Econometrics Journal (2012), volume **01**, pp. **??**–**??**.
Article No. ECTJ-1549

Estimation of Dynamic Latent Variable Models Using Simulated Nonparametric Moments

Michael Creel[†] and Dennis Kristensen[‡]

[†] *Universitat Autònoma de Barcelona, Barcelona Graduate School of Economics and MOVE*
E-mail: michael.creel@uab.es
[‡] *University College London,*
CEMMAP (Centre for Microdata Methods and Practice, Institute of Fiscal Studies)
and CREATES (Center for Research in Econometric Analysis of Time Series, University of Aarhus).
E-mail: d.kristensen@ucl.ac.uk
Received: Feb. 2011

Summary We propose a novel estimation method for dynamic latent variable (DLV) models that combines simulations and nonparametric kernel smoothing techniques to obtain a GMM estimator based on a set of conditional moments. As such it extends the simulated method of moments (SMM) of Duffie and Singleton (1993, *Econometrica*) to allow for the use of conditional moments, instead of unconditional ones. It can also be seen as a generalization of the SMM for static models as proposed in McFadden (1989, *Econometrica*). It is shown that, as the number of simulations diverges and the bandwidth used in the kernel smoothing shrinks, the estimator is consistent and a higher-order expansion reveals the stochastic difference between the infeasible GMM estimator based on exact computation of the conditional moment conditions and the simulated version. In particular, the expansion demonstrates how simulations impact the bias and variance of the proposed estimator. Extensive Monte Carlo results show how the estimator may be applied to a range of DLV models, and that it performs well in comparison to several other estimators that have been proposed in the literature.

Keywords: *dynamic latent variable models, simulation-based estimation, simulated moments, kernel regression, nonparametric estimation*

1. INTRODUCTION

Dynamic latent variables (DLV's) are present in many economic models such as in dynamic stochastic general equilibrium (DSGE), stochastic volatility, and yield curve models. The presence of latent variables complicates the estimation of model parameters, and most existing methods can be quite cumbersome to implement or are model-specific.

We propose a novel, easy-to-implement class of GMM estimators of general DLV models based on conditional moment restrictions implied by the parametric model: For a given candidate value of the parameter of interest, we first simulate a long trajectory of observations from the model. We feed those into a nonparametric kernel regression smoother to obtain an estimate of the conditional mean operator given a set of conditioning variables chosen by the econometrician. Once this simulated estimator of any given set of conditional moment restrictions is available, we can proceed to estimate the parameter using standard GMM methods. The estimator is referred to as the simulated nonparametric moments (SNM) estimator.

Under regularity conditions, we derive the convergence rate of the SNM estimator and establish a higher-order expansion of the estimator relative to the infeasible GMM estimator assuming that the conditional moments can be evaluated exactly. The expansion reveals that the SNM estimator contains additional bias and variance components due to the use of kernel smoothing and simulations. At the same time, the expansion also demonstrates that the SNM estimator does not suffer from the so-called curse of dimensionality normally associated with kernel regression estimators: The order of the variance of a kernel regression estimator in general increases with the dimension of the conditioning variables. However, due to the SNM estimator involving summation over the individual kernel estimates, the resulting variance of the SNM estimator is in contrast invariant to the dimension of the conditioning variables. In particular, it is of the same rate as for unconditional simulated method of moments (SMM). As such we pay no price in terms of first-order variance for using kernel smoothers in our estimation procedure.

Extensive Monte Carlo studies investigate the finite-sample performance of the SNM estimator relative to existing methods across a range of different DLV models, and it is found to perform as well, if not better, compared to competing estimators. An important part of the implementation of the SNM estimator is the choice of bandwidth. We propose a very simple bandwidth selection rule, and find in the Monte Carlo studies that it does well. We also do some limited exploration of the the sensitivity of the estimator to the choice of conditioning variables and to the choice of smoother (instead of kernel regression, any other type of nonparametric regression technique can be employed).

The SNM estimator can be seen as a fairly obvious extension of the SMM method as developed in McFadden (1989) for static models and Duffie and Singleton (1993) for dynamic models. McFadden (1989) shows how one can obtain conditional moments through simulations in static models where the unobserved variables contain no dynamics. His method requires one to be able to directly simulate the dependent variables conditional on the chosen set of conditioning varibles. This is in general not feasible in DLV models since, due to the dynamics in the latent variables, the conditioning variables will be correlated with all past realizations of the latent variables. Duffie and Singleton (1993) develop a SMM estimator for DLV models by also simulating a long trajectory from the model but instead use these to construct unconditional moments. 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).

Given our focus is on fully specified models, efficient estimation of parameters could in principle be done by maximum-likelihood. However, calculation of the likelihood function requires integrating out the DLV's; this involves calculating high-dimensional integrals of the same order as the number of observations. A number of methods to handle computation of these integrals have been developed in recent years, including Expectation-Maximization (EM) algorithms (Fiorentini, Sentana and Shephard, 2004), Markov Chain Monte Carlo (MCMC) methods (Jacquier et al, 2007) and particle filtering (Brownlees, Kristensen and Shin, 2011; Johansen, Doucet and Davy, 2008). However, they are computationally demanding, and their implementation can be quite delicate and require substantial fine-tuning.

Moreover, as is well-known, the likelihood principle is in general quite sensitive to misspecification. So in many applications it is desirable to trade efficiency for robustness and instead calibrate/estimate the model by matching an intelligently selected subset of features of the model with those of the data. The SNM estimator does exactly this and so 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 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, as briefly discussed in Section 4. In this setting, if the DLV model is misspecified, our SNM estimator will often be the better vehicle since it finds the parameter estimates that minimize the forecasting error. In particular, forecasts based on suitably chosen SNM estimates will in general dominate those based on MLE's. For a formal argument of this point, we refer to Weiss (1996).

A number of other non-likelihood based estimation methods for DLV models exist such as indirect inference (Gouriéroux, Monfort and Renault, 1993; Smith, 1993), efficient method of moments (Gallant and Tauchen, 1996), and simulated quasi-likelihood estimators (Altissimo and Mele, 2009; Kristensen and Shin, 2012). In its most general form, indirect inference (II) matches a sample statistic with its model-implied moment - a standard choice of this statistic is an estimator from a so-called auxiliary model. The efficient method of moments (EMM) is closely related to II and also uses an auxiliary model to construct an estimator. Similar to II and EMM, our method relies on choosing an informative set of conditional moments to guarantee identification and efficient estimation of the parameters. In particular, the SNM estimator falls within the framework of II with the statistic chosen as the set of moments generated from the conditional moment restrictions. However, our method does not involve a nested optimization problem which is often the case with II.

Finally, the simulated quasi-maximum likelihood methods of Altissimo and Mele (2009) and Kristensen and Shin (2012) are similar to the SNM except that the simulated data is fed into a kernel density estimate instead of the kernel regression estimator used here.

The remainder of the paper is organized as follows: 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 dis-

cusses some extensions, including forecasting and non-stationary models, while Section 5 concludes. All proofs and lemmas have been relegated to the Appendix.

2. THE SNM ESTIMATOR

2.1. Definition of the estimator

Suppose we have observed y_t, $t = 1, ..., n$, from the following dynamic model:

$$\text{DLV: } \begin{cases} y_t = & r_y\left(y^{t-1}, w^{t-1}, u_t; \theta\right) \\ w_t = & r_w\left(y^{t-1}, w^{t-1}, u_t; \theta\right) \end{cases}, \tag{2.1}$$

where w_t is a vector of dynamic latent variables, and u_t is a vector of independent white noise shocks with 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 $w^{t-1} \equiv \left(w_1', ..., w_{t-1}'\right)'$. The two functions, r_y and r_w are known up to some parameter $\theta \in \Theta$; let $\theta_0 \in \Theta$ denote the true data-generating parameter value. Note that u_t enters the equations for both the observable and latent variables, to allow for potential correlations in the innovations of the two sets of variables.

We first choose a set of "conditioning variables" $x_t = (x_{t,1}, ..., x_{t,d_x})' \in \mathbb{R}^{d_x}$ that are functions of leads and lags of y_t; say, $x_t = x\left(y_{t-q:t+q}\right)$ for some $q \geq 1$ where $y_{s:t} := (y_s, ..., y_t)$ for any $s \leq t$. 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, we choose a collection of "test variables", $\phi_t = (\phi_{t,1}, ..., \phi_{t,L})' \in \mathbb{R}^L$, which are functions of leads and lags of the observations; for example, $\phi_t = \phi\left(y_{t-p:t+p}\right)$ for some $p \geq 1$. There are no restrictions on how ϕ_t is chosen in conjunction with x_t, except that the two variables together should identify the parameter of interest through the corresponding generalized residual functions defined as:

$$\varepsilon_t(\theta) = \phi_t - T\left(\phi\right)\left(x_t; \theta\right) \in \mathbb{R}^L, \quad T\left(\phi\right)\left(x; \theta\right) = E_\theta\left[\phi_t | x_t = x\right], \tag{2.2}$$

where $E_\theta\left[\cdot | x_t\right]$ denotes conditional expectations taken under the model evaluated at θ, $E_\theta\left[\phi_t | x_t\right] = \int \phi\left(y_t, y_{t-1}, ...\right) dP_\theta\left(y_t, y_{t-1}, ... | x_t\right)$. By construction, the residual vector satisfies

$$E_\theta\left[\varepsilon_t\left(\theta\right) | x_t\right] = 0. \tag{2.3}$$

For a set of instruments chosen as functions of the conditioning variables, $z_t = z\left(x_t\right) \in \mathbb{R}^{L \times M}$, moment conditions are now defined by interacting instruments with residuals,

$$g_t\left(\theta\right) := z_t' \varepsilon_t\left(\theta\right) \in \mathbb{R}^M, \tag{2.4}$$

such that $E_\theta\left[g_t\left(\theta\right)\right] = 0$. For these to identify θ_0, we require that $E_{\theta_0}\left[g_t\left(\theta\right)\right] = 0$ if and only if $\theta = \theta_0$. In the following, for notational simplicity, we write $E\left[\cdot\right]$ for $E_{\theta_0}\left[\cdot\right]$.

If the conditional moments $T\left(\phi\right)\left(x_t; \theta\right)$ in equation (**??**) have a known functional form, estimation may proceed using the standard generalized method of moments (GMM): For some sequence of weighting matrices, $W_n \in \mathbb{R}^{M \times M}$, we would compute the GMM estimator as

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} G_n\left(\theta\right)' W_n G_n\left(\theta\right), \quad G_n\left(\theta\right) = \frac{1}{n} \sum_{t=1}^{n} g_t\left(\theta\right).$$

In general, due to the latent dynamic variables, no closed-form functional form of

$T(\phi)(x_t;\theta)$ is available, and we instead propose to approximate the conditional moments by feeding simulations into a nonparametric kernel regression smoother (Li and Racine, 2007, Ch. 2). 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 trajectory of S simulations of generated by Eq. (**??**) at the trial parameter value θ:

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

for $s=1,....S$, where the simulations are initialized at some values $\left(Y^{-1}(\theta), W^{-1}(\theta)\right)$ (for example, the final value of a burn-in period of simulations). Given the simulated values, first compute the corresponding conditioning and test variables, say, $X_s(\theta) = x\left(Y_{s-q:s+q}(\theta)\right)$ and $\Phi_s(\theta) = \phi\left(Y_{s-p:s+p}(\theta)\right)$, and then use these to obtain the following kernel estimator of $T(\phi)(x_t;\theta)$:

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

where $K_h(z) = K(z/h)/h$, $K:\mathbb{R}^{d_x}\mapsto\mathbb{R}$ is a kernel function, and $h>0$ is a bandwidth. To speed up computations, one should not separately fit each of the L test variables, but rather employ a specialized kernel fitting algorithm that saves the weights across variables. Since $d_x = \dim(X_s(\theta))$ 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. In practice, if the individual variables contained in $X_s(\theta)$ are not on the same scale, one should either use different bandwidths for the invidual variables, or rescale $X_s(\theta)$ before implementing the kernel smoother.

A simulated version of the GMM estimator is now obtained by replacing the unknown expectations operator T with its kernel fit $\hat{T}_S$. To be explicit, we first compute the simulated version of the moment functions,

$$\hat{g}_{t,S}(\theta) := z_t'\hat{\varepsilon}_{t,S}(\theta) \in \mathbb{R}^M, \quad \hat{\varepsilon}_{t,S}(\theta) = \phi_t - \hat{T}_S(\phi)(x_t;\theta) \tag{2.7}$$

where we use the observed instruments. The SNM estimator is then defined as

$$\hat{\theta}_{n,S} = \arg\min_{\theta\in\Theta}\hat{G}_{n,S}(\theta)'W_n\hat{G}_{n,S}(\theta), \quad \hat{G}_{n,S}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\hat{g}_{t,S}(\theta), \tag{2.8}$$

The above algorithm is very similar to existing SMM estimators as first proposed in McFadden (1989) for static models and Duffie and Singleton (1993) for dynamic models. The only difference is that we employ kernel smoothers. Duffie and Singleton (1993) also simulate long trajectories, but use these to obtain *un*conditional moments, say $\hat{\Phi}_S(\theta) = \sum_{s=1}^{S}\Phi_s(\theta)/S$. McFadden (1989), as we do, proposes to base the GMM estimator on simulated versions of conditional moments, but his method only applies to static models where it is possible to simulate the dependent variable conditional on past observed conditioning variables. This is not feasible in our setting due to the dynamic nature of the latent variable, w^{t-1}, as we in general do not know the conditional distribution of $w^{t-1}|x_t$. We overcome this issue by relying on nonparametric kernel estimation techniques.

Finally, it is worth pointing out that the kernel smoother could be replaced by other nonparametric regression techniques such as local linear kernel, nearest neighbours or series estimators (Li and Racine, 2007, Ch. 14-15). This should not change the asymptotic

properties of the SNM estimator. As part of the Monte Carlo study (below), we implement a series-based version of it.

2.2. Properties of the SNM estimator

This section deals with the asymptotic properties of the SNM estimator. The results offered here are high level, in the sense that they are based on general assumptions imposed on the DLV model in Eq. (**??**). Given a more concrete formulation of the DLV model, one could provide more primitive conditions that would imply our general assumptions. To demonstrate how the conditions can be verified, we discuss in more detail one particular DLV model.

We will throughout assume that the exact GMM estimator, $\hat{\theta}_n$, is well-behaved. In particular, we impose the following high-level assumption ensuring identification:

ASSUMPTION 2.1. *The moment function $G(\theta) = E[z_t' \varepsilon_t(\theta)]$ satisfies $G(\theta) = 0$ if and only if $\theta = \theta_0$, where $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ and Θ is compact. Moreover, the weighting matrix satisfies $W_n \to^P W_0 > 0$.*

The above conditions are standard for general GMM estimators, see e.g. Newey and McFadden (1994, Section 2) who also give more primitive conditions for them to hold for particular models. Below, we impose additional conditions on the DLV model ensuring that $\sup_{\theta\in\Theta} \|G_n(\theta) - G(\theta)\| \to^P 0$. This property together with Assumption **??** implies that the infeasible GMM estimator is consistent, $\hat{\theta}_n \to^P \theta_0$; see Newey and McFadden (1994, Theorem 2.6).

The goal is now to analyze the simulated version, $\hat{\theta}_{n,S}$, relative to the exact estimator, $\hat{\theta}_n$. As a first step towards such a result, we have to ensure that $\hat{T}_S(\phi)(x_t;\theta)$ 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 kernel estimators using parameter-dependent data. We will impose some fairly high-level assumptions on the DLV model that imply the conditions in Kristensen (2009). In order to state these assumptions, we first introduce some additional notation. Let $Z_s(\theta) = z(X_t(\theta))$ denote the simulated version of the instruments. Also, let $f(x;\theta)$ and $f_{t_0}(x,x';\theta)$ denote the stationary densities of the simulated random variables $X_0(\theta)$ and $(X_0(\theta), X_{t_0}(\theta))$, for some $t_0 \geq 1$, respectively. We then define for any random sequence $V_s(\theta)$, $s = 1,...,t_0$, and for some $\lambda \geq 2$ the following bounds,

$$B_0 = \sup_{x\in\mathbb{R}^{d_x}} \sup_{\theta\in\Theta} f(x;\theta), \quad B_{V,1} = \sup_{x} \sup_{\theta\in\Theta} \|x\|^\lambda E[\|V_0(\theta)\| \,|X_0(\theta) = x] f(x;\theta), \quad (2.9)$$

$$B_{V,2} = \sup_{x,x'\in\mathbb{R}^{d_x}} \sup_{\theta\in\Theta} E[\|V_0(\theta)\| \,\|V_{t_0}(\theta)\| \,|X_0(\theta) = x, X_{t_0}(\theta) = x'] f_{t_0}(x,x';\theta). \quad (2.10)$$

ASSUMPTION 2.2. *The processes $\{X_s(\theta)\}$, $\{\Phi_s(\theta)\}$ and $\{Z_s(\theta)\}$ satisfy:*

1. *For all $\theta \in \Theta$, $\{(\Phi_s(\theta), X_s(\theta), Z_s(\theta))\}$ is stationary with mixing coefficients $\alpha_s(\theta)$ satisfying $\alpha_s(\theta) \leq As^{-\beta}$ for some $0 < A, \beta < \infty$.*
2. *The random variables $\Phi_s(\theta)$ and $X_s(\theta)$ are differentiable w.r.t. θ almost surely with derivatives $\dot{\Phi}_s(\theta)$ and $\dot{X}_s(\theta)$.*

3 For some $\mu \geq 2$ and all $\theta \in \Theta$: $E\left[\|\Phi_s(\theta)\|^{\mu}\right] < \infty$, $E[||\Phi_s(\theta)||^{\mu}||\dot{\Phi}_s(\theta)||^{\mu}] < \infty$, and $E[\|\Phi_s(\theta)\|^{\mu}||\dot{X}_s(\theta)||^{\mu}] < \infty$.

4 For some $\lambda \geq 2$ and $t_0 \geq 1$, the bounds in Eqs. (??)-(??) are finite for $V_s(\theta) = \Phi_s(\theta)$, $V_s(\theta) = \dot{\Phi}_s(\theta)$ and $V_s(\theta) = \Phi_s(\theta)\dot{X}_s(\theta)$.

5 The functions $x \mapsto T(\phi)(x;\theta)$ and $f(x;\theta)$ are $m \geq 2$ times continuously differentiable w.r.t. x.

6 With $d = d_x + d_\theta$, μ and λ given in Assumptions ??(3) and ??(4), the mixing exponent β satisfies $\beta > \frac{1+(\mu-1)(1+d/\lambda+d)}{\mu-2}$.

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

While Assumption **??**(1) assumes that $(\Phi_s(\theta), X_s(\theta), Z_s(\theta))$ is a stationary process, it does not necessarily require $\mathbf{Y}_s(\theta) = (Y_s(\theta), W_s(\theta))$ generated from Eq. (**??**) to be stationary. For example, in case of unit root type behaviour, one could choose the testing functions, conditioning variables and instruments as functions of the differenced process; see Gorodnichenko, Mikusheva and Ng (2012) for more details. We do however implicitly assume that we are able to initialize the simulated processes such that $(\Phi_s(\theta), X_s(\theta), Z_s(\theta))$ is stationary. In practice this is not always possible but due to the assumption of α-mixing, we know that $(\Phi_s(\theta), X_s(\theta), Z_s(\theta))$ will converge towards their stationary solutions as $s \to \infty$. A complete analysis, taking into account incorrect initalization will however not be given here since it will involve longer proofs. For an analysis in the case of unconditional SMM, we refer to Duffie and Singleton (1993) while Kristensen (2009, Theorem 3) give results on kernel estimators when data is not initialized at the stationary distribution. In Section 4.1, we discuss in more detail how our the SNM estimator can be adjusted to handle non-stationary models when it is not obvious how to choose testing, conditioning and instrumental variables that are stationary and at the same time identifies θ.

The moment and smoothness conditions stated in Assumptions **??**(2) and **??**(3) implicitly impose restrictions on the functions $\mathbf{r} = (r_y, r_w)$, ϕ and x. If these three functions are smooth, then the derivative processes are quite easily derived: First, we write the model in Eq. (**??**) more compactly as

$$\mathbf{Y}_s(\theta) = \mathbf{r}\left(\mathbf{Y}^{s-1}(\theta), \mathbf{U}_s; \theta\right), \tag{2.11}$$

and then define the differentiated process $\dot{\mathbf{Y}}_s(\theta) = \left(\dot{Y}_s(\theta), \dot{W}_s(\theta)\right)$ as

$$\dot{\mathbf{Y}}_s(\theta) = \frac{\partial \mathbf{r}\left(\mathbf{Y}^{s-1}(\theta), \mathbf{U}_s; \theta\right)}{\partial \mathbf{Y}^{s-1}(\theta)'}\dot{\mathbf{Y}}^{s-1}(\theta) + \frac{\partial \mathbf{r}\left(\mathbf{Y}^{s-1}(\theta), \mathbf{U}_s; \theta\right)}{\partial \theta}. \tag{2.12}$$

The derivative processes of the conditioning variables and the testing functions are then given as $\dot{X}_s(\theta) = x'\left(Y_{s-q:s+q}(\theta)\right)\dot{Y}_{s-q:s+q}(\theta)$ and $\dot{\Phi}_s(\theta) = \phi'\left(Y_{s-p:s+p}(\theta)\right)\dot{Y}_{s-p:s+p}(\theta)$. The derivative process $\dot{\mathbf{Y}}_s(\theta)$ is well-defined if $\mathbf{r}$ is differentiable. This rules 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, e.g., Pakes and Pollard (1989) to our setting.

Assumptions **??**(4)-**??**(6) are of a more technical nature, and used to verify the conditions in Kristensen (2009); we refer to this paper for further discussion of these assumptions. Finally, the moment condition in Assumption **??**(7) is used to control the impact of the trimming introduced below. It restricts the tail thickness of the distribution of x_t.

To show how the conditions can be verified for specific models, consider, as an example, the linear model $\mathbf{Y}_s(\theta) = \mathbf{A}(\theta)\mathbf{Y}_{s-1}(\theta) + \mathbf{B}(\theta)\mathbf{U}_s$, where $\mathbf{A}(\theta)$ and $\mathbf{B}(\theta)$ are smooth functions of θ while the error term $\mathbf{U}_s$ is i.i.d. with a continuous distribution. This model includes for example the SV model considered in the simulation study by setting $\mathbf{Y}_s = \left(\log\left(y_s^2\right), w_s\right)$ and $\mathbf{U}_s = \left(\log\left(u_{s,1}^2\right), u_{s,2}^2\right)$, c.f. Eq. (**??**). The linear process $\{\mathbf{Y}_s(\theta)\}$ is stationary and geometrically mixing if the eigenvalues of $\mathbf{A}(\theta)$ all lie inside the unit circle. The first-order derivative process, $\dot{\mathbf{Y}}_s(\theta)$, solves

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

where $\dot{\mathbf{A}}(\theta)$ and $\dot{\mathbf{B}}(\theta)$ are derivatives w.r.t. θ. Since $\mathbf{Y}_s(\theta)$ is stationary and mixing and $\mathbf{A}(\theta)$ has eigenvalues inside the unit circle, $\dot{\mathbf{Y}}_s(\theta)$ will also be stationary and mixing. Furthermore, if $E\left[\|\mathbf{U}_s\|^p\right] < \infty$ for some $p \geq 1$, then $E\left[\|\mathbf{Y}_s(\theta)\|^p\right] < \infty$ and $E\left[\left\|\dot{\mathbf{Y}}_s(\theta)\right\|^p\right] < \infty$. So, for example, by choosing x_t as lagged values of y_t, and the test function as bounded by polynomials of an appropriate order, Assumption **??** will hold for this model. Note that even if some of the eigenvalues of $\mathbf{A}(\theta)$ lie outside the unit circle, we can still generate stationary test functions and conditioning variables by taking differences.

We impose the following regularity conditions on the kernel function K:

ASSUMPTION 2.3. *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* $|K(u) - K(u')| \leq \Lambda\|u - u'\|$, *or (ii)* $K(u)$ *is differentiable with* $|\partial K(u)/\partial u| \leq \Lambda$. *For some* $a > 1$, $|\partial^r K(u)/\partial u^r| \leq \Lambda\|u\|^{-a}$ *for* $\|u\| \geq L$ *and* $r = 0, 1$.
2 *For some* $m \geq 1$: $\int K(u)\,u^\alpha du = 0$ *for all* $\alpha \in \{0,1\}^{d_x}$ *with* $|\alpha| = 1, ..., m-1$, *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 the moment conditions that our SNM estimator is based on by trimming away observed values for which $f(x_t;\theta) < a$:

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

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 Eq. (**??**). The trimming function is chosen such that $\tau_{a,t}(\theta) = 0$ when $\hat{f}(x_t;\theta) < a$ where a is a trimming parameter and so allows us to control the behaviour of $\hat{f}(x_t;\theta)$ which appears in the denominator of $\hat{T}_S(\phi)(x_t;\theta)$. Trimming obviously imparts a loss of efficiency so, as $S \to \infty$, we will let $a \to 0$ such that asymptotically it has no impact. We impose the following regularity conditions on $\tau_a(v)$:

ASSUMPTION 2.4. *The trimming function $\tau_a : \mathbb{R} \mapsto [0,1]$, $a > 0$, satisfies $\tau_a(v) = 1$ for $v \geq 2a$ and $\tau_a(v) = 0$ for $v \leq a$. It is continuously differentiable with $|\tau_a'(v)| = O(a)$.*

We here use a smooth trimming function such that the trimmed GMM objective function remains smooth in θ. A simple way of constructing $\tau_a(v)$ is to choose a cdf F with

support $[0,1]$, and define $\tau_a(v) = F((v-a)/a)$ which then in great generality will satisfy Assumption **??**; see also Andrews (1995). We are now ready to state the first main result:

THEOREM 2.1. *Under Assumptions* **??**-**??** *with* $q > 0$ *given in Assumption* **??***.7:*

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

In particular, as $\sqrt{n}a^{-1}h^m \to 0$, $\sqrt{n}a^{-1}\sqrt{\log(S)/(Sh^{d_x})} \to 0$ *and* $\sqrt{n}a^q \to 0$, $\hat{\theta}_{n,S}$ *is first-order equivalent to* $\hat{\theta}_n$.

The above theorem provides an error bound between the SNM and exact estimator. Three errors appear: $O_P\left(a^{-1}h^m\right)$ is the bias component due to kernel smoothing, $O_P\left(a^{-1}\sqrt{\log(S)/(Sh^{d_x})}\right)$ is the variance component from the simulations, and $O_P\left(a^q\right)$ is caused by the trimming. The result shows that as $a, h \to 0$ and $Sh^d/\log(S) \to \infty$ sufficiently fast, the error due to simulations, kernel smoothing and trimming can be made arbitrarily small and so the SNM estimator essentially *is* the infeasible GMM estimator.

Theorem **??** is not completely satisfactory for two reasons: First, as we shall see in the following, the rate of the variance component is not sharp; in particular, the theorem seems to indicate that the SNM estimator suffers from a curse of dimensionality with the variance growing exponentially with $d_x = \dim(x_t)$. We will improve on the above result and show that in fact no curse of dimensionality is present and rather the first-order variance component is of order $O_P\left(1/\sqrt{S}\right)$. Second, one will normally only use a moderate number of simulations and so it is of interest to have a measure of the approximation error incurred by using the SNM estimator for a given choice of S.

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 Salanié (2010) 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 2.5. $\Phi_s(\theta)$ *and* $X_s(\theta)$ *satisfy:*

1. *The two processes* $\Phi_s(\theta)$ *and* $X_s(\theta)$ *are twice continuously differentiable w.r.t.* θ. *Their second order derivatives,* $\ddot{X}_s(\theta)$ *and* $\ddot{\Phi}_s(\theta)$, *satisfy the same mixing and moment conditions as imposed on* $\dot{X}_s(\theta)$ *and* $\dot{\Phi}_s(\theta)$ *in Assumption* **??**.
2. *With* $V(\theta)$ *defined in Assumption* **??***(3), its first derivative w.r.t.* θ *also satisfies Eqs. (***??***)-(***??***).*
3. *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.
4. $E\left[\sup_{\theta\in\Theta} f(x_t;\theta)^{-q} \|\partial T(\phi)(x_t;\theta)/(\partial\theta)\| \|z_t\|\right] < \infty$ *for some* $q > 0$.

ASSUMPTION 2.6. *The kernel* K *is differentiable and its derivative satisfies the same conditions as imposed on* K *in Assumption* **??**.

As with Assumption **??**, the above conditions are quite high-level, but they are normally easy to verify in specific models. In particular, Assumption **??** holds for the linear

model considered earlier under weak moment conditions on the errors. The additional assumptions are used to establish uniform consistency of $\partial \hat{T}_S(\phi)(x;\theta)/(\partial\theta)$. Moreover, together with the mixing conditions imposed earlier, they imply that the exact set of sample moments satisfy a central limit theorem, $\sqrt{n}G_n(\theta_0) \to^d N(0,\Omega_0)$ where

$$\Omega_0 = \sum_{t=-\infty}^{\infty} E\left[g_0(\theta_0)\, g_t(\theta_0)'\right], \tag{2.13}$$

and that the derivatives of the exact sample moments,

$$H_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta), \quad h_t(\theta) = -z_t' \frac{\partial T(\phi_i)(x_t;\theta)}{\partial\theta} \in \mathbb{R}^{M\times d_\theta}, \tag{2.14}$$

are well-defined and satisfy $\sup_{\|\theta-\theta_0\|<\delta} \|H_n(\theta) - H(\theta)\| \to^P 0$ where $H(\theta) = E[h_t(\theta)]$. As a final condition needed for the exact GMM estimator to be asymptotically normally distributed, we impose the following rank condition which is closely related to the identification condition in Assumption **??**:

Assumption 2.7. *With $H_0 = H(\theta_0)$, the matrix $H_0'W_0H_0$ is non-singular.*

Under Assumptions **??** and **??**, the exact GMM estimator is $\sqrt{n}$-asymptotically normally distributed, $\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d N(0,V_0)$, where
$V_0 = (H_0'W_0H_0)^{-1} H_0'W_0\Omega_0W_0H_0 (H_0'W_0H_0)^{-1}$, c.f. Newey and McFadden (1994, Theorem 3.2). We are now ready to state the second main result evaluating the higher-order impact of the simulations on the SNM estimator:

Theorem 2.2. *Assume that Assumptions **??**-**??** 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

$$B = (H_0'W_0H_0)^{-1} H_0'W_0 E\left[\frac{z_t'}{f(x_t;\theta_0)} \sum_{|\alpha|=m} \frac{\partial^{|\alpha|}\left[T(\phi)(x_t;\theta_0) f(x_t;\theta_0)\right]}{\partial x^\alpha}\right],$$

$$D_s = (H_0'W_0H_0)^{-1} H_0'W_0\left\{Z_s(\theta_0)'\Phi_s(\theta_0) - E\left[Z_s(\theta_0)'\Phi_s(\theta_0)\right]\right\}.$$

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:

$$\sqrt{n}(\hat{\theta}_{n,S} - \theta_0) \xrightarrow{d} N\left(0, (H_0'W_0H_0)^{-1} H_0'W_0\{\Omega_0 + \lambda\Sigma_0\}W_0H_0(H_0'W_0H_0)^{-1}\right), \tag{2.15}$$

where $\Sigma_0 = \sum_{s=-\infty}^{\infty} E[D_0D_s']$.

The first part of the theorem gives a bias-variance expansion of the additional estimation errors due to the simulations and kernel smoothing used in the computation of the SNM relative to the exact GMM estimator. There are two leading terms in the expansion: The first term, h^mB, 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 Salanié (2004) and Kristensen and Shin (2012), 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.

Second, in comparison to standard SMM (McFadden, 1989; Duffie and Singleton; 1993), 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: The pointwise variance of the kernel regression estimator $\hat{T}_S(\phi)(x;\theta)$ is of order $1/\left(Sh^d\right)$. 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 $1/S$. The intuition behind this result is the same as for two-step semiparametric estimators (see e.g. Kristensen, 2008, 2010): The computation of $\hat{\theta}_{n,S}$ involves summation over $\hat{T}_S(\phi)(x;\theta)$ which works as a variance reduction device.

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. (**??**), the SNM estimator has an additional variance term, $\lambda\Sigma_0$ where $\lambda \approx n/S$, relative to the exact GMM estimator. This is akin to Duffie and Singleton (1993). It is possible to construct an estimator of Σ_0, akin to a heteroscedasticity and autocorrelation consistent covariance (HAC) estimator, and then adjust standard errors to account for the use of simulations. However, because HAC estimators are often imprecise in small and moderate samples, we advocate the simpler alternative of setting S large enough in relation to n so that the Σ_0 component will be of negligible importance and can be ignored when drawing inference about parameters. That is, in practice, S should be set as large as is computationally feasible so that standard inferential methods for GMM estimators can be employed. In our simulation studies, we find that choosing S to be 5,000-10,000 suffices.

2.3. Discussion

2.3.1. Choice of test functions and instruments An integral part of the proposed estimation procedure is the choice of test functions, ϕ_t, and instruments, z_t. 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 a 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, one chooses (a small number of) test functions that are believed to identify the parameters of interest. 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) and Fourier series (Carrasco et al, 2007). Our examples mix the approaches, using some test functions motivated by the models, supplemented with Fourier-type terms.

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 equa-

tion involving conditional means and variances of the residual function and its Jacobian. 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, or (ii) restrict the instruments to belong to a smaller, tractable class of processes as done in Christensen and Sørensen (2008, Section 4).

2.3.2. Optimal weight matrix In order to conduct any inference, we need estimators of the asymptotic covariance matrices in Theorem **??**, Ω_0 and Σ_0. Also, a consistent estimator of these matrices are needed to obtain an efficient estimate either through two-step GMM or continuous-updating estimation (CUE). 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 HAC estimators (e.g., that of Newey and West, 1987). It is well-known that inferences based upon HAC estimators can be quite unreliable. 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 $\Omega(\theta)$ and $\Sigma(\theta)$ through Monte Carlo which allows for the implementation of a CUE as advocated in Hansen, Heaton and Yaron (1996). We here only discuss the simulation of $\Omega(\theta)$ since $\Sigma(\theta)$ can be treated along the same lines. We first note that $\Omega(\theta) = \lim_{n\to\infty} nE\left[G_n(\theta) G_n(\theta)'\right]$. Given that the model is simulable, so is $G_n(\theta)$. Thus, we may generate $R \geq 1$ such samples of size $m \geq 1$, and for each of them calculate simulated moment conditions as in Eq. (**??**). 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 $\hat{G}_{m,S}^{(r)}(\theta)$ in exactly the same way as $\hat{G}_{n,S}(\theta)$ is computed in Eq. (**??**), except that the simulated data of size m replaces the real sample data. We then define $v_r(\theta) = \hat{G}_{m,S}^{(r)}(\theta) - \bar{G}(\theta)$, where $\bar{G}(\theta) = R^{-1}\sum_{r=1}^R \hat{G}_{m,S}^{(r)}(\theta)$, and obtain the following estimator, $\hat{\Omega}(\theta) = \frac{m}{R}\sum_{r=1}^R v_r(\theta) v_r(\theta)'$.

2.3.3. Choice of the kernel and the bandwidth To implement the SNM estimator, the kernel function $K(\cdot)$ and the bandwidth h must be chosen. There is substantial theoretical and empirical evidence that the choice of the particular kernel function has relatively little effect on the results. 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.

Given the kernel function, the bandwidth must be chosen. Too large a bandwidth over-smooths 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 **??**, where we have a bias term of order h^m and a second-order variance term of order $1/(nSh^{d_x})$. Many methods for choosing the bandwith have been suggested in the literature but these are designed to minimize the MSE of the kernel estimator and as such do not necessarily minimize the MSE of $\hat{\theta}_{n,S}$.

Instead, we treat the bandwidth as an additional parameter to be estimated together with θ 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 h as an additional parameter 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 our implementation are clear in the example code that is provided (see Creel and Kristensen, 2009a). 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. It is outside the scope of this paper to derive the asymptotic properties of the SNM for this bandwidth selection rule, but we conjecture that the arguments of Härdle, Hall and Ichimura (1993) can be carried over to the SNM estimator to show that the selected bandwidth satisfies the conditions of Theorems **??**-**??**; this is supported by our Monte Carlo study which shows that selection rule works very well in practice.

2.3.4. Computational issues Implementation of the SNM estimator can potentially be computationally burdensome if not done carefully since it requires evaluation of long strings of simulations which in turn has to be fed into a kernel regression estimator. However, various techniques for efficient computation of simulated values and corresponding kernel regressions are available; see, e.g., Creel (2005, 2007), Racine (2002) and Yang, Duraiswami, Gumerov and Davis (2003). Moreover, the kernel regression estimator can be replaced by alternative, computationally more efficient, nonparametric methods such as series regression.

Another problem is multiple local minima. As is the case with normal GMM estimators, the SNM objective function is in general not globally convex, so one needs to take care to find the global minimum. Our experience is that gradient-based minimization algorithms are not able to find the global minimum of the SNM objective function. 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 (e.g., 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.

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. 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 SNM software is provided in Creel and Kristensen (2009a).

3. MONTE CARLO RESULTS

This section presents Monte Carlo results that compare the SNM estimator to other estimators. 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. To keep computation time at a reasonable level, we do not attempt to use optimal instruments, and instead just use the conditioning variables, 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. In all cases, we minimize the objective function using simulated annealing, starting from a random point in the parameter space. Except where otherwise noted, we use a simulation length of $S = 10,000$, and in all cases, 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

$$y_t = \rho y_{t-1} + u_t \tag{3.1}$$

where $u_t \sim N(0,1)$. This allows us to compare the exact MLE (OLS) to SNM and so gauge the impact of simulations and kernel smoothing from using SNM. We generate samples using $\rho \sim U(0,1)$. The single test function is $\phi_t = y_t$, and the single conditioning variable is $x_t = y_{t-1}$. The single instrument is the conditioning variable. The GMM estimator based on this set-up is simply the ordinary least squares (OLS) estimator, $\hat{\rho}_{OLS}$. Pretending we cannot compute the conditional moment, we then implement the corresponding SNM estimator yielding $\hat{\rho}_{SNM}$.

We first examine the stochastic difference between the OLS and SNM estimator. To this end, we report the ratio of root mean squared error of the SNM estimator relative to the OLS estimator in Table **??**. According to Theorem **??**, the variance ratio (denoted VR) should satisfy $VR \approx 1 + \Sigma_0/\Omega_0 \times n/S$. Based on the numbers reported in Table **??**, we obtain the following linear fit: $\hat{VR} = 1.16 + 1.55 \times n/S$ with $R^2 = 55\%$. This shows that indeed the asymptotic approximation of Theorem **??** does a reasonably good job in describing the variance of the SNM estimator and that the impact of simulations vanishes very quickly.

Next, we examine the distribution of the SNM estimator. For the OLS estimator, we know that $\sqrt{n/(1-\rho^2)}(\hat{\rho}_{OLS} - \rho) \to^d N(0,1)$. For the SNM estimator, if n/S is sufficiently small so that the Σ_0-term in Theorem **??** may be ignored, we also have $\sqrt{n/(1-\rho^2)}(\hat{\rho}_{SNM} - \rho) \to^d N(0,1)$. By testing the normality of this quantity using the Kolmogorov-Smirnov test, we can explore the reliability of inference when the Σ_0 term is ignored. We do this for samples of size $n \in \{50, 100, 400, 800\}$ and simulation lengths $S \in \{1000, 10000, 50000\}$. Table **??** gives p-values for Kolmogorov-Smirnov tests, first of the hypothesis that $\sqrt{n/(1-\rho^2)}(\hat{\rho}_{SNM} - \rho) \sim N(0,1)$, and secondly of the hypothesis that $\hat{\rho}_{SNM}$ and $\hat{\rho}_{OLS}$ have the same distribution. The hypotheses are never rejected, and the p-values are quite close to 1 when sample sizes and number of simulations are not too small.

In conclusion, for this model, the finite sample distribution of the SNM estimator is not far away from that of the exact GMM (OLS) estimator even for moderate number of simulations. This indicates that, for reasonably large number of simulations, we most likely do not have to take into account simulations when drawing inference based on SNM.

3.2. Stochastic Volatility

This section presents Monte Carlo results for the logarithmic stochastic volatility model of Jacquier, Polson and Rossi (1994) which can be written as:

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

where the white noise $u_t = (u_{t,1}, u_{t,2})'$ is distributed i.i.d. $N(0, I_2)$.

The Monte Carlo design proposed by Sandmann and Koopman (1998) has been adopted in subsequent work by a number of authors, and we adhere to this trend to facilitate comparison with other estimators. Perhaps the most widely used design uses $\theta_0 = (\alpha, \beta, \sigma)' = (-0.736,\ 0.9,\ 0.363)'$, and $n = 500$ observations. We begin with this case.

We here choose the test variables as $\phi_t = \left(|y_t|,\ y_t^2,\ \cos|y_t|,\ \sin|y_t|,\ ...,\ \cos 4\,|y_t|,\ \sin 4\,|y_t|\right)$. 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 $x_t = y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}$. This conditioning variable is intended to capture the recent variability of the series in a parsimonious way, to avoid needing to choose multiple bandwidths.

The SNM estimator as presented uses kernel regression, and this is somewhat computationally demanding. An alternative approach is to use a series regression. The advantage of this idea is that the coefficients can be estimated very quickly by least squares[1]. We implemented this idea using the Fourier flexible form (Gallant, 1981), setting $A = 1$ and $J = 1, 2$ (see Gallant, 1981, page 216), using the same test variables and conditioning variable as above.

Table **??** presents the results. In the first three rows we investigate how the size of S affects the SNM estimator by reporting the RMSE for $S = 1000$, 5000 and 10000. As expected, bias and RMSE decrease as S increases. To further investigate how large the impact of simulations is, we obtain the following linear fits by regressing finite-sample variances of the three parameter estimates on S^{-1}: $\hat{V}_{\hat{\alpha}} = 0.04 + 15.91 \times S^{-1}$, $\hat{V}_{\hat{\beta}} = 0.01 + 0.59 \times S^{-1}$ and $\hat{V}_{\hat{\sigma}} = 0.02 + 4.76 \times S^{-1}$ and with high R^2's (60-70%) in all three cases. This shows that even for moderate sample sizes ($n = 500$), we can expect that the efficiency loss due to simulations is practically zero relative to the finite sample variation for $S = 10000$. For example, for the case of the SNM estimator of α, the fitted contribution of simulations to the over-all variance is $\{15.91/10000\} \,/\, \{0.04 + 15.91/10000\} = 3.74\%$ and so can be safely ignored.

Focusing in the following on the results with $S = 10000$, we see that the SNM estimator of α 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. The series-based version of SNM does not perform as well as the version that uses kernel regression. It is possible that a more careful choice of basis functions could improve the results of the series-based version, and if this were so it would be very useful, as the speedup compared to kernel regression is substantial. Overall, it seems fair to say that the kernel regression version of SNM performs well in general, and better than the only other general purpose estimator, EMM.

Fermanian and Salanié (2004) and Altissimo and Mele (2009) provide results for their estimators with $(\sigma_b, \beta, \sigma) = (0.025,\ 0.95,\ 0.260)$, where $\sigma_b = \exp(\alpha/2)$, and $n = 500$. We apply the SNM estimator design, using the same test variables, conditioning variables and instruments as before. Table **??** 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

[1] We thank an anonymous reviewer for this suggestion

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 non-parametric simulated maximum likelihood (NPSML) estimator. This model, with notation adapted to follow the general DLV model of Eq. (**??**), may be written as:

$$\text{AR Tobit:} \begin{cases} y_t = & \max(0, w_t) \\ w_t = & \alpha + \beta w_{t-1} + \sigma u_t \\ u_t \sim & IIN(0,1) \end{cases} \tag{3.3}$$

This model has one observable variable, y_t, a single latent variable, w_t and a scalar white noise u_t. We use Fermanian and Salanié's Monte Carlo design where $(\alpha, \beta, \sigma) = (0.0, 0.5, 1.0)$ and $n = 150$. The SNM estimator is implemented the same way as for the SV model with the exception that we no longer take absolute values of y_t. Table **??** 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 two-dimensional parameter β has its first element set to 1 for identification. The model, referred to as FA, is

$$\text{FA:} \begin{cases} y_t & = \beta w_t + u_{1t} \\ w_t & = \sqrt{h_t} u_{2t} \\ h_t & = \alpha_1 + \alpha_2 \left(w_{t-1}\right)^2 \end{cases} \tag{3.4}$$

$t = 1, 2, ..., n$, where $u_{1t} \sim N(0, \sigma^2 I_2)$ and $u_{2t} \sim N(0,1)$. The parameter vector design is $(\alpha_1, \alpha_2, \sigma, \beta_2) = (0.2, 0.7, 0.5, -0.5)$, and sample size is $n = 500$. The test variables used are $|y_t|$, $(\cos k\,|y_t|, \sin k\,|y_t|)$, $k = 1, ..., 4$, and $y_{t1}y_{t2}$, while the conditioning variable is $y_{t-1,1} + y_{t-1,2} + y_{t-2,1} + y_{t-2,1}$. 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.

Table **??** 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. Latent network model

The previous examples all use fairly simple models that have application to financial time series. In this section we consider a model that incorporates an unobserved network. 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. For example, 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. 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 this is known to the modeler. The objective is to estimate $p_{CON,}$without directly observing the network.

Suppose that nodes can be in two states, A (e.g., "healthy") and B ("ill"). The probability that a node switches states depends on the number of connections the node has, and the states of the connected nodes. Let $n_{it,}(X)$ be the number of nodes connected to node i that are in state X at time t, where $X \in \{A, B\}$. The probability that node i switches from A to B at time t is

$$p_{it}(A \to B) = 1 - (1 - p_B)^{n_{it}(B)} + p_{SPON}.$$

Here, p_B is the probability that any connected node in state B "infects" the reference node, which is in state A. The probability than none of the connected nodes transmits is $(1 - p_B)^{n_{it}(B)}$. The probability that at least one transmits is the complement of this. Additionally, p_{SPON} is the probability of a spontaneous "illness". Analogously, the probability of a node in state B switching to state A is

$$p_{it}(B \to A) = 1 - (1 - p_A)^{n_{it}(A)},$$

which is symmetric, except for the lack of a spontaneous switch.

We assume that $p_A = 0.01$, $p_B = 0.02$, $p_{SPON} = 0.01$. For this example, these are taken as known, and emphasis is placed on learning about the degree distribution of the network, which depends on p_{CON}. The observed variables are the total number of nodes in state B (total number ill), and the total number of nodes that switch from state B to state A (total number of recoveries), at each point in time. The states of the individual nodes are not observed. Thus, there are two observed variables, and 500 latent variables (the states of the nodes).

We simulate series of length $n = 50$, using univariate and bivariate models. The univariate models use the total number of nodes in state B (illnesses) as y_t in Eq. (**??**). The bivariate models use this variable, plus the number of nodes that switch from state B to state A (recoveries) as y_t. The test functions are y_t and corresponding sine and cosine transformations, as in previous examples. The conditioning variables are either one

or two lags of y_t, which is either univariate or bivariate. Thus, up to four conditioning variables are used, while the previous examples used a single conditioning variable. This feature of this example allows exploration of possible deterioration of performance of the kernel regression estimator as the dimension of the conditioning vector increases. The results are in Table **??**. In this Table we refer to marginal bias and marginal root mean squared error, because the true value of p_{CON} is drawn from a distribution rather than fixed at a given value. We see that is all cases bias and RMSE are small. The bivariate model gives better results, and using two lags instead of one also improves performance[2].

4. EXTENSIONS

A number of extensions of the proposed estimator are available. We here discuss how our method can be adjusted to allow for non-stationarity and reduce biases, and can be used for forecasting.

4.1. Non-Stationary Models

We have worked under the maintained assumption that the testing functions, conditioning variables and instruments all are stationary. If the data-generating process in question is non-stationary, $\Phi_s(\theta)$, $X_s(\theta)$, and $Z_s(\theta)$ have to be chosen with great care to ensure that these remain stationary. 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). One could in this situation instead use differenced moments, $E_\theta[\Delta y_t|\Delta y_{t-1}]$, or other transformations leading to stationary test and conditioning variables; see, e.g., Gorodnichenko, Mikusheva and Ng (2012).

Alternatively, one can use the following alternative simulation scheme that obtain consistent simulated estimates of any given set of conditional moments under non-stationarity: First, simulate S independent trajectories each of length n, $Y_{s,t}(\theta)$, $s = 1, ..., S$ and $t = 1, ..., n$, where the sth trajectory is computed as:

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

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 $X_{s,t}(\theta) = x(Y_{s,t-q:t+q}(\theta))$, $\Phi_{s,t}(\theta) = \phi(Y_{s,t-p:t+p}(\theta))$, and

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

By construction, $(\Phi_{s,t}(\theta), X_{s,t}(\theta))$, $s = 1, ..., S$, 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 $Sh^{d_x} \to 0$, $\hat{T}_{t,S}(\phi)(x;\theta) \to^P T_t(\phi)(x;\theta)$. The estimation now proceeds as in the stationary case.

[2]It is possible to plot the estimated versus true parameter values, and one sees that the result are points close to the 45 degrees line, thoughout the range of true values. This figure is omitted to save space.

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

Instead of simulating unconditionally from the model, one could design an iterative procedure as employed in particle filtering (see e.g. Brownlees et al., 2011). These procedures are however more complicated and are therefore more cumbersome to implement.

The asymptotic properties of the resulting estimator are not covered by Theorems **??**-**??**, since in general non-stationary sequences do not satisfy standard versions of the Law of Large Numbers and Central Limit Theorem. Instead, one would need to combine arguments as developed in, amongst others, Kristensen and Rahbek (2010) and Kristensen and Shin (2012) where asymptotics of estimators in nonlinear, non-stationary models are developed.

4.2. Unbiased Simulator

The current kernel estimator $\hat{T}_S(\phi)(x_t;\theta)$ in Eq. (**??**) 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 $\hat{\varepsilon}_t(\theta) = \tilde{R}(\phi)(x_t) - \hat{R}(\phi)(x_t;\theta)$, where $\tilde{R}(\phi)(x)$ and $\hat{R}(\phi)(x;\theta)$ are kernel estimators using actual and simulated data respectively,

$$\tilde{R}(\phi)(x) = \frac{1}{n}\sum_{t=1}^{n} \phi(y_t) K_h(x_t - x), \quad \hat{R}(\phi)(x;\theta) = \frac{1}{S}\sum_{s=1}^{S} \phi(Y_s(\theta)) K_h(X_s(\theta) - x).$$

$\hat{R}(\phi)(x;\theta)$ is an unbiased estimator of $R(\phi)(x;\theta)$ and we obtain consistency for fixed h and S by the same arguments as used to prove Theorem 1 under the following identification condition: For fixed $h > 0$, $R_h(\phi)(x_t;\theta_0) = R_h(\phi)(x_t;\theta)$ a.s. if and only if $\theta = \theta_0$. 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 $h \to 0$.

4.3. Filtering and Forecasting

In many applications, it is of interest to compute predictions of some function of $\mathbf{Y}_t = (y_t, w_t)$, say $f(\mathbf{Y}_t)$, at some value of θ (in most situations $\theta = \hat{\theta}$, where $\hat{\theta}$ is an estimator). One leading example would be filtering of the latent variable, w_t. A natural filter/prediction would be

$$m(x,\theta) := E_\theta[f(\mathbf{Y}_t) | y_{t-1} = x_1, ..., y_{t-q} = x_q],$$

for some $q \geq 1$. One can use the SNM estimation techniques for this purpose. Let $(Y_s(\theta), W_s(\theta))$, $s = 1, ..., S$, be a time series of S simulations generated by Eq. (**??**) at the parameter value of interest, θ. Given the simulated values, we can compute the kernel regression estimator of $m(x,\theta)$,

$$\hat{m}_S(x,\theta) = \frac{\sum_{s=1}^{S} f(\mathbf{Y}_s(\theta)) K_h(X_s(\theta) - x)}{\sum_{s=1}^{S} K_h(X_s(\theta) - x)}, \quad X_s(\theta) = Y_{s-q:s-1}(\theta),$$

where $K_h(z) = K(z/h)/h$, $K : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ is a kernel function, $d_x = \dim(X_s)$, and $h > 0$ is a bandwidth. A further analysis of the performance of this filtering method in theory and practice is left for future research.

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. We emphasize that the simulation length used in the Monte Carlo work is shorter than would be desirable in empirical applications. When doing empirical work, S should be set as large as is practical, because a larger S leads to a more efficient estimator (see Table **??**) and also because inference that ignores the Σ_0 term will be reliable only when n/S is small.

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 alternative nonparametric regression methods such as nearest neighbors or series regression, and use of high performance algorithms for kernel smoothing (see, e.g., Yang et al., 2003). The analysis of the proposed automated bandwidth selection method would also be of interest. Another interesting possibility is to attempt to use (approximately) optimal instruments and weighting matrices.

APPENDIX

Appendix A. Proofs

Proof of Theorem ??: We first note that $\sup_{\theta\in\Theta} \|G_n(\theta) - G(\theta)\| \to^P 0$ under the assumptions imposed. This follows by standard uniform Law of Large Numbers (LLN) results for stationary and mixing sequences. This implies that the exact GMM estimator is consistent. Theorem **??** will now follow by Kristensen and Shin (2012, Theorem A.5) if we can show that:

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

To prove this claim, write

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

where $A_1 = 2\sup_{\theta\in\Theta}\|G_n(\theta)\|$, and $A_2 = \sup_{\theta\in\Theta}\left\|\hat{G}_{n,S}(\theta) - G_n(\theta)\right\|$. By Assumption **??**, $\|W_n\| = O_P(1)$ while

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

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

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

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

where the last equality follows from Assumption **??**(7) and the LLN. This shows Eq. (**??**). □

Proof of Theorem ??: Following standard arguments for GMM estimators,

$$\sqrt{n}(\hat{\theta}_{n,S} - \theta_0) = \left[\hat{H}_{n,S}(\hat{\theta}_{n,S})'W_n\hat{H}_{n,S}\left(\bar{\theta}_{n,S}\right)\right]^{-1}\hat{H}_{n,S}(\hat{\theta}_{n,S})'W_n\sqrt{n}\hat{G}_{n,S}(\theta_0), \quad \text{(A.2)}$$

for some $\bar{\theta}_{n,S}$ between $\hat{\theta}_{n,S}$ and θ_0, where

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

By the same arguments as in the proof of Theorem **??** combined with Lemmas **??**-**??**, 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 the uniform LLN for stationary and mixing sequences, 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 Eq. (**??**) with the following expan-

sion of the exact GMM estimator,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \left[H_n(\hat{\theta}_n)' W_n H_n(\bar{\theta}_n)\right]^{-1} H_n(\hat{\theta}_n)' W_n \sqrt{n} G_n(\theta_0), \tag{A.3}$$

we obtain

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

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 $\bar{G}_n(\theta_0) = \sum_{t=1}^n \tau_{a,t}(\theta_0) Z_t' T(\phi)(x_i;\theta_0)/n$ is the trimmed version of the true moment conditions, and

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

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

We further decompose the adjustment term into

$$\bigtriangledown \bar{G}_n(\theta_0)\left[\hat{T}_S - T\right] = \bigtriangledown \bar{G}_n(\theta_0)\left[\hat{T}_S - E[\hat{T}_S]\right] + \bigtriangledown \bar{G}_n(\theta_0)\left[E[\hat{T}_S] - T\right], \tag{A.4}$$

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

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

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 LLN,

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

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

$$\bigtriangledown \bar{G}_n(\theta_0)\left[\hat{T}_S - E[\hat{T}_S]\right] = \frac{1}{n}\sum_{t=1}^n \sum_{s=1}^S \Delta(x_t, z_t, V_s) + R_n, \tag{A.5}$$

where $V_s = (\Phi_s(\theta_0), X_s(\theta_0))$, $\Delta_h(x_t, z_t, V_s) := z_t' \psi_h(x_t, V_s)/f(x_t;\theta_0)$, and

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

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. (**??**) as a two-sample U-statistic. Define $D_h(V_s) = E\left[\Delta_h(x_t, z_t, V_s)|V_s\right]$, and $M_{t,S} := \frac{1}{S}\sum_{s=1}^S \left\{\Delta_h(x_t, z_t, V_s) - D_h(V_s)\right\}$. Conditionally on the simulations which we collect in $\mathcal{U}_S = \{U_1, ..., U_S\}$, it is easily seen

that $\{M_{t,S}\}$ satisfies the conditions of Kristensen and Salanié (2010, Lemma 4) such that

$$E\left[\left\|\frac{1}{n}\sum\nolimits_{t=1}^{n} M_{t,S}\right\|^{2} |\mathcal{U}_S\right] \leq C\left(1,A\right) E\left[\left\|M_{t,S}\right\|^{2+\delta} |\mathcal{U}_S\right] n^{-1},$$

where in turn

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

and, by standard arguments,

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

We now have that

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

where, leaving out higher order terms,

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

This proves the first part of the theorem. The second part now follows from Eq. (**??**). □

Appendix B. Lemmas

We state two lemmas without proofs; see Creel and Kristensen (2009b) for these.

LEMMA B.1. *Under Assumptions* **??**-**??**, *the simulated conditional moment estimator satisfies:*

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

LEMMA B.2. *Under Assumptions* **??**-**??** *and* **??**-**??**, *the simulated conditional moment estimator satisfies:*

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

ACKNOWLEDGEMENTS

We thank particants at the 14th International Conference on Computing in Economics and Finance (2008), the NBER-NSF Time Series Conference (2009) and the International

Symposium on Econometric Theory and Applications (2009) for comments and suggestions. Creel gratefully acknowledges the financial support of grants MICINN-ECO2009-11857 and SGR-2009-578. Kristensen gratefully acknowledges financial support of the National Science Foundation (SES-0961596), the Danish Research Foundation through a grant to CREATES, and the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001.

REFERENCES

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

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

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.

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.

Bibby, B. M. and M. Sørensen (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli 1*, 17-39.

Billio, M. and A. Monfort (2003). Kernel-based indirect inference. *Journal of Financial Econometrics 1*, 297-326.

Brownlees, C.T., D. Kristensen and Y. Shin (2011). Smooth filtering and likelihood inference in dynamic latent variables models. Working paper, University College London.

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.

Christensen, B.J. and M. Sørensen (2008). Optimal inference in dynamic models with conditional moment restrictions. CREATES Research Paper No. 2008-51.

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

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

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

Creel, M. and D. Kristensen (2009a). SNM Guide. UFAE and IAE Working Papers 793-09. `http://econpapers.repec.org/paper/aubautbar/793.09.htm`.

Creel, M. and D. Kristensen (2009b). Estimation of dynamic latent variable models using simulated nonparametric moments. UFAE and IAE Working Papers 792-09, `http://econpapers.repec.org/paper/aubautbar/792.09.htm`.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica 61*, 929–952.

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

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

Fridman, M. and L. Harris (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business & Economic Statistics 16*, 284-291.

Gallant, A. R. (1981). On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form. *Journal of Econometrics 15,* 211-245.

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

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

Gorodnichenko, Y., A. Mikusheva and S. Ng. (2012). Estimators for Persistent and Possibly Non-Stationary Data with Classical Properties. *Econometric Theory,* forthcoming.

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

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

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

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

Jackson, M. 2006). The Economics of Social Networks, in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society* (eds. R. Blundell, W. Newey & T. Persson), Ch. 1. Cambridge: Cambridge University Press.

Jacquier, E., M. Johannes and N.G. Polson (2007). MCMC maximum likelihood for latent state models. *Journal of Econometrics 37*, 615-640.

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

Johansen, A. M., A. Doucet and M. Davy (2008). Particle methods for maximum likelihood parameter estimation in latent variable models. *Statistics and Computing 18*, 47-57.

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

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

Kristensen, D. (2010). Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. *Journal of Econometrics 156*, 239-259.

Kristensen, D. and A. Rahbek (2010). Likelihood-based inference for cointegration with nonlinear error-correction. *Journal of Econometrics 158*, 78-94.

Kristensen, D. and B. Salanié (2010). Higher order improvements for approximate estimators. CAM Working Papers 2010-04, University of Copenhagen.

Kristensen, D. and Y. Shin (2012). Estimation of dynamic models with nonparametric simulated maximum likelihood. *Journal of Econometrics 167,* 76–94.

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

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

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.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing, in *Handbook of Econometrics*, Vol IV, ed. R. Engle and D. McFadden, Amsterdam: North-Holland, 2111-2245.

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

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

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

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

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

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

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

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

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.

Table 1. Monte Carlo Results: AR1 model, relative RMSE: $RMSE(\hat{\rho}_{SNM})/RMSE(\hat{\rho}_{OLS})$

	n			
S	50	100	400	800
1000	1.174	1.259	1.597	1.416
10000	1.072	1.034	1.070	1.053
50000	1.065	1.046	1.019	1.018

Table 2. Monte Carlo Results: AR1 model, Kolmogorov-Smirnov p-values

	n			
S	50	100	400	800
1000	0.28	0.54	0.59	0.41
	0.46	0.77	1.00	1.00
10000	0.27	0.49	0.99	0.88
	0.56	1.00	1.00	1.00
50000	0.27	0.30	0.96	0.88
	0.37	1.00	1.00	1.00

Note: For a given S, the top p value is for test of normality of the SNM estimator, the bottom is for the test of equality of SNM and OLS distributions.

Table 3. Monte Carlo Results: Stochastic volatility, SV1 model. $n = 500$. Mean and root mean squared error (in parentheses).

Estimator	$\alpha = -0.736$	$\beta = 0.9$	$\sigma = 0.363$
SNM, $S = 1000$	-0.761 (0.238)	0.865 (0.101)	0.411 (0.150)
SNM, $S = 5000$	-0.744 (0.216)	0.868 (0.120)	0.385 (0.146)
SNM, $S = 10000$	-0.736 (0.201)	0.882 (0.080)	0.380 (0.128)
SNM series ($J = 1$)	-1.02 (0.37)	0.96 (0.14)	0.09 (0.28)
SNM series ($J = 2$)	-1.01 (0.36)	0.95 (0.15)	0.09 (0.28)
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)

Note: 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).

Table 4. Monte Carlo Results: Stochastic volatility SV2 model. $n = 500$. Mean and root mean squared error (in parentheses).

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)

Note: Sources for other estimators: NPSML - Fermanian and Salanié (2004); CD-SNE and J-SNE - Altissimo and Mele (2009).

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

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)

Note: Source for NPSML: Fermanian and Salanié (2004).

Table 6. Monte Carlo Results: Factor ARCH model. Mean and root mean squared error (in parentheses).

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)

Note: 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.

Table 7. Monte Carlo Results: Latent network model. Marginal bias and marginal root mean squared error.

	Marginal bias	Marginal RMSE
Univariate, 1 lag	-0.0008	0.0530
Univariate, 2 lags	0.0027	0.0514
Bivariate, 1 lag	0.0035	0.0187
Bivariate, 2 lags	0.0021	0.0153