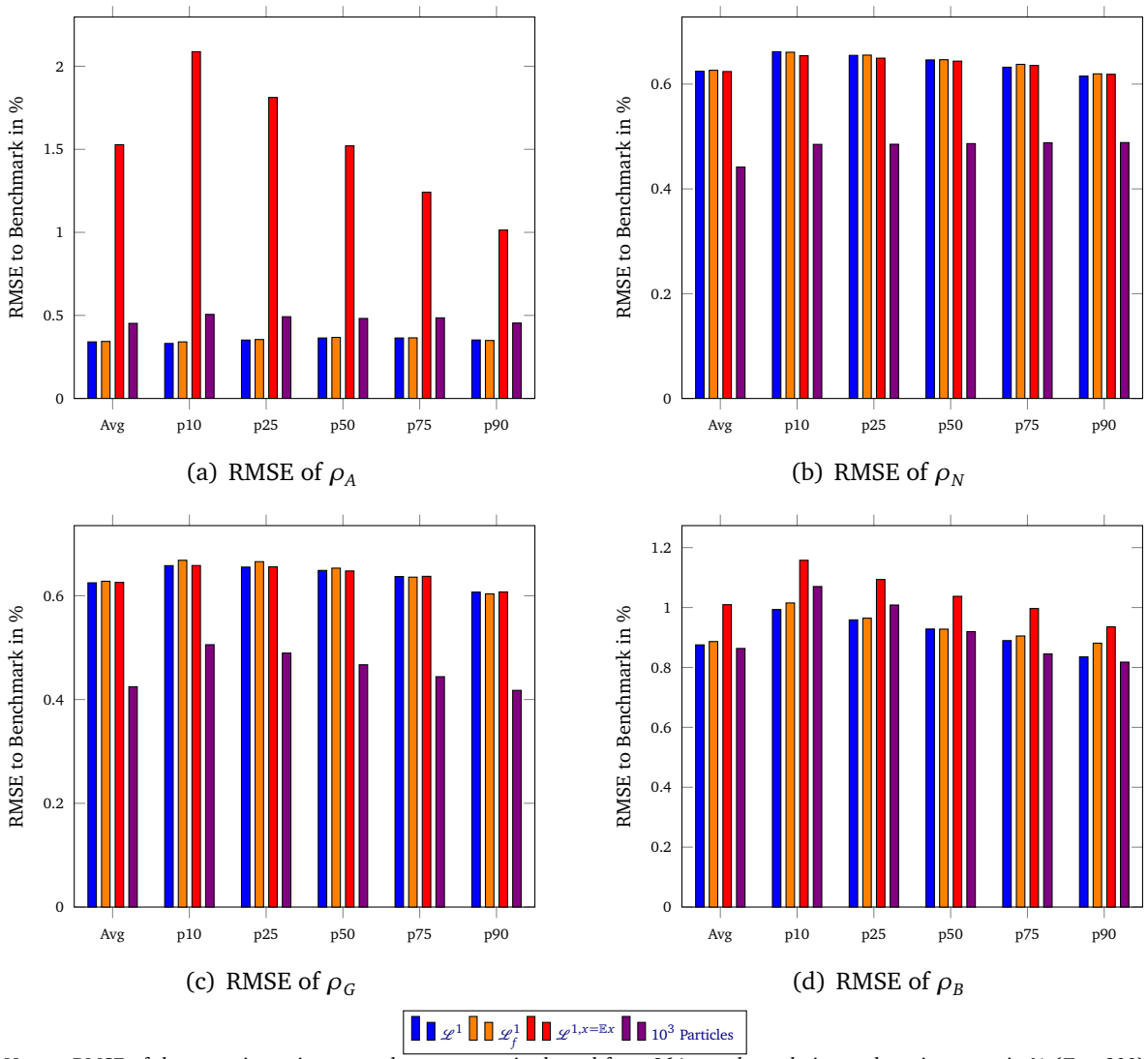
**Notes:** RMSE of the posterior estimates to the true posterior kernel from 864 samples, relative to the prior range in % ($T = 200$). The likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}^1_f$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. A Random-Walk Metropolis-Hastings sampler is used with 150,000 draws, where the first 1/3 of the draws are discarded as burn-in.

**Table 6:** Sampling time per sample

| Kalman | $\mathscr{L}^1$ | $\mathscr{L}^1_f$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|
| 13:34 | 03:33 | 01:55 | 01:46 | 93:04 |

**Notes:** Time on one core of an AMD Epyc 7313 (Milan) (3.0GHz). Format: MM:SS. Sampling time excludes the estimation of the kernel's mode. The likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}^1_f$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution.

**Figure 4:** Posterior errors of estimated autoregressive coefficients (% of prior range)




(a) RMSE of $\rho_A$

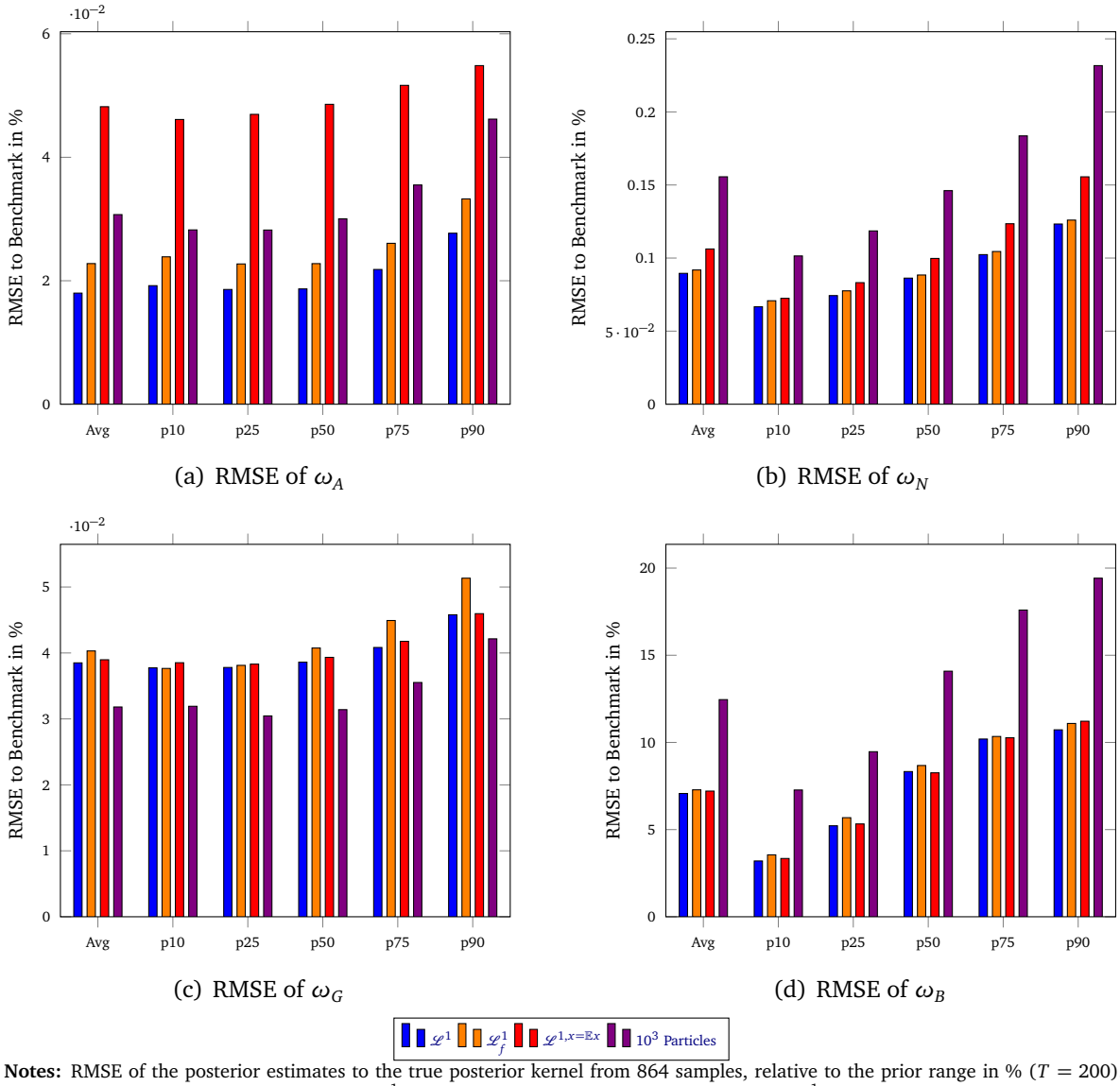
(b) RMSE of $\rho_N$

(c) RMSE of $\rho_G$

(d) RMSE of $\rho_B$

**Notes:** RMSE of the posterior estimates to the true posterior kernel from 864 samples, relative to the prior range in % ($T = 200$). The likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}^1_f$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. A Random-Walk Metropolis-Hastings sampler is used with 150,000 draws, where the first 1/3 of the draws are discarded as burn-in.

**Figure 5:** Posterior errors of estimated innovation's standard deviation (% of prior range)




**Notes:** RMSE of the posterior estimates to the true posterior kernel from 864 samples, relative to the prior range in % ($T = 200$). The likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}^1_f$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. A Random-Walk Metropolis-Hastings sampler is used with 150,000 draws, where the first 1/3 of the draws are discarded as burn-in.

### 2.3 Discussion

In general, the inversion filter has two major limitations: it requires a one-to-one (invertible) policy function, and because of this, the number of observables must equal the number of exogenous states. The initialization proposed here introduces a third condition: the number of endogenous states must not exceed the number of exogenous states. In the following, I will discuss these limitations with reference to economic applications and propose possible remedies.

First, regarding the one-to-one policy function that guarantees the unique mapping from controls to states: under standard conditions—such as strict concavity of the objective with respect to the controls and that the feasible set defined by the constraints is convex—and given that the number of observables equals the number of exogenous states, the policy function is one-to-one. Hence, the inversion filter can be applied to a wide range of problems; however, one should exercise due caution when dealing with non-standard problems. Holden (2023), for example, discuss under which conditions many-to-one mappings can arise due to a occasionally-binding zero lower bound on the policy rates. Cuba-Borda et al. (2019) propose that for many-to-one policy functions, one can consider all roots and then sum over all possible solutions in the likelihood. While this approach can quickly lead to numerical overflow, it remains feasible for occasional many-to-one mappings.

A common problem is spurious roots. Spurious roots can arise when approximating the policy function, particularly in higher-order local expansions. For example, when approximating the policy function of a model with $n$ exogenous states using an $m$-th order Taylor series, the inversion of the local expansion may yield up to $m^n$ candidate roots. If the policy function is one-to-one, only one of these roots corresponds to the true solution of the original system; the others are spurious. In such cases, the Lagrange Inversion Theorem provides a useful tool.

The Lagrange Inversion Theorem gives a formal method for locally inverting an analytic function. Specifically, for a function defined as $y = f(z)$, the theorem allows the construction of a Taylor series for the inverse function $z = f^{-1}(y)$ around a point where $f$ is locally invertible, guaranteeing the correct identification of the true root in its neighborhood.

An additional check for the true root can be obtained from the residuals of the static functions of the model that involve non-observed controls. For the true root, these residuals are close to zero because all equilibrium conditions are satisfied, whereas spurious roots produce non-negligible residuals.

Regarding the second requirement—that the number of observables equals the number of exogenous states—we can either introduce measurement errors to increase the number of observables and integrate over these errors, or, if there are more exogenous states than observables, integrate over the additional exogenous states as we do for the initial endogenous states. Both approaches are feasible, as the distributions of measurement errors and exogenous states are given. Note that when Monte Carlo integration techniques are employed, the former approach resembles a particle filter, with the inversion acting as an importance sampler.[10] However, further research is needed to explore these possibilities.

When the number of endogenous states exceeds the number of exogenous states, the Gram matrix $J^\top J$ is necessarily rank-deficient, precluding its use for volume scaling. However, this specific scenario is rare in applied macroeconomics. Consequently, this limitation does not affect the practical applicability of the inversion filter for a majority of models.

## 3 Applications

### 3.1 Monetary Business Cycle Accounting for the COVID-19 Recession and the post-pandemic inflation in the EMU and US

As a first application, I conduct MBCA, following the approach of Šustek (2011), to analyze the COVID-19 recession and the subsequent post-pandemic inflation in both the EMU and the United States. MBCA extends the standard Business Cycle Accounting framework introduced by Chari et al. (2007) by incorporating inflation, $\pi_t$, into the analysis.

Business Cycle Accounting attributes deviations from a frictionless benchmark model to a set of so-called wedges, which serve as reduced-form distortions of structural frictions and shocks. The real distortions in MBCA correspond to those derived from the distorted first-order conditions presented in the Monte Carlo study (Equations (6)–(11)), with the exception of the Euler equation for capital. Here, the wedge in the capital Euler equation is equivalent to an investment tax, $\exp(z_{It})-1$. Accordingly, the Euler equation for capital, takes the form:

$$\exp(z_{It}) = \beta \mathbb{E}_t \left[ \left( \frac{c_{t+1}}{c_t} \right)^{\sigma} \left( \frac{1-n_{t+1}}{1-n_t} \right)^{\sigma(1-\theta)} \left( (1-\delta)\exp(z_{It+1}) + \alpha \frac{y_{t+1}}{k_{t+1}} \right) \right].$$

[10] Fehrle et al. (2025) approximate the inverse function as importance sampler in a bootstrap filter application. See additionally also the discussion of Herbst and Schorfheide (2015, Chapter 8) on the conditionally optimal importance distribution.

In addition to the real distortions, MBCA introduces a wedge in the Euler equation for bonds $z_{Bt}$. Furthermore, the central bank is assumed to follow a distorted Taylor rule, capturing monetary policy deviations from the monetary policy rule. The distorted bond Euler equation and Taylor rule read

$$\exp(z_{Bt}) = \beta \mathbb{E}_t \left[ \left( \frac{c_{t+1}}{c_t} \right)^{\sigma} \left( \frac{1-n_{t+1}}{1-n_t} \right)^{\sigma(1-\theta)} \frac{r_t}{\pi_{t+1}} \exp(z_{Bt+1}) \right],$$
$$r_t = r^* \left( \frac{1+\pi_t}{1+\pi^*} \right)^{\psi_\pi} \left( \frac{y_t}{y^{ss}} \right)^{\psi_y} \exp(z_{Rt}),$$

where $r_t$ denotes the nominal interest rate, $r^*$, $\pi^*$, and $y^{ss}$ represent the central bank's steady-state or target values for the nominal rate, inflation, and output, respectively, and $z_{Rt}$ captures deviations from the central bank's rule-based policy. These two equations account for the nominal (monetary) distortions in the model. Lastly, to ensure consistency in per capita terms, the household's time preference rate is adjusted for population growth $g_{pop}$, yielding an effective discount factor of $\beta' = \beta / g_{pop}$.

The specification of the evolution of the distortions—the wedges—is more elaborate than for the $z_{it}$ in the Monte Carlo experiment. First, each wedge consists of a long-run (steady-state) component and a time-varying component: $z_{jt} = z_j^{ss} + \bar{z}_{jt}$. Second, the time-varying component $\bar{z}_{jt}$ follows a first-order vectorautoregressive process, implying that the distortions interact over time.

### 3.1.1 Calibration

Table 7 summarizes the parameter values used for the EMU and US economies. The calibration for the US follows Šustek (2011), with one notable exception: I increase the effective discount factor $\beta'$ from 0.99 to 0.995 to better reflect the low interest rate environment during the 2010s. Growth rates, output shares, and the long-run average working time are based on my own calculations.

For the EMU, I depart further from Šustek (2011) by adjusting two key parameters: the capital elasticity of output and the depreciation rate are set to $\alpha = 0.376$ and $\delta = 0.0117$ to align with the empirical estimates reported by ECB (2023) and ECB (2006), respectively.

**Table 7:** Calibration

| Parameter | EMU | US |
|---|---|---|
| Capital elast. $\alpha$ | 0.376 | 0.35 |
| Capital depr. rate $\delta$ | 0.0117 | 0.0118 |
| TR infl. target (p.a.) $\pi^*$ | 2% | 2% |
| TR infl. elast. $\psi_\pi$ | 1.5 | 1.5 |
| TR cycle elast. $\psi_y$ | 0.125 | 0.125 |
| Time preference $\beta'$ | 0.995 | 0.995 |
| Risk aversion $\sigma$ | 1 | 1 |
| St.st. work $n^{ss}$ | 0.125 | 0.151 |
| St.st. infl. $\pi^{ss}$ | 0.125 | 0.151 |
| Population growth $g_{pop}$ | 1.001 | 1.002 |
| Consumpt. growth $g_c$ | 1.007 | 1.005 |
| Invest. growth $g_i$ | 1.008 | 1.006 |
| Avg. consumpt. share $c/y$ | 0.55 | 0.66 |
| Avg. invest. share $i/y$ | 0.22 | 0.18 |

**Notes:** External determined parameters and long-run properties. Remaining parameters are determined endogenously by the model or via ML estimation.

### 3.1.2 Implementation

To align the data with the model, I largely follow Fehrle and Huber (2023). First, I account for differing growth rates among the observables and adjust the model accordingly to ensure internal consistency. The resulting system of equations in stationary variables is presented in Appendix A.2.1. Second, I determine the long-run wedge components endogenously using steady-state relationships and a calibration exercise based on the long-run shares of consumption and investment in output. The parameters of the mean-zero autoregressive process are estimated via ML estimation. Following Šustek (2011), I use output, private consumption, gross fixed investment, hours worked, and the inflation rate as observables. Note that with these observables, the variable $g_t$ captures all expenditures not classified as gross fixed investment or private consumption. As a result, it represents a total residual demand wedge rather than government consumption specifically. The EMU-19 countries (excluding Croatia and Bulgaria) represent the EMU, and Appendix A.2.2 lists the data sources for both the EMU-19 and US data.

Regarding the model solution, I employ a weighted residual method via collocation, as described in Heer and Maußner (2024, Chapter 5) to approximate the policy functions for labor and inflation. Specifically, I use Chebyshev polynomials on a Smolyak sparse grid

of second degree, resulting in a $2 \times 113$ polynomial series. To approximate the agent's expectations, I implement the CUT-4 cubature rule from Adurthi et al. (2018).

I estimate the autoregressive process with a model solution over a grid spanning 10% of the deterministic steady-state capital stock and 1.5 times the unconditional standard deviation of the exogenous states driven from the the autoregressive process. It is important to note that the states of efficiency $z_{At}$, labor wedge $z_{Nt}$, residual wedge $z_{Gt}$, and monetary wedge $z_{Rt}$ can be measured independently of the model's solution and thus are measured independent of the state-space domain. The remaining wedges $z_{It}$ and $z_{Bt}$ are recovered each period using a root-finding algorithm. The Jacobian of the mapping from innovations to the observations is derived analytically from the polynomial series.

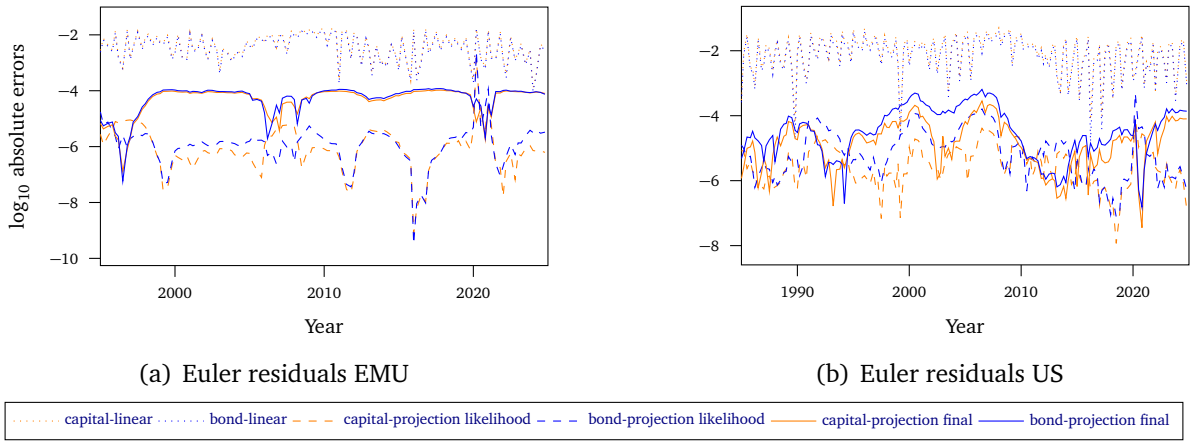
I choose a likelihood that is conditioned solely on the initial observations—$\mathscr{L}^1$. To approximate the integral over the initial endogenous state, I employ Gauss-Hermite quadrature nodes derived from the linear approximation of the endogenous state. The corresponding quadrature weights are obtained from the likelihood conditional on each node-specific endogenous state. The Jacobian required for transforming exogenous to endogenous initial states is obtained via numerical differentiation.

Afterward, I adjust the grid in such way that the extrema of the realized states over the whole sample for the EMU and over the COVID-19 period for the US align with the extrema of the state space. This ensures that the solution is accurately computed during the pandemic period.

Figure 6 plots the Euler residuals at three different solutions: (i) the states measured during likelihood evaluation using the grid applied in that evaluation (dashed line), (ii) the final measured states using the grid employed for the final model solution, and (iii) the final measured states using a first-order perturbation around the deterministic steady state (dotted).

The figure illustrates that, in normal times in the EMU, i.e., for most of the sample period, the solution used for the likelihood evaluation exhibits greater accuracy than the final solution. However, during the peak of the COVID-19 recession, the accuracy of the likelihood-based solution deteriorates, though it still outperforms the average accuracy of the linear solution by half an order of magnitude, and remains over an order of magnitude more accurate than the linear solution during this period. For the US, the model solutions used for the likelihood evaluation outperform the linear solution even more significantly. Resolving the model on the adjusted grid leads to additional improvements in accuracy during the COVID-19 period, resulting in a several orders of magnitude higher accuracy

**Figure 6:** Euler residuals




(a) Euler residuals EMU

(b) Euler residuals US

**Notes:** absolute Euler residuals in $log_{10}$ scale from capital and bond Euler equations. Linear: Perturbation at deterministic steady state.

than the linear perturbation around the deterministic steady state during the COVID-19 recession.

Finally, I assess the influence of each wedge over the period under consideration by calculating the forecast error of the full model and comparing it to versions of the model where all wedges, except one at a time, are held fixed at their steady-state values, while allowing the underlying states to evolve. This approach isolates the impact of individual wedges while preserving the agents' expectations about the wedge's fluctuations.

### 3.1.3 Results

Figure 7 presents the results of the exercise by plotting the forecast errors of the full model alongside model versions in which all wedges, except one, are held fixed at their steady-state values over the period 2020 – 2024. Horizontally, the left-hand panels display results for the EMU, while the right-hand panels correspond to the US. Vertically, the upper panels focus on the business cycle, measured by output $y_t$, and the lower panels on inflation, $\pi_t$.

In both economies, the labor wedge accounts excessively for the COVID-19 recession, while the investment wedge exhibits a strongly countercyclical behavior. The efficiency wedge contributes partially to the recession in the EMU but appears neutral or rather countercyclical in the US. The impact of the residual demand wedge is minor in both economies. According to the classical dichotomy, the bond and monetary wedges do not affect real variables in the model.

The labor wedge also exerts the greatest inflationary pressure on both economies during

the peak of the COVID-19 pandemic. In the EMU, the investment and efficiency wedges additionally act inflationary during this phase, whereas their influence is minor in the US. The residual demand wedge similarly has only a minor impact in both regions during this period.

During the pandemic, the bond and monetary wedges exert strong deflationary effects, particularly in the EMU, where deflationary tendencies dominate. As the economies move through and beyond the recovery phase, inflationary pressure from real economy wedges diminishes, while that from nominal wedges increases, especially from the monetary wedge. The monetary wedge alone accounts excessively for the post-pandemic inflation surge in both economies. Notably, the investment wedge also contributes significantly to the inflation peak in the US.

Table 8 presents the $\Delta$-statistics (Fehrle and Huber, 2023) for output $y_t$ and inflation $\pi_t$. These statistics measure the relative cumulative contribution of each individual wedge to the realized fluctuations, where here fluctuations are continued to be quantified as forecast errors.

The results largely corroborate the previous illustrated results. In the EMU, the labor wedge contributes most significantly to the business cycle, followed by the efficiency wedge, while the investment wedge acts strongly countercyclically. In the US, the labor wedge alone accounts for the bulk of the cycle, with the efficiency wedge playing a countercyclical role.

With respect to inflation, in the EMU the labor and efficiency wedges—both real wedges—are the main contributors, followed by the monetary wedge. In the US, the labor, investment, and monetary wedges contribute similarly to inflation, whereas the efficiency wedge exhibits a deflationary effect.

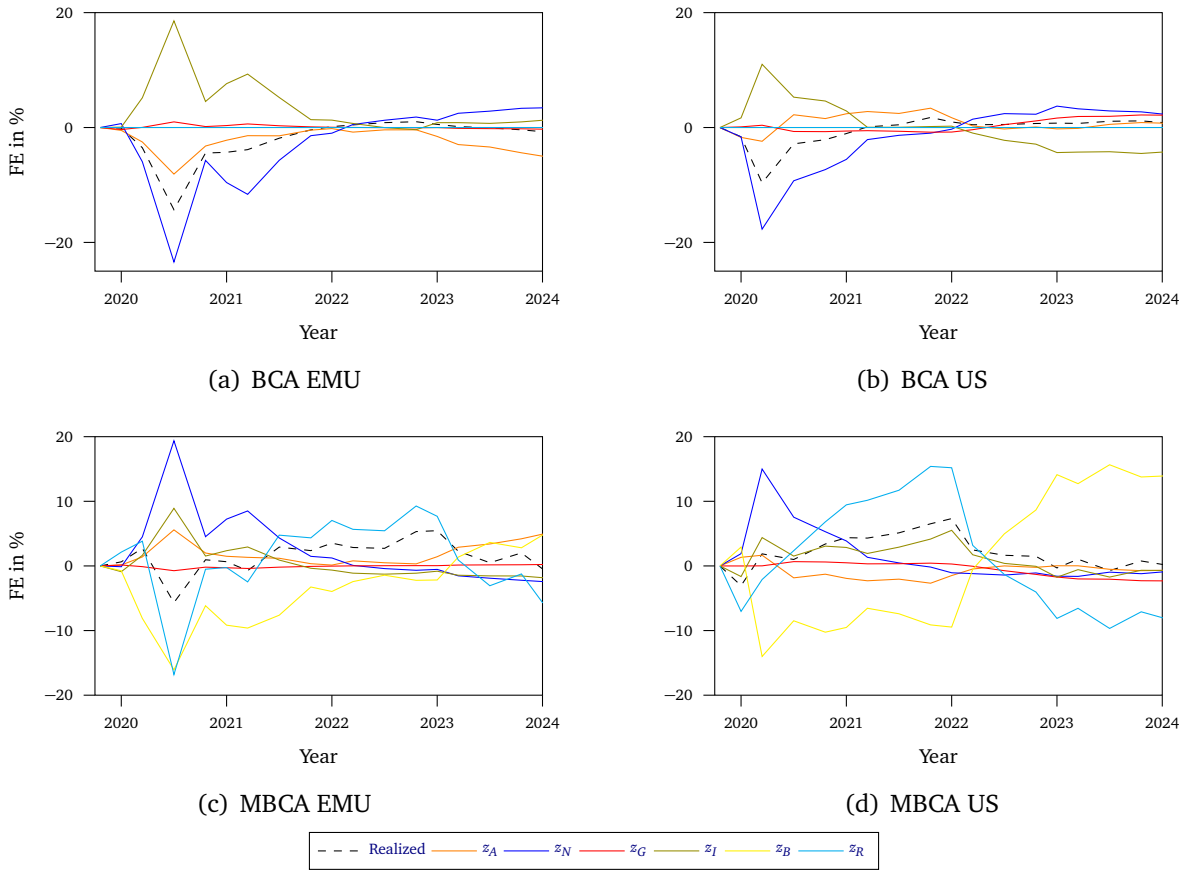
### 3.2 Uncertainty shocks and the Great Moderation

Here, I build on the stochastic general equilibrium model developed by Basu and Bundick (2017). This model produces sizable comoving fluctuations induced by uncertainty shocks and is therefore well suited to analyze their macroeconomic effects.

#### 3.2.1 Framework

The model features optimizing households and firms, and a central bank that follows a Taylor-type monetary policy rule to stabilize inflation and offset adverse demand fluc-

**Figure 7:** Monetary Business Cycle Accounting




**Notes:** FE: forecast errors, Realized: all wedges on, $z_A$: only efficiency wedge on, $z_N$: only labor wedge on, $z_G$: only residual wedge on, $z_I$: only investment wedge on, $z_B$: only bond wedge on, $z_R$: only monetary weddge on.

**Table 8:** MBCA $\Delta$-statistics

| BCA ($y_t$) | | | | | | |
|---|---|---|---|---|---|---|
| | $z_{At}$ | $z_{Nt}$ | $z_{Gt}$ | $z_{It}$ | $z_{Bt}$ | $z_{Rt}$ |
| EMU | 1.26 | 1.52 | -0.04 | -1.89 | 0.00 | 0.00 |
| US | -1.86 | 3.27 | -0.84 | 0.23 | 0.00 | 0.00 |
| **MBCA ($\pi_t$)** | | | | | | |
| | $z_{At}$ | $z_{Nt}$ | $z_{Gt}$ | $z_{It}$ | $z_{Bt}$ | $z_{Rt}$ |
| EMU | 1.14 | 1.48 | -0.04 | 0.20 | -2.15 | 0.75 |
| US | -0.35 | 0.65 | -0.25 | 0.57 | 0.30 | 0.54 |

**Notes:** $\Delta$-statistics report the accumulated forecast error induced by a single wedge $z_{jt}$ in relation to the realized accumulated forecast error.

tuations. Nominal rigidities follow the quadratic price adjustment cost framework of Rotemberg (1982). The baseline model includes two exogenous sources of variation: (i) household discount rate shocks and (ii) technology shocks. Additionally, the discount rate shock's innovations have a time-varying second moment, the uncertainty shock.

Households are modeled with Epstein-Zin recursive preferences, allowing the separation of risk aversion from the intertemporal elasticity of substitution. Additional to price adjustment costs, firms face Jermann (1998) capital adjustment costs. However, firms can choose their capital utilization in addition to the capital and labor input.

I introduce a residual demand shock, representing the combined effects of government consumption and net exports, to capture additional variation in aggregate demand beyond domestic private-sector behavior. Furthermore, I specify all stochastic processes log-normal distributed rather than, as Basu and Bundick (2017), normal, avoiding negative states. Finally, the output gap in the Taylor rule is measured relative to the deterministic trend rather than to previous output, aligning with the model's balanced-growth-path assumption.

### 3.2.2 Estimation

I estimate the same parameters as those used for impulse response function matching of Basu and Bundick (2017). Specifically, I estimate all shock process parameters, namely the (unnormalized) standard deviations of the innovations $\sigma_i$ and the first-order autoregressive coefficients $\rho_i$, with $i = G$ for residual demand shocks, $i = Z$ for technology shocks, $i = A$ for discount rate shocks, and $i = \nu$ for shocks to the standard deviation of the discount rate innovations. In addition, I estimate the elasticity of capital adjustment costs, $\phi_K$.

Column two to four of Table 9 presents my choice of the priors' distribution. All other parameters follow Basu and Bundick (2017), except for the intertemporal elasticity of substitution, which I increase to $\psi = 0.99$. The priors and the calibration ensures a sizable impact of uncertainty shocks at the prior's mean, while producing also only minor impact within likely ranges of the prior.

I use quarterly US data on output, consumption, investment, and hours worked from 1985 – 2019. Appendix A.3.2 lists the data sources. Posterior draws are from a Sequential Monte Carlo Algorithm with $N = 2,500$ particles (Herbst and Schorfheide, 2015, Algorithms 8, 9, and 10.). The likelihood tempering schedule follows $(n/N_\phi)^2$, $n = 1, \ldots, N_\phi$,

**Table 9:** Prior and posterior distribution

| | Prior | | | Posterior | |
|---|---|---|---|---|---|
| Parameter | Prior Distribution | Mean | Standard deviation | Mean | Standard deviation |
| $\phi_K$ | $\mathscr{IG}$ | 5 | Mean/5 | 4.83 | Mean/36 |
| $\omega_Z$ | $\mathscr{IG}$ | 0.0015 | Mean/5 | 0.015 | Mean/10 |
| $\omega_G$ | $\mathscr{IG}$ | 0.001 | Mean/5 | 0.009 | Mean/8 |
| $\omega_A$ | $\mathscr{IG}$ | 0.004 | Mean/5 | 0.0054 | Mean/20 |
| $\omega_v$ | $\mathscr{IG}$ | 0.25 | Mean/5 | 0.1029 | Mean/10 |
| $\rho_Z$ | $\mathscr{B}$ | 0.98 | 0.005 | 0.71 | 0.027 |
| $\rho_G$ | $\mathscr{B}$ | 0.8 | 0.05 | 0.99 | 0.002 |
| $\rho_A$ | $\mathscr{B}$ | 0.8 | 0.05 | 0.95 | 0.003 |
| $\rho_v$ | $\mathscr{B}$ | 0.8 | 0.05 | 0.89 | 0.012 |

**Notes:** $\mathscr{IG}$: Inverted Gamma Distribution, $\mathscr{B}$: Beta Distribution. Posterior draws from a SMC Algorithm with $N = 2,500$ particles (Herbst and Schorfheide, 2015, Algorithms 8, 9, and 10.). The likelihood tempering schedule follows $(n/N_\phi)^2$ with $N_\phi = 200$. Resampling takes place for effective sample sizes lower $N/2$. The sampler includes additional the endogenous states capital and discount rate in a separated block. The states have uninformative priors.

$N_\phi = 200$. Resampling takes place for effective sample sizes lower $N/2$. The sampler treats the endogenous states, namely capital and the discount rate, in a separate block estimated via likelihood function $\mathscr{L}_f^1$ (eq. (4)). These states are assigned non-informative priors. The Jacobian required for transforming exogenous to endogenous initial states is obtained via numerical differentiation. To verify the robustness of the results, I re-estimate the model using the profile likelihood $\mathscr{L}_p^1$ (eq. (5)). Appendix A.3.3 presents the results.

The innovations are identified by inverting the observation equation for given endogenous states using the Lagrange Inversion Theorem. I use a first-order truncated series expansion around the known within-period values of the endogenous states and the expected values of the exogenous states, which guarantees a unique local mapping in the neighborhood of the expected states.[11] The actual solution uses a third-order perturbation solution based on CoRRAM with automatic differentiation (Heer and Maußner, 2024, Chapter 3). The Jacobian of the mapping from innovations to observations is derived analytically from the third-order Taylor series.

[11] While higher-order series expansions are possible without significant computational cost, my experience indicates that for rare but large innovations, the series can diverge far from the expansion point, producing non-convergent terms and less accurate results than the first-order approximation. Moreover, note that the first-order inversion does propagate the effects of stochastic uncertainty. This expansion approach is similar to Kollmann (2017), who guarantee a unique mapping from $y$ to $z$ by replacing higher-order terms in the exogenous states with their expected values.

### 3.2.3 Results

The last two columns of Table 9 present the first two moments of the posterior. Notably, $\omega_Z$ and $\omega_G$ increase from their prior means, while $\omega_\nu$ decreases. Persistence parameters show contrasting revisions: $\rho_Z$ declines markedly, whereas $\rho_G$ and $\rho_A$ move close to one. Across parameters, except for $\rho_z$, posterior standard deviations are smaller than prior ones, highlighting increased precision, indicating, among others, that the data meaningfully inform the model.

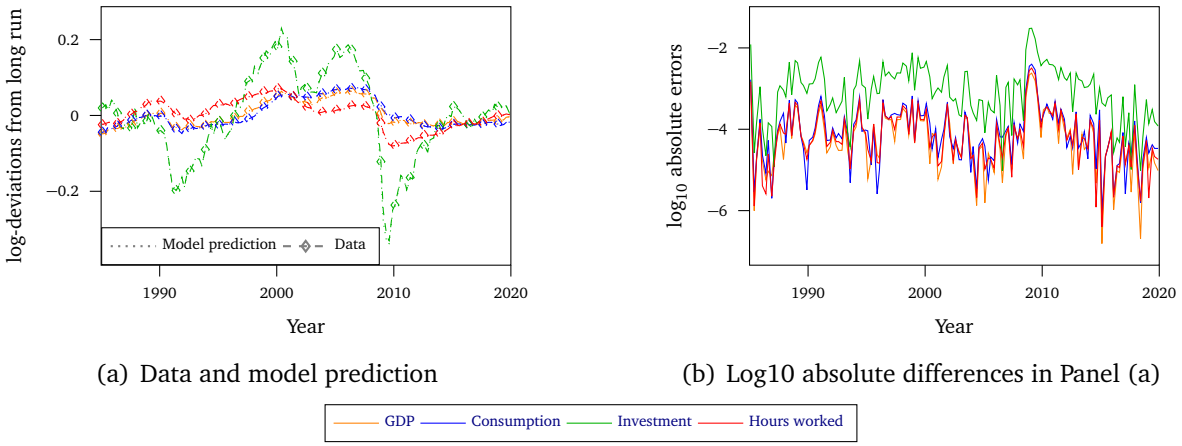
Figure 8 Panel (a) plots the data used as inputs alongside the model predictions at the posterior means based on the filtered states. Since the state estimation relies on a first-order truncation of the Taylor series expansion of the inverse function, the differences between the data and predictions indicate the loss of accuracy due to this approximation. These differences are barely visible in Panel (a). Panel (b) presents the absolute differences on a $\log_{10}$ scale from Panel (a). The time series for output, consumption, and hours worked show differences on the order of $10^{-4}$, while investment exhibits discrepancies an order of magnitude larger.

In Figure 14, I decompose the business cycle, measured as fluctuations in output, into contributions from individual shocks using the model at the posteriors' mean. The decomposition allows one shock to fluctuate at a time, isolating its effect. I also present the interactions between shocks, capturing movements in output that are not explained by any single shock alone. Finally, the sum of the individual shocks and their interactions is presented—the actual data. Table 10 quantifies the decomposition by reporting the respective inverse mean squared error normalized to the sum of all inverse errors.

First and most remarkable, the uncertainty shock accounts for nearly the entire business cycle, whereas the productivity and discount rate shocks mainly contribute to low-frequency fluctuations. Regarding the Great Recession, rising uncertainty strongly depresses output, while productivity and, in particular, the residual demand shock act countercyclically. Notably, a decrease in productivity depresses output in the aftermath of the Great Recession further.

Repeating the exercises in Appendix A.3.3 using the profile likelihood confirms the robustness of the results regarding the likelihood specification.

**Figure 8:** Prediction residuals

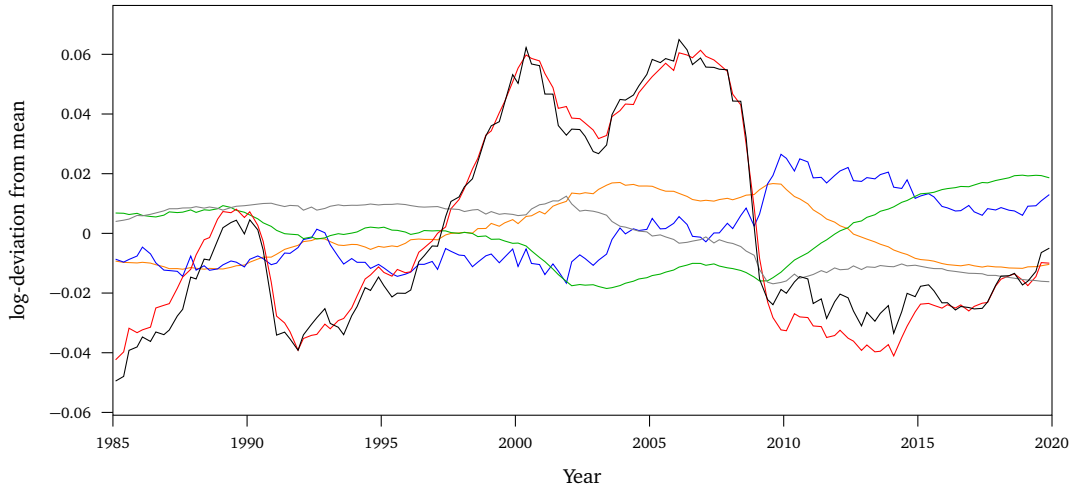



(a) Data and model prediction

(b) Log10 absolute differences in Panel (a)

**Notes:** Data and model long run equals stochastic steady state. Model prediction at posterior's mean.

**Figure 9:** Business cycle decomposition




(a) Data and model prediction

**Notes:** Centered times series from singular shocks on, interactions, and data at the posteriors' mean.

**Table 10:** Normalized inverted mean squared errors

| $Z$ | $G$ | $A$ | $\nu$ | Interaction |
|---|---|---|---|---|
| 0.04 | 0.02 | 0.02 | 0.89 | 0.03 |

**Notes:** Normalized inverted mean squared error: inverted mean squared error normalized to accumulated inverted mean squared errors.

# 4 Conclusion

The initial endogenous state distribution is crucial for inverse filtering in nonlinear state-space and dynamic latent variable models. In this study, I show how this distribution, which is mostly unknown, can be derived under various conditions from initial observations using the change-of-variables theorem. This approach improves the efficiency of state estimation using the inverse filter. Moreover, when the filter is used to construct the likelihood, it also enhances the efficiency of both Frequentist and Bayesian parameter inference. Specifically, it enables the computation of likelihoods that are conditional solely on the initial observations and parameters, that treat the initial endogenous states as parameters, or that profile out the initial endogenous states.

Monte Carlo studies confirm that this approach substantially improves the accuracy of state estimation and, consequently, enhances both Frequentist and Bayesian parameter inference using inversion filters. In addition, the Monte Carlo analysis confirms the literature regarding the superior computational efficiency of inversion filters.

A key limitation of the inversion filter is that it requires stricter conditions: the policy function must be one-to-one, and, thus, the number of observables must equal the number of exogenous states and, with a appropriate initialization, exceed the number of endogenous states. However, a discussion here concludes that inversion filters remain well-suited for many applications and also outlines potential remedies to the mentioned limitations.

Using these insights, a first application—a MBCA Analysis with a global solution—identifies the wedges responsible for the COVID-19-induced recession and the subsequent inflation surge in both the European Monetary Union and the United States. Across both economies, the labor wedge dominates the COVID-19 recession, while the investment wedge behaves countercyclically. The efficiency wedge plays only a limited role. During the pandemic peak, the labor wedge generates the strongest inflationary pressure, with investment and efficiency wedges contributing only modestly. The bond and monetary wedges act inflationary in the aftermath of COVID-19, but suppressed inflation during the pandemic itself.

A second application measures the impact of uncertainty shocks on the business cycle, indicating that uncertainty shocks were the main driver of the business cycle during the Great Moderation (1985 – 2019). In this case, the inversion uses the Lagrange Inversion Theorem, demonstrating its applicability in solving the problem of spurious roots away from the expansion point of local solutions.

Nonlinearities, for example those arising from uncertainty shocks, are discussed as an important feature of all economies, and large shocks will persist in macroeconomic time series even in developed economies due to the COVID-19 recession. The improvements to the inversion filter presented here provide a fast and reliable method for estimating a wide range of nonlinear models with endogenous states, thereby opening up possibilities to address research questions that were previously infeasible. Furthermore, the remedies outlined for the inversion filter's limitations merit further investigation.

## REFERENCES


N. Adurthi, P. Singla, and T. Singh. Conjugate unscented transformation: Applications to estimation and control. *Journal of Dynamic Systems, Measurement, and Control*, 140(3): 030907, 2018.

G. Amisano and O. Tristani. Exact likelihood computation for nonlinear dsge models with heteroskedastic innovations. *Journal of Economic Dynamics and Control*, 35(12):2167–2185, 2011. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2011.08.003. URL https://www.sciencedirect.com/science/article/pii/S0165188911001539. Frontiers in Structural Macroeconomic Modeling.

T. Atkinson, A. W. Richter, and N. A. Throckmorton. The zero lower bound and estimation accuracy. *Journal of Monetary Economics*, 115:249–264, 2020.

S. Basu and B. Bundick. Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3):937–958, 2017.

G. Boehl and F. Strobel. Estimation of dsge models with the effective lower bound. *Journal of Economic Dynamics and Control*, 158:104784, 2024. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2023.104784. URL https://www.sciencedirect.com/science/article/pii/S0165188923001902.

V. V. Chari, P. J. Kehoe, and E. R. McGrattan. Business cycle accounting. *Econometrica*, 75 (3):781–836, 2007.

P. Cuba-Borda, L. Guerrieri, M. Iacoviello, and M. Zhong. Likelihood evaluation of models with occasionally binding constraints. *Journal of Applied Econometrics*, 34(7):1073–1085, 2019.

ECB. Estimates of the euro area capital stock. European Central Bank, Monthly Bulletin, May 2006, Box 4, 2006. URL https://www.ecb.europa.eu/pub/pdf/mobu/mb200605en.pdf.

ECB. The development of the wage share in the euro area since the start of the pandemic, by *Katalin Bodnár and Matthias Mohr*. European Central Bank, Economics Bulletin 4, Box 4, 2023. URL https://www.ecb.europa.eu/pub/pdf/ecbu/eb202304.en.pdf.

R. C. Fair and J. B. Taylor. Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. *Econometrica (pre-1986)*, 51(4):1169, 1983.

L. E. Farmer. The discretization filter: A simple way to estimate nonlinear state space models. *Quantitative Economics*, 12(1):41–76, 2021.

D. Fehrle and J. Huber. Hone the neoclassical lens and zoom in on germany's fiscal stimulus program 2008-2009. 2023.

D. Fehrle, C. Heiberger, and J. Huber. Polynomial chaos expansion: Efficient evaluation and estimation of computational models. *Computational Economics*, pages 1–64, 2025.

L. Guerrieri and M. Iacoviello. Collateral constraints and macroeconomic asymmetries. *Journal of Monetary Economics*, 90:28–49, 2017.

B. Heer and A. Maußner. *Weighted Residuals Methods*, pages 231–310. Springer International Publishing, Cham, 2024. ISBN 978-3-031-51681-8. doi: 10.1007/978-3-031-51681-8_5. URL https://doi.org/10.1007/978-3-031-51681-8_5.

E. P. Herbst and F. Schorfheide. *Bayesian estimation of DSGE models*. Princeton University Press, 2015.

T. D. Holden. Existence and uniqueness of solutions to dynamic models with occasionally binding constraints. *Review of Economics and Statistics*, 105(6):1481–1499, 2023.

J. Huber. An augmented steady-state kalman filter to evaluate the likelihood of linear and time: Invariant state-space models. Technical report, Augsburg University Volkswirtschaftliche Diskussionsreihe, 2022.

U. J. Jermann. Asset pricing in production economies. *Journal of monetary Economics*, 41 (2):257–275, 1998.

J. Kim, S. Kim, E. Schaumburg, and C. A. Sims. Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control*, 32(11):3397–3414, 2008. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2008.02.003. URL https://www.sciencedirect.com/science/article/pii/S0165188908000316.

R. Kollmann. Tractable likelihood-based estimation of non-linear dsge models. *Economics letters*, 161:90–92, 2017.

J. J. Rotemberg. Sticky prices in the united states. *Journal of political economy*, 90(6): 1187–1211, 1982.

R. Šustek. Monetary business cycle accounting. *Review of Economic Dynamics*, 14(4): 592–612, 2011.

# Online Appendix: Not for Publication

# A Appendix

## A.1 Further results of the Monte Carlo experiment

**Table 11:** RMSE of ML estimates from conditional likelihoods to true ML estimates (parameters in % of prior range, T=100)

| | $\mathscr{L}^1$ | $\mathscr{L}_p^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|
| $n^*$ | 4.59 | $5.11^{***}_{***}$ | $5.07^{***}_{***}$ | $7.01^{***}_{***}$ | $8.83^{***}_{***}$ | $11.80^{***}_{***}$ |
| $\sigma$ | 16.14 | 16.04 | $19.43^{***}_{***}$ | $20.94^{***}_{***}$ | $23.45^{***}_{***}$ | $49.55^{***}_{***}$ |
| $\rho_A$ | 1.82 | $2.28^{***}_{***}$ | $4.10^{***}_{***}$ | $4.07^{***}_{***}$ | $4.25^{***}_{***}$ | $2.35^{***}_{***}$ |
| $\rho_N$ | 1.36 | $1.36^{*}_{*}$ | 1.37 | $2.34^{***}_{***}$ | $3.13^{***}_{***}$ | $1.58^{**}_{**}$ |
| $\rho_G$ | 1.54 | 1.54 | 1.54 | $2.06^{***}_{***}$ | $2.73^{***}_{***}$ | $1.24^{***}_{***}$ |
| $\rho_B$ | 2.59 | $3.01^{***}_{***}$ | 2.39 | $3.00^{***}_{***}$ | $3.67^{***}_{***}$ | $4.69^{***}_{***}$ |
| $\omega_A$ | 0.06 | $0.07^{***}_{***}$ | $0.12^{***}_{***}$ | $0.26^{***}_{***}$ | $0.38^{***}_{***}$ | $0.22^{***}_{***}$ |
| $\omega_N$ | 0.40 | $0.45^{***}_{***}$ | $0.43^{***}_{***}$ | $0.64^{***}_{***}$ | $0.80^{***}_{***}$ | $0.87^{***}_{***}$ |
| $\omega_G$ | 0.07 | 0.07 | 0.07 | $0.25^{***}_{***}$ | $0.37^{***}_{***}$ | $0.21^{***}_{***}$ |
| $\omega_B$ | 15.07 | 14.12 | 16.37 | $18.93^{***}_{***}$ | $21.46^{***}_{***}$ | $54.37^{***}_{***}$ |
| $x_1$ | 5.34 | 5.39 | $9.43^{***}_{***}$ | $9.43^{***}_{***}$ | $9.43^{***}_{***}$ | 4.52 |
| $z_{A1}$ | 1.97 | 1.99 | $3.49^{***}_{***}$ | $3.49^{***}_{***}$ | $3.49^{***}_{***}$ | 1.67 |
| $z_{N1}$ | $0.24_{*}$ | $0.26^{*}$ | $0.25^{***}$ | $0.33^{***}_{***}$ | $0.40^{***}_{***}$ | $0.45^{***}_{***}$ |
| $z_{G1}$ | 0.00 | 0.00 | 0.00 | $0.00^{***}_{**}$ | $0.00^{***}_{**}$ | $0.00^{***}_{***}$ |
| $z_{B1}$ | 1.32 | 1.18 | $1.39^{***}_{***}$ | $1.58^{***}_{***}$ | $1.75^{***}_{***}$ | $3.81^{***}_{***}$ |

**Notes:** RMSE of the ML estimates to estimates from the true likelihood from 1032 samples, relative to the prior range in %. Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_p^1$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. Significance levels: $^{*}$, $^{**}$, $^{***}$ denote $p$-values of 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$; $_{*}$, $_{**}$, $_{***}$ indicate the same significance levels for the null that the estimate equals that from $\mathscr{L}_p^1$. Global maximization used 200 stage-one points and 1,000 trial points (default), particle filter estimates are done via swarm particles optimization (100 particles—default).

**Table 12:** RMSE of ML estimates from conditional likelihoods to true ML estimates (parameters in % of prior range, T=400)

| | $\mathscr{L}^1$ | $\mathscr{L}_p^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|
| $n^*$ | 0.73 | 0.75 | $1.29^{***}_{***}$ | $1.54^{***}_{***}$ | $1.77^{***}_{***}$ | $9.27^{***}_{***}$ |
| $\sigma$ | 3.24 | 3.11 | $5.38^{***}_{***}$ | $5.83^{***}_{***}$ | $5.87^{***}_{***}$ | $55.30^{***}_{***}$ |
| $\rho_A$ | 0.36 | $0.46^{***}_{***}$ | $0.92^{***}_{***}$ | $0.85^{***}_{***}$ | $0.81^{***}_{***}$ | $1.10^{***}_{***}$ |
| $\rho_N$ | 0.34 | 0.34 | 0.34 | $0.47^{***}_{***}$ | $0.57^{***}_{***}$ | $0.86^{***}_{***}$ |
| $\rho_G$ | 0.52 | 0.52 | 0.52 | $0.59^{***}_{***}$ | $0.64^{***}_{***}$ | $0.73^{***}_{***}$ |
| $\rho_B$ | 0.48 | $0.51^{**}_{**}$ | $0.56^{***}_{***}$ | $0.63^{***}_{***}$ | $0.67^{***}_{***}$ | $3.11^{***}_{***}$ |
| $\omega_A$ | 0.01 | $0.01^{***}_{***}$ | $0.02^{***}_{***}$ | $0.06^{***}_{***}$ | $0.09^{***}_{***}$ | $0.11^{***}_{***}$ |
| $\omega_N$ | 0.05 | 0.05 | $0.08^{***}_{***}$ | $0.12^{***}_{***}$ | $0.14^{***}_{***}$ | $0.54^{***}_{***}$ |
| $\omega_G$ | 0.02 | 0.02 | 0.02 | $0.06^{***}_{***}$ | $0.08^{***}_{***}$ | $0.11^{***}_{***}$ |
| $\omega_B$ | 2.88 | $2.69^{*}_{*}$ | $4.90^{***}_{***}$ | $5.68^{***}_{***}$ | $5.93^{***}_{***}$ | $54.81^{***}_{***}$ |
| $x_1$ | 3.71 | 3.72 | $8.89^{***}_{***}$ | $8.89^{***}_{***}$ | $8.89^{***}_{***}$ | $4.47^{***}_{***}$ |
| $z_{A1}$ | 1.37 | 1.38 | $3.29^{***}_{***}$ | $3.29^{***}_{***}$ | $3.29^{***}_{***}$ | $1.65^{***}_{***}$ |
| $z_{N1}$ | 0.03 | 0.03 | $0.06^{***}_{***}$ | $0.07^{***}_{***}$ | $0.07^{***}_{***}$ | $0.37^{***}_{***}$ |
| $z_{G1}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $0.00^{***}_{***}$ |
| $z_{B1}$ | 0.33 | 0.32 | $0.67^{***}_{***}$ | $0.70^{***}_{***}$ | $0.75^{***}_{***}$ | $3.83^{***}_{***}$ |

**Notes:** RMSE of the ML estimates to estimates from the true likelihood from 1032 samples, relative to the prior range in %. Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_p^1$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. Significance levels: $^{*}$, $^{**}$, $^{***}$ denote $p$-values of 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$; $_{*}$, $_{***}$, $_{***}$ indicate the same significance levels for the null that the estimate equals that from $\mathscr{L}_p^1$. Global maximization used 200 stage-one points and 1,000 trial points (default), particle filter estimates are done via swarm particles optimization (100 particles—default).

**Table 13:** RMSFE of ML estimates from conditional likelihoods to true ML estimates (in percentage points of the steady state, T=100)

| | $\mathscr{L}^1$ | $\mathscr{L}_p^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|
| $\hat{y}_{T+1}$ | $0.08^{}_{***}$ | $0.10^{***}$ | $0.13^{***}_{***}$ | $0.14^{***}_{***}$ | $0.16^{***}_{***}$ | $0.14^{***}_{***}$ |
| $\hat{c}_{T+1}$ | $0.07^{}_{***}$ | $0.08^{***}$ | $0.09^{***}_{***}$ | $0.10^{***}_{***}$ | $0.12^{***}_{***}$ | $0.15^{***}_{***}$ |
| $\hat{i}_{T+1}$ | $0.38^{}_{***}$ | $0.45^{***}$ | $0.48^{***}_{***}$ | $0.55^{***}_{***}$ | $0.64^{***}_{***}$ | $0.78^{***}_{***}$ |
| $\hat{n}_{T+1}$ | $0.08^{}_{***}$ | $0.10^{***}$ | $0.10^{***}_{*}$ | $0.12^{***}_{***}$ | $0.14^{***}_{***}$ | $0.18^{***}_{***}$ |
| $\hat{y}_{T+4}$ | $0.29^{}_{***}$ | $0.32^{***}$ | $0.41^{***}_{***}$ | $0.45^{***}_{***}$ | $0.50^{***}_{***}$ | $0.45^{***}_{***}$ |
| $\hat{c}_{T+4}$ | $0.23^{}_{***}$ | $0.25^{***}$ | $0.28^{***}_{***}$ | $0.31^{***}_{***}$ | $0.36^{***}_{***}$ | $0.44^{***}_{***}$ |
| $\hat{i}_{T+4}$ | $1.25^{}_{***}$ | $1.41^{***}$ | $1.49^{***}_{***}$ | $1.71^{***}_{***}$ | $1.95^{***}_{***}$ | $2.40^{***}_{***}$ |
| $\hat{n}_{T+4}$ | $0.27^{}_{***}$ | $0.30^{***}$ | $0.31^{***}_{*}$ | $0.37^{***}_{***}$ | $0.43^{***}_{***}$ | $0.54^{***}_{***}$ |

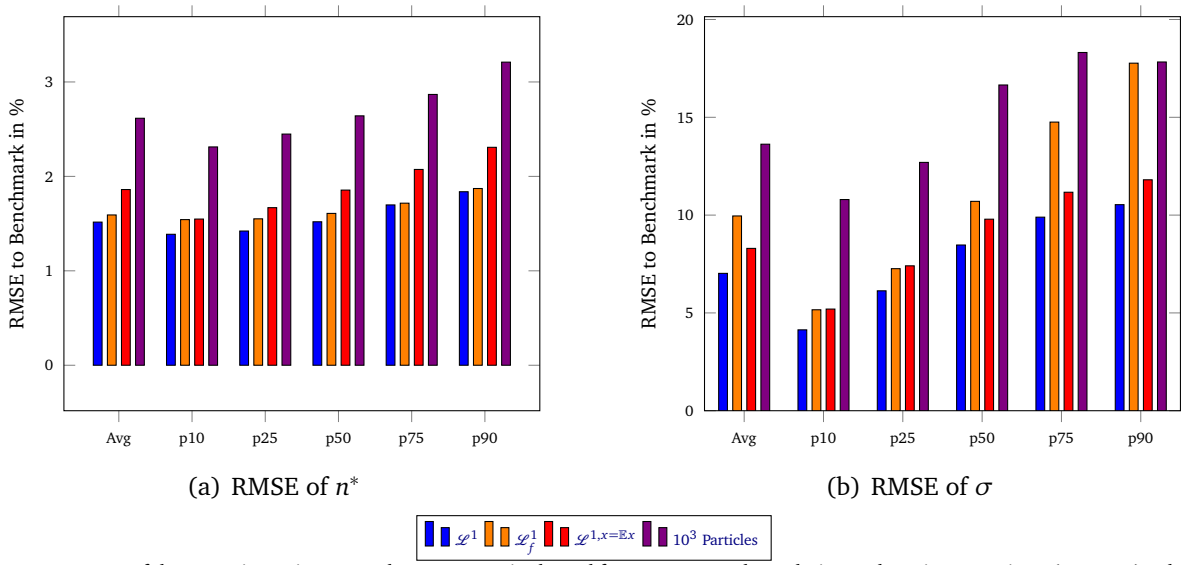
**Notes:** One and four periods RMSFE of the ML estimated models to estimates from the true likelihood estimated model from 1032 samples in percentage points of the respective variable's steady state. Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_p^1$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. Significance levels: $^{*}$, $^{**}$, $^{***}$ denote $p$-values of 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$; $_{*}$, $_{**}$, $_{***}$ indicate the same significance levels for the null that the estimate equals that from $\mathscr{L}_p^1$.

**Table 14:** RMSFE of ML estimates from conditional likelihoods to true ML estimates (in percentage points of the steady state, T=400)

| | $\mathscr{L}^1$ | $\mathscr{L}_p^1$ | $\mathscr{L}^{1,x=\mathbb{E}x}$ | $\mathscr{L}^{10,x=\mathbb{E}x}$ | $\mathscr{L}^{20,x=\mathbb{E}x}$ | $10^3$ Particles |
|---|---|---|---|---|---|---|
| $\hat{y}_{T+1}$ | $0.02^{}_{***}$ | $0.02^{***}$ | $0.04^{***}_{***}$ | $0.03^{***}_{***}$ | $0.03^{***}_{***}$ | $0.08^{***}_{***}$ |
| $\hat{c}_{T+1}$ | $0.02^{}_{*}$ | $0.02^{*}$ | $0.02^{***}_{***}$ | $0.02^{***}_{***}$ | $0.02^{***}_{***}$ | $0.10^{***}_{***}$ |
| $\hat{i}_{T+1}$ | $0.09^{}_{**}$ | $0.10^{**}$ | $0.13^{***}_{***}$ | $0.13^{***}_{***}$ | $0.13^{***}_{***}$ | $0.51^{***}_{***}$ |
| $\hat{n}_{T+1}$ | $0.02^{}_{*}$ | $0.02^{*}$ | $0.03^{***}_{***}$ | $0.03^{***}_{***}$ | $0.03^{***}_{***}$ | $0.12^{***}_{***}$ |
| $\hat{y}_{T+4}$ | $0.07^{}_{***}$ | $0.08^{***}$ | $0.12^{***}_{***}$ | $0.12^{***}_{***}$ | $0.12^{***}_{***}$ | $0.27^{***}_{***}$ |
| $\hat{c}_{T+4}$ | $0.06^{}_{*}$ | $0.07^{*}$ | $0.08^{***}_{***}$ | $0.08^{***}_{***}$ | $0.08^{***}_{***}$ | $0.32^{***}_{***}$ |
| $\hat{i}_{T+4}$ | $0.32^{}_{**}$ | $0.35^{**}$ | $0.43^{***}_{***}$ | $0.44^{***}_{***}$ | $0.45^{***}_{***}$ | $1.70^{***}_{***}$ |
| $\hat{n}_{T+4}$ | 0.07 | 0.08 | $0.09^{***}_{***}$ | $0.09^{***}_{***}$ | $0.10^{***}_{***}$ | $0.39^{***}_{***}$ |

**Notes:** One and four periods RMSFE of the ML estimated models to estimates from the true likelihood estimated model from 1032 samples, in percentage points of the respective variable's steady state. Likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_p^1$ – maximization includes the initial endogenous state; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to the unconditional first moment; $\mathscr{L}^{10,x=\mathbb{E}x}$ and $\mathscr{L}^{20,x=\mathbb{E}x}$ – same, but with the first 10 and 20 periods burned, respectively; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. Significance levels: $^{*}$, $^{**}$, $^{***}$ denote $p$-values of 0.05, 0.01, and 0.001 for the null hypothesis that the estimate equals that from $\mathscr{L}^1$; $_{*}$, $_{**}$, $_{***}$ indicate the same significance levels for the null that the estimate equals that from $\mathscr{L}_p^1$.

**Figure 10:** Posterior errors of estimated steady state and behavioral parameters (% of prior range)—no discards

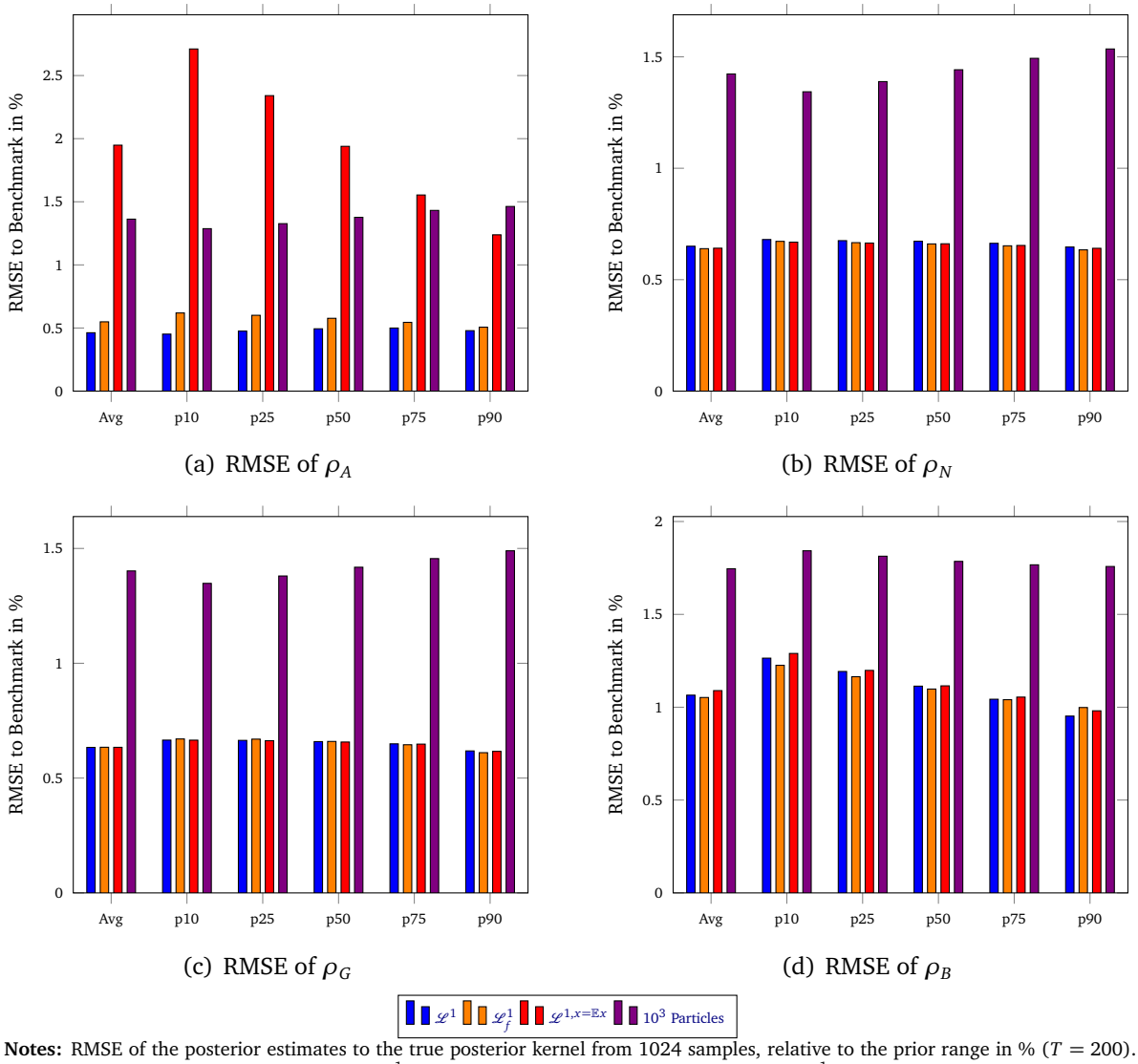



(a) RMSE of $n^*$

(b) RMSE of $\sigma$

**Notes:** RMSE of the posterior estimates to the true posterior kernel from 1024 samples, relative to the prior range in % ($T = 200$). The likelihood specifications are as follows: $\mathbb{E}(\mathscr{L}|Y_1)$ – inverse likelihood conditioned solely on $Y_1$; $\mathscr{L}(X_1|Y_1)$ – posterior draws include the initial endogenous states; $\mathscr{L}(x_1 = \mathbb{E}(x))$ – initial endogenous state set to its unconditional first moment; $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. A Random-Walk Metropolis-Hastings sampler is used with 150,000 draws, where the first 1/3 of the draws are discarded as burn-in.

**Figure 11:** Posterior errors of estimated autoregressive coefficients (% of prior range)—no discards




(a) RMSE of $\rho_A$

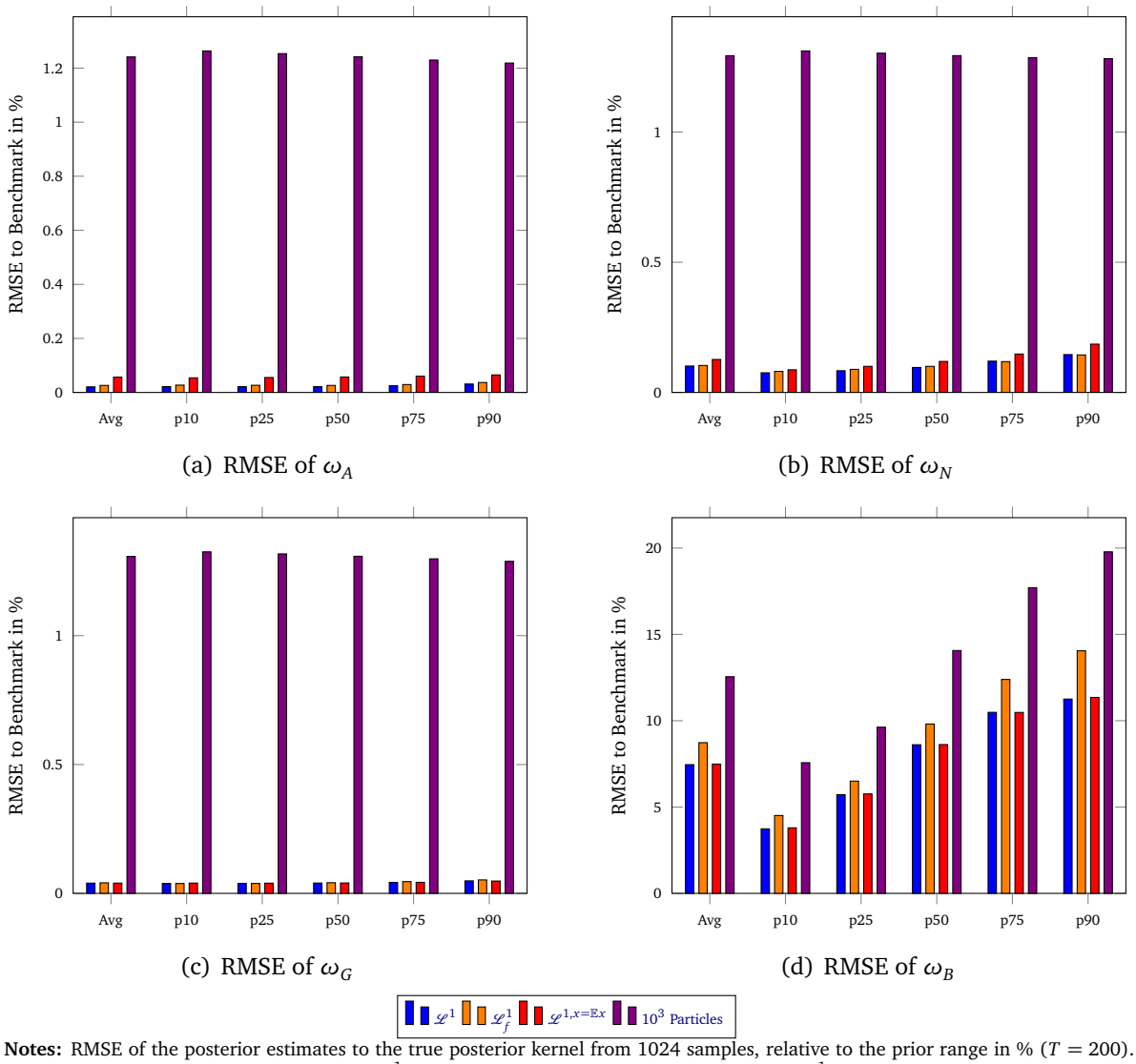
(b) RMSE of $\rho_N$

(c) RMSE of $\rho_G$

(d) RMSE of $\rho_B$

**Notes:** RMSE of the posterior estimates to the true posterior kernel from 1024 samples, relative to the prior range in % ($T = 200$). The likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}_f^1$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. A Random-Walk Metropolis-Hastings sampler is used with 150,000 draws, where the first 1/3 of the draws are discarded as burn-in.

**Figure 12:** Posterior errors of estimated innovation's standard deviation (% of prior range)—no discards




(a) RMSE of $\omega_A$

(b) RMSE of $\omega_N$

(c) RMSE of $\omega_G$

(d) RMSE of $\omega_B$

**Notes:** RMSE of the posterior estimates to the true posterior kernel from 1024 samples, relative to the prior range in % ($T = 200$). The likelihood specifications are as follows: $\mathscr{L}^1$ – inverse likelihood conditioned solely on $y_1$; $\mathscr{L}^1_f$ – posterior draws include the initial endogenous states; $\mathscr{L}^{1,x=\mathbb{E}x}$ – initial endogenous state set to its unconditional first moment, $10^3$ Particles – $10^3$ particles used, drawn from the states' stationary distribution. A Random-Walk Metropolis-Hastings sampler is used with 150,000 draws, where the first 1/3 of the draws are discarded as burn-in.

## A.2 MBCA

### A.2.1 MBCA Equation system

The model in stationary variables $\bar{x}$ read:

$$\alpha \frac{\bar{y}_t}{\bar{n}_t} = \exp(z_{Nt})\theta \frac{\bar{c}_t}{1-\bar{n}_t}, \tag{13}$$

$$\exp(z_{It}) = \beta \mathbb{E}_t \left[ \frac{g_c^{\sigma}}{g_i} \left( \frac{\bar{c}_{t+1}}{\bar{c}_t} \right)^{\sigma} \left( \frac{1-\bar{n}_{t+1}}{1-\bar{n}_t} \right)^{\sigma(1-\theta)} \left( (1-\delta)\exp(z_{It+1}) + \alpha \frac{\bar{y}_{t+1}}{\bar{k}_{t+1}} \right) \right], \quad (14)$$

$$\bar{y}_t = \exp(z_{At}) \bar{k}_t^{\alpha} \bar{n}_t^{1-\alpha}, \quad (15)$$

$$\bar{y}_t = \bar{i}_t + \bar{c}_t + \bar{g}_t, \quad (16)$$

$$g_{pop} g_I \bar{k}_{t+1} = \bar{i}_t + (1-\delta)\bar{k}_t, \quad (17)$$

$$\bar{g}_t = \gamma \bar{y}_t \exp(z_{Gt}), \quad (18)$$

$$\exp(z_{Bt}) = \beta \mathbb{E}_t \left[ \left( \frac{g_c \bar{c}_{t+1}}{\bar{c}_t} \right)^{\sigma} \left( \frac{1-\bar{n}_{t+1}}{1-\bar{n}_t} \right)^{\sigma(1-\theta)} \frac{\bar{r}_t}{\bar{\pi}_{t+1}} \exp(z_{Bt+1}) \right], \quad (19)$$

$$\bar{r}_t = r^* \left( \frac{1+\bar{\pi}_t}{1+\pi^*} \right)^{\psi_\pi} \left( \frac{\bar{y}_t}{y^{ss}} \right)^{\psi_y} \exp(z_{Rt}), \quad (20)$$

$$z_{jt} = z_j^{ss} + \bar{z}_{jt}, \ j \in \{A, N, G, I, B, R\}, \quad (21)$$

$$z_{t+1} = \Pi z_t + \epsilon_{t+1}, \epsilon_{t+1} \sim \mathcal{N}(0_{nz}, \Sigma_{nz \times nz}), \quad (22)$$

$$z_t = [\bar{z}_{At}, \bar{z}_{Nt}, \bar{z}_{Gt}, \bar{z}_{It}, \bar{z}_{Bt}, \bar{z}_{Rt}]'. \quad (23)$$

### A.2.2 MBCA Data source

**EMU-19** All data is on a quarterly frequency.

- Population: Total population national concept Eurostat 30/06/2025 23:00 Thousand persons Seasonally and calendar adjusted data
- Hours worked: Total employment domestic concept Eurostat 16/04/2025 23:00 Thousand hours worked Total - all NACE activities Seasonally and calendar adjusted data
- GDP: Gross domestic product at market prices Eurostat 28/04/2025 23:00 Chain linked volumes (2010), million euro Seasonally and calendar adjusted data
- Consumption: Household and NPISH final consumption expenditure Eurostat 28/-04/2025 23:00 Chain linked volumes (2010), million euro Seasonally and calendar adjusted data
- Investment: Gross fixed capital formation Eurostat 28/04/2025 23:00 Chain linked volumes (2010), million euro Seasonally and calendar adjusted data
- Inflation: Gross domestic product at market prices Eurostat 28/04/2025 23:00 Price index (implicit deflator), 2010=100, euro Seasonally and calendar adjusted data

Gross domestic product at market prices

- Interest rate: 3-month rate Eurostat 17/04/2025 11:00

**US** All data is on a quarterly frequency.

- Population: Population, Thousands, Quarterly, Not Seasonally Adjusted Data, B230-RC0Q173SBEA, Updated: 2025-01-30 Federal Reserve Economic Data, Federal Reserve Bank of St. Louis
- Hours worked: Quarterly hours worked and employment in total U.S. economy and subsectors, Data released March 6, 2025; Bureau of Labor Statistics, Office of Productivity and Technology, (Seasonally Adjusted Data, information requested)
- GDP: Gross domestic product at market prices Table 1.1.6. Real Gross Domestic Product, Chained Dollars [Billions of chained (2017) dollars] Seasonally adjusted at annual rates, Bureau of Economic Analysis, Last Revised on: June 26, 2025
- Consumption: Personal consumption expenditures Table 1.1.6. Real Gross Domestic Product, Chained Dollars [Billions of chained (2017) dollars] Seasonally adjusted at annual rates, Bureau of Economic Analysis, Last Revised on: June 26, 2025
- Investment: Gross private domestic investment Table 1.1.6. Real Gross Domestic Product, Chained Dollars [Billions of chained (2017) dollars] Seasonally adjusted at annual rates, Bureau of Economic Analysis, Last Revised on: June 26, 2025
- Inflation: Gross domestic product at market prices Table 1.1.4. Price Indexes for Gross Domestic Product Chained Dollars[Index numbers, 2017=100] Seasonally adjusted, Bureau of Economic Analysis, Last Revised on: June 26, 2025
- Interest rate: Short-term interest rates, Topic: Economy > Short-term economic statistics OECD.SDD.STES,DSD_STES[at]DF_FINMARK,4.0,filtered,2025-04-29 21-17-08

### A.3 Basu and Bundick (2017)

#### A.3.1 Equation system

The model read:

$$sdf_t - \mathbb{E}_t(\beta((\exp(\rho_a a_{t+1} + \exp(evola_{t+1})e_{a,t+1}))/exp(a_{t+1}))\dots$$

$$\times ((c_{t+1}^{\eta}(1-n_{t+1})^{1-\eta})/(c_t^{\eta}(1-n_t)^{1-\eta}))^{(1-\sigma)/\theta_{vf}}(c_t/c_{t+1})(vf_{t+1}^{1-\sigma}/(vf_{t+1}^{\sigma_1}))^{1-1/\theta_{vf}}) = 0 \tag{24}$$

$$y_t + \text{fixedcost} - \text{productionconstant} \cdot \left(\exp(z_t)n_t\right)^{1-\alpha}(u_t k_t)^{\alpha} = 0 \tag{25}$$

$$n_t \cdot \frac{1-\eta}{\eta}\frac{c_t}{1-n_t} - (1-\alpha)\frac{y_t + \text{fixedcost}}{\mu_t} = 0 \tag{26}$$

$$\frac{1}{1-\phi_k\left(\frac{inv_t}{k_t}-\delta_0\right)}(\delta_1 + \delta_2(u_t-1))u_t k_t - \alpha\frac{y_t+\text{fixedcost}}{\mu_t} = 0 \tag{27}$$

$$\log(r_t) - \Big((1-\rho_r)\big(\log(r_{ss}) + \rho_\pi \log(\pi_t/\pi_{ss}) + \rho_y \log(y_t/1)\big)\Big) = 0 \tag{28}$$

$$c_t + g_y y_t \exp(g_t) - \Big(y_t - inv_t - \frac{\phi_\pi}{2}\big(\frac{\pi_t}{\pi_{ss}}-1\big)^2 y_t\Big) = 0 \tag{29}$$

$$k_{t+1} - \Big(\big(1-(\delta_0 + \delta_1(u_t-1) + \frac{\delta_2}{2}(u_t-1)^2) - \frac{\phi_k}{2}\big(\frac{inv_t}{k_t}-\delta_0\big)^2\big)k_t + inv_t\Big) = 0 \tag{30}$$

$$vf_t - \Big(\text{utilityconstant} \cdot \exp(a_{t+1})(c_t^{\eta}(1-n_t)^{1-\eta})^{\frac{1-\sigma}{\theta_{vf}}} + \beta(vf_{t+1}^{\sigma_1})^{\frac{1}{\theta_{vf}}}\Big)^{\frac{\theta_{vf}}{1-\sigma}} = 0 \tag{31}$$

$$(vf_{t+1}^{\sigma_1}) - \mathbb{E}_t vf_{t+1}^{1-\sigma} = 0 \tag{32}$$

$$1 - rr_t \cdot sdf_t = 0 \tag{33}$$

$$1 - r_t \cdot sdf_t \cdot \pi_{t+1}^{-1} = 0 \tag{34}$$

$$\begin{aligned} 1 - sdf_t \mathbb{E}_t \Bigg( & \frac{u_{t+1}\alpha(y_{t+1}+\text{fixedcost})}{\mu_{t+1}u_{t+1}k_{t+1}} + \frac{1}{1-\phi_k(inv_{t+1}/k_{t+1}-\delta_0)}\Big(1-(\delta_0+\delta_1(u_{t+1}-1)\dots \\ & + \frac{\delta_2}{2}(u_{t+1}-1)^2) - \frac{\phi_k}{2}(inv_{t+1}/k_{t+1}-\delta_0)^2\dots \\ & + \phi_k(inv_{t+1}/k_{t+1}-\delta_0)(inv_{t+1}/k_{t+1})\Big)\Big/\frac{1}{1-\phi_k(inv_t/k_t-\delta_0)}\Bigg) = 0 \end{aligned} \tag{35}$$

$$\phi_\pi\left(\frac{\pi_t}{\pi_{ss}}-1\right)\frac{\pi_t}{\pi_{ss}} - \mathbb{E}_t\Big((1-\theta_\mu) + \frac{\theta_\mu}{\mu_t} + sdf_t\phi_\pi\left(\frac{\pi_{t+1}}{\pi_{ss}}-1\right)\frac{y_{t+1}}{y_t}\frac{\pi_{t+1}}{\pi_{ss}}\Big) = 0 \tag{36}$$

$$a_{t+1} - (\rho_a a_t + \exp(evola_t)e_{a,t}) = 0 \tag{37}$$

### A.3.2 Data source

**US** All data is on a quarterly frequency.

- Population: Population, Thousands, Quarterly, Not Seasonally Adjusted Data, B230-RC0Q173SBEA, Updated: 2025-01-30 Federal Reserve Economic Data, Federal Reserve Bank of St. Louis
- Hours worked: Quarterly hours worked and employment in total U.S. economy and

**Table 15:** Prior and posterior distribution—profile likelihood

| | Prior | | | Posterior | |
|---|---|---|---|---|---|
| Parameter | Prior Distribution | Mean | Standard deviation | Mean | Standard deviation |
| $\phi_K$ | $\mathscr{IG}$ | 5 | Mean/5 | 5.56 | Mean/28 |
| $\omega_Z$ | $\mathscr{IG}$ | 0.0015 | Mean/5 | 0.013 | Mean/11 |
| $\omega_G$ | $\mathscr{IG}$ | 0.001 | Mean/5 | 0.021 | Mean/7 |
| $\omega_A$ | $\mathscr{IG}$ | 0.004 | Mean/5 | 0.0056 | Mean/22 |
| $\omega_\nu$ | $\mathscr{IG}$ | 0.25 | Mean/5 | 0.1213 | Mean/10 |
| $\rho_Z$ | $\mathscr{B}$ | 0.98 | 0.005 | 0.74 | 0.023 |
| $\rho_G$ | $\mathscr{B}$ | 0.8 | 0.05 | 0.99 | 0.003 |
| $\rho_A$ | $\mathscr{B}$ | 0.8 | 0.05 | 0.94 | 0.004 |
| $\rho_\nu$ | $\mathscr{B}$ | 0.8 | 0.05 | 0.89 | 0.01 |

**Notes:** $\mathscr{IG}$: Inverted Gamma Distribution, $\mathscr{B}$: Beta Distribution. Posterior draws from a SMC Algorithm with $N = 1,000$ particles (Herbst and Schorfheide, 2015, Algorithms 8, 9, and 10.). The likelihood tempering schedule follows $(n/N_\phi)^2$ with $N_\phi = 200$. Resampling takes place for effective sample sizes lower $N/2$. The likelihood is a profile likelihood with respect to the endogenous states and conditional on the first observation.

subsectors, Data released March 6, 2025; Bureau of Labor Statistics, Office of Productivity and Technology, (Seasonally Adjusted Data, information requested)

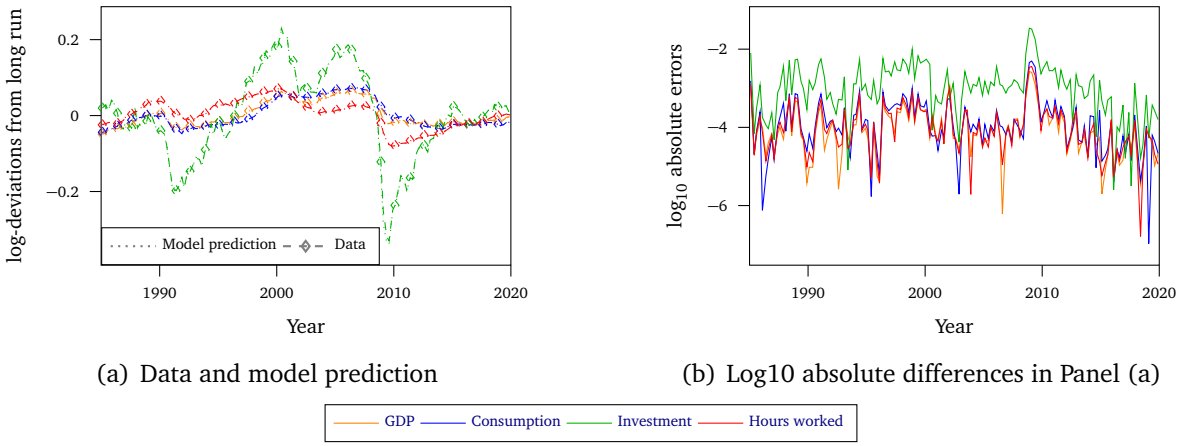
- GDP: Gross domestic product at market prices Table 1.1.6. Real Gross Domestic Product, Chained Dollars [Billions of chained (2017) dollars] Seasonally adjusted at annual rates, Bureau of Economic Analysis, Last Revised on: June 26, 2025

- Consumption: Personal consumption expenditures Table 1.1.6. Real Gross Domestic Product, Chained Dollars [Billions of chained (2017) dollars] Seasonally adjusted at annual rates, Bureau of Economic Analysis, Last Revised on: June 26, 2025

- Investment: Gross private domestic investment Table 1.1.6. Real Gross Domestic Product, Chained Dollars [Billions of chained (2017) dollars] Seasonally adjusted at annual rates, Bureau of Economic Analysis, Last Revised on: June 26, 2025

### A.3.3 Profile likelihood

**Figure 13:** Prediction residuals—profile likelihood

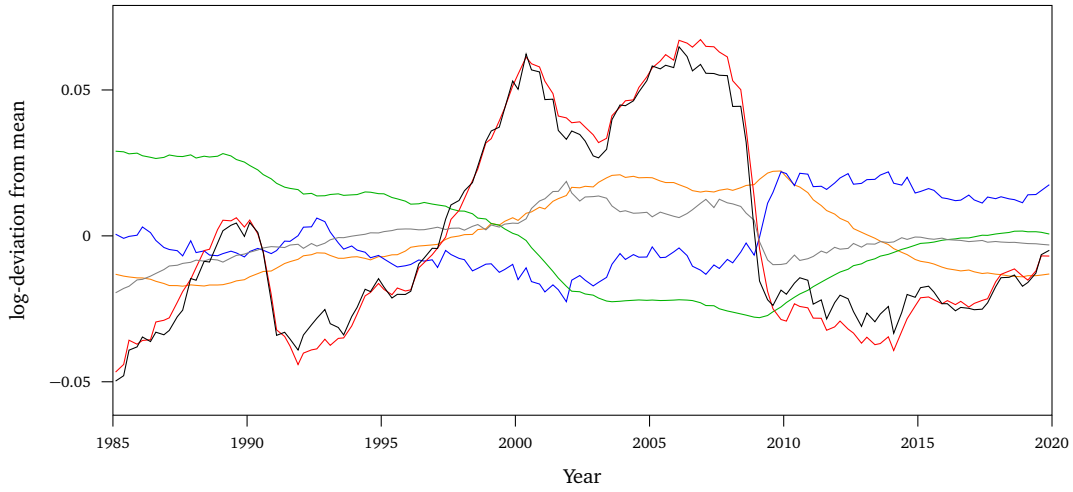



(a) Data and model prediction

(b) Log10 absolute differences in Panel (a)

**Notes:** Data and model long run equals stochastic steady state.

**Figure 14:** Business cycle decomposition—profile likelihood




(a) Data and model prediction

**Notes:** Centered times series from singular shocks on, interactions, and data.

**Table 16:** Normalized inverted mean squared error—profile likelihood

| $Z$ | $G$ | $A$ | $v$ | Interaction |
|---|---|---|---|---|
| 0.01 | 0.03 | 0.01 | 0.91 | 0.03 |

**Notes:** Normalized inverted mean squared error: inverted mean squared error relative to accumulated inverted mean squared errors.