Online Appendix

A Intermediate Theoretical Results

In this section I detail intermediate and auxiliary theoretical results used in the proofs of the main theorems stated in section 5.

Following Douc et al. (2004), I first establish that the distribution of $X_{t,M}$ given a history of observations $\boldsymbol{Y}_r^s$ is itself a uniformly ergodic (inhomogeneous) Markov chain with minorizing constant independent of the parameter $\theta \in \Theta$ and the number of discrete points $M \in \mathbb{Z}^+$. This is the analogous result to Lemma 1 in their paper. Note that a Markov chain with transition kernel P_θ is said to satisfy a uniform minorization condition if there exist a probability measure μ_Q, a positive integer n, and $\epsilon > 0$ such that

$$P_\theta^{(n)}(x, A) \geq \epsilon \mu_Q(A)$$

for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$, where $P_\theta^{(n)}$ is the n-step ahead transition kernel of the Markov chain.

Define $Q_M^- \equiv \inf_{m,m'} P_{\theta,M}(m, m')$ and $Q_M^+ \equiv \sup_{m,m'} P_{\theta,M}(m, m')$ for $M \in \mathbb{Z}^+$. I now state the first lemma

Lemma 1. *Assume (A1) and (B1). Let $s, r \in \mathbb{Z}$, with $r \leq s$, $\theta \in \Theta$, and $M \in \mathbb{Z}^+$. Under $\bar{\mathbb{P}}_\theta$, conditionally on $\boldsymbol{Y}_r^s$, $\{X_{t,M}\}_{t \geq r}$ is an inhomogeneous Markov chain, and for all $t > r$ there exists a function $\mu_{t,M}(\boldsymbol{y}_t^s, A)$ such that:*

(i) for any $A \in \mathcal{B}(\mathcal{X}_M)$, $\boldsymbol{y}_t^s \mapsto \mu_{t,M}(\boldsymbol{y}_t^s, A)$ is a Borel function;

(ii) for any $\boldsymbol{y}_t^s$, $\mu_{t,M}(\boldsymbol{y}_t^s, \cdot)$ is a probability measure on $\mathcal{B}(\mathcal{X}_M)$. In addition, for all $\boldsymbol{y}_t^s$ it holds that $\mu_{t,M}(\boldsymbol{y}_t^s, \cdot) \ll \mu_{c,M}$ (where $\mu_{c,M}$ is counting measure on $\mathcal{X}_M$) and for all $\boldsymbol{Y}_r^s$,

$$\inf_{x \in \mathcal{X}_M} \bar{\mathbb{P}}_\theta(X_{t,M} \in A \,|\, X_{t-1,M} = x, \boldsymbol{Y}_r^s) \geq Q_+^- \mu_{t,M}(\boldsymbol{Y}_t^s, A)$$

The major difference between this Lemma and the one established in Douc et al. (2004) is that for the following results, it will be crucial that the minorizing constant be the same for all M, in order to establish uniform convergence over $M \in \mathbb{Z}^+$ of the approximate likelihood function. Note that although the minorizing measure, $\mu_{t,M}(\boldsymbol{Y}_t^s, \cdot)$, *does* depend on both the number of points, M, and the observations the chain is conditioned on, $\boldsymbol{Y}_t^s$, it doesn't affect the mixing rate. The previous lemma leads to the following corollary, using standard results for uniformly minorized Markov chains (see e.g. Lindvall (1992) Sections III.9-11).

Corollary 1. *Assume (A1) and (B1). Let $r, s \in \mathbb{Z}$ with $r \leq s$, $\theta \in \Theta$, and $M \in \mathbb{Z}^+$. Then for all $t \geq r$, all probability measures μ_1 and μ_2 on $\mathcal{B}(\mathcal{X}_M)$, and all $\boldsymbol{Y}_r^s$,*

$$\left\| \int_{\mathcal{X}_M} \bar{\mathbb{P}}_\theta \left(X_{t,M} \in \cdot \,| X_{r,M} = x, \boldsymbol{Y}_r^s \right) \mu_1 (dx) \right. \\ \left. - \int_{\mathcal{X}_M} \bar{\mathbb{P}}_\theta \left(X_{t,M} \in \cdot \,| X_{r,M} = x, \boldsymbol{Y}_r^s \right) \mu_2 (dx) \right\|_{TV} \leq \rho^{t-r}$$

where $\rho \equiv 1 - Q_+^-$.

This corollary establishes that the Markov chain "uniformly forgets" its history at an exponential rate. That is, no matter where the chain is started, it converges to its ergodic distribution exponentially fast. The fact that the bound is deterministic will be important for establishing strong consistency.

The next step consists of showing that the approximate likelihood function $\ell_{T,M}(\theta, x_{0,M})$ with an arbitrary initial condition $x_{0,M}$ stays within a deterministic bound of $\ell_{T,M}(\theta)$ where $x_{0,M}$ is drawn from its ergodic distribution.

Lemma 2. *Assume (A1)-(A2) and (B1)-(B2). Then, for all $x_{0,M} \in \mathcal{X}_M$ and $M \in \mathbb{Z}^+$,*

$$\sup_{\theta \in \Theta} \left| \ell_{T,M}(\theta, x_{0,M}) - \ell_{T,M}(\theta) \right| \leq 1/(1-\rho)^2, \quad \bar{\mathbb{P}}_{\theta^*}\text{-}a.s.$$

Next I show that $T^{-1}\ell_{T,M}(\theta)$ can be approximated by the sample mean of a $\bar{\mathbb{P}}_{\theta^*}$-stationary ergodic sequence of bounded random variables which has a well defined limit. To this end I first define the quantities:

$$\begin{aligned} \Delta_{t,r,M,x}(\theta) &\equiv \log \bar{p}_{\theta,M} \left(Y_t \left| \boldsymbol{Y}_{-r}^{t-1}, X_{-r,M} = x \right. \right) \\ \Delta_{t,r,M}(\theta) &\equiv \log \bar{p}_{\theta,M} \left(Y_t \left| \boldsymbol{Y}_{-r}^{t-1} \right. \right) \\ &= \int \log \bar{p}_{\theta,M} \left(Y_t \left| \boldsymbol{Y}_{-r}^{t-1}, X_{-r,M} = x \right. \right) \bar{\mathbb{P}}_\theta \left(dx_{-r,M} \left| \boldsymbol{Y}_{-r}^{t-1} \right. \right) \end{aligned}$$

Consider the thought experiment of fixing the number of points M, but letting $T \to \infty$. Define the limiting object as

$$\ell_M(\theta) \equiv \bar{\mathbb{E}}_{\theta^*} \left[\Delta_{0,\infty,M}(\theta) \right]$$

I will show that such a limiting object is well-defined and that the sample analogue converges to this limit almost-surely. In particular, I will show that $\{\Delta_{t,r,M}\}_{r \geq 0}$ and $\{\Delta_{t,r,M,x}\}_{r \geq 0}$ converge uniformly w.r.t. $\theta \in \Theta$ $\bar{\mathbb{P}}_{\theta^*}$-a.s. by showing they are uniform Cauchy sequences.

Lemma 3. *Assume (A1)-(A3) and (B1)-(B3). Then for all $t \geq 1$, $r, r' \geq 0$, and $M \in \mathbb{Z}^+$, $\bar{\mathbb{P}}_{\theta^*}$-a.s.,*

$$\sup_{\theta \in \Theta} \sup_{x, x' \in \mathcal{X}_M} \left| \Delta_{t,r,M,x}(\theta) - \Delta_{t,r',M,x'}(\theta) \right| \leq \rho^{t+\min(r,r')-1} / (1-\rho), \tag{A.1}$$

$$\sup_{\theta \in \Theta} \sup_{x \in \mathcal{X}_M} \left| \Delta_{t,r,M,x}(\theta) - \Delta_{t,r,M}(\theta) \right| \leq \rho^{t+r-1} / (1-\rho), \tag{A.2}$$

$$\sup_{\theta \in \Theta} \sup_{r \geq 0} \sup_{x \in \mathcal{X}_M} \left| \Delta_{t,r,M,x}(\theta) \right| \leq \max\left(\left| \log b_+ \right|, \left| \log c_- (Y_t) \right| \right) \tag{A.3}$$

Equation (A.1) of Lemma 3 shows that $\{\Delta_{t,r,M,x}\}_{r \geq 0}$ is a uniform Cauchy sequence w.r.t. $\theta \in \Theta$ and thus converges $\bar{\mathbb{P}}_{\theta^*}$-a.s. to a limit which does not depend on the initial value x. I label this limit $\Delta_{t,\infty,M}$ and intuitively this can be thought of as $\log \bar{p}_{\theta,M}\left(Y_t \left| \boldsymbol{Y}_{-\infty}^{t-1} \right. \right)$, the marginal likelihood of an observation Y_t given an infinite history of data.

Equation (A.3) of Lemma 3 shows that $\{\Delta_{t,r,M,x}(\theta)\}_{r \geq 0}$ is uniformly bounded in $L^1\left(\bar{\mathbb{P}}_{\theta^*}\right)$ and thus its limit $\Delta_{t,\infty,M}(\theta)$ is also in $L^1\left(\bar{\mathbb{P}}_{\theta^*}\right)$. Furthermore, note that $\{\Delta_{t,\infty,M}(\theta)\}$ is a $\bar{\mathbb{P}}_{\theta^*}$-stationary ergodic process.

By setting $r = 0$ and letting $r' \to \infty$ in equation (A.1), it follows that

$$\sup_{\theta \in \Theta} \left| \Delta_{t,0,M,x}(\theta) - \Delta_{t,\infty,M}(\theta) \right| \leq \rho^{t-1} / (1-\rho)$$

Furthermore, setting $r = 0$ in equation (A.2) implies that

$$\sup_{\theta \in \Theta} \left| \Delta_{t,0,M,x}(\theta) - \Delta_{t,0,M}(\theta) \right| \leq \rho^{t-1} / (1-\rho)$$

By combining these two inequalities, applying the triangle inequality, and summing from 1 to T, I obtain Corollary 2.

Corollary 2. *Assume (A1)-(A2) and (B1)-(B2). Then*

$$\sum_{t=1}^{T} \sup_{M \in \mathbb{Z}^+} \sup_{\theta \in \Theta} \left| \Delta_{t,0,M}(\theta) - \Delta_{t,\infty,M}(\theta) \right| \leq 2 / (1-\rho)^2, \quad \bar{\mathbb{P}}_{\theta^*}\text{-a.s.}$$

Corollary 2 shows that $T^{-1} \ell_{T,M}(\theta)$ can be approximated by the sample mean of a stationary ergodic sequence, uniformly w.r.t. θ. Since $\Delta_{0,\infty,M} \in L^1\left(\bar{\mathbb{P}}_{\theta^*}\right)$, the ergodic theorem implies that $T^{-1} \ell_{T,M}(\theta) \to \ell_M(\theta)$ $\bar{\mathbb{P}}_{\theta^*}$-a.s. and in $L^1\left(\bar{\mathbb{P}}_{\theta^*}\right)$ as $T \to \infty$. Note that this convergence is uniform over $M \in \mathbb{Z}^+$. This will be important when I start considering joint asymptotics in T and M.

Define $\ell(\theta) \equiv \bar{\mathbb{E}}_{\theta^*}\left[\log \bar{p}_\theta \left(Y_0 \left| \boldsymbol{Y}_{-\infty}^{0} \right. \right) \right]$. The next step towards establishing consistency is to show that $\ell_M(\theta) \to \ell(\theta)$ as $M \to \infty$. The difference in these two quantities is related to the difference in the approximate and true filtering distributions for infinite histories of observations, $X_{t,M} \left| \boldsymbol{Y}_{-\infty}^{t} \right.$ and $X_t \left| \boldsymbol{Y}_{-\infty}^{t} \right.$.

I first prove that the ergodic distribution of the approximate discrete Markov chain converges weakly to that of the original continuous Markov chain, i.e. that $X_{t,M} \overset{d}{\to} X_t$ as $M \to \infty$. Proposition 2 establishes this convergence and provides a bound on the difference between the two distributions as a function of the number of points M.

Define $\mathcal{A}$ as the collection of all continuity sets of X_t. I make one further assumption regarding the approximation quality of the sequence of transition kernels $\{P_{\theta,M}\}$.

(BT) The sequence of approximations $P_{\theta,M}$ satisfy

$$\sup_{\theta\in\Theta}\sup_{x\in\mathcal{X}} \left\| P_{\theta,M}(x,\cdot) - P_\theta(x,\cdot) \right\|_{TV} = O\left(h(M)\right) \tag{A.4}$$

where $h(M)$ satisfies $\lim_{M\to\infty} h(M) = 0$.

This assumption allows the practitioner to use *all* of the discretization methods outlined in 4.3 to construct $P_{\theta,M}$. I have chosen to illustrate the case where the Farmer and Toda (2017) method with trapezoidal quadrature rule is used. In this case, assumption (BT) is satisfied with $h(M) = M^{-2/d}$, where d is the dimension of the state space $\mathcal{X}$.[24]

Proposition 2. *Assume (A1)-(A3), (B1)-(B3), and (BT). Then it follows that*

$$\sup_{\theta\in\Theta} \left\| \pi^X_{\theta,M} - \pi^X_\theta \right\|_{TV} = o\left(h^*(M)\right)$$

where $h^*(M)$ satisfies $\lim_{M\to\infty} h^*(M) = 0$. If the transition kernel is approximated as proposed in Farmer and Toda (2017) with a trapezoidal quadrature rule,

$$h^*(M) = M^{-(2-\delta)/d}$$

for any $\delta > 0$.

Note that even faster rates can be achieved through clever choice of the quadrature formula and the assumptions one is willing to make about the smoothness of the likelihood function. By combining Proposition 2 with uniform ergodicity of $X_{t,M}$ and X_t, it can be shown that this approximation error directly translates to probabilities computed under the filtering distributions $X_{t,M} \left| \boldsymbol{Y}_r^t \right.$ and $X_t \left| \boldsymbol{Y}_r^t \right.$.

Lemma 4. *Assume (A1)-(A3), (B1)-(B3), and (BT). Then*

$$\sup_{\theta\in\Theta} \left| \ell_M(\theta) - \ell(\theta) \right| = o\left(h^*(M)\right)$$

Combining Corollary 2, Lemma 2, and Lemma 4 leads to the following pointwise convergence result

[24] For a discussion of error convergence properties see Tanaka and Toda (2015).

Corollary 3. *Assume (A1)-(A3), (B1)-(B3), and (BT). Then for all sequences of initial points* $\{x_{0,M}\}$ *and* $\theta \in \Theta$,

$$\lim_{M,T\to\infty} T^{-1}\ell_{T,M}(\theta, x_{0,M}) = \ell(\theta), \quad \bar{\mathbb{P}}_{\theta^*}\text{-a.s. and in } L^1\left(\bar{\mathbb{P}}_{\theta^*}\right)$$

The final step before I can state the strong consistency result involves showing that $\ell_M(\theta)$ is continuous w.r.t. θ for all $M \in \mathbb{Z}^+$. This will allow me to strengthen Corollary 3 from pointwise convergence to uniform convergence in θ. Note that by (A.3) and the dominated convergence theorem,

$$\ell_M(\theta) = \bar{\mathbb{E}}_{\theta^*}\left[\lim_{r\to\infty} \Delta_{0,r,M,x}(\theta)\right] = \lim_{r\to\infty} \bar{\mathbb{E}}_{\theta^*}\left[\Delta_{0,r,M,x}(\theta)\right]$$

It suffices to show that $\Delta_{0,r,M,x}(\theta)$ is continuous w.r.t θ, since $\{\Delta_{0,r,M,x}(\theta)\}_{r\geq 0}$ is a uniform Cauchy sequence $\bar{\mathbb{P}}_{\theta^*}$-a.s. which is uniformly bounded in $L^1\left(\bar{\mathbb{P}}_{\theta^*}\right)$.

The following additional assumptions are needed to establish continuity

(A4) For all $x, x' \in \mathcal{X}$ and all $y' \in \mathcal{Y}$, $\theta \mapsto q_\theta(x, x')$ and $\theta \mapsto g_\theta(y' | x)$ are continuous.

(B4) For all $M \in \mathbb{Z}^+$, $x \in \mathcal{X}_M$, and $A \in \mathcal{B}(\mathcal{X}_M)$, $\theta \mapsto P_{\theta,M}(x, A)$ is continuous.

Lemma 5. *Assume (A1)-(A4), (B1)-(B4), and (BT), then*

$$\lim_{\delta\to 0} \bar{\mathbb{E}}_{\theta^*}\left[\sup_{M\in\mathbb{Z}^+} \sup_{|\theta'-\theta|\leq\delta} \left|\Delta_{t,\infty,M}(\theta') - \Delta_{t,\infty,M}(\theta)\right|\right] = 0.$$

A direct consequence of Lemma 5 is that the convergence established in Corollary 3 can be strengthened to uniform convergence in $\theta \in \Theta$.

Proposition 3. *Assume (A1)-(A4), (B1)-(B4), and (BT). Then*

$$\lim_{M,T\to\infty} \sup_{\theta\in\Theta} \sup_{x_{0,M}\in\mathcal{X}_M} \left|T^{-1}\ell_{T,M}(\theta, x_{0,M}) - \ell(\theta)\right| = 0, \quad \bar{\mathbb{P}}_{\theta^*}\text{-a.s.}$$

The last assumption needed to establish consistency is an identification assumption guaranteeing that θ^* is a unique maximizer of the likelihood function

(A5) $\theta = \theta^*$ if and only if

$$\bar{\mathbb{E}}_{\theta^*}\left[\log \frac{\bar{p}_{\theta^*}\left(\mathbf{Y}_1^t\right)}{\bar{p}_\theta\left(\mathbf{Y}_1^t\right)}\right] = 0 \qquad \text{for all } t \geq 1. \tag{A.5}$$

This is a high level assumption about the identification of the model. In general, this is a difficult condition to verify because it relies on the ergodic distribution of the joint Markov chain $\{Z_t\}$. For a more thorough discussion on when this assumption is satisfied in the context of HMM, see Douc et al. (2011).

B Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments

This appendix briefly summarizes the method for discretitzing stochastic processes proposed in Farmer and Toda (2017).

Consider the time-homogeneous first-order Markov process

$$\mathbb{P}(X_t \le x' | X_{t-1} = x) = F(x' \,|x\,),$$

where X_t is the random vector of state variables and $F(\cdot\,|x\,)$ is a cumulative distribution function (CDF) that determines the distribution of $X_t = x'$ given $X_{t-1} = x$. The dynamics of any Markov process are completely characterized by its Markov transition kernel. In the case of a discrete state space, this transition kernel is simply a matrix of transition probabilities, where each row corresponds to a conditional distribution. One can discretize the continuous process X_t by applying the Tanaka and Toda (2013) method to each conditional distribution separately.

More concretely, suppose that one has a set of grid points $D_M = \{x_m\}_{m=1}^M$ and an initial coarse approximation $Q = (q_{mm'})$, which is an $M \times M$ probability transition matrix. Additionally, suppose one wants to match some conditional moments of X_t, represented by the moment defining function $T(x)$. The exact conditional moments when the current state is $X_{t-1} = x_m$ are

$$\bar{T}_m = \mathbb{E}\left[T(X_t)\,|X_{t-1} = x_m\,\right] = \int T(x) dF(x\,|x_m\,),$$

where the integral is over x, fixing $X_{t-1} = x_m$. (If these moments do not have explicit expressions, highly accurate quadrature formulas can be used to compute them.) By Theorem 2.1 in Farmer and Toda (2017), these moments can be matched exactly by solving the optimization problem

$$\begin{aligned} &\min_{\{p_{mm'}\}_{m'=1}^M} && \sum_{m'=1}^M p_{mm'} \log \frac{p_{mm'}}{q_{mm'}} \\ &\text{subject to} && \sum_{m'=1}^M p_{mm'} T(x_{m'}) = \bar{T}_m, \quad \sum_{m'=1}^M p_{mm'} = 1, \; p_{mm'} \ge 0 \end{aligned} \tag{B.1}$$

for each $m = 1, 2, \dots, M$, or equivalently the dual problem

$$\min_{\lambda \in \mathbb{R}^L} \sum_{m'=1}^M q_{mm'} \mathrm{e}^{\lambda'(T(x_{m'}) - \bar{T}_m)}. \tag{B.2}$$

(B.2) has a unique solution if and only if the regularity condition

$$\bar{T}_m \in \text{int co } T(D_M) \tag{B.3}$$

holds. Furthermore, if the dual problem has a unique solution λ_m, then the solution to the primal problem (B.1) is given by

$$p_{mm'} = \frac{q_{mm'} e^{\lambda_m' \left(T(x_{m'}) - \bar{T}_m\right)}}{\sum_{m'=1}^{M} q_{mm'} e^{\lambda_m' \left(T(x_{m'}) - \bar{T}_m\right)}} \tag{B.4}$$

Lastly, define the errors associated with the moment matching as:

$$\epsilon_m \equiv \sum_{m'=1}^{M} p_{mm'} T\left(x_{m'}\right) - \bar{T}_m \tag{B.5}$$

The procedure for constructing the finite-state Markov chain approximation to X_t is summarized in Algorithm 2 below.

Algorithm 2: Discretization of Markov processes

1 Select a discrete set of points $D_M = \{x_m\}_{m=1}^{M}$ and an initial approximation $Q = (q_{mm'})$.
2 Select a moment defining function $T(x)$ and corresponding exact conditional moments $\left\{\bar{T}_m\right\}_{m=1}^{M}$. If necessary, approximate the exact conditional moments with highly accurate numerical integrals. Set $m \rightsquigarrow 1$ and define an error tolerance $\kappa > 0$.
3 Solve minimization problem (B.2) and store the resulting solution λ_m.
4 Compute ϵ_m using (B.5). If $\|\epsilon_m\|_\infty < \kappa$, move to step 5. If not, select a smaller set of moments to match and return to step 3.
5 Compute the conditional probabilities corresponding to row m of $P = (p_{mm'})$ using (B.4). Set $m \rightsquigarrow m + 1$. If $m \leq M$, move to step 3, otherwise move to step 6.
6 Collect the computed conditional probability measures in the matrix $P = (p_{mm'})$.

The resulting finite-state Markov chain approximation to X_t takes values in the set D_M and has associated transition matrix P. Since the dual problem (B.2) is an unconstrained convex minimization problem with a typically small number of variables, standard Newton type algorithms can be applied. Furthermore, since the probabilities (B.4) are strictly positive by construction, the transition probability matrix $P = (p_{mm'})$ is a strictly positive matrix, so the resulting Markov chain is stationary and uniformly ergodic by construction.

C Monte Carlo Example Details

In this appendix, I provide additional details and discussion of the Monte Carlo results from section 6.

C.1 Stochastic Volatility: Verifying Theoretical Assumptions

In this section, I verify the theoretical assumptions needed to establish consistency of the MLE in a version of the stochastic volatility model with bounded support. Namely, I consider the following modification to the model

$$\begin{aligned} X_t|X_{t-1} &\sim TN(\mu(1-\rho)+\rho X_{t-1}, \sigma^2, \underline{x}, \bar{x}) \\ Y_t|X_t &\sim N(0, e^{X_t}) \end{aligned}$$

where $\underline{x}$ and $\bar{x}$ are the lower and upper bounds for X_t (the log variance) respectively with $\underline{x} > -\infty$ and $\bar{x} < \infty$. If these bounds are sufficiently below and above the unconditional mean μ the model will generate data that is numerically indistinguishable from the model with unbounded support for X_t. The transition and measurement densities are given by

$$\begin{aligned} q_\theta(x'|x) &= \frac{\phi\left(\frac{x'-\mu(1-\rho)-\rho x}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{\bar{x}-\mu(1-\rho)-\rho x}{\sigma}\right) - \Phi\left(\frac{\underline{x}-\mu(1-\rho)-\rho x}{\sigma}\right)\right)} \\ g_\theta(y|x) &= \frac{1}{e^{x/2}}\phi\left(\frac{y}{e^{x/2}}\right) \end{aligned}$$

Assumption (A1): (a) For all $x \in [\underline{x}, \bar{x}]$ and $\theta \in \Theta$ (for Θ compact), the truncated normal density is strictly bounded away from zero and from infinity. (b) g_θ does not depend on Θ in this example, only on x and thus on the bounds of x. g_θ is a Normal distribution with different variances corresponding to different values of x. Since the variance is bounded away from zero and infinity by the bounds on x, g has a well-defined infimum and supremum which are both strictly positive and less than infinity.

Assumption (A2): Y_t is independent over time conditional on X_t and has a conditional density which is absolutely continuous with respect to Lesbesgue measure. Since X_t is uniformly ergodic and thus trivially positive Harris recurrent, the joint process Z_t is also positive Harris recurrent by example 5.8.2 in Durrett (2019).

Assumption (A3): The quantity b_+ is given by

$$b_+ = \sup_{\theta\in\Theta}\sup_{y_1,x} g_\theta(y_1|x) = \sup_{\theta\in\Theta} g_\theta(0|\bar{x}) < \infty$$

The quantity $b_-(y_1)$ is bounded below by the observation density evaluated at the lowest variance $\underline{x}$

$$b_-(y_1) = \inf_{\theta\in\Theta}\int_{\mathcal{X}} g_\theta(y_1|x)\mu(dx) \geq \inf_{\theta\in\Theta}\inf_{x\in\mathcal{X}} g_\theta(y_1|x) = g_\theta(y_1|\underline{x})$$

and thus it follows that

$$\begin{aligned}\overline{\mathbb{E}}_{\theta^*}[|\log b_-(y_1)|] &= \overline{\mathbb{E}}_{\theta^*}\left[\left|\log \frac{1}{e^{x/2}}\phi\left(\frac{y_1}{e^{x/2}}\right)\right|\right] \\ &= \overline{\mathbb{E}}_{\theta^*}\left[\left|-\frac{x}{2}-\frac{1}{2}\log(2\pi)-\frac{y_1^2}{2\underline{x}}\right|\right] \\ &\leq \frac{x}{2}+\frac{1}{2}\log(2\pi)+\frac{1}{2\underline{x}}\overline{\mathbb{E}}_{\theta^*}\left[y_1^2\right] \\ &< \infty\end{aligned}$$

Assumption (B1): All (B) assumptions will hold for any of the discussed methods for discretization. As a specific example, consider choosing probabilities proportional to the transition density (corresponding to the point mass filter), then the ratio of any two transition probabilities will always be the ratio of the transition density at some two points. By assumption (A1) this ratio will always be strictly positive since the transition density is bounded away from zero below and from infinity above.

Assumption (B3): Since every row of $P_{\theta,M}$ is a valild probability mass function and thus sums to 1, the same argument used to verify the second half of assumption (A3) can be used.

Assumption (BT): The point mass filter's conditional probabilities are a type of Riemann sum integration rule approximation to the exact transition probabilities, which has integration errors of the order M^{-2} where M is the number of points being used in the Riemann sum.

C.2 Stochastic Volatility: Filtered States

Another important dimension for comparison is the accuracy of the filtered states, $\left\{\hat{x}_{t|t}\right\}_{t=1}^{T}$. I provide results on the root mean square error (RMSE) and the mean absolute error (MAE) of all the methods. For a given model specification and method, these are defined as:

$$\text{RMSE} = \left(\frac{1}{T}\sum_{t=1}^{T}\left(\hat{x}_{t|t}-x_t\right)^2\right)^{1/2} \tag{C.1}$$

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{x}_{t|t}-x_t\right| \tag{C.2}$$

I define the average RMSE (ARMSE) and average MAE (AMAE) to be the average of the RMSE and the MAE across simulations for a given method. Table 5 displays the ARMSE and AMAE of each method, where the filtering is done using the corresponding maximum likelihood estimates of the parameters for a given sample.

The DF and BPF perform roughly the same for all sample sizes, with the APF performing slightly worse. However, keep in mind that this is for dramatically

Table 5: Accuracy of Filtered State Estimates

Average Root Mean Squared Error

	Discretization Filter						BPF	APF	EKF
ROT constant *c*	1/2	1	3	5	7	10	-	-	-
T = 100	0.365	0.361	0.359	0.358	0.358	0.358	0.356	0.360	0.481
T = 500	0.373	0.372	0.372	0.372	0.372	0.372	0.375	0.390	0.458
T = 1,000	0.376	0.376	0.375	0.375	0.376	0.375	0.377	0.393	0.453

Average Mean Absolute Error

	Discretization Filter						BPF	APF	EKF
ROT constant *c*	1/2	1	3	5	7	10	-	-	
T = 100	0.296	0.294	0.293	0.292	0.292	0.292	0.290	0.294	0.377
T = 500	0.299	0.298	0.298	0.298	0.298	0.298	0.300	0.312	0.366
T = 1,000	0.300	0.300	0.300	0.300	0.300	0.300	0.301	0.314	0.362

different estimation times for the parameters as discussed in section 6.2. The misspecification of the measurement error distribution using the EKF translates into poor estimates of the unobserved state.

D DSGE Model Estimation

This appendix provides some additional details for the estimation of the New-Keynesian model in section 7. Table 6 lists the parameters that are fixed during the estimation. Most of the parameters are calibrated exactly the same as in Aruoba et al. (2018). The only differences are that r^* and π^* are chosen to reflect the long term averages of real interest rates and inflation incorporating the last ten years of data. R_0, which controls the curvature of the smooth approximation to the kink in the Taylor Rule, is set to $1 + 10^{-20}$, which results in a function which is incredibly close to a kink, but still differentiable.

Table 7 lists the prior distributions for each of the estimated parameters and their corresponding parameters. These are chosen to be exactly the same as in Aruoba et al. (2018).

Table 6: Parameters Fixed During Estimation

Parameter	Value
$400\ln(r^*)$	0.86
$400\ln(\pi^*)$	2.35
$100\ln(\gamma)$	0.38
g	1.54
ν	0.1
η	0.72
α	0.9
R_0	$1 + 10^{-20}$

Table 7: Priors for Estimation

Parameter	Description	Density	$P(1)$	$P(2)$
τ	Inverse IES	$\mathcal{G}$	2.0	0.25
κ	Slope (linearized) Phillips curve	$\mathcal{G}$	0.3	0.1
ψ_1	Taylor rule: weight on inflation	$\mathcal{G}$	1.5	0.3
ψ_2	Taylor rule: weight on output	$\mathcal{G}$	0.5	0.25
ρ_R	Interest rate smoothing	$\mathcal{B}$	0.5	0.2
ρ_z	Persistence: technology shock	$\mathcal{B}$	0.2	0.1
ρ_d	Peristence: discount factor shock	$\mathcal{B}$	0.8	0.1
ρ_g	Peristence: demand shock	$\mathcal{B}$	0.8	0.1
$100\sigma_R$	Std dev: monetary policy shock	$\mathcal{IG}$	0.3	4.0
$100\sigma_z$	Std dev: technology shock	$\mathcal{IG}$	0.4	4.0
$100\sigma_d$	Std dev: discount factor shock	$\mathcal{IG}$	0.4	4.0
$100\sigma_d$	Std dev: demand shock	$\mathcal{IG}$	0.4	4.0