Estimation of Dynamic Latent Variable Models Using Simulated Nonparametric Moments

Michael Creel*

Universitat Autònoma de Barcelona

Version 0.2

PRELIMINARY AND INCOMPLETE

DO NOT CITE WITHOUT PERMISSION

COMMENTS WELCOME

February 2008

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. Since 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. Monte Carlo results show that the estimator performs well in comparison to other estimators that have been proposed for estimation of general DLV models.

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

JEL codes: C13; C14; C15

*Department of Economics and Economic History, Edifici B, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain. michael.creel@uab.es This research was supported by grants SEJ 2006-00712/ECON and CONSOLIDER-INGENIO 2010 (CSD2006-00016).

1 Introduction

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

To fix ideas, consider Billio and Monfort's (2003) definition of the general DLV model:

$$\begin{cases} y_t = & r_t\left(y^{t-1}, y^{*t}, \varepsilon_t; \theta\right) \\ y_t^* = & r_t^*\left(y^{t-1}, y^{*t-1}, \varepsilon_t^*; \theta\right) \end{cases} \tag{1}$$

where $t = 1, ..., n$, y_t is a vector of observable variables, y^{*t} is a vector of latent variables, y^{t-1} is notation for $\left(y_1', ..., y_{t-1}'\right)'$,$\{\varepsilon_t\}$ and $\{\varepsilon_t^*\}$ are two independent white noises with known distributions, and θ is a vector of unknown parameters[1]. Calculation of the likelihood function requires finding the density of y^n, and as Billio and Monfort make clear, this involves calculating an integral of the same order as n, which is in general untractable. Without the density of the observable variables, analytic moments cannot be computed. Thus, maximum likelihood and moment-based estimation methods are not available except in special cases.

A number of econometric methods that deal with the complications that often accompany dynamic latent variable models have been developed over the last two decades. 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) and nonparametric simulated maximum likelihood (Fermanian and Salanié, 2004). These methods have been applied to DLV models in a number of contexts.

[1] The macroeconomic model of the previous paragraph could be formalized by letting y_t^* indicate the vector of all of the agents' actions, and letting y_t be the observed aggregate outcomes. The possible presence of exogenous variables is suppressed for clarity.

Billio and Monfort (2003) provide numerous references for applications.

As noted by Fermanian and Salanié (2004, pg. 702), there often exists a tradeoff between the asymptotic efficiency of a method and its wide applicability. 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, but encounters curse-of-dimensionality problems when used with dynamic models. Solutions based upon lower dimensional marginals of the likelihood function lead to a loss of asymptotic efficiency.

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 Sorensen, 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 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, $\widetilde{y^n}(\theta)$. However, the elements are drawn from their marginal distributions. It is not in general possible to draw from $y_t|y^{t-1};\theta$. To do so, one would need draws from $y^{*t}|y^{t-1};\theta$. If such draws were available, they could be plugged into the first line of the DLV model given in equation 1, which, combined with a draw from ε_t, would give a draw from $y_t|y^{t-1};\theta$. The problem is that the observed value of y^{t-1} is only compatible with certain realizations of the history of the latent variables, y^{*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 methods to simulate from $y^{*t}|y^{t-1};\theta$, 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 $y^{*t}|y^{t-1};\theta$ (Billio and Monfort, 2003, pg. 298).

Indirect inference is generably applicable, but its efficiency depends crucially upon the choice of the auxiliary model. The efficient method of moments 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 generably 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. The choice of the leading parametric model and the degree of the semi-nonparametric part of the model have an important effect on the results that are obtained. Andersen, Chung and Sorensen (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 an 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) 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 parameteterized auxiliary model. The binding functions are conditional moments evaluated at certain points. The parameters of the auxiliary model are the values of the conditional moments, and the parameters are estimated using nonparametric kernel regression methods. The same kernel methods are applied to the simulated data, and minimum distance estimation is used to estimate the parameter of the structural 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.

This paper offers a new implementation of the simulated method of moments that allows for use of conditioning information, for general DLV models. Conditional moments are evaluated using nonparametric kernel smoothing. The proposed estimator is referred to as the simulated nonparametric moments (SNM) estimator. This estimator is quite simple to use, and it is found to perform well in comparison to other estimators that have been proposed for estimation of general DLV models. The next section defines the estimator and discusses specification and hypothesis testing. The third section presents several examples that compare the SNM estimator to other methods, using Monte Carlo. Section 4 applies the estimator to.... and Section 5 concludes.

2 The SNM estimator

2.1 Definition of the estimator

The moment-based estimation framework used in this paper is mostly standard, and is as follows. Let y_t be the current period data on Y_t, a k_Y-dimensional vector of endogenous variables. Let X_t be a k_X-dimensional vector of lagged endogenous and exogenous variables. Define $\phi(x_t;\theta) \equiv E\left[Y_t|X_t = x_t;\theta\right]$. Error functions are of the form

$$\varepsilon(y_t, x_t;\theta) = y_t - \phi(x_t;\theta), \tag{2}$$

Occasionally, an M-estimation approach that downweights extreme errors will be used (Huber, 1964; Gallant, 1987). In this case, error functions are

$$\varepsilon(y_t, x_t;\theta) = \tanh\left(\frac{y_t - \phi(x_t;\theta)}{2}\right) \tag{3}$$

Moment conditions are defined by interacting a vector of instrumental variables $z(x_t)$ with error functions:

$$m(y_t, x_t;\theta) = z(x_t) \otimes \varepsilon(y_t, x_t;\theta) \tag{4}$$

Average moment conditions are

$$m_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(y_t, x_t;\theta) \tag{5}$$

The objective function is

$$s_n(\theta) = m_n'(\theta) W(\hat{\tau}_n) m_n'(\theta) \tag{6}$$

where $W(\hat{\tau}_n)$ is a weighting matrix that may depend upon prior estimates of nuisance parameters.

Often, $\phi(x_t;\theta)$ in equations 2 and 3 has a known functional form, in which case estimation may proceed using the standard generalized method of moments (GMM). When no closed-form functional form is available it may be possible to define an unbiased simulator $\widetilde{\phi}(x_t, u;\theta)$ such that $E\left[\widetilde{\phi}(x_t, u;\theta)\right] = \phi(x_t;\theta)$, where the distribution of u conditional on $X = x_t$ is known. If this is so, a simulated error function can be defined by replacing $\phi(x_t;\theta)$ in equations 2 and 3 with an average of S draws of $\widetilde{\phi}(x_t, u_t^s;\theta)$. Doing so, and then proceeding with normal GMM estimation methods defines the SMM estimator (Gouriéroux and Monfort, 1996, pg. 27). As noted above, it is often not possible to simulate subject to $X_t = x_t$ in the case of general DLV models. If this is the case, the SMM estimator cannot be based upon conditional moments as defined in equation 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, since unconditional moments provide little information on the dynamics of a DLV model.

The basic idea of the simulated nonparametric moments (SNM) estimator proposed in this paper is to replace the expectations $\phi(x_t;\theta)$ that are used to define error functions in equations 2 and 3 with kernel regression fits based on a very long 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; Haerdle, 1991; Li and Racine, 2007). The use here is entirely standard, except for the fact that the data used is simulated rather than real.

In the following, tildes will be used to indicate simulated data or elements that depend upon simulated data. Let $\{(\widetilde{y}_s(\theta), \widetilde{x}_s(\theta))\}_{s=1}^S$ be a simulation of length S from the model at the trial parameter value θ. Kernel regression may be used to fit $\phi(x_t;\theta)$, using simulated data drawn from the model at the parameter value θ:

$$\widetilde{\phi}_S(x_t;\theta) = \sum_{s=1}^{S} \widetilde{w}_s \widetilde{y}_s(\theta) \tag{7}$$

where the weight $\widetilde{w}_s$ is

$$\widetilde{w}_s = \frac{K\left(\frac{x_t - \widetilde{x}_s(\theta)}{h_S}\right)}{\sum_{s=1}^{S} K\left(\frac{x_t - \widetilde{x}_s(\theta)}{h_S}\right)} \tag{8}$$

Note that the same weight $\widetilde{w}_s$ applies to each element of $\widetilde{y}_s$ (a k_Y-vector). To speed up computations, one should not separately fit each of the k_Y endogenous variables, but rather employ a specialized kernel fitting algorithm that saves the weights across variables. Since x_t is of dimension k_x, which is in usually greater than one, the kernel function $K()$ is in general multivariate. The bandwidth parameter is h_S. Note that the kernel regression fit can be evaluated at x_t without requiring that the simulated sequence contain any realizations such that $\widetilde{x}_s = x_t$. What is required for a good fit at x_t is that there there be a large number of realizations that are "close enough" to x_t.

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

To be explicit, the SNM estimator is based on error functions of the form

$$\widetilde{\varepsilon}(y_t, x_t; \theta) = y_t - \widetilde{\phi}_S(x_t; \theta), \tag{9}$$

or

$$\widetilde{\varepsilon}(y_t, x_t; \theta) = \tanh\left(\frac{y_t - \widetilde{\phi}_S(x_t; \theta)}{2}\right) \tag{10}$$

The moment function contribution of an observation is

$$\widetilde{m}(y_t, x_t; \theta) = z(x_t) \otimes \widetilde{\varepsilon}(y_t, x_t; \theta) \tag{11}$$

Average moment conditions are

$$\widetilde{m}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \widetilde{m}(y_t, x_t; \theta) \tag{12}$$

The SNM estimator $\widetilde{\theta}_n$ is defined as the minimizer of the objective function

$$\widetilde{s}_n(\theta) = \widetilde{m}_n'(\theta) W(\hat{\tau}_n) \widetilde{m}_n'(\theta) \tag{13}$$

where $W(\hat{\tau}_n)$ is a weighting matrix that may depend upon prior estimates of nuisance parameters.

2.2 Properties and use

Numerous results exist for the aymptotics of kernel regression estimators. The general flavor of the results is that the kernel regression estimator is consistent as long as the bandwidth tends to zero, but not too quickly. For example, supposing that the model generates a strictly stationary α-mixing sequence, Lu and Cheng (1997) show that $\widetilde{\phi}_S(x_t, \theta) \overset{a.s.}{\to} \phi(x_t, \theta)$, for almost all x_t, as $S \to \infty$. It is important to note in equations 7 and 8 that the kernel simulator $\widetilde{\phi}_S(x_t; \theta)$ does not depend on the sample size of the real data, n, rather, it depends on the sample size of the simulated data, S. This value can be made as large as is desired, and the simulator converges almost surely to the true moment as the simulation length tends to infinity. By making S suitably large, it is possible to make $\widetilde{\phi}_S(x_t; \theta)$ as close as is desired to the true moment $\phi(x_t; \theta)$. In practice, S may be chosen large enough so that the error functions in equations 2 and 9 (or the M-estimation analogues in equations 3 and 10) are essentially identical. If this is the case, the SNM estimator essentially *is* the GMM estimator.

A simple Monte Carlo exercise illustrates this point. Samples of size $n = 30$ were

Table 1: Comparing the SNM and GMM estimators, simple linear model

Parameter	Estimator(s)	Mean	St. Dev.	Min.	Max.
β_1	SNM-GMM	-0.001	0.004	-0.015	0.014
β_1	GMM	0.489	0.473	-0.867	2.143
β_2	SNM-GMM	0.002	0.007	-0.020	0.025
β_2	GMM	0.524	0.723	-1.510	2.964

generated using the classical linear model

$$\begin{cases} y & = \beta_1 + \beta_2 x + \varepsilon \\ x & \sim U(0,1) \\ \varepsilon & \sim N(0,1) \end{cases} \tag{14}$$

The parameters β_1 and β_2 were randomly drawn (separately) from $U(0,1)$ distributions at each of 1000 Monte Carlo replications. The maximum likelihood (ML) estimator is the ordinary least squares (OLS) estimator obtained by regressing y on a constant and x. The ML estimator may be thought of as a GMM estimator that uses the single ($k_y = 1$) error function $e_t = y_t - \beta_1 - \beta_2 x_t$ and the instruments $(1, x_t)$. The SNM estimator was applied, using the endogenous variable y_t, the conditioning variable x_t and instruments $(1, x_t)$. The simulation length was $S = 10^5$, and the bandwidth was chosen using the rule-of-thumb[2] $h_S = S^{-1/(4+k_x)}$, where, as above, k_x is the number of conditioning variables, 1 in this case. A standard Gaussian kernel was used.

Table 1 gives results that compare the distribution of the difference between the SNM and GMM estimators to the distribution of the GMM estimator, over the 1000 Monte Carlo replications. We can see that the difference between the two estimators is distributed tightly around zero, and that the dispersion of the difference is much less than that of the GMM estimator. If the value of the SNM estimator is regressed on a constant, the value of the GMM estimator, and the value of the true parameter, the results are, for the constant, β_1,

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

$$T = 1000 \quad \bar{R}^2 = 0.9999 \quad F(2,997) = 8.5632\text{e+}6 \quad \hat{\sigma} = 0.0036169$$

(standard errors in parentheses)

For the slope, β_2, we obtain

[2] See Li and Racine, 2007, pg. 66.

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

$$T = 1000 \quad \bar{R}^2 = 0.9999 \quad F(2,997) = 6.9636\text{e+}6 \quad \hat{\sigma} = 0.0061481$$

(standard errors in parentheses)

We see that the SNM and GMM estimators are essentially identical, independent of the true parameter value.

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

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

Specification and hypothesis testing Given that the SNM estimator can be made arbitrarily close to the GMM estimator by using a long enough simulation, one can directly apply methods based upon standard asymptotic results for GMM estimators with the same degree of confidence that would be warranted if the GMM estimator were available. An overidentified model's specification may be tested using the familiar χ^2 test based upon $n s_n(\widetilde{\theta}_n)$, (assuming that an optimal weight matrix is used), *etc.* Use of standard asymptotic results is justified as long as the simulation is long enough to render randomness due to simulation negligable in comparison to randomness due to sampling of the real data[3].

It is well-known that inferences based upon asymptotic results for GMM estimators can be quite unreliable (Hansen, Heaton and Yaron, 1996; Windmeijer, 2005), due the difficulty of precisely estimating the covariance matrix of the moment conditions. It may be preferable to conduct Monte Carlo testing or parametric bootstrapping (Davison and Hinkley, 1997, pp. 140-149; Dufour *et al.*, 1998; Dufour and Khalaf, 2002). This is

[3] Gallant and Tauchen (1996) in their presentation of the EMM estimator also assume that the simulation is long enough to warrant ignoring randomness due to simulation.

always feasible any time the SNM estimator can be used, since both SNM and these testing procedures require full simulability of the model, given the model's parameter. Section 4 of this paper gives an example that shows how asymptotic and simulation-based testing may be done.

Choice of the kernel and the bandwidth To implement the SNM estimator, the kernel function $K()$ in equation 8 must be chosen, as must the window width (bandwidth), h_S. 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, as long as the bandwidth parameter is chosen correspondingly to the kernel (Li and Racine, 2007). For this reason, this paper uses Gaussian product kernels exclusively, accompanied by prior rotation of the data to approximate independence of the conditioning variables. Gaussian product kernels lead to error functions that are continuous and relatively smooth in the parameters, which facilitates iterative minimization. Kernels such as the radial symmetric Epanechnikov are relatively inexpensive to compute, but can lead to error functions that are discontinuous in the parameters, which complicates minimization of the objective function that defines the SNM estimator. This paper leaves the possibility of SNM estimation based on local linear or local polynomial kernel methods for future work.

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

Computational issues Estimation of a complicated model using long simulation may become computationally burdensome, especially if the sample size is large and a single CPU or CPU core is used to do all the work. Doing testing by Monte Carlo rather that relying on asymptotic approximations adds another source of computational burden. One may seek to use data-based methods to choose the bandwidth, as well. 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 (Creel, 2007). The widespread availability of multicore processors is an invitation to take advantage of parallelization opportunities in econometric work. All of the results reported in this paper were obtained on computational clusters similar to that described in Creel (2007).

3 Examples and comparison to other estimators

This section presents Monte Carlo results that compare the SNM estimator to other estimators that have been proposed for estimation of DLV models. The intention is to show that the SNM estimator can be used to successfully estimate a variety of DLV models, that the SNM estimator performs well in comparison to alternative estimators, and to give examples of how the moment conditions that define the SNM estimator may be chosen. To limit the computational burden of conducting a number of Monte Carlo exercises, in all cases a simulation length of $S = 5000$ was used to calculate the SNM estimator. It should be noted that use of such a short simulation length penalizes the performance of the SNM estimator.

3.1 Autoregressive Tobit

Consider the autoregressive Tobit model presented by Fermanian and Salanié (2004), with notation adapted to follow the general DLV model of equation 1 of this paper:

$$\begin{cases} y_t = & \max(0, y_t^*) \\ y_t^* = & \alpha + \beta y_{t-1}^* + \sigma \varepsilon_t \\ \varepsilon_t \sim & IIN(0,1) \end{cases} \tag{15}$$

This model has one observable variable, y_t, a single latent variable, and y_t^* and a white noise ε_t. The only available data is the series y^t (and transformations elements of this series). This model as was used by Fermanian and Salanié (2004) to illustrate their non-parametric simulated maximum likelihood (NPSML) estimator. To apply the SNM estimator, four error functions were used. The four endogenous variables used to define error functions were y_t (to provide information on α), $y_t y_{t-1}$ and $y_t y_{t-2}$ (to provide information on β) and $y_t^2 - (\bar{y})^2$ (to provide information on σ). Each of the four error functions used conditioning on y_{t-1} and y_{t-2}. The instruments were the same two conditioning variables, plus a vector of ones. A Monte Carlo study using 1000 replications was done. The sample size was $n = 150$, the same as Fermanian and Salanié used to generate the results they report in their Table 1 (pg. 715). The simulation length $S = 5000$ was used.

Table 2: AR Tobit Results, SNM

Parameter	Mean	Bias	St. Dev.	RMSE	RMSE NPSML
α	0.022	0.022	0.082	0.085	0.215
β	0.599	-0.099	0.111	0.148	0.151
σ	1.119	-0.119	0.166	0.204	0.264

Note that the use of two conditioning variables implies that a multivariate kernel function is needed. A Gaussian product kernel was used, after multiplying the simulated data and the evaluation points by the inverse of the Choleski decomposition of the sample covariance matrix of the real data. This pre-whitening transformation makes the conditioning variables more like independent standard normal random variables, which makes the use of a Gaussian product kernel more appropriate. The simple rule-of-thumb bandwidth $h = S^{-1/(4+k_x)} = 0.2418$ was used.

Estimation was done by minimizing the objective function in equation 13, using the standard error functions as in equation 9, and an identity matrix as the weight. Table 2 reports the results. The last column of this Table reproduces the RMSE of the NPSML estimator, as reported by Fermanian and Salanié (FS, pg. 715) for the same model, using the longer of the two simulation lengths they try. For all three parameters of the model, the SNM estimator has a lower root mean squared error than that of the NPSML estimator.

It is to be noted that the NPSML estimator is not asymptotically efficient in dynamic applications, since it is based upon kernel density fits to a small set of marginals of the likelihood function, rather than to the overall joint density. In general, curse of dimensionality problems make fitting the overall joint density infeasible. For this reason, it is possible for SNM to be more efficient than NPSML. The relative efficiencies will depend upon the selection of marginals used for NPSML estimation and the choice of moments used for estimation by SNM, among other factors. For dynamic problems, use of the NPSML estimator requires decisions to be made regarding how many and which marginals to approximate, how to trim the likelihood function, and other considerations. It may be the case that the bias and variance of the NPSML estimator in the Monte Carlo results for the autoregressive Tobit model reported by Fermanian and Salanié could be reduced by better tuning in this regard, or possibly by using a longer simulation, but the necessity of performing such tuning is a drawback. Use of the SNM estimator requires selection of the endogenous and conditioning variables to use to define moment conditions, but careful thought about the model can give guidance in this regard. The other steps such as selection of the kernel and bandwidth either have little impact on performance, or can be automatized to a large extent.

3.2 Stochastic volatility

The efficient method of moments is well-known means of estimating partially observed nonlinear models. Given a well-chosen auxiliary model, EMM is known on theoretical grounds to be an efficient estimator, and the method has been subjected to diverse Monte Carlo studies. EMM could always be used in any situation where SNM could. The principle advantage of SNM with respect to EMM is simplicity of use, since there is no need for the sometimes problematic step of selecting and estimating the auxiliary model. However, the question arises whether or not the SNM estimator can provide efficiency near that of EMM.

Since SNM is essentially a certain type of GMM estimator, and since previous work suggests that EMM outperforms GMM (Chumacero, 1997; Andersen, Chung and Sorensen, 1999; Gallant and Tauchen, 1999), one may wonder how SNM could possibly rival the efficiency of EMM. The answer is that SNM opens up the possibility of using complex conditional moments that may provide more information than simple unconditional moments. Knowledge of the model, up to the parameter, which is required for any simulation-based estimation method, may point naturally to certain conditional moments which are likely to be informative about the parameter of the model, but which may not have known analytic form. If this is the case, the SNM estimator provides a natural way to take advantage of this information.

Andersen, Chung and Sorensen (1999; henceforth ACS) provide Monte Carlo results comparing EMM with GMM in the context of a simple stochastic volatility model. The model, using the notation of the general DLV model given above, is

$$\begin{cases} y_t & = y_t^* \varepsilon_{t,1} \\ \log\left((y_t^*)^2\right) & = \alpha + \beta \log\left((y_{t-1}^*)^2\right) + \sigma \varepsilon_{t,2} \end{cases} \tag{16}$$

where the white noise $\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2})'$ is distributed i.i.d. $N(0, I_2)$. The latent variable y_t^* is the standard deviation of the observed variable y_t. GMM is applied using a number of simple unconditional moments (see Andersen and Sorensen, 1996, for details), and they implement EMM using a number of auxiliary models. In their conclusions, ACS note that EMM "...provides a very substantial improvement in efficiency relative to simple GMM...." Here, we report Monte Carlo results for SNM estimation of this model, using the design vector $(\alpha, \beta, \sigma) = (-0.736, 0.9, 0.363)$ that ACS focus on, which is intended to be representative of data at a weekly frequency. The sample size is $n = 1000$ observations. The endogenous variables used to define the error functions are $100y_t^2$ and $100y_t^2 y_{t-1}^2$. The first of these seems a natural choice to provide information on α and σ. The second is intended to capture the temporal correlation of the variance, which

Table 3: Stochastic Volatility Results

Parameter	Mean	Bias	St. Dev.	RMSE	Lowest RMSE EMM
α	-0.756	-0.020	0.275	0.276	0.33
β	0.898	0.002	0.036	0.036	0.04
σ	0.400	-0.037	0.122	0.127	0.10

should give information on β. The scaling was done to improve numerical precision in estimation[4]. The conditioning variables are y_{t-1} and y_{t-1}^2. The instruments are the same conditioning variables, plus a vector of ones. Two endogenous variables and three instruments imply a total of six moment conditions. A simulation length of $S = 5000$ was used. A Gaussian product kernel was used, after prewhitening, as explained above. The simple rule-of-thumb bandwidth $h = S^{-1/(4+k_x)} = 0.2418$ was used. Estimation was done by minimizing the objective function in equation 13, using the standard error functions as in equation 9, and an identity matrix as the weight.

Of the 1000 replications, 8 failed to converge to the specified tolerances for the function, gradient and change in parameters within the limiting number of iterations, though none of these runs crashed. Inclusion or exclusion of these 8 replications does not change the results in any important way. The results presented in the Table 3 use the 992 replications that iterated to convergence. These results can be compared to those given in ACS's Table 2 (page 72), which gives results for GMM and EMM estimators, using the same sample size. For purposes of comparison, the last column of Table 3 gives the lowest RMSE from ACS's Table 2. For the α and β parameters, the SNM estimator obtains a RMSE lower than the best of the estimators considered by ACS. In the case of σ, the infeasible GMM estimator and several of the EMM estimators do a little better than the SNM estimator.

Repeating this exercise with a sample size of $n = 2000$, and adjusting the bandwidth parameter accordingly, we obtain results that can be compared with those of ACS's Table 5 (ACS, page 78). In this case, of 1000 Monte Carlo replications, two failed to converge, and the results are dropped. For the remaining 998 replications, the results are summarized in Table 4. ACS's results for the best EMM estimator of those they tried[5] are reproduced in the last column. In the case of the α and β parameters, the SNM and EMM estimators obtain results that are very similar. In the case of σ, the EMM estimator performs better.

In sum, the SNM estimator appears to be more or less as efficient as the EMM estima-

[4]The parameter σ was also scaled to make its gradient of the same order of magnitude as those of the other two parameters. This simple but easily overlooked step greatly facilitates obtaining convergence of the minimization routine.

[5]The auxiliary model contains a GARCH(1,1) leading term and a 4th order Hermite polynomial component.

Table 4: Stochastic Volatility Results

Parameter	Mean	Bias	St. Dev.	RMSE	RMSE EMM (ACS)
α	-0.750	-0.014	0.244	0.244	0.224
β	0.899	0.001	0.028	0.028	0.030
σ	0.407	-0.044	0.129	0.136	0.049

tor for this model and these sample sizes. It is worth noting that the EMM results allowed the auxiliary model to change as the sample size increases, while the SNM results rely on the same set of moment conditions for the two sample sizes. It may be possible to improve the SNM results by including additional moment conditions as the sample size increases or by using an efficient weighting matrix, but these steps were not explored.

3.3 Factor ARCH

Billio and Monfort (2003) illustrate their proposed 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, y_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 is

$$\begin{cases} y_t & = \beta y_t^* + \varepsilon_t \\ h_t & = \alpha_1 + \alpha_2 \left(y_{t-1}^*\right)^2 \\ y_t^* & = \sqrt{h_t}\varepsilon_t^* \end{cases} \tag{17}$$

$t = 1,2,...,n$, where $\varepsilon_t^* \sim N(0,1)$ and $\varepsilon_t \sim N(0,\sigma^2 I_2)$. The parameter vector to estimate is $\theta_0 = (\alpha_1, \alpha_2, \sigma, \beta_2)'$.

This model presented some difficulties when attempting to use the SNM estimator in a typical method of moments framework with error functions as in equation 2. Outlying errors caused intermediate iterations during minimization to wander, and convergence was difficult to obtain. For this reason, an M-estimation strategy was used, based on error functions as in equation 3. The error functions were defined using three endogenous variables: the squares of the two components of y_t, plus the cross product, $y_{t,1}y_{t,2}$. Use of the cross product was found to be helpful for obtaining precise estimates of β_2. These variables were each conditioned on the squares of the two components of y_{t-1} and on the lag of the cross product, $y_{t-1,1}y_{t-1,2}$. The instruments were the same conditioning variables, plus a vector of ones. With four instruments and three endogenous variables, a total of 12 moment conditions were used in estimation. An identity matrix was used to weight the moment conditions. One thousand Monte Carlo replications were done,

Table 5: Factor ARCH Results

Parameter	Mean	Bias	St. Dev.	RMSE	Lowest RMSE alternatives
α_1	0.204	0.004	0.099	0.099	0.132
α_2	0.551	-0.149	0.168	0.224	0.309
σ	0.498	-0.002	0.064	0.064	0.141
β_2	-0.426	0.074	0.291	0.300	0.269

of which 991 satisfied stong converge criteria. The reported results are based upon the 991 fully converged replicates, though the results differ only very slightly and not at all importantly if the remaining replicates are included.

Table 5 reports the 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 three of the four parameters, the SNM estimator has lower RMSE than that of the best of the estimators considered by Billio and Monfort. In the case of β_2, an indirect inference estimator has a slightly lower RMSE than the SNM estimator, but this same indirect inference estimator has RMSEs that are much larger than those of SNM for the other three parameters (see Billio and Monfort's Table 5).

This section has illustrated how the SNM estimator may be applied to several representative DLV models. Moment conditions can be chosen with an eye to the information that they provide about specific parameters. The SNM estimator has been applied in a fairly naive way, without attempting to use an efficient weighting matrix. Nevertheless, the Monte Carlo results show that the SNM estimator performs well in comparison to alternative estimation methods.

4 Application:

5 Conclusion

References

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

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

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

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

[5] Chumacero, Romulo (1997) Finite sample properties of the efficient method of moments, *Studies in Nonlinear Dynamics and Econometrics*, **2**, 35-51.

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

[7] Davison, A.C. and D.V. Hinkley (1997) *Bootstrap Methods and their Application*, Cambridge University Press.

[8] Dufour, J.-M., A. Farhat, L. Gardiol, L. Khalaf (1998) Simulation-based finite sample normality tests in linear regressions, *The Econometrics Journal*, **1**, pp. 154-173.

[9] Dufour, J.-M. and L. Khalaf (2002), Simulation based finite and large sample tests in multivariate regressions, *Journal of Econometrics*, **111**, pp. 303-322.

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

[11] Fiorentini, Gabriele, Enrique Sentana, and Neil Shephard (2004) Likelihood-based estimation of latent generalized ARCH structures, *Econometrica*, **72**, 1481-1517.

[12] Lu, Zudi and Ping Cheng (1997) Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series, *Journal of Statistical Planning and Inference,* **65**, 67-86

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

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

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

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

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

[18] Gallant, A. Ronald, and George 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/`.

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

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

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

[22] Haerdle, W. (1991) *Applied Nonparametric Regression*, Cambridge University Press.

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

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

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

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

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

[28] Lu, Z. and P. Cheng (1997) Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series, *Journal of Statistical Planning and Inference,* **65**, 67-86.

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

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

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

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

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

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

[35] Tauchen, George (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.

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