Estimation of Hidden Markov Models with Nonparametric Simulated Maximum Likelihood

By Dennis Kristensen and Yongseok Shin*

February 2006

Abstract

We propose a nonparametric simulated maximum likelihood estimation (NPSMLE) with built-in nonlinear filtering. By recursively approximating the unknown conditional densities, our method enables a maximum likelihood estimation of general dynamic models with latent variables—including time-inhomogeneous and non-stationary processes. We establish the asymptotic properties of the NPSMLEs for hidden Markov models, and then demonstrate the usefulness of our proposed method with Monte Carlo studies.

*Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA; dkristen@ssc.wisc.edu, yshin@ssc.wisc.edu.

1 Introduction

We propose a general method to estimate dynamic models with latent variables by maximum likelihood. This method incorporates nonlinear filtering into the nonparametric simulated maximum likelihood estimation (NPSMLE) of Kristensen and Shin (2006). The estimator is highly general in the sense that it can be implemented for basically any dynamic model from which one can simulate.

The procedure is a recursion of two steps: The first step is to estimate the unknown conditional densities from simulated observations by kernel methods. The second is to simulate the next set of artificial observations from the estimated densities. The conditional density will converge to the true density as the number of artificial observations (N) goes to infinity, so we can approximate the density arbitrarily well by choosing a sufficiently large N. Given this simulated density, we obtain a simulated maximum likelihood estimator of the model parameters.

We give sufficient conditions for the simulated MLE to be consistent and have the same asymptotic distribution as the infeasible MLE. The latter result appears to be the first of its kind for dynamic latent variable models. Previous studies on simulated MLE of latent dynamic models have at most shown consistency.

Our method is related to simulated methods of moments (SMM) (Duffie and Singleton, 1993) and indirect inference (Gallant and Tauchen, 1996; Gouriéroux et al., 1993; Smith, 1993), but, unlike these approaches, is not subject to the arbitrariness involved in the selection of target moments or auxiliary models. Furthermore, under weak regularity conditions, maximum likelihood estimators enjoy higher efficiency than these estimators, and the SMLEs will inherit these properties.

Related nonparametric simulation-based methods are considered in Fermanian and Salanié (2004) and Altissimo and Mele (2005). The former however only considers the case of fully observable stationary processes, while the latter restricts itself to cross-sectional models where latent variables easily can be dealt with. In comparison, our method allows for potentially non-stationary processes, including time-inhomogeneous dynamics, and utilizes filtering procedures to deal with latent variables. In this regard, our method is closely related to particle filtering (Fernández-Villaverde and Rubio-Ramírez, 2004; Kim et al., 1998; Pitt and Shephard, 1999).

Our method however has several advantages compared to the particle filtering approach: One advantage is that the our simulated likelihood function is smooth in its parameters, which helps us characterize asymptotic properties of the estimators. Furthermore, smoothness is desirable in numerical maximizations since most optimization routines depend on this property. Another is the simplicity of the proposed estimation method, which facilitates the implementation of it; no advanced programming skills are required to apply the method.

On the other hand, one disadvantage of our proposed method is the bias incurred by using kernel methods to estimate the density for a finite number of simulations. However, one can simulate one's way out of this problem by drawing a sufficiently large number of data points—that is, by letting $N \to \infty$.

The paper is organized as follows. In the next section, we set up our framework and present the approximate density and the associated NPSMLE for hidden Markov models. In Section 3, we derive the asymptotics of the NPSMLE under regularity condition. Section 4 presents a specific example where NPSMLE can be used with advantage.

2 NPSMLE for hidden Markov models

Consider a Markov process $w_t = (y_t, z_t) \in \mathbb{R}^d$ where $y_t \in \mathbb{R}^{d_1}$ and $z_t \in \mathbb{R}^{d_2}$ such that $d = d_1 + d_2$. The process belongs to a parametric family given by $\{p_t(v|w;\theta) : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^k$. That is, there exists a true, but unknown, value $\theta_0 \in \Theta$ such that

$$P(w_t \in A | w_{t-1} = w) = \int_A p_t(v|w;\theta_0)\, dv.$$

We are interested in estimating the parameter θ_0. However, only the component y_t is observed such that z_t is a latent variable, and we therefore have to base our estimation method on the observations $\{y_t : t = 1, ..., T\}$.

A natural choice of estimator is the MLE. Suppose that $z_0 \in \mathbb{R}^{d_2}$ is known, although z_t for $t \geq 1$ is unobservable. The (conditional) likelihood can then be written as

$$L_T(\theta) = \log p_T(y_T, y_{T-1}, \ldots, y_1 | y_0, z_0; \theta) = \sum_{t=1}^{T} \log p_t(y_t | x_{t-1}; \theta)$$

where $x_t = (y_t, ..., y_0, z_0)$, and the 2nd equality follows by the standard conditioning argument. The actual MLE is then defined as

$$\tilde{\theta} = \arg\sup_{\theta \in \Theta} L_T(\theta).$$

Unfortunately, the conditional density for the observables is in most cases not available on closed form; $p_t(y_t|x_{t-1};\theta)$ is unknown. Instead, we here propose to combine simulations with nonparametric kernel methods to construct a simulated version of the likelihood. This can then be used to obtain a simulated MLE. Below, we propose three alternative methods to do this, all based on the kernel density method. The two first suffer from certain disadvantages which will lead us to focus on the third one in the following sections.

In the first method, we simulate N trajectories recursively, $\left(Y_{t,i}^{\theta,x_0}, Z_{t,i}^{\theta,x_0}\right) \sim p_t\left(\cdot,\cdot|Y_{t-1,i}^{\theta,x_0}, Z_{t-1,i}^{\theta,x_0};\theta\right)$, $t = 1,...,T$, $i = 1,...,N$, with $(Y_{0,i}^{\theta,x_0}, Z_{0,i}^{\theta,x_0}) = x_0$.This can be done such that the trajectories are mutually independent. Since by construction, $\left(Y_{T,i}^{\theta,x_0},....,Y_{1,i}^{\theta,x_0}\right) \sim p_T(y_T, y_{T-1},\ldots,y_1|x_0;\theta)$, we can use these to calculate a nonparametric kernel density estimator of p_T:

$$\hat{p}_T\left(y_T,\ldots,y_1|x_0;\theta\right) = \frac{1}{N}\sum_{i=1}^{N}\prod_{t=1}^{T}K_h\left(Y_{t,i}^{\theta,x_0} - y_t\right), \tag{1}$$

where $K : \mathbb{R}^{d_1} \mapsto \mathbb{R}$ is a kernel and $h > 0$ a bandwidth; see **?** for an introduction to these. Observe however that this will suffer from a severe curse of dimensionality: Under regularity conditions,

$$\hat{p}_T\left(y_T,\ldots,y_1|x_0;\theta\right) - p_T\left(y_T,\ldots,y_1|x_0;\theta\right) = O_P(1/\sqrt{Nh^{d_1T}}) + O_P\left(h^2\right),$$

where the first component is the variance term and the second the bias term. Note that the variance term grows exponentially with the number of observations T and as such $\hat{p}_T$ will be a not very precise version of the actual density. So a large number of simulations have to be performed to obtain a sufficient degree of accuracy. The deterioration in the convergence rate mirrors the fact that at time t the simulations are performed conditional only on the information available at time 0, x_0, not utilizing the information contained in $(y_{t-1},...,y_1)$.

Instead of using the likelihood for estimation, an alternative could be to use the criterion function given by

$$Q_T(\theta) = \sum_{t=1}^{T}\log p_t(y_t|x_{t-1}^{(q)};\theta), \quad x_{t-1}^{(q)} = (y_{t-1},...,y_{t-q}),$$

for some fixed $q \geq 1$, and then estimate θ_0 by the maximizer of this. The density $p_t(y_t|x_{t-1}^{(q)};\theta)$ could be approximated by simulating N i.i.d. samples of length T as in (1), and then calculate

$$\hat{p}_t\left(y|x;\theta\right) = \frac{\sum_{i=1}^{N}K_h\left(Y_{t,i}^{\theta} - y\right)K_h\left(X_{t-1,i}^{q,\theta} - x\right)}{\sum_{i=1}^{N}K_h\left(X_{t-1,i}^{q,\theta} - x\right)}, \tag{2}$$

where $X_{t-1,i}^{q,\theta} = \left(Y_{t-1,i}^{\theta}, ..., Y_{t-q,i}^{\theta}\right)$. Now the variance component of the simulation error is $O_P(1/\sqrt{Nh^{d_1(1+q)}})$, and as such fixed as $T \to \infty$. But the identification of θ is not clear since we only condition on a subset of the information available at time $t \geq 1$, and the estimator will not in general achieve MLE efficiency. Note that for this estimator, instead of simulating N i.i.d. trajectories, one could simulate one long trajectory, $\left\{\left(Y_t^{\theta}, Z_t^{\theta}\right) : t = 1, ..., \tilde{N}\right\}$ with $\left(Y_t^{\theta}, Z_t^{\theta}\right) \sim p_t\left(\cdot, \cdot | Y_{t-1}^{\theta}, Z_{t-1}^{\theta}\right)$, and then use this to calculate

$$\tilde{p}\left(y|x;\theta\right) = \frac{\sum_{t=1}^{\tilde{N}} K_h\left(Y_t^{\theta} - y\right) K_h\left(X_{t-1,i}^{q,\theta} - x\right)}{\sum_{t=1}^{\tilde{N}} K_h\left(X_{t-1,i}^{q,\theta} - x\right)}. \tag{3}$$

This however requires the process to be stationary, and the resulting density estimator will most likely be less precise; see Kristensen and Shin (2006) for further discussion of this issue.

Given the drawbacks of the two methods outlined above, we now propose a recursive method to construct a simulated version of the likelihood, $\hat{L}_T(\theta)$, which does not suffer from the dimension problem as in the estimator in (1). This is done by *iteratively* calculating kernel-based estimates of $p_t\left(y_t|x_{t-1};\theta\right)$, $t = 1, \ldots, T$. In the following, we suppress the dependence on $\theta \in \Theta$ since it is kept fixed. The procedure works as follows:

1. Given $x_0 = (y_0, z_0)$ and θ, we can simulate $Z_{1,i}^{x_0}$ and $Y_{1,i}^{x_0}$ for $i = 1, \ldots, N$. From these simulated values, we can estimate $p_1(y_1|x_0)$ and $p_1(z_1|y_1, x_0)$ by

$$\hat{p}_1(y_1|x_0) = \frac{1}{N}\sum_{i=1}^{N} K_h(Y_{1,i}^{x_0} - y_1),$$

and

$$\hat{p}_1(z_1|y_1, x_0) = \frac{\sum_{i=1}^{N} K_h(Z_{1,i}^{x_0} - z_1) K_h(Y_{1,i}^{x_0} - y_1)}{\sum_{i=1}^{N} K_h(Y_{1,i}^{x_0} - y_1)}.$$

2. We draw $Z_{1,i}^{x_1}$ for $i = 1, \ldots, N$ from $\hat{p}_1(z_1|x_1) = \hat{p}_1(z_1|y_1, x_0)$. Given the Markov property, for each i, generate $Y_{2,i}^{x_1}$ and $Z_{2,i}^{x_1}$ from $p_2(\cdot, \cdot|y_1, Z_{1,i}^{x_1})$. Kernel estimation gives us:

$$\begin{aligned} \hat{p}_2(y_2|x_1) &= \frac{1}{N}\sum_{i=1}^{N} K_h(Y_{2,i}^{x_1} - y_2), \\ \hat{p}_2(z_2|x_2) &= \frac{\sum_{i=1}^{N} K_h(Z_{2,i}^{x_1} - z_2) K_h(Y_{2,i}^{x_1} - y_2)}{\sum_{i=1}^{N} K_h(Y_{2,i}^{x_1} - y_2)}. \end{aligned}$$

3. We draw $Z_{2,i}^{x_2}$ for $i = 1, \ldots, N$ from $\hat{p}_2(z_2|x_2)$. Given the Markov property, for each i, generate $Y_{3,i}^{x_2}$ and $Z_{3,i}^{x_2}$ from $p_3(\cdot,\cdot|y_2, Z_{2,i}^{x_2})$. Kernel estimation gives us:

$$\begin{aligned}\hat{p}_3(y_3|x_2) &= \frac{1}{N}\sum_{i=1}^{N} K_h(Y_{3,i}^{x_2} - y_3),\\ \hat{p}_3(z_3|x_3) &= \frac{\sum_{i=1}^{N} K_h(Z_{3,i}^{x_2} - z_3)K_h(Y_{3,i}^{x_2} - y_3)}{\sum_{i=1}^{N} K_h(Y_{3,i}^{x_2} - y_3)}.\end{aligned}$$

Once we have obtained the simulated densities, these can be used to calculate a simulated version of the likelihood and then define the NPSMLE as

$$\begin{aligned}\hat{\theta} &= \arg\sup_{\theta\in\Theta} \hat{L}_T(\theta),\\ \hat{L}_T(\theta) &= \sum_{t=1}^{T} \log \hat{p}_t(y_t|x_{t-1};\theta).\end{aligned}$$

The key component in our iterative scheme is that the simulations at time t are done conditional on all (observed) information at time $t-1$, x_{t-1}. That is, we simulate $\left(Y_t^{x_{t-1}}, Z_t^{x_{t-1}}\right)$. This is in contrast to the alternative methods in (1) and (2). In the case of (1), at time t, the draw $(Y_t^{x_0}, Z_t^{x_0})$ is done conditional on the much smaller information set x_0; the iterative scheme thereby increases the precision in our simulations and we can reduce the number of simulations used in each step. For the estimator in (2) only a subset, $x_{t-1}^{(q)} = (y_{t-1}, ..., y_{t-q})$, of the full information is utilized; the iterative scheme utilizes the full set of information and thereby increases the efficiency of the estimator.

On the other hand, in the iterative scheme, each step involves an additional error since we use $\hat{p}_t(z_t|x_t)$ instead of (the unknown) $p_t(z_t|x_t)$ to draw $Z_t^{x_t}$. So in addition to the estimation error incurred by the kernel method, we also face a simulation error. This error will accumulate as $T \to \infty$, but the cumulative approximation error will only grow linearly in T in contrast to the estimator in (1) where the simulation error increases at an exponential rate.

Many latent variables models take the form

$$\begin{aligned}y_t &= g\left(y_{t-1}, z_t, \varepsilon_t\right),\\ z_t &= h\left(z_{t-1}, \eta_t\right),\end{aligned}$$

where $\{(\varepsilon_t, \eta_t)\}$ are i.i.d. and mutually independent, and the density $p_t\left(y_t|z_t, x_{t-1}\right) = p_t\left(y_t|z_t, y_{t-1}\right)$ is known. In this case, an alternative, more precise, estimate of $p_t\left(z_t|x_t\right)$

can be obtained by Bayes' Rule,

$$\hat{p}_t(z_t|x_t) = \frac{p_t(y_t|z_t, y_{t-1})\hat{p}_t(z_t|x_{t-1})}{\hat{p}_t(y_t|x_{t-1})},$$

where

$$\hat{p}_t(z_t|x_{t-1}) = \sum_{i=1}^{N} K_h(Z_{t,i}^{x_{t-1}} - y_t), \quad \hat{p}_t(y_t|x_{t-1}) = \sum_{i=1}^{N} K_h(Y_{t,i}^{x_{t-1}} - y_t).$$

Finally, we note that discrete components, found for example in Markov switching models, can be accommodated for here by modifying the kernel; see Kristensen and Shin (2006) for more details.

3 Asymptotic Properties of the NPSMLE

In this section we show that under regularity conditions on the parametric model, the NPSMLE $\hat{\theta}$ will have the same asymptotic properties as the infeasible estimator $\tilde{\theta}$ for a suitably chosen sequence $N = N(T)$ and $h = h(N)$.

Our proof proceeds in two steps: We first show that $\hat{p}_t(y_t|x_{t-1};\theta) \to^P p_t(y_t|x_{t-1};\theta)$ in a suitable sense as $N \to \infty$ and $h \to 0$. Given this result, we are able to demonstrate that the NPSMLE $\hat{\theta}$ has the same (first order) asymptotic properties as the infeasible estimator $\tilde{\theta}$ for a suitably chosen sequence $N = N(T)$ and $h = h(N)$. The second step is shown by applying the set of general results regarding simulated estimators found in Kristensen and Shin (2006).

In order to show the two above steps, we have to modify the recursive simulation strategy and the simulated version of the likelihood slightly.. In particular, we introduce a trimming device allowing us to control the tail behaviour of the density estimator. Let $\tau_a(\cdot)$ be a continuously differentiable trimming function satisfying $\tau_a(x) = 1$ if $|x| > a$, and 0 if $|x| < a/2$, and $a = a(N) \to 0$ be a trimming sequence. By letting $a \to 0$ sufficiently fast, the trimming will have no effect asymptotically.

We first need to modify our simulation scheme: At step $t \geq 1$, we draw $Z_{t,i}^{x_t}$ from the following trimmed version of $\hat{p}_t(z_t|x_t)$ instead of $\hat{p}_t(z_t|x_t)$ itself,

$$\begin{aligned}
\tilde{p}_t(z_t|x_t;\theta) &= \frac{\hat{p}_t(z_t|x_t)\,\tau_a\{\hat{p}_t(y_t, z_t|x_{t-1})\}}{\int \hat{p}_t(z|x_t)\,\tau_a\{\hat{p}_t(y_t, z_t|x_{t-1})\}\,dz} \qquad (4)\\
&= \frac{\hat{p}_t(y_t, z_t|x_{t-1})\,\tau_a\{\hat{p}_t(y_t, z_t|x_{t-1})\}}{\int \hat{p}_t(y_t, z|x_{t-1})\,\tau_a\{\hat{p}_t(y_t, z|x_{t-1})\}\,dz}.
\end{aligned}$$

This corresponds to first drawing from $\hat{p}_t(z_t|x_t)$, and then discarding all simulated values outside a compact (but growing) set defined in terms of the trimming function. Except for this alteration, the recursive scheme outlined in the previous section remains the same. Second, we also trim the simulated log-likelihood function. So we redefine $\hat{L}_T(\theta)$ as

$$\hat{L}_T(\theta) = \sum_{t=1}^{T} \tau_a\left(\hat{p}(y_t|x_{t-1};\theta)\right) \log \hat{p}(y_t|x_{t-1};\theta),$$

where τ is the same trimming function.

In Kristensen and Shin (2006), a trimming device is also used, but only in the calculation of the simulated likelihood. Here, we employ trimming both in the likelihood evaluation and in the simulation. The latter is due to the fact that we here, in contrast to the fully observed case, cannot simulate $Z_{t,i}^{x_t}$ perfectly, and instead draw from an estimated density. The trimming of the simulated values enables us to control the additional error arising from the imperfect simulation.

Let $W_t^{w,\theta}$ denote a random variable drawn from the transition density of w_t, $p_t(\cdot|w;\theta)$. We then impose the following regularity conditions on the model.

A.1 There exist a Λ_t with $\mathbb{E}\Lambda_t^2 < \infty$ and $\beta_1 \geq 0$ such that $\|W_t^{w,\theta} - W_t^{w',\theta'}\| \leq \Lambda_t\left[\|w-w'\|^{\beta_1} + \|\theta-\theta'\|^{\beta_1}\right]$ for all $\theta, \theta' \in \Theta$ and $w, w' \in \mathbb{R}^l$.

A.2 $\theta \mapsto Y^{x,\theta}$ is differentiable with its derivative, $\dot{Y}^{x,\theta}$, satisfying

$$\|\dot{Y}_t^{x,\theta} - \dot{Y}_t^{x',\theta'}\| \leq \Lambda_t\left[\|x-x'\|^{\beta_1} + \|\theta-\theta'\|^{\beta_1}\right], \quad \|\dot{Y}_t^{x,\theta}\|^2 \leq \Lambda_t\|x\|^{\beta_2},$$

for all $\theta, \theta' \in \Theta$ and $x, x' \in \mathbb{R}^l$, where $\beta_2 \geq 0$.

A.3 The density $p_t(v|w;\theta)$ is $r \geq 2$ times continuously differentiable w.r.t. v with bounded derivatives such that

$$\max_{t\geq 1} \sup_{\theta\in\Theta} \sup_{(v,w)\in\mathbb{R}^{2d}} \sum_{|\lambda|=r} \left|D_v^{\lambda} p_t(v|w;\theta)\right| < \infty.$$

A.4 The density $p_t(v|w;\theta)$ is continuous w.r.t. θ.

A.5 The density $p_t(v|w;\theta)$ is thrice continuously differentiable w.r.t. θ.

Uniform convergence of $\hat{p}$ over x and θ are established under (A.1)-(A.3) using higher-order kernels as defined below. The remaining two assumptions give us a smooth likelihood function which in turn enables us to use standard Taylor expansion arguments.

We impose the following conditions on the actual MLE to obtain consistency of $\hat{\theta}$:

C.1 $\Theta \subseteq \mathbb{R}^d$ is compact.

C.2 $\tilde{\theta} \xrightarrow{P} \theta_0$.

C.3 There exists a sequence $\bar{L}_T > 0$ such that:

1. $L_T(\theta) / \bar{L}_T$ is stochastically equicontinuous.
2. $\sup_{\theta \in \Theta} \sum_{t=1}^{T} |\log p_t(y_t|x_{t-1};\theta)|^{1+\delta} / \bar{L}_T = O_P(1)$, for some $\delta > 0$.
3. $\sum_{t=1}^{T} \|y_t\|^{1+\delta} / \bar{L}_T = O_P(1)$ and $\sum_{t=1}^{T} \mathbb{E}\|\Lambda_t\| / \bar{L}_T = O_P(1)$, for some $\delta > 0$.

The condition (C.3) is used to obtain uniform convergence of the simulated likelihood over Θ. In the stationary case, $\bar{L}_T$ can be chosen as $\bar{L}_T = T$ in which case (C.3.1)-(C.3.3) will follow by the LLN under suitable moment conditions.

In order to show that $\hat{\theta}$ has the same asymptotic distribution as $\tilde{\theta}$, additional assumptions are needed. They basically require that the actual MLE in fact has an asymptotic distribution:

N.1 $\theta_0 \in \text{int}\Theta$.

N.2 $\mathcal{I}_T^{-1} = \mathcal{I}_T^{-1}(\theta_0) \to 0$.

N.3 $W_{j,T}(\theta) = O_P(1)$ uniformly in a neighborhood of θ_0 for $j = 1, ..., d$.

N.4 $(U_T(\theta_0), V_T(\theta_0)) \xrightarrow{d} (U_0, V_0)$ for some random variables (U_0, V_0) with V_0 being non-singular almost surely.

Under these conditions, $\mathcal{I}_T^{1/2}(\tilde{\theta} - \theta_0) \xrightarrow{d} V_0^{-1} U_0$. Regularity conditions under which (C.1)–(C.3) and (N.1)–(N.4) hold can be found in Bickel and Ritov (1996), Bickel et al. (1998), Douc and Matias (2001), Jensen and Petersen (1999) and Leroux (1992).

The kernel K is assumed to belong to the following class of so-called higher-order or bias-reducing kernels:

K.1 The kernel K satisfies $\int_{\mathbb{R}^{d_1}} K(z)dz = 1$; $\int_{\mathbb{R}^{d_1}} z^{\lambda} K(z)dz = 0$, for $1 \leq |\lambda| \leq r-1$; $\int_{\mathbb{R}^{d_1}} \|z\|^r |K(z)| dx < \infty$; $\sup_z [|K(z)| \max(\|z\|, 1)] < \infty$; K is absolutely integrable with a Fourier transform Ψ satisfying $\int_{\mathbb{R}^{d_1}} \left\{(1 + \|z\|) \sup_{b \geq 1} |\Psi(bz)|\right\} dz < \infty$.

Finally, we restrict the class of permissible bandwidths and the number of simulations as follows:

B.L.1 $a^{-T}\sqrt{\log(N)}N^{-1/2}h^{-d_1-1} \to 0$, $h^r a^{-T} \to 0$, and $N^{-2\gamma}\log(a)\,h^{-d_1} \to 0$ for some $\gamma > 0$.

B.L.2 $\bar{L}_T h^{-d_1-1} a^{-T}\sqrt{\log(N)/N} \to 0$, $\bar{L}_T h^r a^{-T} \to 0$, $\bar{L}_T N^{-2\gamma}\log(a)\,h^{-d_1} \to 0$, $\bar{L}_T \log(a)^{-1} \to 0$, $\bar{L}_T N^{-\gamma} \to 0$ for some $\gamma > 0$.

Theorem 1 *Assume that (A.1)–(A.4), (K.1), (C.1)–(C.3) hold. Then $\hat{\theta} \xrightarrow{P} \theta_0$ for any sequences $N \to \infty$, $h \to 0$ satisfying (B.L.1).*

If furthermore (A.5) and (N.1)–(N.4) hold, then $\mathcal{I}_T^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} V_0^{-1}U_0$ for any sequences $N \to \infty$, $h \to 0$ satisfying (B.L.2).

Comparing with the results obtained in Kristensen and Shin (2006), we here require the number of simulations to increase with a higher rate as $T \to \infty$. This is a consequence of the additional error term arising from the imperfect simulation scheme. This additional error seems unavoidable in dynamic models; particle filtering methods suffers from the same drawback. In contrast, this issue is avoided in cross-sectional models due to the independence assumption, c.f. Fermanian and Salanié (2004)

4 Application: stochastic volatility

One of the merits of our approach is generality. In this section, however, we apply NPSMLE to a simple, well-known model for expositional purposes: the discrete-time stochastic volatility model considered in Andersen et al. (1999).

As in Andersen et al. (1999), we consider the following log-normal stochastic autoregressive volatility model for the return series y_t with $t = 1, \ldots, T$:

$$y_t = \sigma_t \varepsilon_t \tag{5}$$

$$\ln \sigma_t^2 = \alpha + \beta \ln \sigma_{t-1}^2 + \sigma_u u_t, \tag{6}$$

where $\theta = (\alpha, \beta, \sigma_u)$ is the parameters of interest and $\{\varepsilon_t, u_t\}$ follows i.i.d. standard normal. It is assumed that $-1 \leq \beta \leq 1$ and $\sigma_u \geq 0$. $\{\sigma_t\}_{t=0}^T$ is the unobserved conditional volatility process. For expositional purposes, we further assume that σ_0 is known. In general, we can treat σ_0 as an unknown parameter to be estimated.

Observe that this model is a hidden Markov model with $z_t = \ln \sigma_{t+1}^2$. Thus, $\hat{L}_T(\theta)$ can constructed as described in Section 2, the only difference being that it is not necessary here to approximate $p(y_t|z_{t-1}, x_{t-1})$ since one can find its analytical form,

$$p(y_t|z_{t-1}, x_{t-1}) = p(y_t|z_{t-1}) = \Phi\left(\frac{y_t}{\sigma_t}\right)/\sigma_t,$$

where Φ is the density of the standard normal distribution. Finally, we maximize $\hat{L}_T(\theta)$ over the relevant parameter space. The mean and root mean squared error of the parameter estimates from 512 Monte Carlo exercises are shown in Table 1, where $T = 1000$. EMM refers to Gallant and Tauchen's efficient method of moments with an auxiliary model of GARCH(1,1) (4 moments). For NPSMLE, $N = 1000$, $h = 0.01$ and $h_\sigma = 0.006$.

	True value	EMM	NPSMLE
α	-0.736	-0.9050 (0.5998)	-0.8705 (0.5332)
β	0.90	0.8808 (0.0792)	0.8817 (0.0602)
σ_u	0.363	0.3786 (0.1782)	0.3732 (0.0895)

Table 1: Estimated parameters of stochastic volatility model

5 Conclusion

We have incorporated an importance sampler into the nonparametric simulated maximum likelihood estimation as proposed by Kristensen and Shin (2006) to explicitly deal with latent variables. Theoretical conditions in terms of the number of simulations and the bandwidth are given ensuring that the NPSMLE inherits the asymptotic properties of the actual MLE. A simulation study demonstrates that the method works well in practice.

A Proof

Proof of Theorem 1 We consider

$$\hat{L}_T(\theta) = \frac{1}{\bar{L}_T}\sum_{t=1}^{T}\tau_{a,t}\log\hat{p}_t(y_t|x_{t-1};\theta), \quad \tilde{L}_T(\theta) = \frac{1}{\bar{L}_T}\sum_{t=1}^{T}\log\bar{p}_t(y_t|x_{t-1};\theta),$$

$$L_T(\theta) = \frac{1}{\bar{L}_T}\sum_{t=1}^{T}\log p_t(y_t|x_{t-1};\theta),$$

where $\bar{L}_T$ is given in (C.3), and $\bar{p}_t(y_t|x_{t-1};\theta)$ is given in Lemma 2. We show that (i) $\sup_{\theta\in\Theta}|\hat{L}_T(\theta) - \tilde{L}_T(\theta)| = o_P(1)$, and (ii) $\sup_{\theta\in\Theta}|\tilde{L}_T(\theta) - L_T(\theta)| = o_P(1)$. To show (i), we split up $\hat{L}_T(\theta) - \tilde{L}_T(\theta)$ into the following four terms,

$$\hat{L}_T(\theta) - \tilde{L}_T(\theta) = \frac{1}{\bar{L}_T}\sum_{t=1}^{T}[\tau_a(\hat{p}_t(y_t|x_{t-1};\theta)) - \tilde{\tau}_{a,t}]\log\hat{p}_t(y_t|x_{t-1};\theta)$$

$$
\begin{aligned}
& +\frac{1}{\bar{L}_T}\sum_{t=1}^{T}\tilde{\tau}_{a,t}\left[\log\hat{p}_t\left(y_t|x_{t-1};\theta\right)-\log\bar{p}_t\left(y_t|x_{t-1};\theta\right)\right] \\
& +\frac{1}{\bar{L}_T}\sum_{t=1}^{T}\left[\tilde{\tau}_{a,t}\log\bar{p}_t\left(y_t|x_{t-1};\theta\right)-\log\bar{p}_t\left(y_t|x_{t-1};\theta\right)\right] \\
& +\frac{1}{\bar{L}_T}\sum_{t=1}^{T}\left[\log\bar{p}_t\left(y_t|x_{t-1};\theta\right)-\log p_t\left(y_t|x_{t-1};\theta\right)\right] \\
:= \quad & B_1\left(\theta\right)+B_2\left(\theta\right)+B_3\left(\theta\right)+B_4\left(\theta\right).
\end{aligned}
$$

Using the same arguments as in the proof of Theorem 2 in Kristensen and Shin (2006),

$$
\begin{aligned}
\left|B_1\left(\theta\right)\right| &\leq \frac{\left|\log\left(a\right)\right|}{\bar{L}_T}\sum_{t=1}^{T}\chi\left\{\left\|x_{t-1}\right\|>N^{\gamma}\right\} \\
&\leq \frac{\left|\log\left(a\right)\right|T\chi\left\{\left\|x_T\right\|>N^{\gamma}\right\}}{\bar{L}_T} \\
&\leq \frac{\sum_{t=1}^{T}\left\|y_t\right\|^{1+\delta}}{\bar{L}_T}\frac{\left|\log\left(a\right)\right|T}{N^{\gamma(1+\delta)}};
\end{aligned}
$$

$$
\begin{aligned}
\left|B_2\left(\theta\right)\right| &\leq \sum_{t=1}^{T}\chi_{A_t(1/4)}\left|\log\hat{p}_t\left(y_t|x_{t-1};\theta\right)-\log\bar{p}_t\left(y_t|x_{t-1};\theta\right)\right| \\
&\leq \frac{T}{\bar{L}_T a}\sup_{\theta\in\Theta}\sup_{y_t\in\mathbb{R}^p}\sup_{\left\|x_{t-1}\right\|\leq N^{\gamma}}\left|\hat{p}_t\left(y_t|x_{t-1};\theta\right)-\bar{p}_t\left(y_t|x_{t-1};\theta\right)\right|;
\end{aligned}
$$

and

$$
\begin{aligned}
\left|B_3\left(\theta\right)\right| &\leq \frac{1}{\bar{L}_T}\sum_{t=1}^{T}\chi\left\{p\left(y_t|x_{t-1};\theta\right)<4a\right\}\left|\log p\left(y_t|x_{t-1};\theta\right)\right| \\
&\quad +\frac{1}{\bar{L}_T}\sum_{t=1}^{T}\chi\left\{p\left(y_{t-1}|x_{t-2};\theta\right)<4a\right\}\left|\log p\left(y_t|x_{t-1};\theta\right)\right| \\
&\quad +\frac{1}{\bar{L}_T}\sum_{t=1}^{T}\chi\left\{\left\|x_{t-1}\right\|>N^{\gamma}\right\}\left|\log p\left(y_t|x_{t-1};\theta\right)\right| \\
&:= B_{3,1}\left(\theta\right)+B_{3,2}\left(\theta\right)+B_{3,3}\left(\theta\right),
\end{aligned}
$$

where

$$
\left|B_{3,i}\left(\theta\right)\right|\leq\frac{\left|\log\left(4a\right)\right|^{-\delta}}{\bar{L}_T}\sum_{t=1}^{T}\left|\log p\left(y_t|x_{t-1};\theta\right)\right|^{1+\delta},\quad i=1,2,
$$

and

$$
\left|B_{3,3}\left(\theta\right)\right|\leq\left\{\frac{\sum_{t=1}^{T}\left\|y_t\right\|^{1+\delta}}{\bar{L}_T}\frac{T}{N^{\gamma(1+\delta)}}\right\}^{\delta/(1+\delta)}\left\{\frac{1}{\bar{L}_T}\sum_{t=1}^{T}\left|\log p\left(y_t|x_{t-1};\theta\right)\right|^{1+\delta}\right\}^{1/\delta}.
$$

All the above bounds are $o_P(1)$ under (C.1)–(C.3) and (B.L.1) together with Lemma 2. This shows (i)

Next, to show (ii), first observe that

$$\begin{aligned}
&\bar{p}(y_t, z_t|x_{t-1};\theta) - p(y_t, z_t|x_{t-1};\theta) \\
= &\int p(y_t, z_t|y_{t-1}, z_{t-1};\theta) \left\{ \frac{\bar{p}(y_{t-1}, z_{t-1}|x_{t-2};\theta) - p(y_{t-1}, z_{t-1}|x_{t-2};\theta)}{\bar{p}(y_{t-1}|x_{t-2};\theta)} \right\} dz_{t-1} \\
&+ \{\bar{p}(y_{t-1}|x_{t-2};\theta) - p(y_{t-1}|x_{t-2};\theta)\} \int \frac{p(y_t, z_t|y_{t-1}, z_{t-1};\theta)\,\bar{p}(y_{t-1}, z_{t-1}|x_{t-2};\theta)}{p(y_{t-1}|x_{t-2};\theta)\,\bar{p}(y_{t-1}|x_{t-2};\theta)} dz_{t-1},
\end{aligned}$$

......[INCOMPLETE]. The consistency result now follows from Theorem 10 of Kristensen and Shin (2006).

The asymptotic distribution result follows by Theorem 12 of Kristensen and Shin (2006) under (N.1)–(N.4) and (B.L.2) since, under these conditions, the above bounds are $o_P\left(\bar{L}_T^{-1}\right)$. <1.5em ∎

B Properties of the Simulated Density

Lemma 2 *Assume that (A.1)–(A.5) and (K.1) hold. Then for any $t \geq 1$, $\hat{p}(y_t, z_t|x_{t-1})$ given in (??) satisfies*

$$\sup_{(y_t,z_t)\in\mathbb{R}^{k+l}} \sup_{\|x_{t-1}\|_\infty \leq N^\gamma} \sup_{\theta\in\Theta} |\hat{p}(y_t, z_t|x_{t-1};\theta) - \bar{p}(y_t, z_t|x_{t-1};\theta)| = \sum_{i=0}^{t-1} a^{-i} R_{t-i},$$

for any $\gamma > 0$ and any $a > 0$, where

$$\begin{aligned}
\bar{p}(y_t, z_t|x_{t-1};\theta) &= \int p(y_t, z_t|y_{t-1}, z_{t-1};\theta)\,\bar{p}(z_{t-1}|x_{t-1};\theta)\,dz_{t-1}, \\
\bar{p}(z_{t-1}|x_{t-1};\theta) &= \frac{p(y_{t-1}, z_{t-1}|x_{t-2};\theta)\,\tau_a\{\hat{p}(y_{t-1}, z_{t-1}|x_{t-2};\theta)\}}{\int p(y_{t-1}, z|x_{t-2};\theta)\,\tau_a\{\hat{p}(y_{t-1}, z|x_{t-2};\theta)\}\,dz},
\end{aligned}$$

and

$$R_t = C_{0,1}\left(\left(1 + a^{-1}\right)\mathbb{E}\left[\Lambda_t^2\right], K\right) O_P\left(\sqrt{\log(N)} N^{-1/2} h^{-d-1}\right) + C_{0,2}\left(K, D_w^r p_t\right) h^r, \quad (7)$$

and $C_{0,i}$, $i = 1, 2$, are given in Lemma 7 of Kristensen and Shin (2006).

Proof We show this recursively. Conditional on all draws prior to time t, we claim that

$$\sup_{(y_t,z_t)\in\mathbb{R}^{k+l}} \sup_{\|x_{t-1}\|_\infty \leq N^\gamma} \sup_{\theta\in\Theta} \left| \hat{p}(y_t, z_t|x_{t-1};\theta) - \int p(y_t, z_t|y_{t-1}, z_{t-1};\theta)\,\tilde{p}(z_{t-1}|x_{t-1};\theta)\,dz_{t-1} \right| = R_t,$$

where R_t is given in the lemma. We show this by verifying the conditions in Lemma 7 of Kristensen and Shin (2006) with $Y_i(\alpha) = \left(Y_{t,i}^{x_{t-1},\theta}, Z_{t,i}^{x_{t-1},\theta}\right)$, $\alpha = (x_{t-1},\theta)$ and $A = \mathbb{R}^{td} \times \Theta$. First, observe that the target density $(y_t, z_t) \mapsto \int p(y_t, z_t|y_{t-1}, z_{t-1};\theta)\,\tilde{p}(z_{t-1}|x_{t-1};\theta)\,dz_{t-1}$, where $\tilde{p}$ is given in (4), is r times continuously differentiable since $p(y_t, z_t|y_{t-1}, z_{t-1};\theta)$ is so. Second, (A.1) combined with

$$\begin{aligned}\left(Y_{t,i}^{x_{t-1},\theta}, Z_{t,i}^{x_{t-1},\theta}\right) &\sim p\left(y_t, z_t|y_{t-1}, Z_{t-1}^{x_{t-1},\theta};\theta\right), \\ Z_{t-1}^{x_{t-1},\theta} &= \tilde{F}_{z_{t-1}|x_{t-1}}^{-1}\left(\varepsilon_{2,t}|x_{t-1};\theta\right),\end{aligned} \tag{8}$$

where

$$\tilde{F}_{z_{t-1}|x_{t-1}}(z|x_{t-1};\theta) = \int_{-\infty}^{z_1}\cdots\int_{-\infty}^{z_{d_2}}\tilde{p}(z_{t-1}|x_{t-1};\theta)\,dz_{t-1},$$

yields

$$\begin{aligned}&\left\|\left(Y_t^{x_{t-1},\theta}, Z_t^{x_{t-1},\theta}\right) - \left(Y_t^{x'_{t-1},\theta\prime}, Z_t^{x'_{t-1},\theta'}\right)\right\| \\ \leq\ & \Lambda_t\left\{\left\|y_{t-1} - y'_{t-1}\right\|^{\beta_1} + \left\|Z_{t-1}^{x_{t-1},\theta} - Z_{t-1}^{x'_{t-1},\theta'}\right\|^{\beta_1} + \left\|\theta-\theta'\right\|^{\beta_1}\right\} \\ \leq\ & \left(1+a^{-1}\right)\Lambda_t\left\{\left\|y_{t-1} - y'_{t-1}\right\|^{\beta_1} + \left\|x_{t-1} - x_{t-1}\right\|^{\beta_1} + \left\|\theta-\theta'\right\|^{\beta_1}\right\},\end{aligned}$$

since, by (8),

$$\begin{aligned}\left\|Z_{t-1}^{x_{t-1},\theta} - Z_{t-1}^{x'_{t-1},\theta'}\right\| &\leq \left\|\frac{\partial\tilde{F}_{z_{t-1}|x_{t-1}}^{-1}\left(\varepsilon_{3t}|\bar{x}_{t-1};\bar{\theta}\right)}{\partial x_{t-1}}\right\|\left\|x_{t-1}-x'_{t-1}\right\| \\ &\quad + \left\|\frac{\partial\tilde{F}_{z_{t-1}|x_{t-1}}^{-1}\left(\varepsilon_{3t}|\bar{x}_{t-1};\bar{\theta}\right)}{\partial\theta}\right\|\left\|\theta-\theta'\right\| \\ &\leq \frac{2}{a}\left\{\left\|x_{t-1}-x'_{t-1}\right\| + \left\|\theta-\theta'\right\|\right\},\end{aligned}$$

where the last inequality is a consequence of the following calculations:

$$\begin{aligned}\frac{\partial\tilde{F}_{z_{t-1}|x_{t-1}}^{-1}\left(\varepsilon_{3t}|x_{t-1};\theta\right)}{\partial\theta} &= -\left[\frac{\partial\tilde{F}_{z_{t-1}|x_{t-1}}\left(z_{t-1}|x_{t-1};\theta\right)}{\partial z_{t-1}}\Big|_{z_{t-1}=Z_t^{x_{t-1},\theta}}\right]^{-1}\frac{\partial\tilde{F}_{z_{t-1}|x_{t-1}}\left(Z_t^{x_{t-1},\theta}|x_{t-1};\theta\right)}{\partial\theta} \\ &= \frac{1}{\tilde{p}\left(Z_t^{x_{t-1},\theta}|x_{t-1};\theta\right)}\int_{-\infty}^{z_1}\cdots\int_{-\infty}^{z_{d_1}}\frac{\partial\tilde{p}\left(z_{t-1}|x_{t-1};\theta\right)}{\partial\theta}dz_{t-1}\Big|_{z=Z_t^{\bar{x}_{t-1},\theta}},\end{aligned}$$

and

$$\frac{\partial\tilde{p}\left(z_{t-1}|x_{t-1};\theta\right)}{\partial\theta} = \frac{1}{\int\hat{p}\left(y_{t-1}, z|x_{t-2};\theta\right)\tau_a\left\{\hat{p}\left(y_{t-1}, z|x_{t-2};\theta\right)\right\}dz}$$

$$\times\left\{\frac{\partial\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)}{\partial\theta}\tau_a\left\{\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)\right\}\right.$$

$$\left.+\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)\frac{\partial\tau_a\left\{\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)\right\}}{\partial\theta}\right\}$$

$$+\tilde{p}\left(z_{t-1}|x_{t-1};\theta\right)\frac{1}{\int\hat{p}\left(y_{t-1},z|x_{t-2};\theta\right)\tau_a\left\{\hat{p}\left(y_{t-1},z|x_{t-1};\theta\right)\right\}dz}$$

$$\times\left\{\int\frac{\partial\hat{p}\left(y_{t-1},z|x_{t-2};\theta\right)}{\partial\theta}\tau_a\left\{\hat{p}\left(y_{t-1},z|x_{t-1};\theta\right)\right\}dz\right.$$

$$\left.+\int\hat{p}\left(y_{t-1},z|x_{t-2};\theta\right)\frac{\partial\tau_a\left\{\hat{p}\left(y_{t-1},z|x_{t-1};\theta\right)\right\}}{\partial\theta}dz\right\}.$$

Thus, with $z_{t-1}=Z_t^{x_{t-1},\theta}$,

$$\begin{aligned}\left\|\frac{\partial\tilde{F}^{-1}_{z_{t-1}|x_{t-1}}\left(\varepsilon_{3t}|x_{t-1};\theta\right)}{\partial\theta}\right\| &\leq \frac{1}{a}\left\{\left\|\frac{\partial\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)}{\partial\theta}\right\|+\left|\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)\right|\right\}\\ &\quad+\frac{\left|p\left(z_{t-1}|x_{t-1};\theta\right)\right|}{a}\int\left\|\frac{\partial\hat{p}\left(y_{t-1},z|x_{t-2};\theta\right)}{\partial\theta}\right\|+\left|\hat{p}\left(y_{t-1},z|x_{t-2};\theta\right)\right|dz\\ &\leq \frac{1}{a}\left\{\left\|\frac{\partial p\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)}{\partial\theta}\right\|+\left|p\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)\right|+o_P(1)\right\}\\ &\quad+\frac{p\left(z_{t-1}|x_{t-1};\theta\right)}{a}\int\left\|\frac{\partial p\left(y_{t-1},z|x_{t-2};\theta\right)}{\partial\theta}\right\|+p\left(y_{t-1},z|x_{t-2};\theta\right)dz+o_P(1)\\ &\leq \frac{C}{a}\end{aligned}$$

uniformly in $(y_{t-1},z_{t-1},x_{t-2},\theta)\in\mathbb{R}^d\times\Theta\times\{\|x_{t-2}\|\leq N^\gamma\}$, since $\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)$ and $\partial\hat{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)/\partial\theta$ converge uniformly on this set. By the same arguments, one can show that $\partial\tilde{F}^{-1}_{z_{t-1}|x_{t-1}}/\partial x_{t-1}$ is uniformly bounded by C/a.

Next,

$$\begin{aligned}&\int p\left(y_t,z_t|y_{t-1},z_{t-1};\theta\right)\left\{\tilde{p}\left(z_{t-1}|x_{t-1};\theta\right)-\bar{p}\left(z_{t-1}|x_{t-1};\theta\right)\right\}dz_{t-1}\\ =&\int p\left(y_t,z_t|y_{t-1},z_{t-1};\theta\right)\left\{\frac{\tilde{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)-\bar{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)}{\bar{p}\left(y_{t-1}|x_{t-2};\theta\right)}\right\}dz_{t-1}\\ &+\left\{\tilde{p}\left(y_{t-1}|x_{t-2};\theta\right)-\bar{p}\left(y_{t-1}|x_{t-2};\theta\right)\right\}\int\frac{p\left(y_t,z_t|y_{t-1},z_{t-1};\theta\right)\bar{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)}{\tilde{p}\left(y_{t-1}|x_{t-2};\theta\right)\bar{p}\left(y_{t-1}|x_{t-2};\theta\right)}dz_{t-1}\\ \leq&\ \frac{C}{a}\sup_{z_{t-1}\in\mathbb{R}^{d_1}}\left|\tilde{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)-\bar{p}\left(y_{t-1},z_{t-1}|x_{t-2};\theta\right)\right|\\ &+\frac{C}{a}\left|\tilde{p}\left(y_{t-1}|x_{t-2};\theta\right)-\bar{p}\left(y_{t-1}|x_{t-2};\theta\right)\right|.\end{aligned}$$

We conclude

$$\sup_{(y_t,z_t)\in\mathbb{R}^d}\sup_{\|x_{t-1}\|\leq N^\gamma}\sup_{\theta\in\Theta}\left|\tilde{p}\left(y_t,z_t|x_{t-1};\theta\right)-\bar{p}\left(y_t,z_t|x_{t-1};\theta\right)\right|$$

$$
\begin{aligned}
&= R_t + a^{-1} \sup_{(y_{t-1}, z_{t-1}) \in \mathbb{R}^d} \sup_{\|x_{t-2}\| \leq N^\gamma} \sup_{\theta \in \Theta} \left| \hat{p}\left(y_{t-1}, z_{t-1} | x_{t-2}; \theta\right) - \bar{p}\left(y_{t-1}, z_{t-1} | x_{t-2}; \theta\right) \right| \\
&\vdots \\
&= \sum_{i=0}^{t-1} a^{-i} R_{t-i}. \qquad \blacksquare
\end{aligned}
$$

References

Altissimo, F. and A. Mele (2005): "Simulated Nonparametric Estimation of Dynamic Models with Applications to Finance," Discussion Paper 539, Financial Markets Group, London School of Economics.

Andersen, T. G., H.-J. Chung, and B. E. Sørensen (1999): "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61–87.

Bickel, P. J. and Y. Ritov (1996): "Inference in Hidden Markov Models I: Local Asymptotic Normality in the Stationary Case," *Bernoulli*, 2, 199–228.

Bickel, P. J., Y. Ritov, and T. Rydén (1998): "Asymptotic Normality of the Maximum-Likelihood Estimator for General Hidden Markov Models," *Annals of Statistics*, 26, 1614–1635.

Douc, R. and C. Matias (2001): "Asymptotics of the Maximum Likelihood Estimator for General Hidden Markov Models," *Bernoulli*, 7, 381–420.

Duffie, D. and K. J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

Fermanian, J.-D. and B. Salanié (2004): "A Nonparametric Simulated Maximum Likelihood Estimation Method," *Econometric Theory*, 20, 701–734.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2004): "Estimating Macroeconomic Models: A Likelihood Approach," Working paper, University of Pennsylvania.

Gallant, A. R. and G. Tauchen (1996): "Which Moments to Match?" *Econometric Theory*, 12, 657–681.

Gouriéroux, C., A. Monfort, and É. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

Jensen, J. L. and N. V. Petersen (1999): "Asymptotic Normality of the Maximum Likelihood Estimator in State Space Models," *Annals of Statistics*, 27, 514–535.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

Kristensen, D. and Y. Shin (2006): "Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood," Manuscript, Department of Economics, University of Wisconsin.

Leroux, B. G. (1992): "Maximum-Likelihood Estimation for Hidden Markov Models," *Stochastic Processes and their Applications*, 40, 127–143.

Pitt, M. K. and N. Shephard (1999): "Filtering Via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599.

Smith, Jr., A. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.