ESTIMATION OF DYNAMIC LATENT VARIABLE MODELS USING SIMULATED NONPARAMETRIC MOMENTS

MICHAEL CREEL AND DENNIS KRISTENSEN

ABSTRACT. Given a model that can be simulated, conditional moments at a trial parameter value can be calculated with high accuracy by applying kernel smoothing methods to a long simulation. With such conditional moments in hand, standard method of moments techniques can be used to estimate the parameter. Because conditional moments are calculated using kernel smoothing rather than simple averaging, it is not necessary that the model be simulable subject to the conditioning information that is used to define the moment conditions. For this reason, the proposed estimator is applicable to general dynamic latent variable models. The estimator is consistent and has the same asymptotic distribution as that of the infeasible GMM estimator based on the same moment conditions. Monte Carlo results show how the estimatod may be applied to a range of dynamic latent variable (DLV) models, and that it performs well in comparison to several other estimators that have been proposed for DLV models. An application to weekly spot exchange rate data further illustrates use of the estimator.

Keywords: dynamic latent variable models; simulation-based estimation; simulated moments; kernel regression; nonparametric estimation
JEL codes: C13; C14; C15

1. INTRODUCTION

Dynamic latent variable (DLV) models are a flexible and often natural way of modeling complex phenomena. As an example, consider a macroeconomic model. A model may specify behavioral rules, learning rules, a social networking structure, and information transmission mechanisms for a large group of possibly heterogeneous agents. If the model is fully specified, it can be used to generate time series data on all of the agents' actions. In attempting to use real world data to estimate the parameters of such model, one finds that real world data is much more aggregated than the data generated by the model. Typically, individual agents' actions are not observed - only macroeconomic aggregates are available. From the econometric point of view, many of the variables generated by the model are latent. In a dynamic, nonlinear context, this can complicate the econometric estimation of the model's parameters.

To fix ideas, consider the general DLV model:

$$\text{DLV: } \begin{cases} y_t = & r_y\left(y^{t-1}, w^{t-1}, u_t; \theta\right) \\ w_t = & r_w\left(y^{t-1}, w^{t-1}, u_t; \theta\right) \end{cases} \tag{1.1}$$

where $t = 1, ..., T$. The observable variables are the d_y-dimensional vector y_t, w_t is a vector of latent variables, and u_t is a vector of errors. The superscript notation is used to indicate a vector of lagged variables up to the time indicated, $y^{t-1} \equiv \left(y_1', ..., y_{t-1}'\right)'$, and $w^{t-1} \equiv \left(w_1', ..., w_{t-1}'\right)'$.[1] There is a vector of independent white noise variables, u_t, with a known distribution. Finally, $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$ is a vector of unknown parameters. This definition closely follows that of Billio and Monfort (2003), with the exception that the same white noise vector enters the equations for both the observable and latent variables, to allow for potential correlations in the innovations of the two sets of variables. Calculation of the likelihood function requires finding the density of y^T, and as Billio and Monfort make clear, this involves calculating an integral of the same order as T, a problem that is in general untractable. Without the density of the observable variables, analytic moments cannot be computed. Thus, maximum likelihood and moment-based estimation methods often are not available.

This paper offers a new estimator that is applicable to general DLV models. It is a new implementation of the simulated method of moments (SMM) that allows use of conditional moments. Conditional moments are evaluated using nonparametric kernel smoothing of-simulated data. The estimator is very simple to use since it is just an ordinary GMM estimator that uses kernel smoothing to evaluate moment conditions. The estimator is referred to as the simulated nonparametric moments (SNM) estimator.

Under regularity conditions, we show that the SNM estimator is consistent and asymptotically normally distributed. Because it is a method of moments estimator, it is not in

[1] The possible presence of observable exogenous variables with known dynamics (for example, static exogenous variables) is suppressed for clarity. The macroeconomic model of the previous paragraph could be formalized by letting w_t indicate the vector of all of the agents' actions, and letting q_t be the observed aggregate outcomes.

general asymptotically efficient when compared to (simulated) maximum-likelihood estimation (MLE). On the other hand, the SNM may be in many cases be more robust towards misspecifications: It is well-known that GMM estimators often remain consistent under departures in certain directions from a given fully specified model (in particular, in terms of the distribution of the errors) while the MLE in contrast becomes inconsistent. Examples of DLV models where GMM-type estimators have proved robust are stochastic volatility (SV) models (Harvey et al, 1994; Ruiz, 1994), DSGE (dynamic stochastic general equilibrium) models (Ruge-Murcia, 2007), and diffusion models (Bibby and Sørensen, 1995).

Furthermore, by choosing moment conditions in a judious manner, method of moments may reach full efficiency. In particular, Carrasco et al. (2007) demonstrate that by using the characteristic function as moments, full ML efficiency can be reached. Finally, Monte Carlo results in our paper show that moment conditions may be chosen such the SNM estimator performs well in comparison to other estimators that have been proposed for estimation of general DLV models.

A number of other econometric methods have been developed over the last two decades to deal with the complications that may accompany DLV models. These include the simulated method of moments (McFadden, 1989; Pakes and Pollard, 1989), indirect inference (Gouriéroux, Monfort and Renault, 1993; Smith, 1993), simulated pseudo-maximum likelihood (Laroque and Salanié, 1993), simulated maximum likelihood (Lee, 1995), the efficient method of moments (Gallant and Tauchen, 1996), the method of simulated scores (Hajivassiliou and McFadden, 1998), kernel-based indirect inference (Billio and Monfort, 2003), the simulated EM algorithm (Fiorentini, Sentana and Shephard, 2004), nonparametric simulated maximum likelihood (Fermanian and Salanié, 2004; Kristensen and Shin, 2006) and simulated nonparametric estimators (Altissimo and Mele, 2007). These methods have been applied to DLV models in a number of contexts. Billio and Monfort (2003) provide numerous references for applications.

As noted by Fermanian and Salanié (2004, pg. 702), there often exists a trade-off between the asymptotic efficiency of a method and its applicability to a wide range of models. Simulated maximum likelihood and the method of simulated scores are asymptotically efficient when they can be applied, but this is not the case when the likelihood function or the score function cannot be expressed as a function of expectations of simulable quantities. Nonparametric simulated maximum likelihood (NPSML) is asymptotically efficient and generally applicable for estimation of static models (Fermanian and Salanié, 2004). Kristensen and Shin (2008) and Altissimo and Mele (2008) generalize this method to handle dynamic models. Their methods are asymptotically efficient when the model is Markovian in the observable variables. This is often an important limitation, since models that are Markovian in all variables are usually not Markovian in a subset of the variables (Florens *et al.* 1993). When the model is not Markovian in the observable variables, the proposed methods are not asymptotically efficient. Furthermore, as mentioned earlier, the NPSML method relies on that the model is correctly specified.

The simulated method of moments (SMM) is generally applicable if unconditional moments are used, but foregoing conditioning information may limit the estimator's ability to capture the dynamics of the model, and can result in poor efficiency (Andersen, Chung and Sørensen, 1999; Michaelides and Ng, 2000; Billio and Monfort, 2003). In the context of DLV models, the usual implementation of SMM that directly averages a simulator normally cannot be based upon conditional moments, since it is not in general possible to simulate from the model subject to the conditioning information. Due to the full specification of the model, it is easy to simulate a path. However, the elements are drawn from their marginal distributions. It is not in general possible to draw from $y_t | \left(y^{t-1}; \theta\right)$. To do so, one would need draws from $w^t | \left(y^{t-1}; \theta\right)$. If such draws were available, they could be inserted into the first line of the DLV model given in equation (1.1), which, combined with a draw from u_t, would give a draw from $y_t | \left(y^{t-1}; \theta\right)$. The problem is that the observed value of y^{t-1} is only compatible with certain realizations of the history of the latent variables, w^{t-1}, but what is the set of compatible realizations is not known. For certain types of model it is possible to circumvent this problem. For example, Fiorentini, Sentana and Shephard (2004) find a way of casting a factor GARCH model as a first-order Markov process, and are then able to use Markov chain Monte Carlo (MCMC) methods to simulate from $w^t | \left(y^{t-1}; \theta\right)$, which is then fed into a simulated EM algorithm to estimate the parameter. However, for DLV models in general, there is no means of simulating from $w^t | \left(y^{t-1}; \theta\right)$ (Billio and Monfort, 2003, pg. 298; Carrasco *et al*., 2007, pg. 544).

Indirect inference is generally applicable, but its efficiency depends crucially upon the choice of the auxiliary model. The efficient method of moments (EMM, Gallant and Tauchen, 1996) is closely related to the indirect inference estimator, and presumes use of an auxiliary model that guarantees good asymptotic efficiency, by closely approximating the structural model. This estimator is both generally applicable and is highly efficient if a good auxiliary model is used, and it is fully asymptotically efficient if the auxiliary model satisfies a smooth embedding condition (see Gallant and Tauchen, 1996, Definition 1). Satisfying this condition is not necessarily an easy thing to achieve. A common practice is to fit a semi-nonparametric (SNP) auxiliary model of the sort proposed by Gallant and Nychka (1987), augmented by a leading parametric model that is known to provide a reasonably good approximation. Andersen, Chung and Sørensen (1999) provide Monte Carlo evidence that shows the importance of the choice of the auxiliary model. They also note that highly parameterized auxiliary models often cannot be successfully fit when the sample size is not large. It is important to keep in mind that a parsimonious parametric auxiliary model may be far from satisfying the smooth embedding condition. This can lead to serious inefficiency and to failure to detect serious misspecifications of the structural model (Tauchen, 1997; Gallant and Tauchen, 2002). In sum, EMM and indirect inference are clearly attractive methods, given that the sample is large enough to use a rich auxiliary model. Even if this is the case, effort and skill are required to successfully use these methods. In the case of EMM, the documentation of the EMM software package (Gallant and Tauchen, 2004, 2007) makes this clear.

The kernel-based indirect inference (KBII) approach suggested by Billio and Monfort (2003) proposes an entirely nonparametric auxiliary model in place of the EMM's highly parameterized auxiliary model. The use of kernel regression methods is considerably simpler than estimation of models based upon a SNP density with a parametric leading term, since software can be written to use data-dependent rules that tune the fitting process to a given data set with little user intervention. The consistency of the kernel regression estimator ensures a good fit to the data. The main drawback with the KBII estimator is that the binding functions are conditional moments of endogenous variables at certain points in the support of the conditioning variables. How many such points to use, and exactly which points to use require decisions on the part of the econometrician. Billio and Monfort recognize this problem and propose a scoring method to choose the binding functions.

The next section defines the estimator and discusses its properties and usage. The third section presents several examples that compare the SNM estimator to other methods, using Monte Carlo. Section 4 applies the estimator to weekly spot market exchange rate data, and Section 5 concludes. All proofs and lemmas have been relegated to the Appendix.

2. THE SNM ESTIMATOR

2.1. **Definition of the estimator.** The moment-based estimation framework used in this paper is as follows. We have observed the sample $\{y_t\}_{t=1}^n$ over the period $t = 1,2,...,n$. The sample is presumed to be a realization of the data generating process defined by Eq. (1.1) at the true parameter value θ_0. For estimation purposes, we introduce a vector $x_t = (x_{t,1},...,x_{t,d_x})'$ of additional variables that are functions of current and lagged values of y_t:

$$x_t = x(y_t, y_{t-1}, y_{t-2}, ...) \in \mathbb{R}^{d_x}.$$

These variables are chosen by the researcher and will function as conditioning variables. In most cases, x_t is simply chosen as the first q lags, $x_t = (y_{t-1},...,y_{t-q})$, but we here allow for more flexibility in their selection.

Likewise, let $\phi_t = (\phi_{t,1},...,\phi_{t,L})'$ be a collection of L "test variables" chosen by the researcher. These are defined as functions of current and future values of y_t:

$$\phi_t = \phi(y_t, y_{t+1}, y_{t+2}, ...) \in \mathbb{R}^L.$$

Given the test variables, we define corresponding generalized residual functions,

$$\varepsilon_t(\theta) = \phi_t - T(\phi)(x_t;\theta) \in \mathbb{R}^L, \tag{2.1}$$

where T denotes the conditional expecations operator w.r.t. x_t,

$$T(\phi)(x;\theta) = E_\theta[\phi_t | x_t = x].$$

Here, $E_\theta[\cdot|x_t]$ denotes conditional expectations taken under the model with θ being the true value, $E_\theta[\phi_t|x_t] = \int \phi(y_t, y_{t+1}, ...)\, dP(y_t, y_{t+1}, ...|x_t;\theta)$. By construction, the residual vector satisfies

$$E_\theta[\varepsilon_t(\theta)|x_t] = 0, \tag{2.2}$$

For a set of instruments that are functions of the conditioning variables,

$$Z_t = Z(x_t) \in \mathbb{R}^{L \times M},$$

moment conditions are now defined by interacting the instruments with the error functions,

$$g_t(\theta) := Z_t' \varepsilon_t(\theta) \in \mathbb{R}^{LM}, \tag{2.3}$$

such that $E_\theta[g_t(\theta)] = 0$. One standard way of choosing the instruments is as follows: For each residual, define a vector of m instrumental variables $z_{t,k} \in \mathbb{R}^m$ as functions of the conditioning variables,

$$z_{t,k} = (z_{k,1}(x_t), ..., z_{k,M}(x_t))', \quad k = 1, ..., L.$$

Here, we use the same number of instruments, m, for each residual, but the instruments may differ across residuals. We then collect the instruments in a block-diagonal $(L \times M)$-matrix,

$$Z_t = \begin{bmatrix} z_{t,1}' & 0 & \cdots & 0 \\ 0 & z_{t,2}' & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & z_{t,L}' \end{bmatrix} \in \mathbb{R}^{L \times M},$$

where $M = Lm$. In principle, one could use a different number of instruments for each residual. This possibility is suppressed here both to simplify notation, and because when optimal instruments are defined below, it will be seen that it is optimal to use the same number of instruments for each residual.

If the conditional moments $T(\phi)(x_t; \theta)$ in equation (2.1) have a known functional form, estimation may proceed using the standard generalized method of moments (GMM): Define the corresponding sample moments as

$$G_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} g_t(\theta). \tag{2.4}$$

For some sequence of $M \times M$ dimensional positive definite weighting matrices $W_n \to^P W > 0$, we would then compute the GMM estimator defined as

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} G_n(\theta)' W_n G_n(\theta).$$

When no closed-form functional form of is available, it may be possible to define an unbiased simulator $\hat{T}(\phi)(x_t; \theta)$. If this is so, simulated moments can be defined by replacing $T(\phi)(x_t; \theta)$ in equation (2.1) with $\hat{T}(\phi)(x_t; \theta)$. Doing so, and then proceeding with normal GMM estimation methods defines the SMM estimator (Gouriéroux and Monfort, 1996, pg. 27; Duffie and Singleton, 1993). However, in the case of general DLV models, it is in general not possible to simulate subject to the conditioning information $x = x_t$. In this case, the standard SMM estimator cannot be based upon conditional moments as defined in equations (2.1)-(2.4). Estimation by SMM using unconditional moments is still feasible, but the Monte Carlo evidence cited above has shown that this approach often has poor efficiency,

due to the fact that unconditional moments provide little information on the dynamics of a DLV model.

The fundamental idea of the simulated nonparametric moments (SNM) estimator proposed here is to replace the expectations $T(\phi)(x_t;\theta)$ that are used to define error functions in equation (2.1) with kernel regression fits based on a simulation from the model. Kernel regression (also known as kernel smoothing) is a well-known nonparametric technique for estimating regression functions of unknown form (Robinson, 1983; Bierens, 1987; Härdle, 1991; Li and Racine, 2007). Its application here is entirely standard, except for the use of simulated data.

In the following, capital letters will be used to indicate simulated data or elements that depend upon simulated data. Let $\{Y_s(\theta),\ s=1,....,S\}$ be a time series of S simulations of generated by Eq. (1.1) at the trial parameter value θ:

$$\begin{cases} Y_s(\theta) = & r_y\left(Y^{s-1}(\theta),W^{s-1}(\theta),U_s;\theta\right) \\ W_s(\theta) = & r_w\left(Y^{s-1}(\theta),W^{s-1}(\theta),U_s;\theta\right) \end{cases}, \tag{2.5}$$

for $s=1,....S$, where the simulations are initialized at some values $\left(Y^{-1}(\theta),W^{-1}(\theta)\right)$. Given the simulated values, first compute the corresponding conditioning and test variables,

$$\begin{aligned} X_s(\theta) &= x\left(Y_{s-1}(\theta),...,Y_{s-q}(\theta)\right), \\ \Phi_s(\theta) &= \phi\left(Y_s(\theta),...,Y_{s+p}(\theta)\right), \end{aligned}$$

and then the kernel estimator of $T(\phi)(x_t;\theta)$,

$$\hat{T}_S(\phi)(x;\theta) = \frac{\sum_{s=1}^{S}\Phi_s(\theta)K_h\left(X_s(\theta)-x\right)}{\sum_{s=1}^{S}K_h\left(X_s(\theta)-x\right)}, \tag{2.6}$$

where $K_h(z)=K(z/h)/h$, $K:\mathbb{R}^{d_x}\mapsto\mathbb{R}$ is a kernel function, and $h>0$ is a bandwidth. To speed up computations, one should not separately fit each of the L test variables, but rather employ a specialized kernel fitting algorithm that saves the weights across variables. Since the dimension of x, d_x, is usually greater than one, the kernel function $K(\cdot)$ is in general multivariate.

We here use the same bandwidth across all variables, mainly to keep the notation simple. In some cases this may not be advisable; in particular, if the individual variables contained in X_t (and thereby in Y_t) are not on the same scale a common bandwidth may lead to a less precise kernel estimator. If one assumes that the kernel function incorporates a "prewhitening" transformation, a common bandwidth may be a reasonable choice. Locally adaptive kernel fitting is an extension that we do not pursue here.

Note that the kernel regression fit can be evaluated at x without requiring that the simulated sequence contains any realizations such that $X_s(\theta)=x$. What is required for a good fit at x is that there there be a large number of realizations that are "close enough" to x.

The SNM estimator now follows the standard moment-based estimation framework, except that the kernel fit $\hat{\mu}(x;\theta)$ is used in place of the expectation of unknown form, $\mu(x;\theta)$.

To be explicit, the SNM estimator is based on estimated residual functions,

$$\hat{\varepsilon}_{t,S}(\theta) = \phi_t - \hat{T}_S(\phi)(x;\theta). \tag{2.7}$$

The moment function contribution of an observation is

$$\hat{g}_{t,S}(\theta) := Z_t' \hat{\varepsilon}_{t,S}(\theta) \in \mathbb{R}^M, , \tag{2.8}$$

where we use the observed instruments. Average moment conditions are now computed as

$$\hat{G}_{n,S}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \hat{g}_{t,S}(\theta), \tag{2.9}$$

and the SNM estimator is the minimizer of corresponding distance function,

$$\hat{\theta}_{n,S} = \arg\min_{\theta\in\Theta} \hat{G}_{n,S}(\theta)' W_n \hat{G}_{n,S}(\theta). \tag{2.10}$$

2.2. **Properties of the SNM estimator.** This section deals with the consistency and asymptotic normality of the SNM estimator. The proof offered here is high level, in the sense that our Assumptions 2- 5 below are made without detailing assumptions on the DLV model in equation (1.1) that would cause them to hold. Given a more concrete formulation of the DLV model, one could provide more low level assumptions that would imply Assumptions 2-5. This is not done here since the intention is not to focus on any particular model. The first assumption defines the true DGP and parameter value:

Assumption 1. *The sample $\{y_t\}_{t=1}^n$ is generated by the DLV model of equations (1.1), at the true parameter value $\theta_0 \in \Theta$.*

Next, assume that the chosen endogenous variables, conditioning variables, and instruments define a GMM estimator that is consistent and distributed asymptotically normally. Of course, this estimator normally is not feasible if the SNM estimator is under consideration, but abstractly, it is assumed to have the usual desirable properties:

Assumption 2. *The following hold:*

(1) *(Compact parameter space) The true parameter value θ_0 lies in the interior of the compact parameter space $\Theta \subseteq \mathbb{R}^{d_\theta}$.*
(2) *(Identification) The moment function $G(\theta) = E[Z_t'\varepsilon_t(\theta)]$ is continuous and satisfies $G(\theta) = 0$ if and only if $\theta = \theta_0$.*
(3) *(Uniform Convergence) $\sup_{\theta\in\Theta} \|G_n(\theta) - G(\theta)\| \to^P 0$.*
(4) *(Asymptotic Normality): $\sqrt{n}G_n(\theta_0) \to^d N(0,\Omega_0)$ where*

$$\Omega_0 = E\left[g_0(\theta_0)g_0(\theta_0)'\right] + 2\sum_{t=1}^{\infty} E\left[g_0(\theta_0)g_t(\theta_0)'\right]. \tag{2.11}$$

(5) *(Derivative) The derivative*

$$H_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta), \quad h_t(\theta) = -Z_t' \frac{\partial T(\phi_i)(x_t;\theta)}{\partial\theta} \in \mathbb{R}^{M\times d_\theta}$$

satisfies $\sup_{\|\theta-\theta_0\|<\delta} \|H_n(\theta) - H(\theta)\| \to^P 0$ where $H(\theta) = E[h_t(\theta)]$.
(6) *(Weighting matrix) $W_n \to^P W > 0$ where $H_0'W_0H_0 > 0$ with $H_0 = H(\theta_0)$.*

Under Assumption 2, it holds that the infeasible GMM estimator satisfies:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d N\left(0, \left(H_0' W_0 H_0\right)^{-1} H_0' W_0 \Omega_0 W_0 H_0 \left(H_0' W_0 H_0\right)^{-1}\right),$$

c.f. Lemma 2 in the Appendix. The goal is now to analyze the simulated version, $\hat{\theta}_{n,S}$, relative to the actual but infeasible one, $\hat{\theta}_n$. In order to do this, we have to ensure that the kernel estimator is consistent uniformly over (x,θ). This is done by verifying the general conditions stated in Kristensen (2008a) where uniform convergence results are obtained for data that are functions of a parameter. We will impose some fairly high-level assumptions on the data-generating model that imply the conditions in Kristensen (2008a). In order to state these assumptions, we first introduce some additional notation. Let $f(x;\theta)$ and $f_T(x,x';\theta)$ denote the stationary densities of $X_0(\theta)$ and $(X_0(\theta), X_T(\theta))$, for some $T \geq 1$, respectively. We then define for any random sequence $V_s(\theta)$, $s = 1,...,T$, and for some $\lambda \geq 2$ the following bounds,

$$B_0 = \sup_{x\in\mathbb{R}^{d_x}} \sup_{\theta\in\Theta} f(x;\theta), \quad B_{V,1} = \sup_x \sup_{\theta\in\Theta} \|x\|^\lambda E\left[\|V_0(\theta)\| \,|X_0(\theta) = x\right] f(x;\theta), \tag{2.12}$$

$$B_{V,2} = \sup_{x,x'\in\mathbb{R}^{d_x}} \sup_{\theta\in\Theta} E\left[\|V_0(\theta)\| \,\|V_T(\theta)\| \,|X_0(\theta) = x, X_T(\theta) = x'\right] f_T\left(x,x';\theta\right). \tag{2.13}$$

We also write the model in equation (2.5) more compactly as

$$Z_s(\theta) = r\left(Z^{s-1}(\theta), U_s; \theta\right)$$

where $Z_s(\theta) = (Y_s(\theta), W_s(\theta))$ and $r = (r_y, r_w)$, and then define the differentiated process $\dot{Z}_s(\theta) = \left(\dot{Y}_s(\theta), W_s(\theta)\right)$ as

$$\dot{Z}_s(\theta) = \frac{\partial r\left(Z^{s-1}(\theta), U_s; \theta\right)}{\partial Z^{s-1}(\theta)'} \dot{Z}^{s-1}(\theta) + \frac{\partial r\left(Z^{s-1}(\theta), U_s; \theta\right)}{\partial \theta}.$$

Assumption 3. *For any given $\theta \in \Theta$, there exists a stationary solution $\{(Y_s(\theta), W_s(\theta))\}$ to the model in eq. (2.5). This solution is α-mixing with mixing coefficients $\alpha_s(\theta)$ satisfying $\alpha_s(\gamma) \leq As^{-\beta}$ for some $0 < A, \beta < \infty$ (which do not depend on θ).*

Assumption 4. *The function $r(z,u;\theta)$ is $m \geq 2$ times continuously differentiable w.r.t. (z,θ) and the set of test functions, $\phi(y_t,...,y_{t+p})$, is continuously differentiable.*

Assumption 5. *For some $\mu \geq 2$,*

$$E\left[\|\phi(Y_s(\theta))\|^\mu\right] < \infty, \quad E[\|\phi(Y_s(\theta))\,\phi'(Y_s(\theta))\,\dot{Y}_s(\theta)\|^\mu] < \infty, \quad E[\|\phi(Y_s(\theta))\|^p \|\dot{X}_s(\theta)\|^\mu] < \infty.$$

With $\bar{d} = d_x + d_y$, the mixing exponent β in assumption (3) satisfies:

$$\beta > \frac{1 + (\mu - 1)\left(1 + \bar{d}/\lambda + \bar{d}\right)}{\mu - 2}.$$

The bounds defined in Eq. (2.12)-(2.13) are finite for $V = Y$, $V = \dot{Y}$ and $V = Y\dot{X}$.

Assumption 3 says that the simulated path $\{(Y_s(\theta), W_s(\theta))\}$ generated from equation (2.5) is stationary for any given value of θ. Thus, we implicitly assume that we are able to initialize the process at its stationary distribution. In practice this is not possible but due

to the assumption of α-mixing, we know that $(Y_s(\theta), W_s(\theta))$ will converge towards the stationary solution as $s \to \infty$. Thus, if we simulate a long enough trajectory ($S \to \infty$), we expect that the kernel estimator based on the non-stationary solution will be asymptotically equivalent to the one based on stationary one. A complete analysis, taking into account the discrepancy between the non-stationary and stationary simulated solution in finite samples, will however not be given here since it will involve much longer proofs and more complicated assumptions on the model. For an analysis in the case of unconditional SMM, we refer to Duffie and Singleton (1993) where it is demonstrated that the estimator remains consistent without having to initialize at the stationary distribution.

For a set of sufficient conditions for assumption 3 to hold for Markov models, we refer to Kristensen (2008b). The condition could probably be weakened to assume that the process is recurrent (and therefore not stationary) since it is possible to show pointwise convergence of kernel estimators of conditional moments in this setting (see, for example, Karlsen and Tjoestheim, 2001). However, just showing pointwise convergence gets markedly more difficult in this setting; furthermore, to our knowledge, no uniform convergence results for non-stationary processes are currently available. The smoothness conditions in assumption 4 hold for most dynamic models; they do however rule out discontinuous models such as threshold models. Finally, the moment conditions imposed in assumption 5 can also be verified through the general results of Kristensen (2008b).

We impose regularity conditions on the kernel function K:

Assumption 6. *The kernel K satisfies:*

(1) $|K(u)| \leq \bar{K} < \infty$ *and* $\int |K(u)|\,du \leq \mu < \infty$. *There exist* $\Lambda, L < \infty$ *such that either (i)* $K(u) = 0$ *for* $\|u\| > L$ *and* $|K(u) - K(u')| \leq \Lambda \|u - u'\|$, *or (ii)* $K(u)$ *is differentiable with* $|\partial K(u)/\partial u| \leq \Lambda$. *For some* $a > 1$, $\left|\partial^i K(u)/\partial u^i\right| \leq \Lambda \|u\|^{-a}$ *for* $\|u\| \geq L$ *and* $i = 0, 1$.

(2) *For some* $m \geq 1$: $\int K(u)\, u^i du = 0$, $i = 1, \ldots, m-1$, *and* $\int K(u) \|u\|^m du < \infty$.

This class of kernel allows for higher-order kernels ($m > 2$) and standard kernels ($m = 2$) such as the Gaussian one.

Finally, we need to redefine our SNM estimator in order to derive the desired theoretical results. We introduce a trimming sequence $a > 0$ in order to handle that the density of $X_t(\theta)$, $f(x;\theta)$, in general is not bounded away from zero. For our theoretical results, we then need to redefine our SNM estimator to trim away observed values for which $f(x;\theta) \geq a$. That is, we redefine $\hat{g}_{t,S}(\theta)$ as

$$\hat{g}_{t,S}(\theta) = \mathbb{I}\left\{\hat{f}(x_t;\theta) \geq a\right\} Z_t' \hat{\varepsilon}_t(\theta),$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function and $\hat{f}(x;\theta) = \sum_{s=1}^{S} K_h(X_s(\theta) - x)/S$ is the simulated kernel estimator of $f(x;\theta)$. Replacing $\hat{g}_{t,S}(\theta)$ with the above new definition, the SNM estimator is still given by equations (2.9) and (2.10). In practice, for S sufficiently large, trimming is most likely not required since it is used to control deviations between the simulated and actual version of $T(\phi)(x_t;\theta)$.

Theorem 1. *Assume that assumptions 1-6 hold. Then the SNM estimator satisfies:*

$$||\hat{\theta}_{n,S} - \hat{\theta}_n|| = O_P\left(a^{-1}h^m\right) + O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{d_x})}\right) + O_P\left(a^{-q}\right). \tag{2.14}$$

In particular, if

$$\sqrt{n}a^{-1}h^m \to 0, \quad \sqrt{n}a^{-1}\sqrt{\log(S)/(Sh^{d_x})} \to 0, \quad \sqrt{n}a^{-q} \to 0,$$

then the SNM estimator is first-order equivalent to $\hat{\theta}_n$ *and:*

$$\sqrt{n}(\hat{\theta}_{n,S} - \theta_0) \xrightarrow{d} N\left(0, \left(H_0'W_0H_0\right)^{-1} H_0'W_0\Omega_0W_0H_0\left(H_0'W_0H_0\right)^{-1}\right). \tag{2.15}$$

Proof. See the Appendix. □

By making S suitably large, it is possible to make $\hat{T}(\phi)(x;\theta)$ as close as is desired to the true moment $T(\phi)(x;\theta)$. In principle, S could be chosen large enough so that the differences between the error functions in equations (2.2) and (2.7) are smaller than the machine precision of a digital computer. If this is the case, the SNM estimator essentially *is* the infeasible GMM estimator.

2.3. **Discussion.**

2.3.1. *Illustration of equivalence of SNM and infeasible GMM.* A simple Monte Carlo exercise serves to illustrate the equivalence of the SNM and GMM estimators, when the GMM estimator is feasible. Samples of the observable variables $\{(y_{t,1}, y_{t,2})\}$, $t = 1, 2, ..., 30$, were generated using the classical linear model (CLM)

$$\text{CLM:} \begin{cases} y_{t,1} & = \beta_1 + \beta_2 y_{t,2} + u_t \\ y_{t,2} & \sim U(0,1) \\ u_t & \sim N(0,1) \end{cases}. \tag{2.16}$$

The parameters β_1 and β_2 were randomly drawn (separately) from $U(0,1)$ distributions at each of 1000 Monte Carlo replications. The single test function is chosen as $\phi_1(y_t) = y_t$, and the single conditioning variable is $x_t = y_{t,2}$. For this DGP, the maximum likelihood (ML) estimator is the ordinary least squares (OLS) estimator obtained by regressing y on a constant and x. The ML estimator may be thought of as a GMM estimator that uses the single ($k_y = 1$) error function $\varepsilon_t = y_t - \beta_1 - \beta_2 x_t$ and the instruments $(1, x_t)$. The SNM estimator was applied, using the endogenous variable y_t, the conditioning variable x_t and instruments $(1, x_t)$. The simulation length was $S = 500000$, and the $h_S = S^{-1/(4+k_x)}$ is chosen using a simple rule-of-thumb procedure.[2] A standard Gaussian kernel was used.

- **DENNIS:** Wouldn't a (slightly) more interesting example be

$$y_t = \mu + \rho y_{t-1} + u_t.$$

? At least this is an actual time series model where data is correlated.

[2] See Li and Racine, 2007, pg. 66. Recall that k_x is the number of conditioning variables ($k_x = 1$ in the present case).

Table **??** gives results that compare the distribution of the difference between the SNM and GMM estimators to the distribution of the GMM estimator, over the 1000 Monte Carlo replications. We can see that the difference between the two estimators is distributed tightly around zero, and that the dispersion of the difference is much less than that of the GMM estimator. If the value of the SNM estimator is regressed on a constant, the value of the GMM estimator, and the value of the true parameter, the results are (estimated standard errors in parentheses), for the constant, β_1:

$$\widehat{\beta}_1(SNM) = \underset{(0.00023012)}{-0.00106912} + \underset{(0.00030566)}{1.00292}\,\widehat{\beta}_1(GMM) - \underset{(0.00050332)}{0.00267236}\,\beta_1$$

For the slope, β_2, we obtain

$$\widehat{\beta}_2(SNM) = \underset{(0.00038392)}{2.50475\text{e-}5} + \underset{(0.00029626)}{1.00389}\,\widehat{\beta}_2(GMM) - \underset{(0.00073023)}{0.000178451}\,\beta_2$$

In both cases, R^2 is higher than 0.999. We see that the SNM and GMM estimators are essentially identical, independent of the true parameter value.

Recall that the GMM estimator is fully asymptotically efficient for this model. Comparing root mean squared error (RMSE) over the 1000 Monte Carlo replications, the RMSE of the SNM estimator relative to RMSE of the fully efficient GMM estimator is 1.003 in the case of β_1, and 1.004 in the case of β_2. Since the estimators are essentially the same, so are their efficiencies. The SNM estimator can be very efficient if moment conditions are well-chosen.

These results illustrate the fact that when a long enough simulation is used the SNM estimator essentially *is* the GMM estimator that uses the same endogenous variables and the same conditioning variables. The GMM estimator adds information about the functional form of the moment condition, while the SNM estimator fits it nonparametrically. When S is large enough, the nonparametric fit is so good that the SNM estimator is practically identical to the GMM estimator. Of course, one would only use the SNM estimator when the functional form of $T(\phi)(x;\theta)$ is unknown, so that the GMM estimator is infeasible.

2.3.2. *Choice of Test Functions and Instruments.* An integral part of the proposed estimation procedure is the set of test functions, $\phi_1, ..., \phi_L$, and the instruments. We here discuss in turn how these can be chosen.

Regarding the test functions, these can either be chosen in a model-specific manner or in non-model-based way. In the model-specific procedure for choosing test functions, the researcher chooses different test functions depending on the model. For a given model, he/she chooses (a small number of) test functions that he/she believes identify the parameters of interest. An example of this approach can be found in our simulation study where we consider a stochastic volatility model; the parameters in this model describe the dynamics of the conditional second moment of the observed process and we therefore choose test functions mirroring this fact.

In the non-model based method, the researcher uses (a relatively large number) test functions that (approximately) span the unknown score function. Examples of test functions within this approach are Hermite polynomials (Bansal et al, 1994; Gallant and Tauchen, 2002) and the characteristic function (Carrasco et al, 2007; Chacko and Viceira, 2005). To illustrate how this approach can be implemented within our setting, consider the characteristic function approach: We would in this case choose

$$\phi_k(y_t) = \exp\left(i\tau_k' y_t\right), \quad k = 1, ..., L,$$

where $\tau_k \in \mathbb{R}^{d_y}$, $k = 1, ..., L$, is a collection of points chosen by the researcher. As L grows larger, the GMM estimator will converge towards the (quasi-)maximum likelihood estimator based on $f(y_t|x_t;\theta)$, c.f. Carrasco et al. (2007).

A representation of optimal instruments within our setting can be found in Anatolyev (2003) where it is shown that the optimal instruments solve a stochastic recursion equation involving conditional means and variances of the residual function and its Jacobian; see also Hansen et al. (1988). However, solving this recursion equation is infeasible in practice, except in a few special cases. A feasible method is either to (i) approximate the optimal instruments (Anatolyev, 2002), or (ii) restrict the instruments to belong to a smaller, tractable class of processes.

An example of (ii) can be found in Christensen and Sørensen (2008, Section 4): They derive optimal instruments within the following finite-dimensional class of instruments: For a given set of functions $\xi_j : \mathbb{R}^{d_x} \mapsto \mathbb{R}$, $k = 1, ..., N$, the instruments are restricted to be on the form

$$Z_t = \sum_{k=1}^{N} a_k \xi_k(x_t) \in \mathbb{R}^{L \times d_\theta},$$

for some coefficients $a_k \in \mathbb{R}^{L \times d_\theta}$. The functions could for example be chosen as indicator functions, $\xi_k(x) = \mathbb{I}\left\{x \leq x_k^*\right\}$, $k = 1, , ..., N$, where $x_1^*, ..., x_N^*$ are given (non-stochastic) points (see e.g. Domínguez and Lobato, 2004), or as Hermite polynomials. Define $A = [a_1'|\cdots|a_N'] \in \mathbb{R}^{d_\theta \times LN}$. The optimal weights are then shown to be on the form (Christensen and Sørensen, 2008, Theorem 4.1),

$$A^*(\theta) = s(\theta) V(\theta)^{-1},$$

where $s(\theta) = (s_1(\theta), ..., s_N(\theta)) \in \mathbb{R}^{d_\theta \times LN}$ and $V(\theta) = E\left[H_n(\theta) H_n(\theta)'\right] \mathbb{R}^{LN \times LN}$ with

$$s_k(\theta) = E\left[\xi_k(x_t) \frac{\partial T(\phi)(x_t;\theta)}{\partial \theta}\right] \in \mathbb{R}^{d_\theta \times L}, \quad H_{n,k}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \xi_k(x_t)\varepsilon_t(\theta) \in \mathbb{R}^L.$$

These two sets of moments, $s(\theta)$ and $V(\theta)$, can easily be implemented using simulations.

2.3.3. *Optimal weight matrix.* Given that the SNM estimator has the same asymptotic distribution as the infeasible GMM estimator, one can use standard methods and asymptotic results for GMM estimators to make statistical inferences with the SNM estimator. For example, supposing that an estimated optimal weight matrix is used, an overidentified model's

specification may be tested using the familiar χ^2 test based upon $nG_n(\theta)$ where $G_n(\theta)$ is defined in equation (2.9).

In order to conduct any inference, we need an estimator of the asymptotic covariance matrix of the moment conditions $\Omega_0 = \lim_{n\to\infty} nE\left[G_n(\theta_0)\,G_n(\theta_0)'\right]$, as given in eq. (2.11). Also, a consistent estimator of this matrix is needed if one wishes to use an efficient weight matrix to estimate θ_0. In the ordinary GMM setting without a fully simulable model, this covariance matrix must be estimated using only the sample data, which requires use of one of the kernel-based heteroscedasticity and autocorrelation-consistent covariance matrix estimators (for example, that of Newey and West, 1987). It is well-known that inferences based upon such covariance estimators can be quite unreliable (Hansen, Heaton and Yaron, 1996; Windmeijer, 2005).

In the context of the SNM estimator, or any other moment-based estimator that relies on a fully simulable model, it is possible to estimate Ω_0 though Monte Carlo, in much the same way as was proposed of estimation of Φ_0 above. We first note that $\Omega_0 = \lim_{n\to\infty} nE\left[G_n(\theta)\,G_n(\theta)'\right]$. The idea now is to estimate the moment on the right hand side of this equation using an average of R random draws of $G_n(\theta)\,G_n(\theta)'$, replacing the real sample test and conditioning variables with independent simulations from the model, given an initial consistent estimate of the model's parameter, $\hat{\theta}_{n,S}$, as was discussed in detail in the previous section. We may generate $R \geq 1$ such samples of size n, and for each of them calculate simulated moment conditions as in equation (2.9). Given the rth such replication of the test and conditioning variables, $(\hat{Y}_t^{(r)}, \hat{X}_t^{(r)})\}_{t=1}^n$ $(r = 1, 2, ..., R)$, we then compute $\hat{G}_{n,S}^{(r)}(\hat{\theta}_{n,S})$ in exactly the same way as $\hat{G}_{n,S}(\theta)$ is computed in Eq. (2.4), except that the simulated data at $\hat{\theta}_{n,S}$ replaces the real sample data. We then define $v_r = \hat{G}_{n,S}^{(r)}(\hat{\theta}_{n,S}) - \bar{G}$, where $\bar{G} = R^{-1}\sum_{r=1}^R \hat{G}_{n,S}^{(r)}(\hat{\theta}_{n,S})$, and obtain the following estimator of Ω_0:

$$\hat{\Omega}_0 = \frac{n}{R}\sum_{r=1}^{R} v_r v_r'. \tag{2.17}$$

This procedure requires that the average moments be simulated many times. The average moments depend in part upon the instruments that have been chosen. When the instruments are simple functions of the $\hat{X}_t^{(r)}$, they may be calculated quickly for each replication of the simulated data, and the proposed procedure is not unduly demanding, in terms of computational resources. If one wishes to use optimal instruments as was discussed in subsection (2.3.2), then in principle it would be necessary to nest a second simulation loop inside the first, to calculate the instruments that are optimal for each of the R simulated samples. This double loop entails too much computation for this proposal to be of practical, given the computational resources available to most researchers.

A second possibility for estimation of Ω_0 is to recognize that an alternative representation of the covariance matrix is

$$\Omega_0 = \lim_{n\to\infty} E\left[\frac{1}{n}\mathbf{Z}'\varepsilon(\theta)\varepsilon(\theta)'\mathbf{Z}\right],$$

where $\varepsilon(\theta)$ and $\mathbf{Z}$ are defined in the previous subsection. This suggests using the following estimator for Ω_0:

$$\tilde{\Omega}_0 = \frac{1}{n}\mathbf{Z}'\hat{\Phi}_0\mathbf{Z}, \tag{2.18}$$

where $\hat{\Phi}_0$ is an estimator of the covariance matrix of $\varepsilon(\theta)$. This estimator requires that the optimal instruments be calculated only a single time, for the real sample data. Monte Carlo estimation is used only for the covariance of the residuals, which does not involve the instruments. This estimator could of course be used for any choice of instruments. When naive instruments are used, either $\hat{\Omega}_0$ or $\tilde{\Omega}_0$ could be used to define an estimated optimal weight matrix. When optimal instruments are used, only $\tilde{\Omega}_0$ will be feasible in most cases.

Once the covariance of the moments is estimated, hypothesis testing may then be done using standard results for GMM estimators with an inefficient weight matrix, or a the estimate of Ω_0 may be used in a second round of estimation to compute a more efficient estimate of θ_0, as is standard practice.

- **DENNIS:** I'm not convinced about the validity of the estimator $\tilde{\Omega}_0 = \frac{1}{n}\mathbf{Z}'\hat{\Phi}_0\mathbf{Z}$ above.

  It seems as if you implicitly assume that the errors are homoskedastic: $E\left[\varepsilon(\theta)\varepsilon(\theta)'|\mathbf{Z}\right] = \Phi_0$ - or am I missing something?

- These methods have the advantage that they obviate the need for decisions regarding lag lengths, pre-whitening and so forth that attend the use of kernel-based covariance matrix estimators that use only the sample data. However, for use of $\hat{\Omega}_0$ we probably want to drop high order autocovariances, both to speed up calculations and to obtain a stable positive definite estimate. Need to explore this further using simulations.
- Since G is continuous in θ, this estimator of Ω should be consistent, if R goes to infinite. If R is finite, then it's asymptotically unbiased, due to consistency of $\hat{\theta}$, but inconsistent since the simulation variance doesn't go to zero. The noise in the estimator will upwardly bias the χ^2 specifiation test, so true models will be over-rejected. I think that in practise, the noise due to finite R is likely to be small compared to the noise due to the estimate of θ, but who knows? In the examples below I am using $R = 1000$, which seems quite safe to me. Maybe we want a formal proof of consistency?
- [DROP THIS?] To provide some rudimentary evidence of this covariance estimator's performance, a Monte Carlo study of 1000 replications was done. Data was generated using the classical linear model of equations (2.16). The SNM estimator was applied using a sample size $n = 30$, a simulated sample size $S = 10000$, and $R = 1000$ draws were used to estimate Ω for each of the 1000 Monte Carlo replications. The true value of Ω for this model is $\Omega_{11} = 1$, $\Omega_{12} = 1/2$, $\Omega_{22} = 1/3$. Over the 1000 Monte Carlo replications, the mean and standard errors (in parentheses) of the

replications of $\hat{\Omega}$ are $\Omega_{11} : 1.036\,(0.048)$, $\Omega_{12} : 0.517 : (0.025)$, $\Omega_{22} : 0.343 : (0.015)$. In none of the three cases is the Monte Carlo mean signigicantly different from the true value.

2.3.4. *Choice of the kernel and the bandwidth.* To implement the SNM estimator, the kernel function $K(\cdot)$ must be chosen, as must the bandwidth, h. Regarding the kernel, in this paper attention is restricted to local constant kernel regression estimators (Li and Racine, 2007). In this context, much theoretical and empirical evidence shows that the choice of the particular kernel function has relatively little effect on the results for a given bandwidth. For this reason, this paper uses Gaussian product kernels exclusively, accompanied by prior rotation of the data to approximate independence of the conditioning variables. Gaussian product kernels lead to error functions that are continuous and relatively smooth in the parameters, which facilitates iterative minimization. Kernels such as the radial symmetric Epanechnikov are relatively inexpensive to compute, but lead to error functions that can be discontinuous in the parameters, which complicates minimization of the objective function that defines the SNM estimator. This paper leaves the possibility of SNM estimation based on local linear or local polynomial kernel methods for future work.

Given the kernel function, the bandwidth must be chosen. The bandwidth does have an important effect upon the quality of the kernel regression fit. Too large a bandwidth oversmooths the data, and induces a fit with low variance but high bias. Too small a bandwidth has the opposite effect. The bandwidth may be chosen using data-driven methods such as leave-one-out cross validation, or by using rule-of-thumb methods that are known to work well in certain circumstances but may perhaps perform poorly in others. In this paper, a simple rule-of-thumb method is used throughout, since investigation of data-driven methods would add substantially to the computational burden of the Monte Carlo work presented below. It is expected that use of a data-driven method would improve the performance of the SNM estimator. Future work will address this issue more carefully.

2.3.5. *Computational issues.* Estimation of a complicated model using long simulation may become computationally burdensome, since kernel smoothing is a computationally intensive procedure. In common with normal GMM estimators (Chernozhukov and Hong, 2003, especially pp. 296-298), the SNM objective function is not globally convex, so one needs to take care to find the global minimum by using estimation methods such as simulated annealing (Goffe *et al.,* 1994). One may seek to use data-based methods to choose the bandwidth, as well. These factors imply that use of the SNM estimator is computationally intensive. However, kernel regression fitting, which is at the heart of the SNM estimator, is easily parallelized (Racine, 2002; Creel, 2005), as is Monte Carlo work (Creel, 2007). The widespread availability of multicore processors is an invitation to take advantage of parallelization opportunities in econometric work. All of the results reported in this paper were obtained on a computational cluster that provided a total of 16 CPU cores, running

the PelicanHPC distribution of GNU/Linux[3]. A special version of PelicanHPC that contains examples, documentation and all software needed to compute the SNM estimator, on a single computer or on a cluster, is available on request from the authors.

3. Monte Carlo Results for the Stochastic Volatility Model

This section presents Monte Carlo results for the logarithmic stochastic volatility model of Jacquier, Polson and Rossi (1994) that has widely been used as a test bed for estimators. Adapting the notation to conform with the general DLV model of equation (1.1), the model is

$$\text{SV:} \begin{cases} q_t &= \exp(w_t/2)\, u_{t,1} \\ w_t &= \alpha + \beta w_{t-1} + \sigma u_{t,2} \end{cases} \tag{3.1}$$

where the white noise $u_t = (u_{t,1}, u_{t,2})'$ is distributed i.i.d. $N(0, I_2)$. The stochastic volatility model of equation (3.1) will be referred to as the SV model. A slightly different parameterization is used by many authors:

$$\begin{cases} q_t &= \sigma_b \exp(w_t/2)\, u_{t,1} \\ w_t &= \beta w_{t-1} + \sigma u_{t,2} \end{cases} \tag{3.2}$$

The parameters of the two models are the same, except $\sigma_b = \exp(\alpha/2)$.

The Monte Carlo design proposed by Sandmann and Koopman (1998) has been adopted in subsequent work by a number of authors, and we adhere to this trend to facilitate comparison with other estimators. Perhaps the most widely used design uses the parameter values $\theta = (\alpha, \beta, \sigma)' = (-0.736,\, 0.9,\, 0.363)'$, so this is the case we focus on.

To apply the SNM estimator, we must choose moment conditions by specifying the test variables y_t and the conditioning variables x_t. The score function of log likelihood function would provide the best moments, so it can be a guide. Conditional on latent variables, the score vector for an observation t is just that of a $N(0, \exp y_t^*)$ random variable, which is

$$\frac{\partial \ln L(y_t;\theta)}{\partial \theta} = \left(\frac{-1}{2} + \frac{q_t^2}{2 \exp w_t} \right) \begin{bmatrix} 1 \\ w_{t-1} \\ \varepsilon_{2t} \end{bmatrix}$$

The score vector depends upon the quantities q_t^2, $\exp w_t$, w_{t-1} and ε_{2t}. Only the first of these is observable. Our idea is to define proxies for the other quantities. Note that $E q_t^2 = \exp w_t$, so q_t^2 is an unbiased proxy for $\exp w_t$. This proxy for the first latent quantity adds no information, since it is the same as the observable quantity. However, it suggests that $\log q_t^2$ can be used as a proxy for w_t. Furthermore, we can use $\log q_{t-1}^2$ as a proxy for w_{t-1}, the second of the latent quantities in the score vector.

For ε_{2t}, note that

$$\varepsilon_{2t} = \frac{w_t - \alpha - \beta w_{t-1}}{\sigma}$$

[3]PelicanHPC is described at http://pareto.uab.es/mcreel/PelicanHPC. It is the evolution of the ParallelKnoppix distribution of GNU/Linux, which was described in Creel (2007).

Substituting the proxies $\log q_t^2$ and $\log q_{t-1}^2$ in place of w_t and w_{t-1}, respectively, we obtain

$$\widehat{\varepsilon_{2t}} = \frac{\left(\log q_t^2 - \alpha - \beta \log q_{t-1}^2\right)}{\sigma}$$

as a proxy to use in place of ε_{2t}.

To make clear the relationship with the notation of subsection (2.1), the test variables we use to define the SNM are $y_t = \left(q_t^2,\ \log q_{t-1}^2\right),\ \left(\log q_t^2 - \alpha - \beta \log q_{t-1}^2\right)/\sigma\big)$. The single conditioning variable ($d_x = 1$) is $x_t = q_{t-1}^2$. This gives an example of how test functions and conditioning functions can both depend upon transformations involving current period variables and their lags (leads are not used in this example, but they could be). We experimented informally with other test and conditioning functions, but the simple setup used here gives results that are attractive enough to justify moving attention to other aspects of the modeling process.

It may appear odd that one of the test variables that is fitted using kernel regression, $\log q_{t-1}^2$, is a simple transformation of the conditioning variable, q_{t-1}^2. To explain how this can be useful, image the more extreme case of fitting the test variable q_{t-1}^2 using itself as a conditioning variable. What could be the advantage of fitting a variable conditional on its own value? In a parametric estimation context with real data, this would not make sense - we would be fitting an identity. In present context of kernel-based fitting using simulated data, the simultaneous use a variable both as a test variable and as a conditioning variable causes the objective function put a premium on values of θ that are able to make the kernel regression fit to this variable resemble a 45° line. Use of the same variable as a test variable and as a conditioning variable can help to impose stability and internal consistency on the simulations. Monte Carlo results that omitted $\log q_{t-1}^2$ as a test variable (available on request) resulted in higher root mean squared errors (RMSEs) that those reported below.

We will consider two sets of instrumental variables. The first is the simple naive set $\{1, x_t\}$. Thus we have $d_z = 2$. The same instruments are used for each of the $d_\phi = 3$ test functions. The second is the set of approximately optimal instruments that results from estimating the quantities in Eq. (**??**). We estimate only the diagonal of Φ_0, and the other elements are set to zero. Thus, the resulting instruments are only approximately optimal. Some simplification of this form is required to limit the computational burden of performing Monte Carlo replications, and it would probably be needed even in a pure estimation context. Exploration of the possibility of obtaining better approximately optimal instruments is left to future work. When approximately optimal instruments are used, we have $d_z = 3$ instruments for each of the $d_\phi = 3$ test functions, and the instruments differ for each test function.

We calculate three versions of the SNM estimator using: (a) naive instruments and naive (identity) weight matrix; (b) naive instruments and optimal weight matrix; and (c) approximately optimal instruments and corresponding optimal weight matrix. Results for case (b) use the estimate of θ_0 from case (a) to estimate the optimal weight matrix using Eq. (2.17). Case (c) uses the estimate of θ_0 resulting from case (b) to go on to estimate optimal instruments. For case (c), the optimal weight matrix estimated using Eq. (2.18).

We use sample sizes of $n = 500$ and $n = 2000$ observations, and simulation lengths of $S = 1000$ and $S = 5000$. The kernel is a simple Gaussian product kernel. The simulated and real test functions and conditioning functions are each scaled by corresponding sample standard deviation of the real sample data. The window width is set based on the simulation length using the rule-of-thumb $h = S^{-1/(d_x+4)}$, where the dimension of the conditioning functions is $d_\pi = 1$ in all cases. To estimate the covariance matrices Ω_0 and Φ_0 using the Monte Carlo procedure described above, we use 1000 random draws in each case (this is nested inside the Monte Carlo study of the SNM estimator).

The objective function that defines the SNM estimator can exhibit multiple local minima. While it is not extremely irregular, the existence of multiple local minima in a Monte Carlo context requires a programmable means of avoiding acceptance of false global minimizers. Our solution is to use a global minimization algorithm and a careful search over the plausible parameter space. We use a simulated annealing (Goffe *et al.*, 1994) routine, searching over $\theta = (\alpha,\ \beta,\ \sigma) \in \Theta = (-10.6,\ 0) \times (0.005,\ 0.995) \times (0.001,\ 1.0)$. This is a large enough parmeter space so that the final estimate is always a safe distance from the boundary. The starting point for each Monte Carlo replication is a random point drawn from a uniform density over Θ. The cooling schedule is conservative: $T_{m+1} = 0.75T_m$, where m indexes the temperature changes. Convergence criteria require that the criterion function change less than 10^{-6}, and that the width of the active search interval for each parameter be less than 10^{-3}. This tolerance for convergence is weaker than what would be typical in an estimation context with a single sample, but it is perfectly adequate in a Monte Carlo context, where the slight noise is averaged out over the many Monte Carlo replications. For all combinations of n,S and choices of instruments and weight matrices, a total of 500 Monte Carlo replications is used in each case.

Table **??** presents results for the sample size $n = 500$. Results for the SNM estimator are given in lines 1-5, while results for other estimators are given in the remaining lines, to facilitate comparison. Comparing lines 1, 2, and three, we see that use of the approximately optimal instruments (case (c), results in line 3) is not advisable, since RMSEs are uniformly smaller in lines 1 and 2. The relatively crude approximation to Φ_0 that we use, estimating only the main diagonal, appears not to be very successful. For this reason, when we increase the simulation length to $S = 5000$, in lines 4 and 5, we drop case (c). Comparing lines 1 and 4, and 2 and 5, we see that increasing the simulation length uniformly reduces RMSEs over all parameters, for both cases (a) and (b). Comparing lines 4 and 5, we see that use of the optimal weight matrix substantially improves the precision of estimation of α and σ, while it leads to a slight increase in RMSE in the case of β. Overall, the best version of the SNM estimator is that using naive instruments and the optimal weight matrix, with the longer simulation (results in line 5).

Next we compare the results for the preferred version of SNM to those for a number of alternative estimators (given in lines 6 - 10). For the parameter α, SNM does considerably better than GMM, EMM and MCMC, while SMHD does somewhat better than SNM and MCL is a great deal better than all other estimator. For β, SNM is a little better than GMM,

EMM and MCL. The MCMC estimator is a good deal better than SNM, and SMHD is much better than all competitors. For σ, SNM has a RMSE that is less than half that of MCMC, the closest rival. Overall, the performance of SNM is consistently good, and it clearly dominates the two other moment-based approaches (GMM and EMM).

Table **??** presents results for the sample size $n = 2000$.

4. EXTENSIONS

A number of extensions of the proposed estimator are available. We here discuss how our method can be extended to allow for non-stationarity and how it can be adjusted to obtain a consistent estimator without requiring $S \to \infty$.

4.1. **Non-Stationary Models.** We have worked under the maintained assumption that the process is stationary. If the the data-generating process in question is non-stationary, our simulated estimator will no longer work. Suppose for example, we wish to base our estimation on the following conditional moment, $E_\theta[y_t|y_{t-1}]$. Under non-stationarity, the distribution of (y_t, y_{t-1}) will in general change over time and so the conditional moment $E_\theta[y_t|y_{t-1}]$ is no longer time-invariant (in contrast to the stationary case). To handle this case, we now introduce an alternative simulation scheme: First, simulate S independent trajectories each of length n, $\{Y_{s,t}(\theta) : t = 1,...,n\}$, $s = 1,....,S$, where the sth trajectory is computed as:

$$\begin{cases} Y_{s,t}(\theta) = & r_y\left(Y_s^{t-1}(\theta), W_s^{t-1}(\theta), U_{s,t}; \theta\right) \\ W_{s,t}(\theta) = & r_w\left(Y_s^{t-1}(\theta), W_s^{t-1}(\theta), U_{s,t}; \theta\right) \end{cases}, \quad t = 1,...,n. \tag{4.1}$$

We here assume that we have observed the initial values (y_0, w_0) and then start the simulations there, $(Y_{s,0}(\theta), W_{s,0}(\theta)) = (y_0, w_0)$.[4] We then compute

$$\begin{aligned} X_{s,t}(\theta) &= x(Y_{s,t-1}(\theta),...,Y_{s,t-q}(\theta)), \\ \Phi_{s,t}(\theta) &= \phi(Y_{s,t}(\theta),...,Y_{s,t+p}(\theta)), \end{aligned}$$

and

$$\hat{T}_{t,S}(\phi)(x;\theta) = \frac{\sum_{s=1}^S \Phi_{s,t}(\theta) K_h(X_{s,t}(\theta) - x)}{\sum_{s=1}^S K_h(X_{s,t}(\theta) - x)}, \quad t = 1,...,n. \tag{4.2}$$

By construction, $(\Phi_{s,t}(\theta), X_{s,t}(\theta))$ are i.i.d. simulations from the target distribution at time t, $(\Phi_{s,t}(\theta), X_{s,t}(\theta)) \sim f_t(\phi, x; \theta)$, $s = 1,...,S$. Thus, as $h \to 0$ and $Sh^{d_x} \to 0$,

$$\hat{T}_{t,S}(\phi)(x;\theta) \to^P T_t(\phi)(x;\theta) = E_\theta[\phi_t|x_t = x] = \int \phi f_t(\phi|x;\theta).$$

We now proceed as in the stationary case and define $\hat{\varepsilon}_{t,S}(\theta) = \phi_t - \hat{T}_{t,S}(\phi)(x_t;\theta)$, and $\hat{g}_{t,S}(\theta) = Z_t'\hat{\varepsilon}_{t,S}(\theta)$. The asymptotic properties of the resulting estimator are however not covered by Theorem 1, since in general the infeasible GMM-estimator does not satisfy Assumption 2. For theoretical results in non-stationary environments, see Kitamura

[4] Alternatively, one can impose a prior on w_0 and simulate from this.

and Phillips (1997) for some results for linear models and Kristensen and Shin (2008) for simulation-based likelihood-inference.

4.2. **Unbiased Simulator.** The current kernel estimator $\hat{T}_S(\phi)(x_t;\theta)$ in eq. (2.6) has a bias of order h^m. An alternative specification of the SNM estimator that leads to an unbiased simulator can be constructed by following the main idea of Altissimo and Mele (2009); see also Billio and Monfort (2003). The estimator takes as starting point the following redefined residual functions,

$$\varepsilon_{t,h}(\theta) = R_h(\phi)(x_t) - R_h(\phi)(x_t;\theta), \tag{4.3}$$

where

$$R_h(\phi)(x_t;\theta) := E_\theta\left[\phi(y_t) K_h(x_t - x)\right],$$

and $R_h(\phi)(x_t) = R_h(\phi)(x_t;\theta_0)$. A simulated version can then be obtained as

$$\hat{\varepsilon}_t(\theta) = \tilde{R}(\phi)(x_t) - \hat{R}(\phi)(x_t;\theta), \tag{4.4}$$

where $\tilde{R}(\phi)(x)$ and $\hat{R}(\phi)(x;\theta)$ are kernel estimators using actual and simulated data respectively,

$$\tilde{R}(\phi)(x) = \frac{1}{n}\sum_{t=1}^{n} \phi(y_t) K_h(x_t - x), \tag{4.5}$$

$$\hat{R}(\phi)(x;\theta) = \frac{1}{S}\sum_{s=1}^{S} \phi(Y_s(\theta)) K_h(X_s(\theta) - x). \tag{4.6}$$

We now have that $\hat{R}(\phi)(x;\theta)$ is an unbiased estimator of $R(\phi)(x;\theta)$ such that we will obtain consistency for fixed h and S by the same arguments as in Duffie and Singleton (1993). The identification condition for fixed h now becomes

$$R_h(\phi)(x_t;\theta_0) = R_h(\phi)(x_t;\theta) \text{ a.s. } \Leftrightarrow \theta = \theta_0;$$

see Altissimo and Mele (2009) for a further discussion. Furthermore, as $h \to 0$ and $Sh^{d_x} \to \infty$,

$$\hat{R}(\phi)(x;\theta) \to^P T(\phi)(x;\theta) f(x;\theta),$$

where $f(x;\theta)$ is the density of $f(x;\theta)$. Thus, by choosing the instruments appropriately, the estimator based on $\hat{R}(\phi)(x;\theta)$ is equivalent to the SNM estimator as $h \to 0$. Finally, note that the above estimator shares some similarities with the one of Domínguez and Lobato (2004).

5. Conclusion

This paper has proposed a simulated method of moments estimator that allows use of conditional moments, in the case of general dynamic latent variable models. The estimator is consistent and asymptotically normally distributed, with the same asymptotic distribution as that of the infeasible GMM estimator defined by the same moment conditions. The Monte Carlo results show that use of conditional moments allows the proposed simulated method

of moments estimator to obtain efficiency that is very competitive with other estimation methods.

The SNM estimator relies on the user specifying the moment conditions to use in estimation, as is the case with any method of moments estimator, but the rest of the process can be automatized in software to a high degree. In the present implementation, the kernel function is a Gaussian product kernel, and the bandwidth is chosen using a given rule that depends only on the number of conditioning variables and on the simulation length. One can use the proposed Monte Carlo estimator of the efficient weight matrix that requires no tuning or pre-whitening decisions. Some of the other estimators to which the SNM estimator is compared in this paper require much more active decision making on the part of the modeler. An example is the newer version of the EMM estimator that uses MCMC methods, as presented in Gallant and Tauchen (2007). This version of EMM requires estimation of a SNP density augmented by a leading parametric model to define the score generator. Selection of the parameterization of the score generator is complicated by the fact that it involves many parameters. After estimation of the score generator, the model is estimated using MCMC methods that also require judgement about proper tuning of the Markov chain. Another example is the KBII estimator proposed by Billio and Monfort. Selection of the points at which the binding functions are evaluated is a non-trivial issue which requires judgement. The Monte Carlo results reported here suggest that the SNM estimator can give good performance without requiring the modeler to make any decisions other than the set of moment conditions to use.

The Monte Carlo results provided in this paper show that the SNM estimator achieves root mean squared errors that are often better than those of alternative estimators, and are rarely worse. These results are quite acceptable as they stand, but it is anticipated that they may be improved upon in the future, for two reasons. First, use of an estimated optimal weight matrix is likely to improve efficiency of estimation. Preliminary results suggest that the covariance matrix of the moment conditions can be estimated quite reliably using a Monte Carlo estimator. Future work will investigate the performance of the SNM estimator using an estimated optimal weight matrix. Secondly, a data-based method of choosing the smoothing parameter could improve the fit of the kernel smoother to the true conditional expectations, which would likely improve the results of the SNM estimator. These are simple, obvious extensions to expore. Additional topics for further research include methods to obtain a high precision fit to the conditional moments that define the estimator while using less computational time. Possibilities include the use of sieve estimation methods instead of kernel smoothing, use of approximate nearest neighbors, and use of high performance algorithms for kernel smoothing, such as the improved fast Gauss transform (Yang *et al.*, 2003). Use of an optimal bandwidth may also be helpful for this purpose, since it may be possible to obtain the same quality of fit to $T(\phi)(x;\theta)$ while using a shorter simulation length. Another interesting possibility is to attempt to use optimal or approximately optimal instruments. Use of a local linear kernel function instead of the local constant kernel used in this paper would automatically provide estimates of the derivatives $\partial T(\phi)(x;\theta)/\partial x'$ (see

equation 2.2) of conditional moments with respect to each of the conditioning variables, which could be of use in attempting to approximate optimal instruments.

A CDROM image file that provides the current implementation of the SNM estimator with documentation and all code needed to replicate the Monte Carlo results for the SV model is available on request from the authors.

APPENDIX

APPENDIX A. PROOFS

Proof of Theorem 1. We claim that

$$\begin{aligned}(A.1)\qquad &\sup_{\theta\in\Theta}\left|\hat{G}_{n,S}(\theta)'W_n\hat{G}_{n,S}(\theta)-G_n(\theta)'W_nG_n(\theta)\right| \\ &= O_P\left(a^{-1}h^r\right)+O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{k_x})}\right)+O_P\left(a^{-q}\right).\end{aligned}$$

If this holds, it follows by standard results, see e.g. Kristensen and Salanié (2008, Theorem 1), that the asserted result in eq. (2.14) is true. To prove the claim, write

$$\begin{aligned}&\sup_{\theta\in\Theta}\left|\hat{G}_{n,S}(\theta)'W_n\hat{G}_{n,S}(\theta)-G_n(\theta)'W_nG_n(\theta)\right| \\ =\;&\sup_{\theta\in\Theta}\left|\left[\hat{G}_{n,S}(\theta)-G_n(\theta)\right]'W_n\left[\hat{G}_{n,S}(\theta)+G_n(\theta)\right]\right| \\ \leq\;&\sup_{\theta\in\Theta}\left\|\hat{G}_{n,S}(\theta)-G_n(\theta)\right\|\times\sup_{\theta\in\Theta}\left\{\left\|\hat{G}_{n,S}(\theta)\right\|+\|G_n(\theta)\|\right\}\times\|W_n\| \\ \leq\;&A_2\times(A_1+A_2)\times\|W_n\|,\end{aligned}$$

where

$$A_1=\sup_{\theta\in\Theta}\|G_n(\theta)\|,\quad A_2=\sup_{\theta\in\Theta}\left\|\hat{G}_{n,S}(\theta)-G_n(\theta)\right\|.$$

First, by assumption 2, $\|W_n\|=O_P(1)$ while

$$A_1\leq\sup_{\theta\in\Theta}\|G_n(\theta)-G(\theta)\|+\sup_{\theta\in\Theta}\|G(\theta)\|=O_P(1)+O(1).$$

Second,

$$\begin{aligned}A_2\;&\leq\;\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{L}\left|\hat{T}(\phi_i)(x_t;\theta)-T(\phi_i)(x_t;\theta)\right|\|z_t\| \\ &\leq\;\frac{1}{n}\sum_{t=1}^{n}\|z_t\|\times\sum_{i=1}^{L}\sup_{\theta\in\Theta}\sup_{x:\hat{f}(x)\geq a}\left|\hat{T}(\phi_i)(x_t;\theta)-T(\phi_i)(x_t;\theta)\right| \\ &\quad+\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{L}\mathbb{I}\{\hat{f}(x_t)<a\}\left|T(\phi_i)(x_t;\theta)\right|\|z_t\| \\ &=\;A_{2,1}+A_{2,2}.\end{aligned}$$

Here, $A_{2,1}=O_P\left(a^{-1}h^m\right)+O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{k_x})}\right)$ by Lemma 3 together with the fact that $n^{-1}\sum_{t=1}^{n}\|z_t\|=O_P(1)$, while the second term satisfies:

$$\begin{aligned}A_{2,2}\;&\leq\;\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}\left\{\frac{a^q}{f(x_t)^q}>1\right\}\left|T(\phi_i)(x_t;\theta)\right|\|z_t\| \\ &\leq\;a^q\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}f(x_t)^{-q}\left|T(\phi_i)(x_t;\theta)\right|\|z_t\| \\ &=\;O_P(a^q).\end{aligned}$$

This completes the proof of eq. (A.1) and thereby (2.14).

To show eq. (2.15), write

$$\sqrt{n}(\hat{\theta}_{n,S}-\theta_0)=\sqrt{n}(\hat{\theta}_{n,S}-\hat{\theta}_n)+\sqrt{n}(\hat{\theta}_n-\theta_0).$$

where the first term is $o_P(1)$ by (2.14) in conjunction with the conditions imposed on a,h and S, while the second term converges in distribution by Lemma 2. □

APPENDIX B. LEMMAS

Lemma 2. *Under Assumptions 1-2,*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\rightarrow^d N\left(0,\left(H_0'W_0H_0\right)^{-1}H_0'W_0\Omega_0W_0H_0\left(H_0'W_0H_0\right)^{-1}\right),$$

where H_0, Ω_0 and W_0 are given in Assumption 2.

Proof. First, consistency follows from Newey and McFadden (1994, Theorem 2.1). To show asymptotic normality, we first note that $\hat{\theta}_n$ satisfies the following first order condition:

$$H_n(\hat{\theta}_n)'W_nG_n(\hat{\theta}_n)=0, \tag{B.1}$$

where $H_n(\theta)$ is given in Assumption 2. A Taylor expansion of $G_n(\hat{\theta}_n)$ around θ_0 yields

$$G_n(\hat{\theta}_n)=G_n(\theta_0)+H_n(\bar{\theta}_n)(\hat{\theta}_n-\theta_0)$$

where $\bar{\theta}_{n,i}\in[\hat{\theta}_{n,i},\theta_{0,i}]$, $i=1,...,\dim(\theta)$. Substituting this into (B.1), and rearranging the terms yields

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=\left[H_n(\hat{\theta}_n)'W_nH_n(\bar{\theta}_n)\right]^{-1}H_n(\hat{\theta}_n)'W_n\sqrt{n}G_n(\theta_0),$$

which converges in distribution towards the desired normal distribution by the Assumption 2. □

Lemma 3. *Under Assumptions 3-6, the simulated conditional moment estimator satisfies:*

$$\sup_{\theta\in\Theta}\sup_{x:\hat{f}(x)\geq a}\left|\hat{T}(\phi_i)(x;\theta)-T(\phi_i)(x;\theta)\right|=O_P\left(a^{-1}h^r\right)+O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{d_x})}\right).$$

Proof. For a given test function, ϕ_i, we write $\hat{T}(\phi_i)(x;\theta)=\hat{g}(x;\theta)/\hat{f}(x;\theta)$, where

$$\hat{g}(x;\theta)=\frac{1}{S}\sum_{s=1}^{S}\phi_i(Y_s(\theta))K_h(X_s(\theta)-x),\quad \hat{f}(x;\theta)=\frac{1}{S}\sum_{s=1}^{S}K_h(X_s(\theta)-x).$$

It is easily checked that Assumptions 2-6 imply the conditions (A.1)-(A.5) in Kristensen (2008a). His Theorem 2 combined with standard bias expansions of higher-order kernel estimators now yield that

$$\begin{aligned}\sup_{\theta\in\Theta}\sup_x|\hat{g}(x;\theta)-g(x;\theta)| &= O_P(h^m)+O_P\left(\sqrt{\log(S)/(Sh^{d_x})}\right),\\ \sup_{\theta\in\Theta}\sup_x|\hat{f}(x;\theta)-f(x;\theta)| &= O_P(h^m)+O_P\left(\sqrt{\log(S)/(Sh^{d_x})}\right),\end{aligned}$$

where $g(x;\theta)=T(\phi_i)(x;\theta)f(x;\theta)$. It now follows by using the same arguments as in Andrews (1995, Proof of Theorem 1) that the claimed result holds. □

APPENDIX C. TABLES

REFERENCES

[1] Altissimo, F. and A. Mele (2009) Simulated Nonparametric Estimation of Dynamic Models, forthcoming in *Review of Economic Studies*.

[2] Anatolyev, S. (2002) Approximately optimal instrument for multiperiod conditional moment restrictions. Manuscript, New Economic School.There and

[3] Anatolyev, S. (2003) The form of the optimal nonlinear instrument for multiperiod conditional moment restrictions, *Econometric Theory* 19, 602-609.

[4] Andersen, T. and B. Sørensen (1996) GMM estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Business & Economic Statistics*, **14**, 328-352.

[5] Andersen, T., H.-J. Chung and B. Sørensen (1999) Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Econometrics,* **91**, 61-87.

[6] Andrews, D.W.K. (1995) Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, **11**, 560-596.

[7] Bansal, R., A.R. Gallant, R. Hussey, and G. Tauchen (1994) Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data. *Journal of Econometrics,* **66**, 251-287.

[8] Bierens, H. (1987) Kernel estimators of regression functions, in *Advances in Econometrics*, Vol 1 (ed. T. Bewley), Cambridge University Press, 99-144.

[9] Bibby, B. M. and M. Sørensen (1995). Martingale Estimation Functions for Discretely Observed Diffusion Processes, *Bernoulli*, **1**, 17-39.

[10] Billio, Monica and Alain Monfort (2003) Kernel-based indirect inference, *Journal of Financial Econometrics,* **1**, 297-326.

[11] Carrasco, M., M. Chernov, J.-P. Florens and E. Ghysels (2007) Efficient estimation of general dynamic models with a continuum of moment conditions, *Journal of Econometrics* **140**, 529-573.

[12] Chacko, G. and L. Viceira (2005) Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, **116**, 259–292.

[13] Chernozhukov, V. and H. Hong (2003) An MCMC approach to classical estimation, *Journal of Econometrics*, **115**, 293-346.

[14] Christensen, B.J. and M. Sørensen (2008) Optimal Inference in Dynamic Models with Conditional Moment Restrictions. CREATES Research Paper No. 2008-51.

[15] Creel, M. (2005) User-friendly parallel computations with econometric examples, *Computational Economics*, **26**, 107-128.

[16] Creel, M. (2007) I ran four million probits last night: HPC clustering with ParallelKnoppix, *Journal of Applied Econometrics*, **22**, 215-223.

[17] Domínguez, M.A. and I.N. Lobato (2004) Consistent Estimation of Models Defined by Conditional Moment Restrictions. *Econometrica*, **72**, 1601-1615.

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

[19] Fermanian, J.-D. and B. Salanié (2004) A nonparametric simulated maximum likelihood estimation method, *Econometric Theory*, **20**, 701-734.

[20] Fiorentini, G., E. Sentana, and N. Shephard (2004) Likelihood-based estimation of latent generalized ARCH structures, *Econometrica*, **72**, 1481-1517.

[21] Florens, J.P., M. Mouchart and J. M. Rolin (1993) Noncausality and marginalization of Markov processes, *Econometric Theory*, **9**, 241-262.

[22] Gallant, A.R. (1987) *Nonlinear Statistical Models*, John Wiley & Sons.

[23] Gallant, A. R., and D. Nychka (1987) Seminonparametric maximum likelihood estimation, *Econometrica,* **55**, 363–390.

[24] Gallant, A. R. and G. Tauchen (1996) Which moments to match? *Econometric Theory* **12**, 657-681.

[25] Gallant, A.R., and G. Tauchen (1999) The relative efficiency of method of moments estimators, *Journal of Econometrics*, **92**, 149-172.

[26] Gallant, A.R., and G. Tauchen (2002) Simulated score methods and indirect inference for continuous-time models, Chapter in preparation for the *Handbook of Financial Econometrics*.

[27] Gallant, A.R., and G. Tauchen (2004) EMM: A program for efficient method of moments estimation: User's Guide Version 1.7, http://econ.duke.edu/webfiles/get/emm/ver17/guide/.

[28] Gallant, A.R., and G. Tauchen (2007) EMM: A program for efficient method of moments estimation: User's Guide Version 2.5, http://econ.duke.edu/webfiles/arg/emm/emm.tar.

[29] Goffe, W.L., G. D. Ferrier and J. Rogers (1994) Global optimization of statistical functions with simulated annealing, *Journal of Econometrics,* **60**, 65-99.

[30] Gouriéroux, C. and A. Monfort (1996) Simulation-based econometric methods, *Oxford University Press*.

[31] Gouriéroux, C., A. Monfort, and E. Renault (1993), Indirect inference, *Journal of Applied Econometrics,* **8**, S85-S118.

[32] Hajivassiliou, V. and D. McFadden (1998) The method of simulated scores for the estimation of limited-dependent variable models, *Econometrica,* **66**, 863-896.

[33] Hansen, L.P., J.C. Heaton and M. Ogaki (1988) Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association,* **83**, 863–871+

[34] Härdle, W. (1991) *Applied Nonparametric Regression*, Cambridge University Press.

[35] Hansen, L., J. Heaton and A. Yaron (1996) Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, **14**, 262-280.

[36] Harvey A.C., E. Ruiz and N. Shephard (1994) Multivariate stochastic variance models, *Review of Economic Studies,* **61**, 247-264.

[37] Huber, P. (1964) Robust estimation of a location parameter, *Annal of Mathematical Statistics,* **35**, 73-101.

[38] Karlsen, H.A. and D. Tjøstheim (2001) Nonparametric Estimation in Null Recurrent Time Series, *Annals of Statistics*, **29**, 372-416.

[39] Kitamura, Y. and P.C.B. Phillips (1997) Fully modified IV, GIVE and GMM estimation with possibly non-stationary regressors and instruments. *Journal of Econometrics*, **80**, 85-123.

[40] Kristensen, D. (2008a) Uniform Convergence Rates of Kernel Estimators with Heterogenous, Dependent Data, working paper, http://www.columbia.edu/~dk2313/workingpapers/wp_files/hetero-unif-conv-22-05-08.pdf.

[41] Kristensen, D (2008b) Uniform Ergodicity of a Class of Markov Chains with Applications to Time Series Models, working paper, http://www.columbia.edu/~dk2313/workingpapers/wp_files/MC-uniform-14-01-07.pdf.

[42] Kristensen, D. and B. Salanié (2008) Higher Order Improvements for Approximate Estimators, working paper, Department of Economics, Columbia University.

[43] Kristensen, D. and Y. Shin (2008) Estimation of dynamic models with nonparametric simulated maximum likelihood, working paper, http://www.columbia.edu/~dk2313/workingpapers/wp_files/fo-smle-18-01-06.pdf.

[44] Laroque, G. and B. Salanié (1993) Simulation-based estimation of models with lagged latent variables, *Journal of Applied Econometrics,* **8**, S119–S133.

[45] Lee, L.F. (1995) Asymptotic bias in simulated maximum likelihood estimation of discrete choice models, *Econometric Theory,* **11**, 437– 483.

[46] Li, Q. and J. Racine (2007), *Nonparametric econometrics*, Princeton University Press.

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

[48] Michaelides, A. and S. Ng (2000) Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators, *Journal of Econometrics,* **96**, 231-66.

[49] Newey, W.K. and K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703-08.

[50] Pakes, A. and D. Pollard (1989) Simulation and the asymptotics of optimization estimators, *Econometrica,* **57**, 1027–1057.

[51] Racine, J. (2002) Parallel distributed kernel estimation, *Computational Statistics & Data Analysis*, **40**, 293-302.

[52] Robinson, P. (1983) Nonparametric estimators for time series, *Journal of Time Series Analysis*, **4**, 185-207.

[53] Ruge-Murcia, F.J. (2007) Methods to estimate dynamic stochastic general equilibrium models, *Journal of Economic Dynamics and Control*, **31**,2599-2636.

[54] Ruiz, E. (1994) Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics,* **63**, 289-306

[55] Smith, A. (1993) Estimating nonlinear time series models using simulated vector autoregressions, *Journal of Applied Econometrics,* **8**, S63-S84.

[56] Tauchen, G. (1997) New minimum chi-square methods in empirical finance, in *Advances in Econometrics*, Seventh World Congress, eds. D. Kreps, and K. Wallis, Cambridge UK: Cambridge University Press, 279–317.

[57] Windmeijer, F. (2005) A finite sample correction for the variance of linear efficient two-step GMM estimators, *Journal of Econometrics,* **126**, 25-51.

[58] Yang, C., R. Duraiswami, N. Gumerov and L. Davis (2003) Improved fast Gauss transform and efficient kernel density estimation, Ninth IEEE International Conference on Computer Vision (ICCV'03), **1**, 464-471.

[59] Yu, J. (2005) On leverage in a stochastic volatility model, *Journal of Econometrics*, **127**, 165-178.

UNIVERSITAT AUTÒNOMA DE BARCELONA

COLUMBIA UNIVERSITY AND CREATES (CENTER FOR RESEARCH IN ECONOMETRIC ANALYSIS OF TIME SERIES, UNIVERSITY OF AARHUS).