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. It is shown that as the number of simulations diverges, the estimator is consistent and a higher-order expansion reveals the stochastic difference between the infeasible GMM estimator based on the same moment conditions and the simulated version. In particular, we show how to adjust standard errors to account for the simulations. Monte Carlo results show how the estimator 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.

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 an 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:

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

where $t = 1, ..., T$. The observable variables are the vector y_t, w_t is a vector of latent variables, and u_t is a vector of independent white noise shocks, with a known distribution. 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] Finally, $\theta \in \Theta$ 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 establish a higher-order expansion describing the stochastic differences between the infeasible GMM estimator assuming that the conditional moments can be evaluated exactly and our simulated

---

[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 y_t be the observed aggregate outcomes.

version. The expansion reveals that the SNM estimator suffers from the usual bias component due to the kernel smoothers. On the other hand, we demonstrate that there is no so-called "curse of dimensionality": Kernel regression estimators are well-known to perform poorly when a large number of conditioning variables are included, and our conditional moment estimator will have a pointwise variance of order $1/\left(Sh^d\right)$ where S is number of simulations and d is the number of conditioning variables. However, in the computation of the SNM estimator we sum over the kernel estimators, which reduces the variance of the resulting parametric estimator to the order $1/S$. This is the same rate as for unconditional SMM, and as such we pay no price in terms of first-order variance for using conditional moments instead of unconditional ones in our simulated estimation procedure. In particular, the SNM estimator will be asymptotically normally distributed but with an additional variance term of order $1/S$ relative to the exact GMM estimator due to the simulations.

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, 2008) and simulated nonparametric estimators (Altissimo and Mele, 2009). 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 (2009) generalize this method to handle dynamic models. However, by choosing moment conditions in a judious manner, the 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. This is supported by our Monte Carlo results which show that moment conditions may be chosen such the SNM estimator performs well in comparison to other estimators, including NPSMLE, that have been proposed for estimation of general DLV models.

Even if the (simulated) GMM estimator does not reach full efficiency, there are situations where it may be preferable to the (simulated) MLE: The aforementioned efficiency gain of the MLE is often off-set by the MLE's high sensitivity to misspecifications. In contrast, the SNM will in many situations be 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). This issue is particularly important if the main goal with estimating the DLV model is to use it for forecasting. In this setting, if the DLV model is misspecified, our SNM estimator will often be the better vehicle since this finds the parameter estimates that minimize the forecasting error. In particular, the SNM estimates can still be used for forecasting purposes under misspecification. In contrast, the MLE's will in general lead to suboptimal forecasts. For a formal argument of this point, we refer to Weiss (1996).

Finally, in comparison to simulated MLE methods, we will demonstrate that the SNM estimator will suffer from fewer biases due to simulations. This is due to the fact that the simulated conditional moments enter the GMM objective function linearly, while simulated densities enter the log-likelihood nonlinearly. As such any asymptotic efficiency gains of the SMLE may be offset by these additional biases.

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) of 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.

We now turn to a presentation of the SNM estimator. 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 discusses some extensions, 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:

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

These variables are chosen by the researcher and will function as conditioning variables. A natural choice in many situations would be to use the first q lags of y_t, that is, $x_t = (y_{t-1}, ..., y_{t-q})$, 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 leads and lags of the observations,

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

There are no restrictions on how these functions are chosen in conjunction with the conditioning variables, except that they jointly should identify the parameter of interest through the corresponding generalized residual functions. These are defined as

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

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

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

and $E_\theta[\cdot | x_t]$ is 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

$$(2.2) \qquad\qquad E_\theta[\varepsilon_t(\theta) | x_t] = 0.$$

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,

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

such that $E_\theta[g_t(\theta)] = 0$. Our identifying restriction is then that $E_{\theta_0}[g_t(\theta)] = 0$ if and only if $\theta = \theta_0$. 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

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

For some sequence of $M \times M$ dimensional positive definite weighting matrices, W_n, 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 $T(\phi)(x_t; \theta)$ 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 simulated method of moments (SMM) estimator (Gouriéroux and Monfort, 1996, pg. 27; Duffie and Singleton, 1993). A unbiased estimator can normally be obtained in static models, see e.g. McFadden (1989). 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:

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

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

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

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

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

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.

For notational ease, we use the same bandwidth across all variables. It should be emphasized that this may not always be advisable; in particular, if the individual variables contained in X_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 "pre-whitening" 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 $\hat{T}_S(\phi)(x; \theta)$ is used in place of the expectation of unknown form, $T(\phi)(x_t; \theta)$. To be explicit, the SNM estimator is based on estimated residual functions,

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

The moment function contribution of an observation is

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

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

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

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

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

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 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 our general assumptions. To demonstrate how the conditions can be verified, we discuss in more detail one particular DLV model.

We 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 1. *The exact GMM estimator satisfies:*

(1) *The moment function $G(\theta) = E[Z_t' \varepsilon_t(\theta)]$ is continuous and satisfies $G(\theta) = 0$ if and only if $\theta = \theta_0$, where θ_0 lies in the interior of $\Theta. \in \Theta \subseteq \mathbb{R}^{d_\theta}$ and Θ is compact.*
(2) $\sup_{\theta \in \Theta} \|G_n(\theta) - G(\theta)\| \to^P 0.$
(3) $W_n \to^P W > 0.$

The above conditions are standard for GMM estimators, see e.g. Newey and McFadden (1994) who also give more primitive conditions for them to hold for particular models. Under Assumptions 1, the infeasible GMM estimator is consistent, $\hat{\theta}_n \to^P \theta_0$.

The goal is now to analyze the simulated version, $\hat{\theta}_{n,S}$, relative to the actual but infeasible one, $\hat{\theta}_n$. As a first step towards such a result, 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 (2009) 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 (2009). 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 the simulated random variables $X_0(\theta)$ 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,

$$(2.11) \qquad B_0 = \sup_{x \in \mathbb{R}^{d_x}} \sup_{\theta \in \Theta} f(x; \theta), \quad B_{V,1} = \sup_x \sup_{\theta \in \Theta} \|x\|^{\lambda} E\left[\|V_0(\theta)\| \,|X_0(\theta) = x\right] f(x; \theta),$$

$$(2.12) \qquad B_{V,2} = \sup_{x,x' \in \mathbb{R}^{d_x}} \sup_{\theta \in \Theta} E\left[\|V_0(\theta)\| \|V_T(\theta)\| \,|X_0(\theta) = x, X_T(\theta) = x'\right] f_T(x, x'; \theta).$$

We write the model in Eq. (2.5) more compactly as

$$(2.13) \qquad \mathbf{Y}_s(\theta) = r\left(\mathbf{Y}^{s-1}(\theta), U_s; \theta\right),$$

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

$$(2.14) \qquad \dot{\mathbf{Y}}_s(\theta) = \frac{\partial r\left(\mathbf{Y}^{s-1}(\theta), U_s; \theta\right)}{\partial \mathbf{Y}^{s-1}(\theta)'} \dot{\mathbf{Y}}^{s-1}(\theta) + \frac{\partial r\left(\mathbf{Y}^{s-1}(\theta), U_s; \theta\right)}{\partial \theta}.$$

Assumption 2. *The test and moment functions satisfy:*

(1) *The test functions, $\phi(y_t, ..., y_{t+p})$, are continuously differentiable and for some $\mu \geq 2$ and with $\underline{Y}_s(\theta) = (Y_{s+p}(\theta), ..., Y_{s-p}(\theta))$: $E\left[\|\phi(\underline{Y}_s(\theta))\|^{\mu}\right] < \infty$,*

$$E[\|\phi(\underline{Y}_s(\theta)) \phi'(\underline{Y}_s(\theta)) \dot{\underline{Y}}_s(\theta)\|^{\mu}] < \infty \text{ and } E[\|\phi(\underline{Y}_s(\theta))\|^p \|\dot{X}_s(\theta)\|^{\mu}] < \infty.$$

(2) *The functions $x \mapsto T(\phi)(x; \theta)$ and $f(x; \theta)$ are $m \geq 2$ times continuously differentiable w.r.t. x. The function $r(z, u; \theta) = (r_y(z, u; \theta), r_w(z, u; \theta))$ is twice differentiable in z.*

(3) *$E\left[\sup_{\theta \in \Theta} f(x_t; \theta)^{-q} \|T(\phi)(x_t; \theta)\| \|z_t\|\right] < \infty$ for some $q > 0$.*

Assumption 3. *The process $\{\mathbf{Y}_s(\theta)\}$ satisfies:*

(1) *For any given $\theta \in \Theta$, there exists a stationary solution $\{\mathbf{Y}_s(\theta)\}$ to the model in Eq. (2.13). This solution is α-mixing with mixing coefficients $\alpha_s(\theta)$ satisfying $\alpha_s(\theta) \leq As^{-\beta}$ for some $0 < A, \beta < \infty$ which do not depend on θ.*

(2) *The bounds defined in Eqs. (2.11)-(2.12) are finite for $V_s(\theta) = \phi(\underline{Y}_s(\theta))$, $V_s(\theta) = \partial \phi(\underline{Y}_s(\theta)) / (\partial \underline{Y}_s(\theta)) \dot{\underline{Y}}_s(\theta)$ and $V_s(\theta) = \phi(\underline{Y}_s(\theta)) \dot{X}_s(\theta)$.*

(3) *With $d = d_x + d_\theta$, λ given in Eq. (2.11), and μ in A.3.1, the mixing exponent β satisfies*

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

Assumption 2 requires relevant moments of the test functions to exist and that the model defined through the function r is sufficiently smooth in the state variables. For a set of sufficient conditions for Assumption 2.1 to hold for Markov models, we refer to Kristensen (2007). The smoothness conditions in 2.2 hold for most dynamic models; they do however rule out discontinuous models such as threshold models. The smoothness restriction on r is however only imposed for technical convenience, and we conjecture that our results also go through for models with discontinuous dynamics by adapting the techniques developed in, for example,

Pakes and Pollard (1989) to our setting. For some discontinuous models such as limited dependent variable models, one can alternatively employ the trick proposed in Fermanian and Salanie (2004, Section 3) to deal with discontinuities. The moment condition in 2.3. is used to ensure that the trimming has no effect asymptotically. It is implicitly a requirement regarding the tail thickness of the distribution of x_t.

Assumption 3 says that the simulated path $\{\mathbf{Y}_s(\theta)\}$ 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 $\{\mathbf{Y}_s(\theta)\}$ 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 the stationary one. In practice, observations belonging to a burn-in period may be discarded to reduce the effect of initializing at a non-stationary solution. 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 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; see also Kristensen (2009, Theorem 3) for some results on kernel estimators when data is not initialized at the stationary distribution. The stationarity and mixing conditions could probably be weakened to assume that the process is recurrent (and therefore potentially non-stationary) since it is possible to show pointwise convergence of kernel estimators of conditional moments in this setting (see, for example, Karlsen and Tjøstheim, 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.

As a specific example, consider the linear model,

$$\mathbf{Y}_s(\theta) = A\mathbf{Y}_{s-1}(\theta) + BU_s,$$

where $\theta = (A, B)$, and U_s i.i.d. with a continuous distribution. This model includes for example the standard SV model considered in the simulation study,

$$
\begin{aligned}
\log\left(Y_t^2(\theta)\right) &= \log\left(\sigma_t^2(\theta)\right) + \log\left(u_{t,1}^2\right), \\
\log\left(\sigma_t^2(\theta)\right) &= \alpha + \beta \log\left(\sigma_{t-1}^2(\theta)\right) + u_{t,2}.
\end{aligned}
$$

The linear process $\{\mathbf{Y}_s(\theta)\}$ is stationary and geometrically mixing for all θ such that A has eigenvalues inside the unit circle, c.f. Kristensen (2007). The first-order derivative w.r.t. (A, B),

$\dot{\mathbf{Y}}_s(\theta) = \left(\dot{\mathbf{Y}}_{A,s}(\theta), \dot{\mathbf{Y}}_{B,s}(\theta)\right)$, solves

$$
\begin{aligned}
\dot{\mathbf{Y}}_{A,s}(\theta) &= \mathbf{Y}_s(\theta) + A\dot{\mathbf{Y}}_{A,s}(\theta) + BU_s, \\
\dot{\mathbf{Y}}_{B,s}(\theta) &= A\dot{\mathbf{Y}}_{B,s}(\theta) + U_s.
\end{aligned}
$$

Since $\mathbf{Y}_s(\theta)$ is stationary and mixing, $\dot{\mathbf{Y}}_s(\theta)$ will also be stationary and mixing for all A with eigenvalues inside the unitcircle. Furthermore, if $E\left[\|U_s\|^p\right] < \infty$, then $E\left[\|\mathbf{Y}_s(\theta)\|^p\right] < \infty$ and $E\left[\|\dot{\mathbf{Y}}_s(\theta)\|^p\right] < \infty$. So, for example, by choosing X_t as simply lagged values of y_t, and the test function as bounded by polynomials of an appropriate order, Assumptions 2-3 will hold for this model.

We impose regularity conditions on the kernel function K:

Assumption 4. *The kernel* $K : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ *satisfies:*

(1) $\sup_{u \in \mathbb{R}^{d_x}} |K(u)| < \infty$ *and* $\int |K(u)| \, du < \infty$. *There exist* $\Lambda, L < \infty$ *such that either (i)* $K(u) = 0$ *for* $\|u\| > L$ *and* $|K(u) - K(u')| \le \Lambda \|u - u'\|$, *or (ii)* $K(u)$ *is differentiable with* $|\partial K(u)/\partial u| \le \Lambda$. *For some* $a > 1$, $|\partial^r K(u)/\partial u^r| \le \Lambda \|u\|^{-a}$ *for* $\|u\| \ge L$ *and* $r = 0, 1$.

(2) *For some* $m \ge 1$: $\int K(u) u^\alpha du = 0$ *for all* $\alpha \in \{0,1\}^{d_x}$ *with* $|\alpha| = 1, \ldots, m - 1$, *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 redefine our SNM estimator to trim away observed values for which $f(x_t;\theta) < a$. That is, we redefine $\hat{g}_{t,S}(\theta)$ as

$$
\hat{g}_{t,S}(\theta) = \tau_{a,t}(\theta) Z_t' \hat{\varepsilon}_t(\theta),
$$

where $\tau_{a,t}(\theta) = \tau_a(\hat{f}(x_t;\theta))$ is a trimming 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). The trimming function is chosen such that $\tau_{a,t}(\theta) = 0$ when $\hat{f}(x_t;\theta)$ is "close to zero". The trimming is introduced for technical reasons to handle the well-known denominator problem with kernel regression estimators: We are only able to show uniform consistency of $\hat{T}_S(\phi)(x;\theta)$ for x on a compact interval of the form $\{x : \hat{f}(x;\theta) \ge a\}$ for any lower bound $a > 0$. The trimming function $\tau_{a,t}(\theta)$ enforces this bound in the estimation. The use of trimming obviously imparts a loss of efficiency so as $S \to \infty$, we will let $a \to 0$ such that asymptotically the trimming has no impact. 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 $T(\phi)(x_t;\theta)$. We impose the following regularity conditions on the trimming function:

Assumption 5. *The trimming function* $\tau_a : \mathbb{R} \mapsto [0,1]$, $a > 0$, *satisfies* $\tau_a(z) = 1$ *for* $z \geq a$ *and* $\tau_a(z) = 0$ *for* $z \leq a/2$. *It is continuously differentiable with* $|\tau'_a(z)| = O(a)$.

We here use a smooth trimming function, $\tau_a(z)$ such that the trimmed GMM objective function remains smooth in θ. A simple way of constructing $\tau_a(z)$ is to choose a cdf F with support $[0,1]$, and define $\tau_a(z) = F((2z-a)/a)$ which then in great generality will satisfy Assumption 5; see also Andrews (1995) and Ai (1997). We are now ready to state the first main result:

Theorem 1. *Assume that Assumptions 1-5 hold. As* $a^{-1}h^m \to 0$, $a^{-1}\sqrt{\log(S)/(Sh^{d_x})} \to 0$ *and* $a \to 0$, *then the SNM estimator is consistent:* $\hat{\theta}_{n,S} \to^P \theta_0$.

Proof. See the Appendix. □

As $h \to 0$ and $Sh^d \to \infty$, $\hat{T}_S(\phi)(x;\theta)$ converges towards the true moment $T(\phi)(x;\theta)$. In principle, S could be chosen so large so that the differences between the error functions in equations (2.1) 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. The above result formalizes this intuition and shows that under suitable conditions on the bandwidth and the trimming parameters, the SNM estimator and the GMM estimator are asymptotically indistinguishable as $S \to \infty$.

However, in practice, one will normally only use a more moderate number of simulations and so it is still of interest to have a measure of the approximation error incurred by using the SNM estimator. In particular, in order to compute correct standard errors, one should account for the additional sampling error introduced by the simulations. To this end, we will now analyze in further detail the stochastic difference between the exact and simulated GMM estimator. We follow the same strategy as in Kristensen and Salanie (2009) and use a functional Taylor expansion of $\hat{G}_{n,S}(\theta)$ w.r.t. $\hat{T}$ to evaluate the higher-order properties of the SNM estimator. For this higher-order analysis to be formally correct, we need to strengthen our assumptions.

Assumption 6. *The exact GMM estimator satisfies:*

(1) $\sqrt{n}G_n(\theta_0) \to^d N(0, \Omega_0)$ *where*

$$(2.15) \qquad \Omega_0 = E\left[g_0(\theta_0)g_0(\theta_0)'\right] + 2\sum_{t=1}^{\infty} E\left[g_0(\theta_0)g_t(\theta_0)'\right].$$

(2) *The derivative*

$$(2.16) \qquad H_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta), \quad h_t(\theta) = -Z_t'\frac{\partial T(\phi_i)(x_t;\theta)}{\partial \theta} \in \mathbb{R}^{M \times d_\theta}$$

satisfies $\sup_{\|\theta - \theta_0\| < \delta} \|H_n(\theta) - H(\theta)\| \to^P 0$ *where* $H(\theta) = E[h_t(\theta)]$.

(3) *With* $H_0 = H(\theta_0)$, *the matrix* $H_0'W_0H_0$ *is non-singular.*

Under Assumptions 1 and 6, the exact GMM estimator is $\sqrt{n}$-asymptotically normally distributed,

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

c.f. Lemma 3 in the Appendix. As with Assumption 1, the above conditions are completely standard conditions for GMM-estimators, c.f. Newey and McFadden (1994).

Next, we assume that the DLV model is sufficiently smooth in θ such that derivatives of the simulated trajectories exist and are well-behaved:

Assumption 7. *The test and moment functions satisfy:*
(1) *The test functions, $\phi(y_t, ..., y_{t+p})$, are twice continuously differentiable. The first order derivatives satisfy the same moment conditions as imposed on ϕ in Assumption 2.*
(2) *The functions $x \mapsto T(\phi)(x; \theta)$ and $f(x; \theta)$ are continuously differentiable w.r.t θ, and their derivatives are $m \geq 2$ times continuously differentiable w.r.t. x.*
(3) *$E\left[\sup_{\theta \in \Theta} f(x_t; \theta)^{-q} \|\partial T(\phi)(x_t; \theta)/(\partial \theta)\| \|z_t\|\right] < \infty$ for some $q > 0$.*

Assumption 8. *The process $\{\mathbf{Y}_s(\theta)\}$ satisfies:*
(1) *$\{\mathbf{Y}_s(\theta)\}$ is twice continuously differentiable w.r.t. to θ, and the derivatives satisfy the same mixing conditions as $\{\mathbf{Y}_s(\theta)\}$ in Assumption 3.*
(2) *With $V(\theta)$ defined in Assumption 3.2, its first derivative w.r.t. θ also satisfies Eqs. (2.11)-(2.12).*

Assumption 9. *The kernel K is differentiable and its derivative satisfies the same conditions as imposed on K in Assumption 4.*

These conditions are used to establish uniform consistency of $\partial \hat{T}_S(\phi)(x; \theta)/(\partial \theta)$, which in turn gives us uniform consistency of the simulated version of $H_n(\theta)$. As with Assumptions 2 and 3, the conditions are quite high-level, but they are normally easy to verify in specific models. It is for example possible to verify Assumptions 7 and 8 for the linear model considered earlier. We are now ready to state the second main result evaluating the higher-order impact of the simulations on the SNM estimator:

Theorem 2. *Assume that Assumptions 1-9 hold, and that $a^{-1}h^m \to 0$, $a^{-1}\sqrt{\log(S)/(Sh^{d_x+4})} \to 0$ and $a \to 0$. Then the SNM estimator satisfies:*

$$\hat{\theta}_{n,S} - \hat{\theta}_n = h^m B + \frac{1}{S}\sum_{s=1}^{S} D_s + O_P(a^q) + O_P\left(a^{-1}h^m\right) + O_P\left(\frac{1}{\sqrt{nSh^{d_x+\delta}}}\right),$$

for some $\delta > 0$ where

$$B = \left(H_0' W_0 H_0\right)^{-1} H_0' W_0 E\left[\frac{Z_t'}{f(x_t; \theta_0)}\sum_{|\alpha|=m}\frac{\partial^{|\alpha|}[T(\phi)(x_t; \theta_0)f(x_t; \theta_0)]}{\partial x^\alpha}\right],$$

$$D_s = \left(H_0' W_0 H_0\right)^{-1} H_0' W_0 \left\{Z(X_s(\theta_0))' \phi(Y_s(\theta_0)) - E\left[Z(X_s(\theta_0))' \phi(Y_s(\theta_0))\right]\right\}.$$

In particular, if $\sqrt{n}a^{-1}h^m \to 0$, $\sqrt{n}a^q \to 0$, $Sh^{d+\delta} \to \infty$, and $n/S \to \lambda \geq 0$, then the SNM estimator satisfies:

$$(2.17) \qquad \sqrt{n}(\hat{\theta}_{n,S} - \theta_0) \xrightarrow{d} N\left(0, \left(H_0'W_0H_0\right)^{-1}H_0'W_0\left\{\Omega_0 + \lambda\Sigma_0\right\}W_0H_0\left(H_0'W_0H_0\right)^{-1}\right),$$

where

$$\Sigma_0 = E\left[D_0D_0'\right] + 2\sum_{s=1}^{\infty} E\left[D_0D_s'\right].$$

Proof. See the Appendix. □

The first part of the theorem gives a bias-variance expansion of the estimation error from using the SNM instead of the exact GMM estimator. There are two leading terms in the expansion: The first term, $h^m B$, is the bias due to the use of kernel smoothers and the second term, $\sum_{s=1}^{S} D_s/S$ where $E[D_s] = 0$, is an additional variance component due to the use of simulations. A number of points should be emphasized here:

First, in comparison to the NPSMLE of Fermanian and Salanie (2004) and Kristensen and Shin (2008), the SNM suffers from fewer biases. For the NPSMLE, an additional bias term of order $1/S$ appears due to the simulator entering the objective function nonlinearly. This is not the case here, as in SMM, since $\hat{T}_S(\phi)(x;\theta)$ enters $\hat{G}_{n,S}(\theta)$ linearly. This will in general imply that the SNM estimator will be less biased than the NPSMLE. One could potentially combine the NPSML and SNM estimators in the spirit of Hajivassiliou (2000) to obtain an efficient two-step estimator with reduced bias.

Second, in comparison to standard SMM (Duffie and Singleton; 1993; McFadden, 1989), we here have a bias component, $h^m B$, due to the use of kernel smoothers. On the other hand, there is no first-order curse of dimensionality: It is well-known that the pointwise variance of a kernel estimator is of order $1/(Sh^d)$. One could fear that this would lead to a first-order variance component of the SNM of the same order. This is however not the case; rather the first order variance component is of order $1/S$. To understand the intuition behind this result, it is useful to think of the SNM estimator as a semiparametric two-step estimator: In the first step, a nonparametric smoother is used to estimate the conditional means, $T(\phi)(x_t;\theta)$, $t = 1,...,n$, These are in turn used to compute $\hat{\theta}_{n,S}$. The nonparametric error in the first step is integrated out in the second step due to us summing over $T(\phi)(x_t;\theta)$, $t = 1,...,n$, in the computation of $\hat{\theta}_{n,S}$. The summation is asymptotically equivalent to integrating over the kernel smoother which reduces its variance as is well-known from the literature on semiparametric estimators, see e.g. Kristensen (2008, 2010).

The second part of the result states how, under suitable choices of bandwidths and trimming parameters, the simulations impact the standard errors of the SNM estimator: As can be seen from Eq. (2.17), the SNM estimator has an additional variance term, $\lambda\Sigma$ where $\lambda \approx n/S$, relative to the exact GMM estimator. This is akin to Duffie and Singleton (1993). It is easy to construct

an estimator of Σ, and so the second result can be used to adjust standard errors to account for the use of simulations.

2.3. **Discussion.**

2.3.1. *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

$$(2.18) \qquad \phi_k(y_t) = \exp\left(i\tau_k' y_t\right), \quad k = 1, ..., L,$$

where $\tau_k \in \mathbb{R}^{d_y}$, $k = 1, ..., L$, is a collection of points chosen by the researcher. As shown in Carrasco et al. (2007), the GMM estimator using no conditional information will converge towards the (quasi-)maximum likelihood estimator as L grows larger. We conjecture that this still holds true when we condition on x_t such that our SNM estimator based on $\phi(y_t)$ defined in eq. (2.18) will be asymptotically equivalent to the (quasi-)maximum likelihood estimator associated with the conditional density $f(y_t | x_t; \theta)$.

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, $\xi_k(x) = \mathbb{I}\{x \leq x_k^*\}$, $k = 1,,...,N$, 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),

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

where $J(\theta) = (j_1(\theta), ..., j_N(\theta)) \in \mathbb{R}^{d_\theta \times LN}$ and $V(\theta) = E\left[s_n(\theta) s_n(\theta)'\right] \in \mathbb{R}^{LN \times LN}$ with

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

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

2.3.2. *Optimal weight matrix.* 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 $n\hat{G}_{n,S}(\theta)$ where $\hat{G}_{n,S}(\theta)$ is defined in Eq. (2.9).

In order to conduct any inference, we need estimators of the asymptotic covariance matrices in Theorem 2, Ω_0 and Σ_0. Also, a consistent estimator of these matrices are needed if one wishes to use an efficient weight matrix to estimate θ_0. In the ordinary GMM setting without a fully simulable model, these covariance matrices 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). However, 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 and Σ_0 through Monte Carlo, in much the same way as was proposed of estimation of Φ_0 above. We here only discuss the estimation of Ω_0, since an estimator of Σ_0 can be constructed along the same lines.

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 Eq. (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:

$$(2.19) \qquad\qquad \hat{\Omega}_0 = \frac{n}{R} \sum_{r=1}^{R} v_r v_r'.$$

Once Ω_0 and Σ_0 are estimated, hypothesis testing may then be done based on equation 2.17, which is analogous to the standard result for GMM estimators with an inefficient weight matrix, or the estimate of Ω_0 and Σ_0 may be used in a second round of estimation to compute a more efficient estimate of θ_0, as is common practice.

2.3.3. *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 radially symmetric Epanechnikov product kernels exclusively, accompanied by prior rotation of the data to approximate independence of the conditioning variables. The possibility of SNM estimation based on local linear or local polynomial kernel methods is left for future work.

Given the kernel function, the bandwidth must be chosen. 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. This bias-variance trade-off is clear from Theorem 2, where we have a bias term of order h^m and a variance term of order nSh^{d_x}. Many methods for choosing the bandwidth have been suggested in the literature, ranging from a simple rule-of-thumb bandwidth common to all test variables and conditioning variables, to individual bandwidths by test variable and conditioning variable, chosen using a data-dependent rule such as cross validation. It should be noted though that these methods are designed to minimize the MSE of the kernel estimator and as such do not necessarily minimize the MSE of $\hat{\theta}_{n,S}$. Thus, standard bandwidth selection methods may not work particularly well in our setting; see Kristensen and Salanie (2009) for a further discussion.

In the examples that follow, we use a single bandwidth for all test variables, and we use only a single conditioning variable, so there is only one bandwidth to be chosen. We treat the bandwidth as an additional parameter to estimate, in the manner of Härdle, Hall and Ichimura (1993). The real sample data, y_t, is out-of-sample from the point of view of the simulated data, $Y_s(\theta)$. By including the bandwidth as a parameter to estimate in the econometric objective function, we are effectively using an out-of-sample cross validation procedure. Use of a single bandwidth that is chosen in a data-dependent way gives a balance between computational

convenience and reliable nonparametric fit. The details of out implementation are clear in the example code that is provided. This strategy has the important advantage that it is automatic, which frees the modeler's attention to deal with more fundamental issues such as choice of test variables, conditioning variables, and instruments. There is of course plenty of scope for experimentation which could possibly improve on our results.

2.3.4. *Computational issues.* Estimation of a complicated model using long simulation may become computationally burdensome, since kernel smoothing is a computationally intensive procedure. Another problem is multiple local minima. As is the case 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. This problem becomes more severe if an Epanechnikov kernel is used, because it is kinked. Our experience is that gradient-based minimization algorithms are not able to find the global minimum of the SNM objective function regardless of the kernel that is selected. The existence of multiple local minima in a Monte Carlo context requires a means of finding the global minimizer with a high degree of confidence, yet with minimal user intervention. Our solution is to use a simulated annealing algorithm (Goffe *et al.*, 1994). This is a heuristic minimizer that searches over a parameter space defined by reasonable bounds (for example, we impose stationarity and non-negativity of variances), gradually contracting the region of search. The starting point for each Monte Carlo replication is a random point drawn from a uniform density over the parameter space, to avoid the possibility that over-rapid contraction of the search region could bias the result towards the starting point. Given that we use a robust global minimizer, we use the Epanechnikov kernel, in spite of its kinks, for its computational advantages.

All of 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. While the embarrassingly parallel nature of Monte Carlo is well known, Creel (2007) provides discussion of the methods we use here. The widespread availability of multicore processors is an invitation to take advantage of parallelization opportunities in econometric work. Once the step to parallelization is made, ambitious projects with high computational demands become much more tractible. All of the results reported in the next section of this paper were obtained on a computational cluster that provided a total of 32 CPU cores, running the PelicanHPC distribution of GNU/Linux[2]. PelicanHPC (Creel, 2009) contains all software and scripts needed to replicate the results reported in the next section, on a single computer or on a cluster. Documentation for the software is provided by Creel and Kristensen (2009).

---

[2]PelicanHPC is the evolution of the ParallelKnoppix distribution of GNU/Linux, which was described in Creel (2007).

3. MONTE CARLO RESULTS

This section presents Monte Carlo results that compare the SNM estimator to other estimators. We also briefly explore the use of SNM to estimate a latent network model. The intention is to show that the SNM estimator can be used to successfully estimate a variety of DLV models, and that the SNM estimator can perform well in comparison to alternative estimators.

As discussed above, the SNM estimator is computationally demanding, especially in the context of Monte Carlo work. Because of this, we limit the directions that we explore. We do not attempt to use optimal instruments, and instead just use the single conditioning variable (explained below), augmented with a vector of ones. Nor do we attempt to use an optimal weight matrix, and instead we just use an identity matrix as the weight. Informal Monte Carlo work with some of the examples and a small number of replications suggested that use of an estimated optimal weight matrix would not improve the results that are presented in this paper. We leave exploration of the use of approximately optimal instruments and/or an optimal weight matrix for future work, possibly when code can be written to use graphical processing units (GPUs).

In all cases, we minimize the objective function using simulated annealing, starting from a random point in the parameter space. All of the Monte Carlo results use a simulation length of $S = 10000$, and in all cases except the first (the AR1 model), the 2 percent of observations with the lowest value of a kernel density fit to the conditioning variable are trimmed. In all cases we generate 500 Monte Carlo replications. Other details are found in the computer code that we offer, as noted above.

3.1. **AR1.** The first model we consider is a simple autoregressive model

$$(3.1) \qquad y_t = \alpha + \beta y_{t-1} + u_t$$

where $u_t \sim N(0,1)$. The sample size is $n = 50$, and we generate samples using $\alpha = 0, \beta = 0.9$. The single test function is $\phi_t = y_t$, and the single conditioning variable is $x_t = y_{t-1}$. The instruments are a vector of ones, plus the conditioning variable. This implementation of SNM mimics the moment conditions that define the ordinary least squares (OLS) estimator of α and β, which is also the maximum likelihood estimator. As such, comparison of SNM with OLS serves as a benchmark for how well SNM can perform in comparison to a fully efficient estimator. Given that the data generating process does not create outlying observations and that the sample size is small ($n = 50$), no trimming is done.

Table 1 contains the results. We see that the SNM estimator has performance similar to the OLS estimator, and in fact has a lower root mean squared error for the slope parameter. The results illustrate that the SNM estimator that uses the same moment conditions as a feasible GMM estimator will lead to estimation results close to those of the feasible GMM estimator.

3.2. **Stochastic Volatility.** 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 Eq. (1.1), the model is

$$(3.2) \qquad \text{SV1:} \begin{cases} y_t & = \exp\left(w_t/2\right) u_{t,1} \\ w_t & = \alpha + \beta w_{t-1} + \sigma u_{t,2} \end{cases}$$

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 Eq. (3.1) will be referred to as the SV1 model. A slightly different parameterization (referred to here as the SV2 model) is used by some authors:

$$(3.3) \qquad \text{SV2:} \begin{cases} y_t & = \sigma_b \exp\left(w_t/2\right) u_{t,1} \\ w_t & = \beta w_{t-1} + \sigma u_{t,2} \end{cases}$$

The parameters of the two versions are the same, except that $\sigma_b = \exp(\alpha/2)$. Our code is written for the SV2 model, and estimates σ_b. To recover an estimate of α, we use $\hat{\alpha} = 2 \log \hat{\sigma}_b$.

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 $\theta_0 = (\alpha, \beta, \sigma)' = (-0.736, 0.9, 0.363)'$, for the SV1 model, and a sample size of $n = 500$ observations. We begin with this case.

To apply the SNM estimator, we must choose moment conditions by specifying the test variables ϕ_t and the conditioning variables x_t. For the SV1 and SV2 models, the sign of y_t gives no information. As such, we work with $z_t = |y_t|$. The test variables we use are $\phi_t = \left(z_t, z_t^2, \cos z_t \sin z_t, ..., \cos 4z_t \sin 4z_t\right)$. The first two test variables are clearly related to the variance of y_t, while the remaining test variables are motivated by the characteristic function approach to defining moment conditions. The single conditioning variable is $x_t = z_{t-1} + z_{t-2} + z_{t-3} + z_{t-4}$. This conditioning variable is intended to capture the recent variability of the series in a parsimonious way, to avoid needing to choose multiple bandwidths.

Table 3 presents the results. For the parameter α, the SNM estimator has a root mean squared error (RMSE) that is less than that of all alternatives except quasi-maximum likelihood (QML) and the Monte Carlo Likelihood (MCL) estimator. For β, the SNM estimator has RMSE equal to that of EMM, and higher than ML and MCMC. For σ, the ML, MCMC and MCL estimators outperform SNM, which achieves lower RMSE than EMM and QML. Overall, it seems fair to say that SNM performs well in general, and better than the only other general purpose estimator, EMM.

Fermanian and Salanié (2004) and Altissimo and Mele (2009) illustrate the estimators they propose using the SV2 model with the design $(\sigma_b, \beta, \sigma) = (0.025, 0.95, 0.260)$. The sample

size in both cases is $n = 500$. We apply the SNM estimator to data from this model and design, using the same test variables, conditioning variables and instruments as were used for the SV1 model. Table 4 provides the results, along with those of the two cited papers for comparison. For σ_b, the estimators all have very good precision, with the SNM estimator doing best. For β, the SNM estimator achieves an RMSE that is about 20% lower than that of the alternatives. For σ, the SNM estimator suffers from more bias than the alternatives, yet still achieves a relatively low RMSE.

3.3. **Autoregressive Tobit.** Fermanian and Salanié (2004) used an autoregressive Tobit model to illustrate the nonparametric simulated maximum likelihood (NPSML) estimator. This model, with notation adapted to follow the general DLV model of Eq. (1.1), may be written as:

$$
(3.4) \qquad \text{AR Tobit:} \begin{cases} y_t = & \max(0, w_t) \\ w_t = & \alpha + \beta w_{t-1} + \sigma u_t \\ u_t \sim & IIN(0, 1) \end{cases}
$$

This model has one observable variable, y_t, a single latent variable, w_t and a scalar white noise u_t. Fermanian and Salanié's Monte Carlo example used the true parameter values $(\alpha, \beta, \sigma) = (0.0,: 0.5,: 1.0)$ and the sample size $n = 150$. This same design is used here. To apply the SNM estimator, the same procedure (test variables, conditioning variable, etc.) as was used for the SV1 and SV2 models is repeated exactly, with the exception that we directly use y_t rather than its absolute value. Table 5 reports the results, along with Fermanian and Salanié's results for comparison. The SNM estimator has considerably lower RMSEs for the parameters α and σ, while for β the RMSE of SNM is a little higher than that of NPSML. The SNM estimator is considerably less biased for σ.

3.4. **Factor ARCH.** Billio and Monfort (2003) illustrate the kernel-based indirect inference (KBII) estimator with several Monte Carlo examples, one of which is a simple factor ARCH model. The model has a scalar common latent factor, w_t, and two observed endogenous variables, $y_t = (y_{t1}, y_{t2})'$. The 2×1 dimensional parameter β has its first element set to 1, for identification. The model, referred to as FA, is

$$
(3.5) \qquad \text{FA:} \begin{cases} y_t & = \beta w_t + u_{1t} \\ w_t & = \sqrt{h_t} u_{2t} \\ h_t & = \alpha_1 + \alpha_2 (w_{t-1})^2 \end{cases}
$$

$t = 1, 2, ..., n$, where $u_{1t} \sim N(0, \sigma^2 I_2)$ and $u_{2t} \sim N(0, 1)$. The parameter vector design is $(\alpha_1, \alpha_2, \sigma, \beta_2) = (0.2, 0.7, 0.5, -0.5)$. The sample size is $n = 500$.

Let $z_t = |y_t|$. The test variables used are $z_t, \cos z_t \sin z_t, \cos 2z_t, \sin 2z_t, ..., \cos 4z_t \sin 4z_t$, along with $y_{t1} y_{t2}$, to identify the sign of β_2. The conditioning variable is $z_{t-1,1} + z_{t-1,2} + z_{t-2,1} +$

$z_{t-2,1}$. The motivation for summing over the two observable variables is that variation in either of the two observable variables is an indication of variation in the latent variable. This also keeps the dimensionality low, which helps to avoid excessive computational time. The instruments are again the same conditioning variable augmented with a vector of ones.

Table 6 reports the Monte Carlo results, together with the lowest RMSE that Billio and Monfort obtain using several versions of kernel-based indirect inference, indirect inference, and simulated method of moments (see Billio and Monfort, 2003, Table 5, page 317). For all four parameters, the SNM estimator outperforms the estimators considered by Billio and Monfort, in terms of bias and RMSE.

3.5. **Degree distribution of a latent network.** The previous examples all use fairly simple models that have application to financial series. In this section we consider a model that incorporates an unobserved network. The results of this section are only intended as a sketch of possibilities, to show that the SNM estimator could be of use for the estimation of more ambitious models that include networks.

Much theoretical and empirical work is currently being done to incorporate network effects into economic models (Jackson, 2006). In many cases where one would suspect that network effects could be important, the actual network is not observed. The chances of a new graduate obtaining a given job could depend upon his or her social contacts, but these contacts are private information. The traffic received by a web site depends in part on the number of sites that link to it, but it may be costly to learn the complete list of sites that link to a give site. The information flow through a network of financial agents no doubt affects the price and volatility of financial assets, yet agents will certainly attempt to hide at least some contacts. We investigate the possibility of learning about a latent network from observed outcomes generated by the network. A distinct problem, studied by Copic, Jackson and Kirman (2009), is how an observed network comes to be formed from latent groups of nodes that share some characteristics (homophily). In their paper, the network is observed, and the question is what latent structure led to the formation of the network. Here we do not observe the network, and attempt to learn about its structure though an observed output.

Consider a latent random graph network, of the Bernoulli type, as described by Jackson (2006, Section 3.1.1). There are N nodes. Connections between nodes are assumed to be the outcome of independent Bernoulli trials. The probability that two nodes are connected is p_{CON}. The degree of a node is the number of links that it has. The frequency distribution of the degrees of the nodes is known as the degree distribution of the network. For the Bernoulli network, the degree distribution is binomial($N-1$, p_{CON}), and as N becomes large, the degree distribution is approximately Poisson. Here, we work with $N = 500$, and is this is known to the modeler. The objective is to estimate $p_{CON,}$ without directly observing the network.

Suppose that a disease is transmitted though the network. Analogously, we could think of transmission of information or new innovations. Jackson and Rogers (2007) provide a theoretical background and additional references. The disease will provide the signal that will be used to learn about the network structure. Suppose that a disease exists such that, in each period t:

- the probability of transmission from an infected node to a connected uninfected node is $p_{TRANS} = 0.01$
- the probability that an infected node recovers is $p_{REC} = 0.9$
- the probability of an uninfected node becoming infected spontaneously, independent of the nodes to which it is connected, is $p_{SPON} = 0.01$

We assume that enough medical research has been done so that these probabilities may be taken as known. Emphasis is therefore on learning about the network (p_{CON}) rather than learning about the disease.

Let w_{it} indicate if node i is healthy ($w_{it} = 0$) or infected ($w_{it} = 1$) in period t. The states of the individual nodes are not observed. What is observed is the overall infection rate $y_t = N^{-1} \sum_{i=1}^{N} w_{it}$. Figure 5.1 shows a typical series generated from the model. Note that the series exhibits autocorrelation, leptokurtosis, and volatility clustering, all of which are features associated with financial and some macroeconomic series.

We do Monte Carlo work using SNM applied to y_t to estimate p_{CON}. The time series length is $n = 50$. Because our intention here is only to sketch possibilities for use of SNM to estimate this sort of model, we will not burden the reader with the details of the test functions, conditioning variables, and so on. They may be found in the example code that is provided. We performed 100 replications, where at each trial the true p_{CON} is drawn from $U(0, 0.5)$. Figure 5.2 shows the results. We can see that the estimated p_{CON} closely tracks the true value, with good precision.

Some comments are in order. First, the model used here is highly structured: everything about the data generating process is known except the degree distribution and the specific network that generates the data. The size of the network and the characteristics of the disease are known. It would clearly be of interest to attempt to relax these assumptions. If we were modeling diffusion of an innovation, estimation of p_{TRANS} could be of central interest. Relaxation of the assumption of known network size would be needed in many applications. Second, the portion of the parameter space on which p_{CON} is well identified depends on the characteristics of the disease. If the disease is too infectious (p_{TRANS} too high), it will be difficult to identify p_{CON} when it is large, because in this case the probability that a node is infected will depend almost entirely on p_{REC}, and it will be insensitive to small changes in p_{TRANS} and p_{CON}. The results here are merely suggestive of the potential for estimating models with a latent network structure using SNM or other simulation-based estimators. For the highly specific case we examine, good results can be obtained. Further work would need to seriously examine identifiability, especially as assumptions are relaxed to allow for more interesting models.

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 in general 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:

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

where $U_{s,t}$, $s = 1, ..., S$, $t = 1, ..., n$, are i.i.d. draws. We here assume that we have observed the initial values (y_0, w_0) and then start the simulations there, $(Y_{s,0}(\theta), W_{s,0}(\theta)) = (y_0, w_0)$.[3] 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

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

By construction, $(\Phi_{s,t}(\theta), X_{s,t}(\theta))$, $s = 1, ..., S$, 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 Assumptions A.2 and A.7. For theoretical results in non-stationary environments, see Kitamura and Phillips (1997) for some results for linear models and Kristensen and Shin (2008) for simulation-based likelihood-inference.

---

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

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,

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

where

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

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

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

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

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

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

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 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. Our Monte Carlo results use moment

conditions based on test variables that are motivated in a large part by the characteristic function approach. A less obvious choice is the selection of the conditioning variables. It is desirable to capture the information contained in the history of the system, but there is also the need to economize on dimensionality, to avoid deterioration of the kernel fit without requiring an excessively long simulation, which is computationally burdensome.

The rest of the estimation process can be automatized in software to a high degree. In the present implementation, the kernel function is an Epanechnikov product kernel. The quality of a kernel fit is largely insensitive to the particular kernel that is chosen, as long as the bandwidth is chosen accordingly (Li and Racine, 2007. pg. 15). For this reason, we choose the Epanechnikov kernel for its computational advantages. The bandwidth is automatically chosen, in a data dependent way. Future work could explore other ways of choosing the bandwith, but at least for the applications we have explored in this paper, this strategy is both convenient and effective.

Some of the other estimators to which the SNM estimator is compared in this paper require a considerable degree of 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.

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 much worse. Generally speaking, the estimator is not excessively biased. 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 could improve efficiency of estimation, though exploratory work with some of the examples considered in this paper suggests that this improvement may not be great, at least for moment conditions of the type we use. Probably more important would be to use optimal or approximately optimal instruments.

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.

As was previously noted, the code for the SNM estimator, as well as all supporting software and the scripts to replicate the Monte Carlo examples given in this paper are available on the

PelicanHPC CD image (Creel, 2009), with documentation provided by Creel and Kristensen (2009).

APPENDIX

Appendix A. Proofs.

Proof. [Proof of Theorem 1] We claim that

(5.1)
$$\sup_{\theta \in \Theta} \left| \hat{G}_{n,S}(\theta)' W_n \hat{G}_{n,S}(\theta) - G_n(\theta)' W_n G_n(\theta) \right|$$

$$= O_P\left(a^{-1} h^m\right) + O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{d_x})}\right) + O_P\left(a^{-q}\right).$$

If this holds, it follows by standard results, see e.g. Newey and McFadden (1994, Theorem 2.1) that under Assumption 1 $\hat{\theta}_{n,S}$ is consistent if the three right hand terms go to zero. To prove the claim, write

$$\sup_{\theta \in \Theta} \left| \hat{G}_{n,S}(\theta)' W_n \hat{G}_{n,S}(\theta) - G_n(\theta)' W_n G_n(\theta) \right|$$

$$= \sup_{\theta \in \Theta} \left| \left[\hat{G}_{n,S}(\theta) - G_n(\theta) \right]' W_n \left[\hat{G}_{n,S}(\theta) + G_n(\theta) \right] \right|$$

$$\leq \sup_{\theta \in \Theta} \left\| \hat{G}_{n,S}(\theta) - G_n(\theta) \right\| \times \sup_{\theta \in \Theta} \left\{ \left\| \hat{G}_{n,S}(\theta) \right\| + \left\| G_n(\theta) \right\| \right\} \times \|W_n\|$$

$$\leq A_2 \times (A_1 + A_2) \times \|W_n\|,$$

where

$$A_1 = 2 \sup_{\theta \in \Theta} \|G_n(\theta)\|, \quad A_2 = \sup_{\theta \in \Theta} \left\| \hat{G}_{n,S}(\theta) - G_n(\theta) \right\|.$$

First, by Assumption 1, $\|W_n\| = O_P(1)$ and

$$A_1 \leq 2 \sup_{\theta \in \Theta} \|G_n(\theta) - G(\theta)\| + 2 \sup_{\theta \in \Theta} \|G(\theta)\| = O_P(1).$$

Second,

$$A_2 \leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \left\| \tau_{a,t}(\theta) \hat{T}(\phi)(x_t;\theta) - T(\phi)(x_t;\theta) \right\| \|z_t\|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \|z_t\| \times \sup_{\theta \in \Theta} \sup_{\hat{f}(x;\theta) \geq a/2} \left\| \hat{T}(\phi)(x;\theta) - T(\phi)(x;\theta) \right\|$$

$$+ \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}\left\{ \hat{f}(x_t;\theta) < a/2 \right\} \|T(\phi)(x_t;\theta)\| \|z_t\|$$

$$= A_{2,1} + A_{2,2}.$$

Here, $A_{2,1} = O_P\left(a^{-1}h^m\right) + O_P(a^{-1}\sqrt{\log(S)/(Sh^{d_x})})$ by Lemma 4 together with the fact that $n^{-1}\sum_{t=1}^{n}\|z_t\| = O_P(1)$, while the second term satisfies:

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

where the last equality follows from Assumption 2.3 and the LLN. This completes the proof of Eq. (5.1). $\qquad\square$

Proof. [Proof of Theorem 2] We have that $\hat{\theta}_{n,S}$ is consistent. So with probability approaching one, $\hat{\theta}_{n,S}$ lies in a (small) neighbourhood of θ_0 which in turn lies in the interior of Θ. Thus, the SNM estimator solves the following first-order condition,

$$
\hat{H}_{n,S}(\hat{\theta}_{n,S})'W_n\hat{G}_{n,S}(\hat{\theta}_{n,S}) = 0,
$$

where

$$
\hat{H}_{n,S}(\theta) = -\frac{1}{n}\sum_{t=1}^{n}\tau_{a,t}(\theta)Z_t'\frac{\partial\hat{T}_S(\phi)(x_t;\theta)}{\partial\theta} + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tau_{a,t}(\theta)}{\partial\theta}Z_t'\left[\phi_t - \hat{T}_S(\phi)(x_t;\theta)\right].
$$

By a standard Taylor expansion of $\hat{G}_{n,S}(\hat{\theta}_{n,S})$ around θ_0,

$$
\hat{G}_{n,S}(\hat{\theta}_{n,S}) = \hat{G}_{n,S}(\theta_0) + \hat{H}_{n,S}\left(\bar{\theta}_{n,S}\right)(\hat{\theta} - \theta_0),
$$

where $\bar{\theta}_{n,S}$ lies between $\hat{\theta}_{n,S}$ and θ_0. Substituting this back into the first-order condition and rearranging the terms yields

$$
(5.2)\qquad \sqrt{n}(\hat{\theta}_{n,S} - \theta_0) = \left[\hat{H}_{n,S}(\hat{\theta}_{n,S})'W_n\hat{H}_{n,S}\left(\bar{\theta}_{n,S}\right)\right]^{-1}\hat{H}_{n,S}(\hat{\theta}_{n,S})'W_n\sqrt{n}\hat{G}_{n,S}(\theta_0).
$$

We have

$$
\begin{aligned}
\hat{H}_{n,S}(\theta) - H_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n}\tau_{a,t}(\theta)Z_t'\left[\frac{\partial\hat{T}_S(\phi)(x_t;\theta)}{\partial\theta} - \frac{\partial T_S(\phi)(x_t;\theta)}{\partial\theta}\right] \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}(\tau_{a,t}(\theta) - 1)Z_t'\frac{\partial T_S(\phi)(x_t;\theta)}{\partial\theta} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tau_{a,t}(\theta)}{\partial\theta}Z_t'\left[\phi_t - \hat{T}_S(\phi)(x_t;\theta)\right].
\end{aligned}
$$

By the same arguments as in the proof of Theorem 1 combined with Lemmas 4-5, we obtain that each of the right hand side terms converge to zero uniformly over $\{\theta : \|\theta - \theta_0\| < \delta\}$ under the conditions given in the theorem. This combined with Assumption A.7.3, yields that $\hat{H}_{n,S}(\hat{\theta}_{n,S})$

and $\hat{H}_{n,S}\left(\bar{\theta}_{n,S}\right)$ both converge in probability towards H_0. Combining Eqs. (5.2) and (5.6),

$$\hat{\theta}_{n,S} - \hat{\theta}_n = \left(\left[H_0'W_0H_0\right]^{-1}H_0'W_0 + o_P\left(1\right)\right)\left[\hat{G}_{n,S}(\theta_0) - G_n(\theta_0)\right].$$

We now analyze $\hat{G}_{n,S}(\theta_0) - G_n(\theta_0)$ in further detail: Write

$$\hat{G}_{n,S}(\theta_0) = \bar{G}_n(\theta_0) + \bigtriangledown\bar{G}_n(\theta_0)\left[\hat{T}_S - T\right],$$

where

$$\bar{G}_n(\theta_0) = \frac{1}{n}\sum_{t=1}^{n}\tau_{a,t}\left(\theta_0\right)Z_t'T\left(\phi\right)\left(x_i;\theta_0\right),$$

is the trimmed version of the true moment conditions, and

$$\bigtriangledown\bar{G}_n(\theta_0)\left[\hat{T}_S - T\right] = \frac{1}{n}\sum_{t=1}^{n}\tau_{a,t}\left(\theta_0\right)Z_t'\left[\hat{T}_S\left(\phi\right)\left(x_i;\theta_0\right) - T\left(\phi\right)\left(x_i;\theta_0\right)\right]$$

is an adjustment term measuring the impact of the simulations. As shown in the proof of Theorem 1, $\bar{G}_n(\theta_0) = G_n(\theta_0) + O_P\left(a^q\right)$.

We further decompose the adjustment term into

$$(5.3) \qquad \bigtriangledown\bar{G}_n(\theta_0)\left[\hat{T}_S - T\right] = \bigtriangledown\bar{G}_n(\theta_0)\left[\hat{T}_S - E[\hat{T}_S]\right] + \bigtriangledown\bar{G}_n(\theta_0)\left[E[\hat{T}_S] - T\right],$$

where $E[\hat{T}_S\left(\phi\right)\left(x;\theta_0\right)]$ is the (conditional on x) expectation w.r.t. the simulations. By standard results for bias of kernel regression estimators,

$$E[\hat{T}_S\left(\phi\right)\left(x;\theta_0\right)] = T\left(\phi\right)\left(x;\theta_0\right) + h^m\frac{1}{f\left(x;\theta_0\right)}\frac{\partial^r\left[T\left(\phi\right)\left(x;\theta_0\right)f\left(x;\theta_0\right)\right]}{\partial x^r} + o\left(h^m\right),$$

uniformly over x. Plugging this expression into the expression of $\bigtriangledown\bar{G}_n(\theta_0)\left[E\left[\hat{T}_S\right] - T\right]$ and appealing to the Law of Large Numbers,

$$\begin{aligned}\bigtriangledown\bar{G}_n(\theta_0)\left[E[\hat{T}_S] - T\right] &= h^m\frac{1}{n}\sum_{t=1}^{n}\tau_{a,t}\left(\theta_0\right)\frac{Z_t'}{f\left(x_t;\theta_0\right)}\frac{\partial^r\left[T\left(\phi\right)\left(x_t;\theta_0\right)f\left(x_t;\theta_0\right)\right]}{\partial x^r} + o\left(h^m\right)\\ &= h^m E\left[\sum_{|\alpha|=m}\frac{\partial^{|\alpha|}\left[T\left(\phi\right)\left(x_t;\theta_0\right)f\left(x_t;\theta_0\right)\right]}{\partial x^\alpha}\right] + o\left(h^m\right).\end{aligned}$$

Next, the first term in Eq. (5.3) can be written as

$$(5.4) \qquad \bigtriangledown\bar{G}_n(\theta_0)\left[\hat{T}_S - E[\hat{T}_S]\right] = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{S}\Delta\left(x_t, Z_t, V_s\right) + R_n,$$

where $V_s = \left(Y_s\left(\theta_0\right), X_s\left(\theta_0\right)\right)$,

$$\Delta_h\left(x_t, Z_t, V_s\right) := \frac{Z_t'}{f\left(x_t;\theta_0\right)}\psi_h\left(x_t, V_s\right),$$

$$\psi_h\left(x_t, V_s\right) := \phi\left(Y_s\left(\theta_0\right)\right)K_h\left(X_s\left(\theta_0\right) - x\right) - E\left[\phi\left(Y_s\left(\theta_0\right)\right)K_h\left(X_s\left(\theta_0\right) - x\right)\right].$$

Here, $R_{n,S}$ is a higher-order term containing the effects of trimming which can be ignored. We recognize the sum on right hand side of Eq. (5.4) as a U-statistic. Define

$$D_h(V_s) = E[\Delta_h(x_t, Z_t, V_s) | V_s],$$

and

$$W_{t,S} := \frac{1}{S} \sum_{s=1}^{S} \{\Delta_h(x_t, Z_t, V_s) - D_h(V_s)\}.$$

Conditionally on the simulations which we collect in $\mathscr{E}_S$, it is easily seen that $\{W_{t,S}\}$ satisfies the conditions of Lemma 6 such that

$$E\left[\left\|\frac{1}{n}\sum_{t=1}^{n} W_{t,S}\right\|^2 | \mathscr{E}_S\right] \leq C(1,A) E\left[\|W_{t,S}\|^{2+\delta} | \mathscr{E}_S\right] n^{-1},$$

where in turn

$$E\left[\|W_{t,S}\|^{2+\delta}\right] \leq \frac{1}{S^{1+\delta/2}} C(1,A) E\left[\left\|\frac{\tau_t Z_t'}{f(x_t;\theta_0)}\psi_h(x_t, V_s)\right\|^{2+2\delta}\right],$$

and, by standard arguments,

$$E\left[\left\|\frac{\tau_{a,t}(\theta_0) Z_t'}{f(x_t;\theta_0)}\psi_h(x_t, V_s)\right\|^{2+2\delta}\right] = O\left(\frac{1}{h^{d+\delta}}\right).$$

We now have that

$$\bigtriangledown\bar{G}_n(\theta_0)\left[\hat{T}_S - E[\hat{T}_S]\right] = \frac{1}{S}\sum_{s=1}^{S} D_h(V_s) + O_P\left(\frac{1}{\sqrt{nSh^{d+\delta}}}\right),$$

where, leaving out higher order terms,

$$\begin{aligned}
D_h(V_s) &\simeq E\left[\frac{Z_t'}{f(x_t;\theta_0)}\{\phi(Y_s(\theta_0)) K_h(X_s(\theta_0) - x_t) - E[\phi(Y_s(\theta_0)) K_h(X_s(\theta_0) - x_t)]\} | V_s\right] \\
&\simeq \int \frac{Z(x)'}{f(x;\theta_0)}\phi(Y_s(\theta_0)) K_h(X_s(\theta_0) - x_t) f(x) dx - \int Z(x)' T(\phi)(x_t;\theta_0) f(x) dx \\
&\simeq Z_s(\theta_0)' \phi(Y_s(\theta_0)) - E[Z_s(\theta_0)' \phi(Y_s(\theta_0))].
\end{aligned}$$

This proves the first part of the theorem.

The second part follows by combining the first part with Eq. (5.6). $\qquad\square$

Appendix B. Lemmas.

Lemma 3. *Under Assumption 1, $\hat{\theta}_n$ is consistent. If in addition Assumption 6 holds,*

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

where H_0, Ω_0 and W_0 are given in Assumptions 1 and 6.

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:

$$(5.5) \qquad H_n(\hat{\theta}_n)'W_nG_n(\hat{\theta}_n) = 0,$$

where $H_n(\theta)$ is given in Assumption . 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 $\bar{\theta}_{n,i} \in [\hat{\theta}_{n,i}, \theta_{0,i}]$, $i = 1, ..., d_\theta$. Substituting this into Eq. (5.5), and rearranging the terms yields

$$(5.6) \qquad \sqrt{n}(\hat{\theta}_n - \theta_0) = \left[H_n(\hat{\theta}_n)'W_nH_n(\bar{\theta}_n) \right]^{-1} H_n(\hat{\theta}_n)'W_n\sqrt{n}G_n(\theta_0),$$

which converges in distribution towards the claimed normal distribution by Assumption 6. □

Lemma 4. *Under Assumptions 2-4, the simulated conditional moment estimator satisfies:*

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

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

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

It is easily checked that our Assumptions A.3-A.5 imply the conditions (A.1)-(A.5) in Kristensen (2009). His Theorem 2 combined with standard bias expansions of higher-order kernel estimators now yield that

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

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

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

Lemma 5. *Under Assumptions 2-4 and 7-9, the simulated conditional moment estimator satisfies:*

$$\sup_{\theta \in \Theta} \sup_{x: \hat{f}(x) \geq a} \left\| \frac{\partial \hat{T}(\phi)(x; \theta)}{\partial \theta} - \frac{\partial T(\phi)(x; \theta)}{\partial \theta} \right\| = O_P\left(a^{-1}h^m \right) + O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{d_x+2})} \right),$$

Proof. With $\hat{g}(x; \theta)$ and $\hat{f}(x; \theta)$ given in the proof of Lemma 4,

$$\frac{\partial \hat{T}(\phi_k)(x; \theta)}{\partial \theta} = \frac{\partial \left[\hat{g}(x; \theta)/\hat{f}(x; \theta) \right]}{\partial \theta} = \frac{\partial \hat{g}(x; \theta)/\partial \theta}{\hat{f}(x; \theta)} - \frac{\hat{g}(x; \theta)}{\hat{f}(x; \theta)} \frac{\partial \hat{f}(x; \theta)/\partial \theta}{\hat{f}(x; \theta)},$$

where

$$(5.7) \qquad \frac{\partial \hat{g}(x;\theta)}{\partial \theta} = \frac{1}{S}\sum_{s=1}^{S} \frac{\partial \phi_k(Y_s(\theta))}{\partial y} \dot{Y}_s(\theta) K_h(X_s(\theta)-x)$$

$$+ \frac{1}{S}\sum_{s=1}^{S} \phi_k(Y_s(\theta)) \dot{X}_s(\theta) K'_h(X_s(\theta)-x)$$

and

$$(5.8) \qquad \frac{\partial \hat{f}(x;\theta)}{\partial \theta} = \frac{1}{S}\sum_{s=1}^{S} \dot{X}_s(\theta) K'_h(X_s(\theta)-x).$$

We first note that

$$E\left[\frac{\partial \hat{g}(x;\theta)}{\partial \theta}\right] = \frac{\partial g(x;\theta)}{\partial \theta} + O(h^m), \quad E\left[\frac{\partial \hat{f}(x;\theta)}{\partial \theta}\right] = \frac{\partial f(x;\theta)}{\partial \theta} + O(h^m),$$

uniformly over $(x;\theta)$. Next, we see that Assumptions 2-4 and 7-9 imply that the conditions (A.1)-(A.5) in Kristensen (2009) are met for each of the three kernel averages in Eqs. (5.7)-(5.8). His Theorem 2 then yields that the variance components of the first kernel average in Eq. (5.7) is of order $O(\log(S)/(Sh^{d_x}))$ while the two remaining ones both are of order $O(\log(S)/(Sh^{d_x+2}))$. Using the same arguments as in Andrews (1995, Proof of Theorem 1), the result now follows. $\qquad\qquad\square$

Lemma 6. *Assume that $\{W_t\}$ is a stationary and α-mixing sequence with mixing coefficients α_t, $t = 1,2,...$, that satisfy $\alpha_t \le At^{-a}$ for some $A,a > 0$. Also, assume that $E[W_t] = 0$ and $E\left[\|W_t\|^{2r+\delta}\right] < \infty$ for some $r,\delta > 0$. Then, there exists a constant $C(r,A) < \infty$ such that:*

$$E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}W_t\right\|^{2r}\right] \le C(r,A) E\left[\|W_t\|^{2r+\delta}\right]^{2/(2r+\delta)} n^{r(q-1)},$$

where

$$q = \frac{2r}{2r+r_0} < 1, \quad r_0 = \frac{a\delta}{2r+\delta} > 0.$$

In particular, if the sequence is geometrically mixing, we can choose $q = 0$.

Proof. From Hahn and Kuersteiner (2004, Lemma 7), there exists a constant $c(r) < \infty$ such that for any $1 \le m \le c(r)n$:

$$E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}W_t\right\|^{2r}\right] \le c(r) E\left[\|W_t\|^{2r+\delta}\right]^{2/(2r+\delta)} \left[n^{-r}m^{2r} + \alpha_m^{\delta/(2r+\delta)}\right]$$

$$= c(r) E\left[\|W_t\|^{2r+\delta}\right]^{2/(2r+\delta)} \left[n^{-r}m^{2r} + Am^{-a\delta/(2r+\delta)}\right]$$

By choosing $m = c(r) n^{q/2}$, we obtain

$$E\left[\left\|\frac{1}{n}\sum_{t=1}^{n} W_t\right\|^{2r}\right] \leq c(r)\left[1 + Ac(r)\right] E\left[\|W_t\|^{2r+\delta}\right]^{2/(2r+\delta)} n^{r(q-1)}.$$

$\square$

References

[1] Ai, C. (1997) A semiparametric maximum likelihood estimator, *Econometrica*, **65**, 933-963.

[2] Altissimo, F. and A. Mele (2009) Simulated nonparametric estimation of dynamic models, *Review of Economic Studies,* **76**, 413-450.

[3] Anatolyev, S. (2002) Approximately optimal instrument for multiperiod conditional moment restrictions, working paper, New Economic School.

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

[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] Copic, J., M. Jackson and A. Kirman (2009) Identifying community structures from network data via maximum likelihood methods, *The B.E. Journal of Theoretical Economics*, **9** : Iss. 1 (Contributions), Article 30.

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

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

[18] Creel, M. (2009) PelicanHPC, GRECS computer code,
http://econpapers.repec.org/software/aubgrecss/005.09.htm

[19] Creel, M. and D. Kristensen (2009) SNM Guide, UFAE-IAE working paper,
http://econpapers.repec.org/paper/aubautbar/793.09.htm

[20] Domínguez, M.A. and I.N. Lobato (2004) Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, **72**, 1601-1615.

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

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

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

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

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

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

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

[28] 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*.

[29] 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/`.

[30] 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`.

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

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

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

[34] Hahn, J. and G. Kuersteiner (2004) Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects, manuscript, Department of Economics, MIT.

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

[36] Hajivassiliou, V. (2000) Some practical issues in maximum simulated likelihood. In *Simulation-based Inference in Econometrics,* (eds. R. Mariano, T. Schuermann and M.J. Weeks), 71-99. Cambridge: Cambridge University Press.

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

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

[39] Härdle, W, P. Hall and H. Ichimura (1993) Optimal smoothing in single-index models, *The Annals of Statistics*, **21**, 157-178.

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

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

[42] Jackson, M. and L. Yariv (in press) Diffusion, strategic interaction, and social structure, forthcoming in *Handbook of Social Economics* (edited by Benhabib, Bisin, Jackson), Elsevier.

[43] Jacquier, E., Polson, N.G., and P. Rossi (1994) Bayesian analysis of stochastic volatility models, (with discussion), *Journal of Business and Economic Statistics,* **12**, 371-417.

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

[45] 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.

[46] Kristensen, D. (2008) Estimation of partial differential equations with applications in finance. *Journal of Econometrics*, **144**, 392-408.

[47] Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogenous, dependent data, *Econometric Theory,* **25**, 1433-1445.

[48] Kristensen, D. (2010) Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. Forthcoming in *Journal of Econometrics*.

[49] Kristensen, D (2007) 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.

[50] Kristensen, D. and B. Salanié (2009) Higher order improvements for approximate estimators, working paper, Department of Economics, Columbia University.

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

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

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

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

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

[56] 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.

[57] Newey, W. and D. McFadden (1994), Large sample estimation and hypothesis testing, in *Handbook of Econometrics*, Vol IV, ed. R. Engle and D. McFadden, Elsevier Science, 2111-2245.

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

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

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

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

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

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

[64] Sandman, G. and Koopman, S. (1998) Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics*, **67**, 271-301.

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

[66] 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.

[67] Weiss, A. (1996) Estimating time series models using the relevant cost function, *Journal of Applied Econometrics*, **11**, 539-560.

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

[69] 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.

TABLES

Table 1: Monte Carlo Results: AR1 model. $n = 50$. Mean and root mean squared error (in parentheses).

Estimator	$\alpha = 0.0$	$\beta = 0.9$
SNM	-0.011 (0.288)	0.861 (0.091)
OLS	0.003 (0.275)	0.823 (0.117)

Table 2: Monte Carlo Results: Classical linear model. SNM-GMM is the difference between the SNM and the GMM estimators.

	$\hat{\beta}_1$				$\hat{\beta}_2$			
	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max
SNM-GMM	-0.001	0.004	-0.015	0.014	0.002	0.007	-0.020	0.025
GMM	0.489	0.473	-0.867	2.143	0.524	0.723	-1.510	2.964

Table 3: Monte Carlo Results: Stochastic volatility, SV1 model. $n = 500$. Mean and root mean squared error (in parentheses). Sources for other estimators: ML - Fridman and Harris (1996); EMM - Andersen, *et al.* (1999); MCMC - Jacquier *et al.* (1994); QML and MCL - Sandmann and Koopman (1998).

Estimator	$\alpha = -0.736$	$\beta = 0.9$	$\sigma = 0.363$
SNM	-0.736 (0.201)	0.882 (0.080)	0.380 (0.128)
ML	-0.87 (0.43)	0.88 (0.05)	0.37 (0.08)
EMM	-0.91 (0.60)	0.88 (0.08)	0.38 (0.20)
MCMC	-0.87 (0.34)	0.88 (0.046)	0.35 (0.067)
QML	-0.736 (0.02)	0.845 (0.18)	0.417 (0.21)
MCL	-0.745 (0.02)	0.897 (0.10)	0.325 (0.07)

Table 4: Monte Carlo Results: Stochastic volatility SV2 model. $n = 500$. Mean and root mean squared error (in parentheses). Sources for other estimators: NPSML - Fermanian and Salanié (2004); CD-SNE and J-SNE - Altissimo and Mele (in press).

Estimator	$\sigma_b = 0.025$	$\beta = 0.95$	$\sigma = 0.26$
SNM	0.025 (0.003)	0.907 (0.079)	0.331 (0.135)
NPSML	0.022 (0.004)	0.913 (0.107)	0.318 (0.180)
CD-SNE	0.024 (0.003)	0.909 (0.110)	0.229 (0.134)
J-SNE	0.027 (0.005)	0.942 (0.095)	0.297 (0.149)

Table 5: Monte Carlo Results: AR Tobit model. $n = 150$. Mean and root mean squared error (in parentheses). Source for NPSML: Fermanian and Salanié (2004).

Estimator	$\alpha = 0.0$	$\beta = 0.5$	$\sigma = 1.0$
SNM	-0.001 (0.094)	0.523 (0.159)	0.966 (0.140)
NPSML	-0.010 (0.215)	0.510 (0.151)	0.810 (0.264)

Table 6: Monte Carlo Results: Factor ARCH model. Mean and root mean squared error (in parentheses). Source for Other estimator: Billio and Monfort (2003, Tables 3, 4 and 5, pp. 313-317). The Other estimator is that with the lowest RMSE for the given parameter.

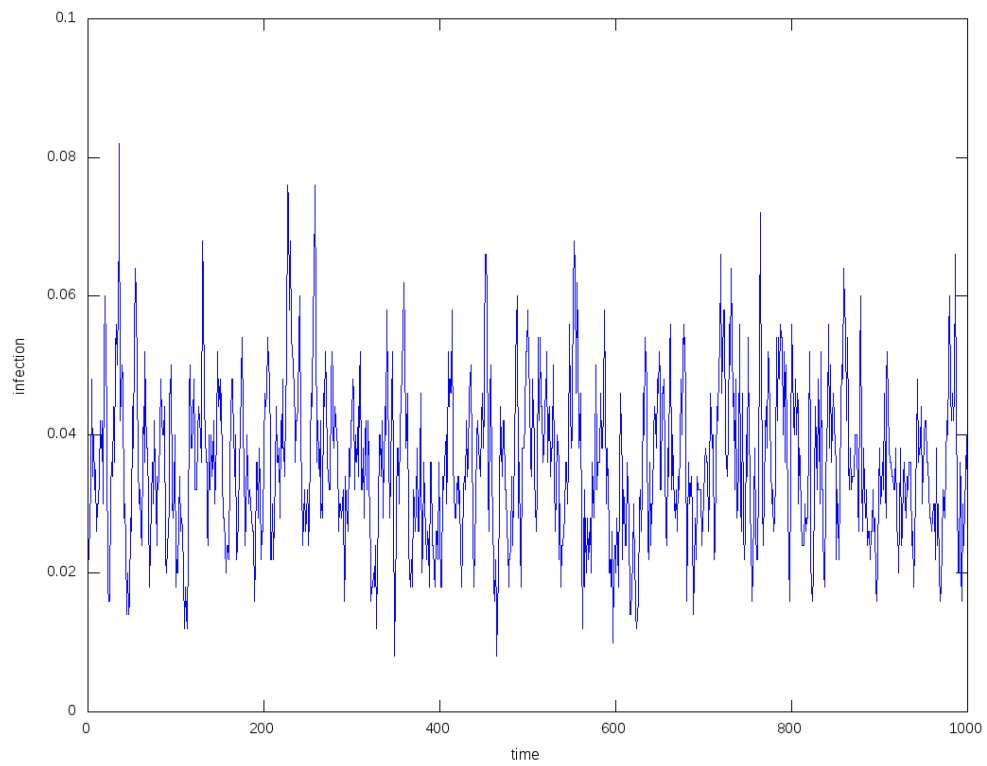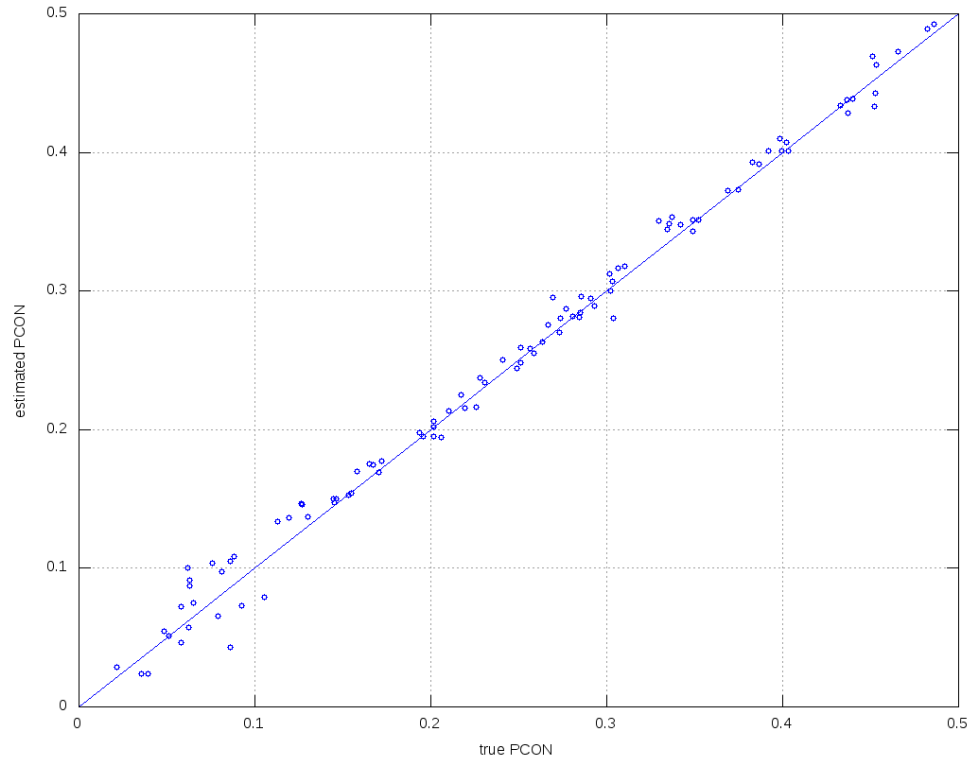| Estimator	$\alpha_1 = 0.2$	$\alpha_2 = 0.7$	$\sigma_0 = 0.5$	$\beta_{20} = -0.5$
SNM	0.223 (0.069)	0.681 (0.192)	0.480 (0.048)	-0.522 (0.069)
Other	0.244 (0.132)	0.659 (0.309)	0.461 (0.141)	-0.445 (0.269)

FIGURES

Figure 5.1: Overall infection rate y_t, latent network model

Figure 5.2: Latent network model, true and estimated p_{CON}, with 45º line for reference

UNIVERSITAT AUTÒNOMA DE BARCELONA AND MOVE

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