Proofs for Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

Jonathan R. Long
Department of Economics
North Carolina State University
Raleigh NC 27695-8109 USA

March 1995
Revised August 1996

Summary

This document contains proofs and the algorithm display that were omitted from

> Gallant, A. Ronald, and Jonathan R. Long (1996), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi Square," *Biometrika* 84, 125−−141

In addition, definitions and statements of theorems are reproduced. The following is a summary of the complete paper.

We propose a minimum chi-square estimator for the parameters of an ergodic system of stochastic differential equations with partially observed state. We prove that the efficiency of the estimator approaches that of maximum likelihood as the number of moment functions entering the chi-square criterion increases and as the number of past observations entering each moment function increases. The minimized criterion is asymptotically chi-squared and can be used to test system adequacy. When a fitted system is rejected, inspecting studentized moments suggests how the fitted system might be modified to improve the fit. The method and diagnostic tests are applied to daily observations on the U.S. dollar to Deutschmark exchange rate from 1977 to 1992.

1. The estimation problem

We wish to estimate the parameter ρ in the system of stochastic differential equations

$$dU_t = A(t, U_t, \rho)dt + B(t, U_t, \rho)dW_t \qquad (0 \le t < \infty). \tag{1}$$

The parameter ρ has dimension p_ρ, the state vector U_t has dimension d, W_t is a k-dimensional vector of independent Wiener processes, $A(\cdot, \cdot, \rho)$ maps $[0, \infty) \times \Re^d$ into $\Re^d$, and $B(\cdot, \cdot, \rho)$ is a $d \times k$ matrix comprised of the column vectors $B_1(\cdot, \cdot, \rho), \cdots, B_k(\cdot, \cdot, \rho)$, each of which maps $[0, \infty) \times \Re^d$ into $\Re^d$. The state U_t is interpreted as the solution of the integral equations

$$U_t = U_0 + \int_0^t A(s, U_s, \rho)\, ds + \sum_{i=1}^{k} \int_0^t B_i(s, U_s, \rho)\, dW_{is}$$

where $\int_0^t B_i(s, U_s, \rho)\, dW_{is}$ denotes the Itô stochastic integral (Karatzas & Shreve, 1991).

The system is observed at equally spaced time intervals $t = 0, 1, \ldots$ and selected characteristics

$$y_t = T(U_{t+L}) \qquad (t = -L, -L+1, \ldots) \tag{2}$$

of the state are recorded, where y_t is an M-dimensional vector and $L \ge 0$ is the number of lagged variables. Throughout data are denoted by $\{\tilde{y}_t\}$, simulations by $\{\hat{y}_t\}$, and the random variables to which they correspond by $\{y_t\}$.

We assume that U_t, and hence y_t, is stationary and ergodic for $\rho \in \mathcal{R} \subset \Re^{p_\rho}$. We further assume that the stationary distribution of y_t is absolutely continuous. Thus, for each setting of parameter ρ and lag length L, there exists a time-invariant density $p(y_{-L}, \ldots, y_0|\rho)$ such that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N} g(\hat{y}_{t-L}, \ldots, \hat{y}_t) = \int \cdots \int g(y_{-L}, \ldots, y_0)\, p(y_{-L}, \ldots, y_0|\rho)\, dy_{-L} \cdots dy_0 \tag{3}$$

where $\{\hat{y}_t : t = -L, \ldots, N\}$ is realization of length $N + L + 1$ from the system. This assumes that g is integrable and that either U_0 is a sample from the stationary distribution of U_t or that a longer realization was observed and enough initial observations were discarded for transients to have dissipated.

2. The auxiliary model

2·1. *A dense class of smooth densities*

Denote a partial derivative of a function $f(\zeta)$ on $\Re^\ell$ by

$$D^\lambda f(\zeta) = \Big(\frac{\partial^{\lambda_1}}{\partial \zeta_1^{\lambda_1}}\Big) \cdots \Big(\frac{\partial^{\lambda_\ell}}{\partial \zeta_\ell^{\lambda_\ell}}\Big) f(\zeta),$$

where $\lambda = (\lambda_1, \ldots, \lambda_\ell)$. Letting $|\lambda| = |\lambda_1| + \cdots + |\lambda_\ell|$, the Sobolev norm of f with respect to a weight function μ is

$$\|f\|_{m,p,\mu} = \Big\{ \sum_{|\lambda| \le m} \int |D^\lambda f(\zeta)|^p \mu(\zeta)\, d\zeta \Big\}^{1/p} \qquad (1 \le p < \infty)$$

$$\|f\|_{m,\infty,\mu} = \max_{|\lambda| \le m} \sup_{\zeta \in \Re^\ell} |D^\lambda f(\zeta)| \mu(\zeta) \qquad (p = \infty).$$

Gallant & Nychka (1987) proposed a nonparametric density estimator for densities in a class $\mathcal{H}$ of smooth densities on $\Re^\ell$ $(\ell \ge 1)$:

Given an integer $m_0 > \ell/2$, a bound $\mathcal{B}_0$, some small $\epsilon_0 > 0$, some $\delta_0 > \ell/2$, $\mathcal{H}$ consists of those density functions h that have the form $h(\zeta) = e^2(\zeta) + \epsilon_0 \phi(\zeta)$, with $\|e\|_{m_0,2,\mu_0} < \mathcal{B}_0$, where $\mu_0(\zeta) = (1 + \zeta'\zeta)^{\delta_0}$ and $\phi(\zeta) = (2\pi)^{-\ell/2} e^{-\zeta'\zeta/2}$. We shall take $\mathcal{H}$ as the parameter space for the density estimation problem.

Writing $\zeta^\lambda = \zeta_1^{\lambda_1} \cdots \zeta_\ell^{\lambda_\ell}$, $h \in \mathcal{H}$ has the representation $h(\zeta) = (\sum_{|\lambda|<\infty} a_\lambda \zeta^\lambda)^2 \phi(\zeta) + \epsilon_0 \phi(\zeta)$. Convergence of $(\sum_{|\lambda|<\infty} a_\lambda \zeta^\lambda)^2 \phi(\zeta)$ is with respect to the norm $\| \cdot \|_{\lfloor m_o - \ell/2 \rfloor, \infty, \mu}$ where $\mu(\zeta) = (1 + \zeta'\zeta)^\delta$ for some δ that satisfies $\ell/2 < \delta < \delta_0$ and $\lfloor \cdot \rfloor$ denotes the integer part.

Let $P_K(\zeta) = \sum_{|\lambda|<K} a_\lambda \zeta^\lambda$ denote the truncation of $\sum_{|\lambda|<\infty} a_\lambda \zeta^\lambda$ above to a polynomial of degree K on $\Re^\ell$. Put $\bar{a}_\lambda = a_\lambda / [\int \{P_K(\zeta)\}^2 \phi(\zeta)\, d\zeta + \epsilon_0]^{1/2}$ and let $\theta = (\theta_1, \ldots, \theta_{p_K})$ represent the normalized coefficients $\{\bar{a}_\lambda : 1 \le |\lambda| \le K\}$. A truncated expansion of h is $h_K(\zeta|\theta) =: \{g_K(\zeta|\theta)\}^2 \phi(\zeta) + \epsilon_0 \phi(\zeta)$ where $g_K(\zeta|\theta) = \sum_{|\lambda|<K} \bar{a}_\lambda \zeta^\lambda$. These truncations are dense in $\mathcal{H}$.

2.2. Regularity conditions

Define $x = x_{-1} = (y_{-L}, \ldots, y_{-1})$, $y = y_0$, $p(x, y|\rho) = p(y_{-L}, \ldots,\ y_0|\rho)$, $p(x|\rho) = \int p(y_{-L}, \ldots, y_0|\rho)\, dy_0$, $p(y|x, \rho) = p(x, y|\rho)/p(x|\rho)$, and $\ell = M(L + 1)$. Note that if ρ^o denotes the true value of the parameters in the system (1), then ρ^o is also the true value of ρ in the density $p(x, y|\rho)$ $(\rho \in \mathcal{R} \subset \Re^{p_\rho})$ induced by the system according to (2) and (3).

ASSUMPTION 1 *The process $\{V_t : t = -L, \ldots, \infty\}$ obtained by sampling the system (1) at times $t = 0, 1, \ldots$ and putting $V_{t-L} = U_t$ is strong mixing of size $-4r/(r - 4)$ for some $r > 4$. The conditional density $p(y|x, \rho)$ of $p(x, y|\rho)$ in (3) is defined on the closure $\bar{\mathcal{R}}^o$ of an open ball $\mathcal{R}^o$ that contains ρ^o. The true value ρ^o is an isolated minimum of $\bar{s}(\rho) := \int\int \log p(y|x, \rho) p(x, y|\rho^o)\, dy dx$ and the matrix $\int\int \{(\partial/\partial\rho) \log p(y|x, \rho^o)\}$*

$\{(\partial/\partial\rho)\log p(y|x,\rho^o)\}'p(x,y|\rho^o)\,dydx$ *is nonsingular. The derivatives* $(\partial/\partial\rho_i)\log p(y|x,\rho)$ *and* $(\partial^2/\partial\rho_i\partial\rho_j)\log p(y|x,\rho)$ *are continuous in* ρ *over* $\bar{\mathcal{R}}^o$ *and are dominated by a function d(x,y) that has finite r-th moment with respect to* $p(x,y|\rho^o)$.

Assumption 1 is the specialization to the present situation of Assumptions 1 to 6 of Gallant (1987, Chapter 7), which are standard regularity conditions for estimation of the parameters of a dynamic model by quasi maximum likelihood; see also Potscher & Prucha (1991a, 1991b). Because the process $\{V_t\}$ is strong mixing, so is $\{(x_t, y_t)\}$. Ergodicity of $\{(x_t, y_t)\}$ is then a conclusion of the strong law of large numbers (Gallant, 1987, Chapter 7, Theorem 1).

ASSUMPTION 2 *For some integer* $m_0 > \ell/2$, *some bound* $\mathcal{B}_0$, *some small* $\epsilon_0 > 0$, *and some* $\delta_0 > \max(2, \ell/2)$, $p(x,y|\rho^o)$ *has the form* $p(x,y|\rho^o) = e^2(x,y) + \epsilon_0\phi(x,y)$ *with* $\|e\|_{m_0,2,\mu_0} < \mathcal{B}_0$, *where* $\mu_0(x,y) = (1 + x'x + y'y)^{\delta_0}$ *and* $\phi(x,y) = (2\pi)^{-\ell/2}e^{-x'x/2+y'y/2}$. *Similarly for* $p(x|\rho^o)$.

Assumption 2 is sufficient to imply the Gallant-Nychka result stated in Subsection 2·1 for $\zeta = (x,y)$ and $h^o(\zeta) = p(x,y|\rho^o)$. Thus, $p(x,y|\rho^o)$ has the representation

$$\begin{aligned} p(x,y|\rho^o) &= \{g(x,y|\rho)\}^2\phi(x,y) + \epsilon_0\phi(x,y) \\ g(\zeta|\rho) &= \sum_{|\lambda|<\infty} a_\lambda \zeta^\lambda. \end{aligned} \tag{4}$$

We shall use the truncated expansion

$$\begin{aligned} f_K(x,y|\theta) &= \{g_K(x,y|\theta)\}^2\phi(x,y) + \epsilon_0\phi(x,y) \\ g_K(\zeta|\theta) &= \sum_{|\lambda|<K} \bar{a}_\lambda \zeta^\lambda \end{aligned} \tag{5}$$

described in Subsection 2·1 as the auxiliary model f_K for the theoretical results in Section 4. For applications we shall make some algebraic modifications as described in Subsection 2·3 below.

Let $\bar\theta = (\bar\theta_1, \ldots, \bar\theta_{p_K})$ denote the coefficients given by the Gallant-Nychka construction. If $p(x,y|\rho^o)$ is not itself a truncated expansion, then these differ from the coefficients $\theta^o = (\theta^o_1, \ldots, \theta^o_{p_K})$ that solve the moment equations $m_K(\rho^o, \theta) = 0$, where

$$m_K(\rho^o,\theta) = \int\int \frac{\partial}{\partial\theta}\log\{f_K(y|x,\theta)\}\,p(x,y|\rho^o)\,dydx,$$

$$f_K(y|x,\theta) = f_K(x,y|\theta)/f_K(x|\theta), \quad f_K(x|\theta) = \int f_K(x,y|\theta)dy.$$

Nonetheless, the coefficients θ^o will serve as well as the coefficients $\bar{\theta}$ as stated in Theorem 1 below. Note that here, and throughout, $f_K(x|\theta)$ is not the truncated expansion of $p(x|\rho^o)$ but rather the integral of the truncated expansion of $p(x,y|\rho^o)$.

THEOREM 1 *Under Assumptions 1 and 2*

$$\lim_{K\to\infty} \|f_K(y,x|\theta^o) - p(y,x|\rho^o)\|_{\lfloor m_o-\ell/2\rfloor,\infty,\mu} = 0.$$

With respect to the representation (4), the partial derivative of $\log p(x,y|\rho)$ is

$$\frac{\partial}{\partial\rho}\log p(x,y|\rho) = 2\Big\{\frac{g(x,y|\rho)}{g^2(x,y|\rho)+\epsilon}\Big\}\frac{\partial}{\partial\rho}g(x,y|\rho).$$

ASSUMPTION 3 *Both $p(x,y|\rho^o)$ and $p(x|\rho^o)$ possess moment generating functions, and*

$$\int\int\Big\{a'\frac{\partial}{\partial\rho}g(x,y|\rho^o)\Big\}^2\Big\{p(x,y|\rho^o)+\phi(y)p(x|\rho^o)\Big\}dy\,dx < \infty$$

for every $a \in \Re^{p_\rho}$.

Assumptions 1 and 2 are technical conditions that imply standard properties of quasi maximum likelihood estimators and some intuitively plausible properties of estimators based on Hermite expansions. For stochastic differential equations where $p(x,y|\rho)$ is known in closed form, they can often be verified at sight. For instance, the Ornstein-Uhlenbeck process $dU_t = (\rho_1+\rho_2U_t)dt + \rho_3 dW_t$ generates a Gaussian density for observations; the process $dU_t = (\rho_1+\rho_2U_t)dt + \rho_3(U_t)^{1/2}dW_t$ generates a gamma marginal and non-central chi-square conditional. Restrictions on the parameters ρ_i of these two processes that imply ergodicity are given in Aït-Sahalia (1996). Assumption 3 is central because it allows approximation of the score $(\partial/\partial\rho)\log p(x,y|\rho^o)$ by a polynomial (Gallant, 1980).

When the entire state vector U_t is not observed $\{y_t\}$ may not be Markov. In this case, the asymptotic variance of the quasi maximum likelihood estimator of ρ based upon L lags is $(\mathcal{V}^o_{L,0})^{-1}(\mathcal{V}^o_L)(\mathcal{V}^o_{L,0})^{-1}$, where

$$\mathcal{V}^o_L = \mathcal{V}^o_{L,0} + \sum_{\tau=1}^{\infty}\mathcal{V}^o_{L,\tau} + (\sum_{\tau=1}^{\infty}\mathcal{V}^o_{L,\tau})', \tag{6}$$

$$\mathcal{V}^o_{L,\tau} = E\Big\{\frac{\partial}{\partial\rho}\log p(y_\tau|x_{\tau-1},\rho^o)\Big\}\Big\{\frac{\partial}{\partial\rho}\log p(y_0|x_{-1},\rho^o)\Big\}', \tag{7}$$

$$E(W) = \int\cdots\int W(y_{-L},\ldots,y_\tau)p(y_{-L},\ldots,y_\tau|\rho^o)\,dy_{-L}\cdots dy_\tau;$$

and the asymptotic variance of the maximum likelihood estimator is $(\mathcal{V}^o)^{-1}$, where

$$\begin{aligned}\mathcal{V}^o &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E\Big\{\frac{\partial}{\partial\rho}\log p(y_t|y_0,\ldots,y_{t-1},\rho^o)\Big\}\Big\{\frac{\partial}{\partial\rho}\log p(y_t|y_0,\ldots,y_{t-1},\rho^o)\Big\}', \quad (8)\\ E(W) &= \int\cdots\int W(y_0,\ldots,y_t)p(y_0,\ldots,y_t|\rho^o)\,dy_0\cdots dy_t\end{aligned}$$

(Gallant, 1987, Chapter 7, Theorem 6). Define

$$M_K^o = \int\int\Big\{\frac{\partial}{\partial\theta}\log f_K(y|x,\theta^o)\Big\}\Big\{\frac{\partial}{\partial\rho}\log p(y|x,\rho^o)\Big\}'\,p(x,y|\rho^o)\,dydx, \tag{9}$$

$$\mathcal{I}_{K,0}^o = \int\int\Big\{\frac{\partial}{\partial\theta}\log f_K(y|x,\theta^o)\Big\}\Big\{\frac{\partial}{\partial\theta}\log f_K(y|x,\theta^o)\Big\}'\,p(x,y|\rho^o)\,dydx. \tag{10}$$

We assume that an analysis based on L lags is valid for large L.

Assumption 4 *We require*

$$\lim_{L\to\infty}\mathcal{V}_{L,0}^o = \lim_{L\to\infty}\mathcal{V}_L^o = \mathcal{V}^o.$$

In addition, $\hat{b}'(\partial/\partial\theta)\log f_K(y|x,\theta^o)$ with $\hat{b} = (\mathcal{I}_{K,0}^o)^{-1}(M_K^o)a$, which is the L_2 projection of $a'(\partial/\partial\rho)\log p(y|x,\rho^o)$ onto the linear span of $(\partial/\partial\theta)\log f_K(y|x,\theta^o)$, has fourth moment bounded uniformly in K for each $a\in\Re^{p_\rho}$.

The last requirement is plausible because $a'(\partial/\partial\rho)\log p(y|x,\rho^o)$ has finite fourth moment under the assumptions in place. Since Lemma 2 implies a uniform bound on the second moment, the bound on the fourth moment can be relaxed by imposing a more stringent mixing requirement in Assumption 1. See Hall & Heyde (1980, p. 20) and the proof of Lemma 3.

2·3. A form suitable for applications

There is no need to retain the term $\epsilon_0\phi(\zeta)$ in the truncated expansion $h_K(\zeta|\theta) = \{g_K(\zeta|\theta)\}^2\phi(\zeta)+\epsilon_0\phi(\zeta)$ provided that those $\log\{g_K(\zeta|\theta)\}^2$ that become too small to have machine representation during an optimization of the log likelihood are set to the smallest number that the machine can represent. In tests, optimizations using this strategy have been more efficient and stable than methods that retained the term $\epsilon_0\phi(\zeta)$. With the term deleted, a change of location and scale prior to conditioning, and some rearrangement of polynomial coefficients, the conditional density can be put in the form

$$h_K(y_t|x_{t-1},\theta) = \frac{[P\{R^{-1}(y-\mu_{x_{t-1}}),x_{t-1}\}]^2\phi\{R^{-1}(y-\mu_{x_{t-1}})\}}{|\det(R)|^{1/2}\int\{P(z,x_{t-1})\}^2\phi(z)\,dz},$$

where $x_{t-1} = (y_{t-L}, \dots, y_{t-1})$,

$$\mu_{x_{t-1}} = b_0 + Bx_{t-1}, \tag{11}$$

$$P(z, x) = \sum_{\alpha=0}^{K_z} \sum_{\beta=0}^{K_x} a_{\beta\alpha} x^\beta z^\alpha, \tag{12}$$

and R is an upper triangular matrix. We refer to μ_x as the mean function and to $P^2(z,x)\phi(z)$ as the Hermite polynomial.

The ability of the model to approximate conditionally heteroscedastic data is much improved by replacing R above by $R_{x_{t-1}}$, where

$$\text{vech}(R_{x_{t-1}}) = \rho_0 + P|x_{t-1} - \mu_{x_{t-2}}|, \tag{13}$$

with vech(R) denoting a vector of length $M(M+1)/2$ containing the elements of the upper triangle of R, and $|x|$ denoting elementwise absolute value. We refer to R_x as the variance function. This yields the auxiliary model used in our applications:

$$f(y_t|x_{t-1}, \theta) = \frac{[P\{R_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}}), x_{t-1}\}]^2 \phi\{R_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}})\}}{|\det(R_{x_{t-1}})|^{1/2} \int \{P(z, x_{t-1})\}^2 \phi(z)\, dz}. \tag{14}$$

The vector θ contains the coefficients $A = (a_{\beta\alpha})$ of the polynomial (12), the coefficients (b_0, B) of the mean function (11), and the coefficients (ρ_0, P) of the variance function(13). To achieve identification, the coefficient $a_{0,0}$ is set to 1.

There is no need for the number of lagged values of y_t in the polynomial $P(z,x)$, the mean function μ_x, or the variance function R_x to be the same. Accordingly, denote them by L_p, L_μ, and L_r, respectively. This can by accomplished within the notational scheme above by putting $L = \max(L_p, L_\mu, L_r + L_\mu)$ and setting certain elements of A, (b_0, B), and (ρ_0, P) to zero. Also, when M is large, coefficients $a_{\beta\alpha}$ corresponding to monomials z^α that represent high order interactions can be set to zero with little effect on the adequacy of approximations. Let $I_z = 0$ indicate that no interaction coefficients are set to zero, $I_z = 1$ indicate that coefficients corresponding to interactions z^α of order larger than $K_z - 1$ are set to zero, and so on; similarly for x^β and I_x. We also find that setting to zero the elements of P in (13) that correspond to the off-diagonal elements of R_x can improve the stability of optimizations with little effect on the adequacy of approximations.

3. The minimum chi-square estimator

The minimum chi-square estimator $\hat{\rho}_n$ that we propose is computed as follows. Use the auxiliary model

$$f(y_t|y_{t-L},\ldots,y_{t-1},\theta) \qquad (\theta \in \Theta \subset \mathcal{R}^{p_\theta}) \tag{15}$$

given by (14) and the data $\{\tilde{y}_t : t = -L,\ldots,n\}$ to compute the maximum likelihood estimate

$$\tilde{\theta}_n := \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{t=0}^{n} \log\{f(\tilde{y}_t|\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1},\theta)\} \tag{16}$$

and the corresponding estimate of the information matrix

$$\tilde{\mathcal{I}}_n := \frac{1}{n}\sum_{t=0}^{n}\Big\{\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1},\tilde{\theta}_n)\Big\}\Big\{\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1},\tilde{\theta}_n)\Big\}'. \tag{17}$$

Define

$$m(\rho,\theta) := \int\cdots\int \frac{\partial}{\partial\theta}\log\{f(y_0|y_{-L},\ldots,y_{-1},\theta)\}\, p(y_{-L},\ldots,y_0|\rho)\, dy_{-L}\cdots dy_0 \tag{18}$$

which, for given $\rho \in \mathcal{R} \subset \Re^{p_\rho}$, is computed as an average

$$m(\rho,\theta) \doteq \frac{1}{N}\sum_{t=0}^{N}\frac{\partial}{\partial\theta}\log\{f(\hat{y}_t|\hat{y}_{t-L},\ldots,\hat{y}_{t-1},\theta)\} \tag{19}$$

over a long simulation $\{\hat{y}_t\}$ generated from the system (1) by means of simulation methods. The proposed minimum chi-square estimator is

$$\hat{\rho}_n := \underset{\rho\in R}{\operatorname{argmin}}\;\; m'(\rho,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\rho,\tilde{\theta}_n). \tag{20}$$

Under regularity conditions implied by the restrictions placed on $p(y_{-L},\ldots,y_0|\rho)$ ($\rho \in \mathcal{R} \subset \Re^{p_\rho}$) by Assumption 1, Gallant & Tauchen (1996) have investigated the asymptotics of $\hat{\rho}_n$ for auxiliary models $f(y_{-L},\ldots,y_0|\theta)$ ($\theta \in \Theta \subset \Re^{p_\theta}$) that satisfy standard regularity conditions for quasi maximum likelihood estimation such as Assumptions 1 to 6 of Gallant (1987, Chapter 7), or those listed in Potscher & Prucha (1991a, 1991b), which are similar to our Assumption 1. Their results are as follows. If θ^o is an isolated solution of the equations $m(\rho^o,\theta) = 0$ and $p_\rho < p_\theta$, then $\hat{\rho}_n$ converges almost surely to ρ^o and $n^{\frac{1}{2}}(\hat{\rho}_n - \rho^o)$ converges in distribution to $N\Big[0,\{(M^o)'(\mathcal{I}^o)^{-1}(M^o)\}^{-1}\Big]$, where

$$\mathcal{I}^o = \int\cdots\int\Big\{\frac{\partial}{\partial\theta}\log f(y_0|x_{-1},\theta^o)\Big\}\Big\{\frac{\partial}{\partial\theta}\log f(y_0|x_{-1},\theta^o)\Big\}' p(y_{-L},\ldots,y_0,|\rho^o)\, dy_{-L}\cdots dy_0,$$

$x_{-1} = (y_{-L}, \ldots, y_{-1})$, $M^o = M(\rho^o, \theta^o)$, and $M(\rho, \theta) = (\partial/\partial\rho')m(\rho, \theta)$. Further, $\lim_{n\to\infty} \hat{M}_n = M^o$ and $\lim_{n\to\infty} \tilde{\mathcal{I}}_n = \mathcal{I}^o$ almost surely, where $\hat{M}_n = M(\hat{\rho}_n, \tilde{\theta}_n)$.

Under the null hypothesis that the system (1) is the correct model,

$$L_0 := n\, m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n) \tag{21}$$

is asymptotically chi-squared on $p_\theta - p_\rho$ degrees freedom. Under the null hypothesis that $h(\rho^o) = 0$, where h maps $\mathcal{R}$ into $\Re^q$,

$$L_h := n\Big\{m'(\hat{\rho}^*, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}^*, \tilde{\theta}_n) - m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n)\Big\}$$

is asymptotically chi-squared on q degrees freedom where

$$\hat{\rho}^* = \underset{h(\rho)=0}{\operatorname{argmin}} \;\; m'(\rho, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\rho, \tilde{\theta}_n).$$

Because $N\Big[0, \mathcal{I}^o - (M^o)\{(M^o)'(\mathcal{I}^o)^{-1}(M^o)\}^{-1}(M^o)'\Big]$ is the limiting distribution of $n^{\frac{1}{2}} m(\hat{\rho}_n, \tilde{\theta}_n)$, when L_0 exceeds the chi-square critical point inspection of the t-ratios $T_n = S_n^{-1} n^{\frac{1}{2}} m(\hat{\rho}_n, \tilde{\theta}_n)$, where $S_n = \Big(\operatorname{diag}[\tilde{\mathcal{I}}_n - (\hat{M}_n)\{(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\hat{M}_n)\}^{-1}(\hat{M}_n)']\Big)^{1/2}$, can suggest reasons for failure. Different elements of the score correspond to different characteristics of the data. For this purpose, the quasi-t-ratios

$$\hat{T}_n := \{(\operatorname{diag} \tilde{\mathcal{I}}_n)^{1/2}\}^{-1} n^{\frac{1}{2}} m(\hat{\rho}_n, \tilde{\theta}_n), \tag{22}$$

which are under-estimates, are usually adequate and are cheaper to compute because they avoid computation of $\hat{M}_n$, which must be done numerically.

4. Efficiency

In this section we state our main result, Theorem 2, which is established by means of three lemmas that are stated here. Proofs are in the Appendix.

Theorem 2 *Assumptions 1, 2, 3, and 4 imply*

$$\lim_{L\to\infty} \lim_{K\to\infty} \lim_{n\to\infty} \operatorname{var}\{\sqrt{n}(\hat{\rho}_n - \rho^o)\} = (\mathcal{V}^o)^{-1}$$

where $(\mathcal{V}^o)^{-1}$ *is the asymptotic variance of the maximum likelihood estimator of* ρ *given by (8).*

The implication of this result is that the asymptotic variance of the proposed minimum chi-square estimator (20) can be made as close as desired to the asymptotic variance of the maximum likelihood estimator by taking L and K suitably large. The result is general in that it applies to any sequence $\{V_t\}_{t=-L}^{\infty}$ and density $p(x,y|\rho)$ that satisfy Assumptions 1 through 4. That is, the sequence $\{V_t\}_{t=-L}^{\infty}$ does not have to be generated by a stochastic differential equation.

In the remainder of this section, notation is as in Subsection 2·2: $p(x,y|\rho) = p(y_{-L},\ldots, y_0|\rho)$, $p(x|\rho) = \int p(y_{-L},\ldots,y_0|\rho)dy_0$, $x = x_{-1} = (y_{-L},\ldots,y_{-1})$, $y = y_0$, $f_K(y|x,\theta)$ is given by (5), M_K^o is given by (9), $\mathcal{I}_{K,0}^o$ is given by (10), etc.

Our first lemma states that if the scores of $p(y|x,\rho^o)$ are in the linear span of the scores of $f_K(y|x,\theta^o)$, then $(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)$ converges to $\mathcal{V}_{L,0}^o$ as given by (7).

LEMMA 1
$$\lim_{K\to\infty}(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(M_K^o) = \mathcal{V}_{L,0}^o$$
if and only if
$$\lim_{K\to\infty}\min_b \int\int\Big\{a'\frac{\partial}{\partial\rho}\log p(y|x,\rho^o) - b'\frac{\partial}{\partial\theta}\log f_K(y|x,\theta^o)\Big\}^2 p(x,y|\rho^o)\,dy\,dx = 0$$
for every $a \in \Re^{p_\rho}$.

The second lemma states that the scores of $p(y|x,\rho^o)$ are in the linear span of the scores of $f_K(y|x,\theta^o)$.

LEMMA 2 *Assumptions 1, 2, and 3 imply*
$$\lim_{K\to\infty}\min_b \int\int\Big\{a'\frac{\partial}{\partial\rho}\log p(y|x,\rho^o) - b'\frac{\partial}{\partial\theta}\log f_K(y|x,\theta^o)\Big\}^2 p(x,y|\rho^o)\,dy\,dx = 0$$
for every $a \in \Re^{p_\rho}$.

The third lemma is of some interest in its own right because it states that the asymptotic variance of the proposed minimum chi-square estimator can be made as close as desired to the asymptotic variance $(\mathcal{V}_{L,0}^o)^{-1}(\mathcal{V}_L^o)(\mathcal{V}_{L,0}^o)^{-1}$ of the quasi maximum likelihood estimator of ρ based on L lags by taking K suitably large. Thus, if the process $\{y_t\}_{t=-L}^{\infty}$ is Markov, then Lemma 3, itself, implies efficiency for large K.

Lemma 3 *Assumptions 1, 2, 3, and 4 imply*

$$\lim_{K\to\infty}\lim_{n\to\infty} \text{var}\sqrt{n}(\hat\rho_n - \rho^o)\} = (\mathcal{V}^o_{L,0})^{-1}(\mathcal{V}^o_L)(\mathcal{V}^o_{L,0})^{-1}$$

where $\mathcal{V}^o_L$ *is given by (6) and* $\mathcal{V}^o_{L,0}$ *by (7).*

Appendix

Proofs

Proof of Theorem 1. Put $s_n(h) = (1/n)\sum\{\log\int h(y_{t-L},\ldots,y_t)dy_t - \log h(y_{t-L},\ldots,y_t)\}$. By an argument analogous to Section 3 of Gallant & Nychka (1987), $s_n(h)$ converges uniformly to $\bar s(h,p) = -\int\int \log h(y|x)\; p(y|x,\rho^o)\,dy\,p(x|\rho^o)\,dx$ which is minimized at $h(y,x) = p(y,x|\rho^o)$. By Theorem 0 of Gallant & Nychka (1987), $\lim_{n\to\infty}\|f_{K_n}(y,x|\tilde\theta_{K_n}) - p(y,x|\rho^o)\|_{[m_o-\ell/2],\infty,\mu} = 0$ almost surely for any random sequence K_n that satisfies $\lim_{n\to\infty} K_n = \infty$ almost surely, where $\theta_K = (\theta_1,\ldots,\theta_{p_K})$. The auxiliary model satisfies Assumptions 4 through 6 of (Gallant, 1987, Chapter 7). By Theorem 4 of (Gallant, 1987, Chapter 7) we can choose N_K for each K such that $n > N_K$ implies that $\|f_K(y,x|\tilde\theta_K) - f_K(y,x|\theta^o_K)\|_{[m_o-\ell/2],\infty,\mu} < K^{-1}$ because $\|f_K(y,x|\tilde\theta_K) - f_K(y,x|\theta^o_K)\|_{[m_o-\ell/2],\infty,\mu}$ is continuous in $(\tilde\theta_K,\theta^o_K)$. We can also impose $N_K < N_{K+1}$. Define $K_n = K$ for $N_K < n \le N_{K+1}$. Then

$$\begin{aligned}&\lim_{n\to\infty}\|f_{K_n}(y,x|\theta^o_{K_n}) - p(y,x|\rho^o)\|_{[m_o-\ell/2],\infty,\mu}\\ &\quad\le \lim_{n\to\infty} K_n^{-1} + \lim_{n\to\infty}\|f_{K_n}(y,x|\tilde\theta_{K_n}) - p(y,x|\rho^o)\|_{[m_o-\ell/2],\infty,\mu} = 0.\end{aligned}$$

Proof of Theorem 2. By Lemma 3, $\lim_{K\to\infty}\lim_{n\to\infty}\text{var}\{\sqrt{n}(\hat\rho_n-\rho^o)\} = (\mathcal{V}^o_{L,0})^{-1}\mathcal{V}^o_L(\mathcal{V}^o_{L,0})^{-1}$. By Assumption 4, $\lim_{L\to\infty}\mathcal{V}^o_L = \lim_{L\to\infty}\mathcal{V}^o_{L,0} = \mathcal{V}^o$.

Proof of Lemma 1. Put $\hat b = (\mathcal{I}^o_{K,0})^{-1}(M^o_K)a$ and note that

$$\begin{aligned}&a'\Big\{\mathcal{V}^o_{L,0} - (M^o_K)'(\mathcal{I}^o_{K,0})^{-1}(M^o_K)\Big\}a\\ &\quad= a'(\mathcal{V}^o_{L,0})a - 2\hat b'(M^o_K)a + \hat b'(\mathcal{I}^o_{K,0})\hat b\\ &\quad= \min_b \int\int\Big\{a'\frac{\partial}{\partial\rho}\log p(y|x,\rho^o) - b'\frac{\partial}{\partial\theta}\log f_K(y|x,\theta^o)\Big\}^2 p(x,y|\rho^o)\,dydx\end{aligned}$$

Proof of Lemma 2. Abbreviate as follows: $g = g(x,y|\rho^o)$, $G = (\partial/\partial\rho)g(y,x|\rho^o)$, $g_K = g_K(x,y|\theta^o)$, $G_K = (\partial/\partial\theta)g_K(y,x|\theta^o)$, $p = p(x,y|\rho^o)$, $p_x = p(x|\rho^o)$, $\phi = \phi(x,y)$, $\phi_x = \phi(x)$,

$\phi_y = \phi(y)$, and $\epsilon = \epsilon_0$. Break the sign ambiguity as follows. There is a point (x^o, y^o) where $g^2\phi - \epsilon\phi$ is positive. Theorem 1 implies that $g_K^2\phi - \epsilon\phi$ must be positive at (x^o, y^o) for large K. Let g and g_K have the same sign at (x^o, y^o) for large K. Using $(A+B)^2 \le 2A^2 + 2B^2$, $(\int AB dy)^2 \le \int A^2 dy \int B^2 dy$, and letting $\mathcal{B}$ denote an upper bound that does not depend on K we have for any b that

$$
\begin{aligned}
&\min_b \frac{1}{16} \int\int \Big\{ a' \frac{\partial}{\partial\rho} \log p(y|x,\rho^o) - b' \frac{\partial}{\partial\theta} \log f_K(y|x,\theta^o) \Big\}^2 p(x,y|\rho^o)\, dy dx \\
= \quad & \min_b \frac{1}{4} \int\int \Big(\frac{a'Gg\phi}{g^2\phi + \epsilon\phi} - \frac{b'G_K g_K \phi}{g_K^2\phi + \epsilon\phi} + \frac{\int a'Gg\phi dy}{\int g^2\phi + \epsilon\phi dy} - \frac{\int b'G_K g_K \phi dy}{\int g_K^2 \phi + \epsilon\phi dy} \Big)^2 p\, dy dx \\
\le \quad & \frac{1}{2} \int\int \Big(\frac{a'Gg\phi}{g^2\phi + \epsilon\phi} - \frac{b'G_K g_K \phi}{g_K^2\phi + \epsilon\phi} \Big)^2 + \Big(\frac{\int a'Gg\phi dy}{\int g^2\phi + \epsilon\phi dy} - \frac{\int b'G_K g_K \phi dy}{\int g_K^2 \phi + \epsilon\phi dy} \Big)^2 p\, dy dx \\
\le \quad & \int\int \Big(\frac{g\phi}{g^2\phi + \epsilon\phi} - \frac{g_K\phi}{g_K^2\phi + \epsilon\phi} \Big)^2 (a'G)^2 p\, dy dx \\
& + \int \Big\{ \int \Big(\frac{g}{\int g^2 \phi_y dy + \epsilon} - \frac{g_K}{\int g_K^2 \phi_y dy + \epsilon} \Big)^2 \phi_y dy \Big\} \Big\{ \int (a'G)^2 \phi_y dy \Big\} p_x dx \\
& + \int\int \Big\{ \frac{g_K \phi}{g_K^2\phi + \epsilon\phi} \Big\}^2 (a'G - b'G_K)^2 p\, dy dx \\
& + \int \Big\{ \frac{\int g_K^2 \phi_y dy}{(\int g_K^2 \phi_y dy + \epsilon)^2} \Big\} \Big\{ \int \Big(a'G - b'G_K \Big)^2 \phi_y dy \Big\} p_x dx \\
\le \quad & \int\int \Big(\frac{g\phi}{g^2\phi + \epsilon\phi} - \frac{g_K\phi}{g_K^2\phi + \epsilon\phi} \Big)^2 (a'G)^2 p\, dy dx \\
& + \int \Big\{ \int \Big(\frac{g}{\int g^2 \phi_y dy + \epsilon} - \frac{g_K}{\int g_K^2 \phi_y dy + \epsilon} \Big)^2 \phi_y dy \Big\} \Big\{ \int (a'G)^2 \phi_y dy \Big\} p_x dx \\
& + \mathcal{B} \int\int (a'G - b'G_K)^2 (p + \phi_y p_x) dy dx.
\end{aligned}
$$

By Assumption 3, the density $(p + \phi_y p_x)/2$ has a moment generating function. The polynomials are dense in an L_2 probability space whose probability density has a moment generating function (Gallant, 1980). The vector G_K contains all monomials in (x, y) up to degree K. Therefore, we can choose a b for each K such that $\lim_{K\to\infty} \mathcal{B} \int\int (a'G - b'G_K)^2 (p + \phi_y p_x) dy dx = 0$. Theorem 1 states that $\lim_{K\to\infty} \| f_K(y, x|\theta^o) - p(y, x|\rho^o) \|_{[m_o - \ell/2], \infty, \mu} = 0$ which implies that $\lim_{K\to\infty} \sup_{x,y} |g\phi - g_K\phi| = 0$. Since $\{ g\phi/(g^2\phi + \epsilon\phi) - g_K\phi/(g_K^2\phi + \epsilon\phi) \} = \{ g/(g^2 + \epsilon) - g_K/(g_K^2 + \epsilon) \}$ is a bounded, continuous function of $(g\phi, g_K\phi)$ we have $\lim_{K-\infty} | g\phi/(g^2\phi + \epsilon\phi) - g_K\phi/(g_K^2\phi + \epsilon\phi) | = 0$ pointwise in (x, y). By the Dominated Convergence Theorem

$$
\lim_{K\to\infty} \int\int \Big(\frac{g\phi}{g^2\phi + \epsilon\phi} - \frac{g_K\phi}{g_K^2\phi + \epsilon\phi} \Big)^2 (a'G)^2 p\, dy dx = 0.
$$

Similarly, Theorem 1 implies that for each fixed x

$$\lim_{K\to\infty} \int \Big(\frac{g}{\int g^2 \phi_y dy + \epsilon} - \frac{g_K}{\int g_K^2 \phi_y dy + \epsilon} \Big)^2 \phi_y dy = 0.$$

Moreover,

$$\int \Big(\frac{g}{\int g^2 \phi_y dy + \epsilon} - \frac{g_K}{\int g_K^2 \phi_y dy + \epsilon} \Big)^2 \phi_y dy \le 2 \Big(\frac{\int g^2 \phi_y dy}{(\int g^2 \phi_y dy + \epsilon)^2} + \frac{\int g_K^2 \phi_y dy}{(\int g_K^2 \phi_y dy + \epsilon)^2} \Big)$$

and is therefore bounded uniformly in K. By the Dominated Convergence Theorem

$$\lim_{K\to\infty} \int \Big\{ \int \Big(\frac{g}{\int g^2 \phi_y dy + \epsilon} - \frac{g_K}{\int g_K^2 \phi_y dy + \epsilon} \Big)^2 \phi_y dy \Big\} \Big\{ \int (a'G)^2 \phi_y dy \Big\} p_x dx = 0.$$

Proof of Lemma 3. Let $\mathcal{J}_K^o = \int\int (\partial^2/\partial\theta\partial\theta') \log f_K(y|x,\theta^o) p(y,x|\rho^o)\, dy dx$ and let $o_s(1)$ denote a matrix or vector whose elements converge almost surely to zero and similarly $o_p(1)$ for convergence in probability. Assumptions 4 through 6 of Gallant (1987, Chapter 7) are satisfied by f_K permitting application of Theorems 1 through 6 of Gallant (1987, Chapter 7) which justifies the following Taylor expansion of the first order conditions.

$$\begin{aligned}
0 &= (\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1} \sqrt{n}\, m(\hat\rho_n, \tilde\theta_n) \\
&= \{M_K^o + o_s(1)\}' \{\mathcal{I}_{K,0}^o + o_s(1)\}^{-1} [\sqrt{n}\, m(\rho^o, \tilde\theta_n) + \{M_K^o + o_s(1)\}\sqrt{n}(\hat\rho_n - \rho^o)] \\
&= (M_K^o)'(\mathcal{I}_{K,0}^o)^{-1} \{\sqrt{n}\, m(\rho^o, \tilde\theta_n) + (M_K^o)\sqrt{n}(\hat\rho_n - \rho^o)\} + o_p(1) \\
&= (M_K^o)'(\mathcal{I}_{K,0}^o)^{-1} \{\sqrt{n}\, m(\rho^o, \theta^o) + (\mathcal{J}_K^o)\sqrt{n}(\tilde\theta_n - \theta^o) + (M_K^o)\sqrt{n}(\hat\rho_n - \rho^o)\} + o_p(1) \\
&= (M_K^o)'(\mathcal{I}_{K,0}^o)^{-1} \{(\mathcal{J}_K^o)\sqrt{n}(\tilde\theta_n - \theta^o) + (M_K^o)\sqrt{n}(\hat\rho_n - \rho^o)\} + o_p(1)
\end{aligned}$$

Application of Theorem 6 of Gallant (1987, Chapter 7) to

$$(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)\sqrt{n}(\hat\rho_n - \rho^o) = -(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(\mathcal{J}_K^o)\sqrt{n}(\tilde\theta_n - \theta^o) - o_p(1)$$

gives

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{var}\{\sqrt{n}(\hat\rho_n - \rho^o)\} \\
&\quad = \Big\{(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)\Big\}^{-1} \Big\{(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}\mathcal{I}_K^o(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)\Big\} \Big\{(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)\Big\}^{-1}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{I}_K^o &= \mathcal{I}_{K,0}^o + \sum_{\tau=1}^{\infty} \mathcal{I}_{K,\tau}^o + (\sum_{\tau=1}^{\infty} \mathcal{I}_{K,\tau}^o)' \\
\mathcal{I}_{K,\tau}^o &= E\Big\{\frac{\partial}{\partial\theta} \log f_K(y_{t+\tau}|x_{t+\tau-1}, \theta^o)\Big\}\Big\{\frac{\partial}{\partial\theta} \log f_K(y_t|x_{t-1}, \theta^o)\Big\}'
\end{aligned}$$

For $a \in \Re^{p_\rho}$ let $\hat{b}$ be as in the proof of Lemma 1. Then

$$
\begin{aligned}
& a'(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(\sum_{\tau=1}^{\infty}\mathcal{I}_{K,\tau}^o)(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)a \\
&\quad = \sum_{\tau=1}^{\infty} E\Big\{\hat{b}'\frac{\partial}{\partial\theta}\log f_K(y_{t+\tau}|x_{t+\tau-1},\theta^o)\Big\}\Big\{\hat{b}'\frac{\partial}{\partial\theta}\log f_K(y_t|x_{t-1},\theta^o)\Big\} \\
&\quad = \sum_{\tau=1}^{\infty} E(X_{t+\tau}X_t) \;=\; \sum_{\tau=1}^{\infty} E\{X_t E(X_{t+\tau}|\mathcal{F}_t)\}
\end{aligned}
$$

where $\mathcal{F}_t$ is the smallest σ-algebra such that the random variables $\{V_{t-\tau}\}_{\tau\geq 0}$ as defined in Assumption 1 are measurable. The first order conditions of quasi maximum likelihood estimation imply $E(X_{t+\tau}) = 0$ so by Proposition 2 of Gallant (1987, Chapter 7) we have

$$\|E(X_{t+\tau}|\mathcal{F}_t)\|_2 \leq 2(2^{1/p}+1)\|X_{t+\tau}\|_4\mathcal{O}\Big\{(\tau^{-4r/(r-4)})^{1/2-1/4}\Big\}$$

for some $r > 4$. By Assumption 4, for some bound $\mathcal{B}$ and some $\delta > 0$ we have

$$\left| E\Big\{X_t E(X_{t+\tau}|\mathcal{F}_t)\Big\}\right| \leq \|X_t\|_2\|E(X_{t+\tau}|\mathcal{F}_t)\|_2 \leq \mathcal{B}\mathcal{O}(\tau^{-1-\delta})$$

Hence

$$a'(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(\sum_{\tau=1}^{\infty}\mathcal{I}_{K,\tau}^o)(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)a = \sum_{\tau=1}^{T} E(X_{t+\tau}X_t) + \mathcal{O}(T^{-\delta})$$

where $\mathcal{O}(T^{-\delta})$ does not depend on K. Now $\lim_{K\to\infty}\|X_t - a'(\partial/\partial\rho)\log p(y_t|x_{t-1},\rho^o)\|_2 = 0$ implies $\lim_{K\to\infty}|E(X_{t+\tau}X_t)| = |a'(\mathcal{V}_{L,\tau})a|$. Thus, we have

$$\lim_{K\to\infty}\Big\{(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(\mathcal{I}_K^o)(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)\Big\} = \mathcal{V}_{L,0}^o + \sum_{\tau=1}^{\infty}\mathcal{V}_{L,\tau}^o + \Big(\sum_{\tau=1}^{\infty}\mathcal{V}_{L,\tau}^o\Big)'$$

$$\lim_{K\to\infty}\Big\{(M_K^o)'(\mathcal{I}_{K,0}^o)^{-1}(M_K^o)\Big\} = \mathcal{V}_{L,0}^o.$$

6. Addendum: Simulation Schemes

The formulas implemented by weak2.f and stng1.f are as follows.

6·1. *Explicit Order 2 Weak Scheme*

Recursion

$$
\begin{aligned}
\hat{U}_{t+\Delta} &= \hat{U}_t + \frac{1}{2}\Big[A(\Upsilon,\rho) + A(\hat{U}_t,\rho)\Big]\Delta \\
&+ \frac{1}{4}\sum_{j=1}^{k}\Big\{\Big[B_j(R_j^+,\rho) + B_j(R_j^-,\rho) + 2B_j(\hat{U}_t,\rho)\Big]\Delta W_j \\
&\quad + \sum_{\substack{r=1\\ r\neq j}}^{k}\Big[B_j(Y_r^+,\rho) + B_j(Y_r^-,\rho) - 2B_j(\hat{U}_t,\rho)\Big]\Delta W_j \Delta^{-1/2}\Big\} \\
&+ \frac{1}{2}\sum_{j=1}^{k}\Big\{\Big[B_j(R_j^+,\rho) - B_j(R_j^-,\rho)\Big]I_{jj} + \sum_{\substack{r=1\\ r\neq j}}^{k}\Big[B_j(Y_r^+,\rho) - B_j(Y_r^-,\rho)\Big]I_{rj}\Big\}\Delta^{-1/2}
\end{aligned}
$$

Supporting values

$$
\begin{aligned}
\Upsilon &= \hat{U}_t + A(\hat{U}_t,\rho)\Delta + \sum_{j=1}^{k} B_j(\hat{U}_t,\rho)\Delta W_j \\
R_j^{\pm} &= \hat{U}_t + A(\hat{U}_t,\rho)\Delta \pm B_j(\hat{U}_t,\rho)\Delta^{1/2} \\
Y_j^{\pm} &= \hat{U}_t \pm B_j(\hat{U}_t,\rho)\Delta^{1/2}
\end{aligned}
$$

Integral approximation

$$
\begin{aligned}
I_{rj} &= (1/2)\Big[\Delta W_j \Delta W_r + V_{rj}\Big] \\
V_{rj} &= -\Delta I_{(0,\frac{1}{2}]}(U_{rj}) + \Delta I_{(\frac{1}{2},1]}(U_{rj}) && r < j \\
V_{rj} &= -\Delta && r = j \\
V_{rj} &= -V_{jr} && r > j
\end{aligned}
$$

Independent random variables

$$
\begin{aligned}
\Delta W_j &\sim N(0,\Delta) && j = 1,\ldots,k \\
U_{rj} &\sim U(0,1] && r = 1,\ldots,j-1, \quad j = 1,\ldots,k
\end{aligned}
$$

6·2. *Explicit Order 1 Strong Scheme*

Recursion

$$\begin{aligned}\hat{U}_{t+\Delta} &= \hat{U}_t + A(t,\hat{U}_t,\rho)\Delta + \sum_{j=1}^{k} B_j(t,\hat{U}_t,\rho)\Delta W_j \\ &\quad + \frac{1}{\sqrt{\Delta}}\sum_{j=1}^{k}\sum_{r=1}^{k}\Big[B_j(t,\Upsilon_r,\rho) - B_j(t,\hat{U}_t,\rho)\Big] I_{rj}\end{aligned}$$

Supporting values

$$\Upsilon_j = \hat{U}_t + A(t,\hat{U}_t,\rho)\Delta + B_j(t,\hat{U}_t,\rho)\sqrt{\Delta}$$

Integral approximation

$$\begin{aligned}I_{rj} &= (1/2)\Big[(\Delta W_j)^2 - \Delta\Big] && r = j \\ I_{rj} &= (1/2)\Delta W_r \Delta W_j + (\Delta C)^{1/2}(\mu_r \Delta W_j - \mu_j \Delta W_r) && r \neq j \\ &\quad + \frac{\Delta}{2\pi}\sum_{\ell=1}^{p}\frac{1}{\ell}\Big[\zeta_{r\ell}\Big(\frac{\Delta W_j}{\sqrt{\Delta/2}} + \eta_{j\ell}\Big) - \zeta_{j\ell}\Big(\frac{\Delta W_r}{\sqrt{\Delta/2}} + \eta_{r\ell}\Big)\Big] \\ C &= \frac{1}{12} - \frac{1}{2\pi^2}\sum_{\ell=1}^{p}\frac{1}{l^2} && p = 50\end{aligned}$$

Independent random variables

$$\begin{aligned}\Delta W_j &\sim N(0,\Delta) && j = 1,\ldots,k \\ \mu_j &\sim N(0,1) && j = 1,\ldots,k \\ \eta_{j\ell} &\sim N(0,1) && j = 1,\ldots,k, \quad \ell = 1,\ldots,p \\ \zeta_{j\ell} &\sim N(0,1) && j = 1,\ldots,k, \quad \ell = 1,\ldots,p\end{aligned}$$

References

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Rev. of Financ. Stud.* **9**, 385–426.

Andersen, T. G., & Lund, J. (1996). Estimating continuous time stochastic volatility models of the short term interest rate. *J. Econometrics*, in press.

Bansal, R., Gallant, A. R., Hussey, R., & Tauchen, G. (1993). Computational aspects of nonparametric simulation estimation. In *Computational Techniques for Econometrics and Economic Analysis*, Ed. D. A. Belsley, pp. 3–22. Boston: Kluwer Academic Publishers.

Bansal, R., Gallant, A. R., Hussey, R., & Tauchen, G. (1995). Nonparametric estimation of structural models for high-frequency currency market data. *J. Econometrics* **66**, 251–87.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**, 135–56.

Duffie, D., & Glynn, P. (1995). Efficient Monte Carlo simulation of security prices. *Ann. Appl. Probab.* **5**, 897–906.

Duffie, D., & Singleton, K. J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica* **61**, 929–52.

Gallant, A. R. (1980). Explicit estimators of parametric functions in nonlinear regression. *J. Am. Statist. Assoc.* **75**, 182–93.

Gallant, A. R. (1987). *Nonlinear Statistical Models.* New York: Wiley.

Gallant, A. R., & Nychka, D. W. (1987). Seminonparametric maximum likelihood estimation. *Econometrica* **55**, 363–90.

Gallant, A. R., & Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* **57**, 1091–120.

Gallant, A. R., & Tauchen, G. (1996). Which moments to match? *Econometric Theory,* in press.

Gallant, A. R., Rossi, P. E., & Tauchen, G. (1992). Stock prices and volume. *Rev. of Financ. Stud.* **5**, 199–242.

GILL, P. E., MURRAY, W., SAUNDERS, M. A. & WHITE, M. H. (1983). *User's Guide for SOL/NPSOL: A Fortran Package for Nonlinear Programming.* Palo Alto, California: Stanford University.

GOURIEROUX, C., MONFORT, A., & RENAULT, E. (1993). Indirect inference. *J. Appl. Econometrics* **8**, S85–S118.

HALL, P., & HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications*, New York: Academic Press.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. & STAHEL, W. A. (1986). *Robust Statistics, The Approach Based on Influence Functions*, New York: Wiley.

INGRAM, B. F., & LEE, B. S. (1991). Simulation estimation of time series models. *J. Econometrics* 47, 197–250.

KARATZAS, I., & SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus.* 2nd Edition. Berlin: Springer-Verlag.

KLOEDEN, P. E., & PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations.* Berlin: Springer-Verlag.

LO, A. (1988). Maximum likelihood estimation of generalized Itô process with discretely sampled data. *Econometric Theory* 4, 231–47.

POTSCHER, B. M. & PRUCHA, I. R. (1991a). Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: consistency and approximation concepts. *Econometric Reviews* 10, 125–216.

POTSCHER, B. M. & PRUCHA, I. R. (1991b). Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: asymptotic normality. *Econometric Reviews* 10, 253–325.

TAUCHEN, G., & PITTS, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica* 51, 485–505.