NCER Working Paper Series

A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions

Stan Hurn
Andrew McClelland
Kenneth Lindsay

Working Paper #65
October 2010
A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions

A. S. Hurn and A. J. McClelland
School of Economics and Finance, Queensland University of Technology

K.A. Lindsay
Department of Mathematics, University of Glasgow

Abstract
This paper develops a quasi-maximum likelihood (QML) procedure for estimating the parameters of multi-dimensional stochastic differential equations. The transitional density is taken to be a time-varying multivariate Gaussian where the first two moments of the distribution are approximately the true moments of the unknown transitional density. For affine drift and diffusion functions, the moments are shown to be exactly those of the true transitional density and for nonlinear drift and diffusion functions the approximation is extremely good. The estimation procedure is easily generalizable to models with latent factors, such as the stochastic volatility class of model. The QML method is as effective as alternative methods when proxy variables are used for unobserved states. A conditioning estimation procedure is also developed that allows parameter estimation in the absence of proxies.

Keywords
stochastic differential equations, parameter estimation, quasi-maximum likelihood, moments.

JEL Classification C22, C52

Corresponding author
Stan Hurn School of Economics and Finance
Queensland University of Technology
Brisbane, 4001, Australia
email s.hurn@qut.edu.au

Acknowledgments
We are particularly grateful to Yacine Aït-Sahalia for very helpful discussions on earlier versions of this paper. The financial assistance of the Carnegie Trust for the Universities of Scotland is also gratefully acknowledged.
1 Introduction

Estimation of the parameters of systems of stochastic differential equations (SDEs) by maximum likelihood poses a number of considerable challenges. First and foremost among these is that the likelihood function is seldom known in closed-form. Exact maximum likelihood (EML) estimation, particularly in multi-dimensions, is therefore infeasible for all practical purposes apart from a few trivial cases. One possible way of maintaining a likelihood framework is to use the closed-form approximation to the true likelihood function developed for the univariate case by Aït-Sahalia (2002). Aït-Sahalia (2008) generalized this method to the multi-dimensional case and Aït-Sahalia and Kimmel (2007) apply the method to the stochastic volatility model using proxy variables for the unobserved volatility. The closed-form approximation works well in the cases in which it has been applied (see for example, Jensen and Poulsen (2002) and Hurn, Jeisman and Lindsay (2007) for univariate comparisons), but it becomes more difficult to implement as the dimension of the problem increases.

The aim of this paper is to develop a simple yet comprehensive quasi-maximum likelihood (QML) framework for estimating the parameters of multivariate SDEs. The strategy is to trade-off estimating the shape of the true transitional density function for a procedure that is able to handle multi-dimensions and unobservable state variables in a straightforward and robust way. Consequently, the transitional probability density function from which the log-likelihood function is constructed, is approximated by the Gaussian distribution which, although not the true transitional density function, is an excellent approximation to this density for intervals of short duration.\(^1\)

The proposed procedure takes advantage of the simplicity of the Gaussian approximation to the transitional density function, but improves the quality of the approximation by identifying accurately the first and second moments of the true transitional density.\(^3\) It will be shown that for stochastic differential equations in which the drift and diffusion functions are affine in the state variables, the first and second moments of the true but unknown transitional density may be computed exactly and the resultant QML parameter estimates are consistent. This result does not generalize to the case of non-affine specification of drift and diffusion, but an analytical bound on

---

\(^1\)Another possible approach, loosely based on the likelihood principle, recognizes that the characteristic function of the transitional density can be computed in closed form if the drift and diffusion are affine functions of the state (Duffie and Kan, 1996). In these cases parameter estimation by a method of moments procedure may be possible (Singleton 2001; Jiang and Knight, 2002; Chacko and Viciera, 2003) where the moments are obtained by selective differentiation of the characteristic function with respect to the parameter vector.

\(^2\)Indeed, it is the quality of this approximation that underlies the success of the Aït-Sahalia (2002, 2008) expansions.

\(^3\)There is some precedence for the QML procedure developed within this paper. Indeed, Fisher and Gilles (1996) and Duffee (2002) propose a similar procedure for estimating the parameters of affine multi-factor term structure models. However, these papers are concerned solely with affine processes.
the approximation error is derived which demonstrates that the size of this error is related to the lowest moment condition that is not satisfied precisely. Simulation experiments are presented to demonstrate the efficacy of QML in dealing with both univariate and multivariate diffusions and the method is applied to the the problem of estimating the parameters of the Heston stochastic volatility model.

Even if the likelihood function for the diffusion under consideration were known, it is only useful in its primitive form if all the state variables are observed. In multivariate problems, a major challenge arises because some of the state variables may be unobserved. A potential advantage of the QML procedure proposed here is that it can handle unobserved state variables by conditioning the multivariate Gaussian density function on the unobserved states. This aspect of QML is investigated within the context of the stochastic volatility application. The results suggest that this QML procedure has potential as a method for quick estimation of models with latent variables in situations where no observable proxies are available.

2 The QML Procedure

Suppose the $N$-dimensional process $X(s) = (X_1, \cdots, X_N)^T$ with sample space $S$ satisfies the stochastic differential equation

$$dX = \alpha(X; \theta) \, ds + \sigma(X; \theta) \, dW,$$  \hspace{1cm} (1)

where $\theta$ is a vector of model parameters, $\sigma(X; \theta)$ is an array of dimension $N \times M$ with $M \leq N$, and $dW$ is the vector of increments in the $M$ dimensional vector Wiener process $W$ with $M \times M$ covariance matrix $Q$ defined by $Q \, ds = \mathbb{E}[dW \otimes dW]$.

Let $f_0(X, s \mid X_t, \theta)$ denote the transitional probability density function of the process $X$ for $s \in [0, \infty)$, then the probability flux vector associated with SDE (1) is

$$J = \alpha(X; \theta) f_0(X, s \mid X_t, \theta) - \frac{1}{2} \operatorname{div} (G(X; \theta) f_0(X, s \mid X_t, \theta)),$$  \hspace{1cm} (2)

where $G$ is the $N \times N$ diffusion matrix given by $\sigma Q \sigma^T$. Conservation of probability density requires that $f_0(X, s \mid X_t, \theta)$ satisfies the Fokker-Planck equation

$$\frac{\partial f_0(X, s \mid X_t, \theta)}{\partial s} + \operatorname{div} J = 0, \quad (X, s) \in S \times (0, \infty)$$  \hspace{1cm} (3)

with boundary and initial conditions

$$J^T n = 0, \quad (X, s) \in \partial S \times (0, \infty),$$

$$f_0(X, 0 \mid X_t, \theta) = \delta(X - X_t), \quad X \in S.$$  \hspace{1cm} (4)
where \( n \) is the unit outward normal to \( \partial S \). When the state \( X_t \) is fully observed, the initial density will be a product of delta functions, but otherwise it will be a product of delta functions and the conditional density of the unobserved state variables.

The true values of the parameters, denoted henceforth by \( \theta_0 \), are to be estimated from data typically consisting of a sequence of observations \( X_0, \ldots, X_T \) of the system at discrete times \( t_0, \ldots, t_T \). In particular, the maximum-likelihood (ML) estimate, \( \theta_{ML} \), of \( \theta_0 \) is obtained by maximizing the conditional log-likelihood function of the observed sample, namely

\[
\log L_T(\theta) = \sum_{t=1}^{T} \log f_0(X_t, \Delta \mid X_{t-1}, \theta),
\]

with respect to the parameters \( \theta \), where \( \Delta \) is the duration of the interval between observations. The associated score function is

\[
S_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \log f_0(X_t, \Delta \mid X_{t-1}, \theta)}{\partial \theta},
\]

which will be zero at the maximum likelihood estimate, \( \theta_{ML} \), where \( \theta_{ML} \) is well known to provide a consistent estimator of \( \theta_0 \).

The primary difficulty with this approach is that the log-likelihood function in equation (5) is seldom known in closed-form, with the vast majority of known cases relating to univariate models. Consequently EMIL estimation in multi-dimensions is infeasible for all practical purposes. The central idea of this avenue of research is to replace the true transitional density by the multivariate Gaussian density

\[
f_1(X, s \mid X_t, \theta) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\Sigma_1|^{1/2}} \exp \left[ -\frac{1}{2} (X - \mu_1)^T \Sigma_1^{-1} (X - \mu_1) \right],
\]

where \( \mu_1 = \mu_1(X_t, s, \theta) \) and \( \Sigma_1 = \Sigma_1(X_t, s, \theta) \) are respectively the conditional mean and conditional covariance of the distribution. The primary reason for using the Gaussian distribution is that the true transitional density is asymptotically Gaussian for the short time intervals, \( \Delta \), typically encountered in financial applications\(^4\). Nevertheless, the multivariate Gaussian density is not the true transitional density and therefore the log-likelihood function based on this density will be misspecified. Proposition 1 establishes conditions under which the QML estimator of \( \theta \) is a consistent estimator.

**Proposition 1**

*Consider the continuous multivariate process satisfying equation (1), with drift specification \( \alpha(X; \theta) \) and diffusion specification \( G(X; \theta) \) and true parameters \( \theta_0 \). Let \( f_1(X, \Delta \mid X_t, \theta) \) be the multivariate density

\[^4\]This recognition is the basis of the approximation to the true transitional density function developed by Ait-Sahalia (2002, 2008) using Hermite polynomials. \*
Gaussian approximating density for $f_0$ with conditional mean $\mu_1(X_t, \Delta, \theta)$ and conditional covariance $\Sigma_1(X_t, \Delta, \theta)$. The quasi-maximum likelihood estimator $\theta_{QML}$ satisfies

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \log f_1(X_t, \Delta | X_{t-1}, \theta_{QML})}{\partial \theta} = 0. \quad (7)$$

The quasi-maximum likelihood estimator is consistent, that is

$$\lim_{T \to \infty} \theta_{QML} = \theta_0,$$

provided

$$\lim_{T \to \infty} \mu_1(X_t, \Delta, \theta_{QML}) = \mu_0(X_t, \Delta, \theta_0)$$

and

$$\lim_{T \to \infty} \Sigma_1(X_t, \Delta, \theta_{QML}) = \Sigma_0(X_t, \Delta, \theta_0).$$

**Proof** The proof of Proposition 1 is in the Appendix.

Proposition 1 establishes that if the conditional mean and conditional covariance of the approximating multivariate Gaussian distribution are the conditional mean and conditional covariance of the true transitional density function, then the QML parameter estimates will be consistent notwithstanding the fact that the transitional density is misspecified. Furthermore, consistent standard errors may be computed using standard methods for misspecified likelihood functions.

### 3 Evolution of Moments

As established in Proposition 1, the effective implementation of the QML procedure relies crucially on the accurate estimation of the first two moments of the true transitional probability density function. The simplest approach is known as discrete maximum likelihood (DML). The system in equation (1) is discretized using either an Euler or a Milstein approximation and the discrete drift and diffusion functions so obtained are used as approximations to the mean and variance of the true transitional distribution. As these moments are determined by the initial point of each transition and do not change as the process evolves, DML is generally not a consistent estimation method. Kessler (1997) develops a more sophisticated approximation to the mean and variance of the approximating Gaussian transitional density using series expansion based on the infinitesimal operator. This is an improvement on simple DML but the moments are still approximations. The distinguishing aspect of this research is that for affine models the first and second moments of the true transitional density can be found exactly. The conditions for Proposition 1 to hold are therefore satisfied. For non-affine models, the moments cannot be computed exactly, but the error in the approximation can be computed and the order of this error related to the first moment condition that is not satisfied exactly.
3.1 Affine Models

The most general affine specification of drift and diffusion as functions of state are respectively

\[
\alpha = a(s) + A(s)X, \quad G = B(s) + C(s)X, \tag{8}
\]

where \(a(s)\) is a column vector of dimension \(N\), \(A(s)\) and \(B(s)\) are matrices of type \(N \times N\) and \(C(s)\) is a vector of \(N \times N\) matrices (or a tensorial array of type \(N \times N \times N\)). Let the conditional mean \(\mu_0\) and the conditional covariance \(\Sigma_0\) of the true transitional probability density be denoted by

\[
\begin{align*}
\mu_0(X_t, s, \theta) &= \int_S X f_0(X, s | X_t, \theta) \, dX, \\
\Sigma_0(X_t, s, \theta) &= \int_S (X \otimes X) f_0(X, s | X_t, \theta) \, dX - \mu_0(X_t, s, \theta) \otimes \mu_0(X_t, s, \theta).
\end{align*} \tag{9}
\]

The evolution of the conditional first and second moments of \(X\) are governed by the equations

\[
\begin{align*}
\frac{d}{ds} \int_S X f_0(X, s | X_t, \theta) \, dX &= - \int_S X \text{div } J \, dX, \\
\frac{d}{ds} \int_S (X \otimes X) f_0(X, s | X_t, \theta) \, dX &= - \int_S (X \otimes X) \text{div } J \, dX. \tag{10}
\end{align*}
\]

The divergence theorem applied to these integrals, in combination with the zero probability flux boundary condition (4), gives respectively

\[
\begin{align*}
\frac{d}{ds} \int_S X f_0(X, s | X_t, \theta) \, dX &= \int_S \alpha(X; \theta) f_0(X, s | X_t, \theta) \, dX, \\
\frac{d}{ds} \int_S (X \otimes X) f_0(X, s | X_t, \theta) \, dX &= \int_S (\alpha(X; \theta) \otimes X + X \otimes \alpha(X; \theta)) f_0(X, s | X_t, \theta) \, dX + \int_S G(X; \theta) f_0(X, s | X_t, \theta) \, dX. \tag{11}
\end{align*}
\]

where the calculation has assumed that surface terms involving the diffusion \(G\) vanish on \(\partial S\).

When \(\alpha(X; \theta) = a(s) + A(s)X\) then the first of equations (11) becomes

\[
\frac{d\mu_0}{ds} = a(s) \int_S f_0(X, s | X_t, \theta) \, dX + A(s) \int_S X f_0(X, s | X_t, \theta) \, dX = a(s) + A(s) \mu_0. \tag{12}
\]

Now replace \(\alpha(X; \theta)\) and \(G(X; \theta)\) by expressions (8) in the second of equations (11) to get

\[
\begin{align*}
\frac{d}{ds} \int_S (X \otimes X) f_0(X, s | X_t, \theta) \, dX &= \int_S \left[(a \otimes X + AX \otimes X + X \otimes a + X \otimes (AX)) + B(s) + C(s)X \right] f_0(X, s | X_t, \theta) \, dX.
\end{align*}
\]
When the definitions of $\mu_0(Y_t, s, \theta)$ and $\Sigma_0(Y_t, s, \theta)$ are used to replace the first and second moments of $X$, the previous equation becomes

\[
\frac{d}{ds} \left( \Sigma_0 + \mu_0 \otimes \mu_0 \right) = a(s) \otimes \mu_0 + A(s)(\Sigma_0 + \mu_0 \otimes \mu_0) + \mu_0 \otimes a(s) + (\Sigma_0 + \mu_0 \otimes \mu_0) A^T(s)s + B(s) + C(s)\mu_0. \tag{13}
\]

Rearranging this equation and removing the time derivatives of $\mu_0$ using property (12) yields the ordinary differential equations

\[
\begin{align*}
\frac{d\mu_0}{ds} &= a(s) + A(s)\mu_0, \\
\frac{d\Sigma_0}{ds} &= A(s)\Sigma_0 + \Sigma_0 A^T(s) + B(s) + C(s)\mu_0,
\end{align*}
\tag{14}
\]

for the conditional mean and conditional covariance of the process described by equation (1). Thus the conditional mean value and conditional covariance of the transitional probability density function satisfy exactly a system of $N(N + 3)/2$ linear ordinary differential equations with coefficients which may be functions of time. These equations have closed-form solutions that depend only upon the parameters $\theta$ and do not require knowledge of the true transitional probability density function.\(^5\) This leads naturally to Corollary 1 for the QML estimates of the parameters of SDE (1) with affine drift and diffusion functions.

**Corollary 1**

Given affine specifications for the drift and diffusion functions of SDE (1), then the use of the multivariate Gaussian density with conditional mean and conditional covariance

\[
\begin{align*}
\mu_1(X_t, s, \theta) &= \mu_0(X_t, s, \theta_{\text{QML}}), \\
\Sigma_1(X_t, s, \theta) &= \Sigma_0(X_t, s, \theta_{\text{QML}}),
\end{align*}
\]

to approximate the true transitional density, will yield consistent parameter estimates of $\theta_0$.

**Proof**

By the weak law of large numbers $\mu_1(X_t, s, \theta_{\text{QML}})$ converges to $\mu_0(X_t, s, \theta_0)$ and $\Sigma_1(X_t, s, \theta_{\text{QML}})$ converges to $\Sigma_0(X_t, s, \theta_0)$. Consequently the QML score function in equation (7) will converge to the true score function as $T \to \infty$. \(\blacksquare\)

**Example: The CIR Model**

Consider the Cox, Ingersol and Ross (1985) equation

\[
dX = \lambda(\beta - X) ds + \sigma \sqrt{X} dW \tag{15}
\]

\(^5\)More generally, closed from solutions can be constructed whenever $G(X; \theta)$ is a quadratic form.
with drift specification \( \mu(X; \theta) = \lambda(\beta - X) \), diffusion specification \( G(X; \theta) = \sigma^2 X \) and sample space \( S = (0, \infty) \). In this process

\[
a(s) = \lambda \beta, \quad A(s) = -\lambda, \quad B(s) = 0, \quad C(s) = \sigma^2.
\]

Assuming that \( X = X_t \) at \( s = 0 \), then the initial value of the true transitional probability density function is \( f_0(X, 0 | X_t, \theta) = \delta(X - X_t) \) and the initial value of the conditional mean process and conditional variance are respectively

\[
\mu_0(X_t, 0, \theta) = X_t, \quad \Sigma_0(X_t, 0, \theta) = 0.
\] (16)

The solution of the first of equations (14) gives immediately

\[
\mu_0(X_t, s, \theta) = \beta + (X_t - \beta) e^{-\lambda s},
\]

which in turn allows the second of equations (14) to be integrated to obtain

\[
\Sigma_0(X_t, s, \theta) = \sigma^2 \frac{(1 - e^{-\lambda s})}{2\lambda} \left[ \beta (1 - e^{-\lambda s}) + 2X_t e^{-\lambda s} \right].
\]

Thus the conditional mean and conditional variance of the square-root process evolving from observation \( X_t \) at time \( t \) are known exactly without requiring a detailed specification of the form of the true transitional probability density function.

### 3.2 Non-Affine Models

In the case of non-affine models, closed-form solutions for the moments of the true transitional density are not available and therefore QML parameter estimates will not be consistent. Even when the conditional mean of the process can be found in closed form, Proposition 1 indicates that consistency is only guaranteed provided the conditional covariance is also known in closed form. This will never happen unless the specification of the diffusion function \( G(X; \theta) \) is a quadratic form.

Recall that the conditional first and second moments of the process evolve according to equations (11). In the non-affine case the computation of the right hand side of these equations requires knowledge of \( f_0(X, s | X_t, \theta) \) because these integrals can no longer be expressed as combinations of the conditional first and second moments of the process. In this case the QML procedure introduces two sources of error. First, error necessarily occurs because the approximating transitional density, \( f_1(X, s | X_t, \theta) \), is a misspecification of the true transitional density. Second, additional error may arise if it becomes necessary to approximate numerically the integrals in equations (11).

Suppose that the solution of equations (11) requires the computation of

\[
\int_S h(X, s, \theta) f_1(X, s | X_t, \theta) dX,
\] (17)
where \( h(X, s, \theta) \) is a non-affine function of state. At each instant \( s \), let \( \Sigma_1(X_t, s, \theta) = L_1L_1^T \) be the Choleski decomposition of the conditional covariance matrix, then the transformation \( X = \mu_1(X_t, s, \theta) + L_1Z \), where \( Z \sim N(0_N, I_N) \), may be used to reorganize (17) into the form

\[
\frac{1}{(2\pi)^{N/2}} \int_S h(\mu_1(X_t, s, \theta) + L_1Z, s, \theta) \exp\left(-\frac{1}{2}Z^TZ\right) dZ.
\]

This integral may be expressed as an \( N \)-dimensional repeated integral in which the integration over each dimension may be represented generically in the form

\[
\int_{-\infty}^{\infty} \xi(z)e^{-z^2/2} dz,
\]

where \( \xi(z) \) depends on the current variable of integration, namely \( z \), and variables still to be integrated although the presence of the latter has been suppressed for representational convenience.

In the context of integrals with integrands of generic form \( e^{-z^2/2}\xi(z) \), the provable theorem in connection with Gauss-Hermite quadrature is that

\[
\int_{-\infty}^{\infty} e^{-z^2/2}\xi(z) dz = \sum_{k=0}^{n} a_k\xi(z_k) + \frac{\xi_{2n+2}(\eta)}{(2n+2)!} \int_{-\infty}^{\infty} e^{-z^2/2}H_{n+1}^2(z) dz, \tag{18}
\]

provided \( \xi \in C^{2n}(\mathbb{R}) \) and where \( H_{n+1}(z) \) is the (probabilists’) Hermite polynomial of order \( (n+1) \) with roots \( z_0, \ldots, z_n \). It is clear from equation (18) that the quadrature is exact for all polynomials of degree less than or equal to \( (2n+1) \), which means that this quadrature has maximum precision.

Bearing in mind that what is required is a parsimonious expression for the value of the integral (17), only Gauss-Hermite quadratures of low order are of interest. In particular, the Gauss-Hermite quadratures of precision three, five and seven are respectively

\[
\begin{align*}
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2}\xi(z) dz & \approx \frac{\xi(-1)}{2} + \frac{\xi(1)}{2}, \\
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2}\xi(z) dz & \approx \frac{\xi(-\sqrt{3}) + 4\xi(0) + \xi(\sqrt{3})}{6}, \\
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2}\xi(z) dz & \approx \frac{\beta^2\xi(-\alpha) + \alpha^2\xi(-\beta) + \alpha^2\xi(\beta) + \beta^2\xi(\alpha)}{12},
\end{align*}
\]

where \( \alpha = \sqrt{3} + \sqrt{6} \) and \( \beta = \sqrt{3} - \sqrt{6} \). The nodes of each quadrature in (19) are the respective zeros of \( H_2(z) = z^2 - 1 \), \( H_3(z) = z^3 - 3z \) and \( H_4(z) = z^4 - 6z^2 + 3 \).

**Example: The CKLS Model**

Consider now the generalization of the CIR model

\[
dX = \lambda(\beta - X) ds + \sigma X^{\gamma/2} dW, \tag{20}
\]
where \( G(X; \theta) = \sigma^2 x^\gamma \) and \( \gamma \) is the levels-effect parameter. Because the drift specification of the CKLS model remains an affine function of state, then the conditional mean of the process is precisely
\[
\mu_0(X_t, s, \theta) = \beta + (X_t - \beta)e^{-\lambda s}.
\]
On the other hand the conditional covariance of the CKLS process is the solution of the initial value problem
\[
\frac{d\Sigma_0}{ds} + 2\lambda \Sigma_0 = \sigma^2 \int_0^\infty X^\gamma f_0(X, s | X_t, \theta) dX, \quad \Sigma_0(0) = 0. \tag{21}
\]
The integral arising on the right hand side of this equation cannot be expressed in terms of known moments of the process and therefore further progress requires the approximation of \( f_0(X, s | X_t, \theta) \). When \( f_0 \) is approximated by the univariate Gaussian density function \( f_1(X, s | X_t, \theta) \) with mean value \( \mu_1 = \mu_0 \) and variance \( \Sigma_1 \) determined by solving the initial value problem
\[
\frac{d\Sigma_1}{ds} = -2\lambda \Sigma_1 + \frac{\sigma^2}{\sqrt{2\pi} \Sigma_1} \int_0^\infty X^\gamma \exp\left[-\frac{(X - \beta - (X_t - \beta)e^{-\lambda s})^2}{2\Sigma_1}\right] dX, \quad \Sigma_1(0) = 0. \tag{22}
\]
Under the change of variable of integration from \( X \) to \( Z \) where \( X = \beta + (X_t - \beta)e^{-\lambda s} + \sqrt{\Sigma_1} Z \), equation (22) becomes
\[
\frac{d\Sigma_1}{ds} = -2\lambda \Sigma_1 + \frac{\sigma^2}{\sqrt{2\pi} \Sigma_1} \int_{-\mu_1/\sqrt{\Sigma_1}}^{\infty} \left(\beta + (X_t - \beta)e^{-\lambda s} + Z \sqrt{\Sigma_1}\right)^\gamma e^{-Z^2/2} dZ, \tag{23}
\]
with initial condition \( \Sigma_1(0) = 0 \). The integral in this equation is estimated by a low-order Gaussian quadrature, noting in passing that because \( \Sigma_1(0) = 0 \), it is anticipated that the range of integration in (23) is in practice \( \mathbb{R} \) for small values of \( \Delta \).

To summarize, Gaussian quadratures of varying degrees of precision enable the second source of error identified previously to be eliminated\(^6\) for all practical purposes leaving the main source of error as the misspecification of the shape of the true transitional density. The extent of the error in the approximating density for non-affine models is taken up in the following section. An analysis of this approximation error is undertaken to explore the extent to which the approximating transitional density differs from the true transitional density.

### 4 Analysis of Approximation Error

The essence of the QMLE procedure is to approximate an unknown transitional density function with the multivariate Gaussian distribution satisfying the correct initial conditions and boundary

---

\(^6\) Another approach, not described in detail here, chooses an approximating probability density function with the property that the integrals arising from non-affine specifications can be computed in closed-form. For example, in the case of the CKLS model the offending integral can be computed exactly for both the gamma and lognormal approximating probability densities. Of course, parameter estimates of \( \theta_0 \) using this approach are not consistent.
conditions. The previous sections have demonstrated that this procedure yields consistent estimates if the drift and diffusion functions of the SDE are affine but that this consistency result does not generalize to non-affine specifications of drift and diffusion. The analysis of this section provides an upper bound for the growth of the mean square error of the approximation and applies to both affine and non-affine cases.

**Proposition 2**

Let \( f_0(X, s | X_t, \theta) \) denote the true transitional probability density function of the process \( X(t) \) satisfying equation (1). Let \( f_1(X, s | X_t, \theta) \) be an approximating probability density function satisfying the same boundary conditions as \( f_0 \), then the mean square error of the approximation, namely

\[
\text{MSE}(s) = \int_S (f_0 - f_1)^2 dX,
\]

satisfies the differential inequality

\[
\frac{d(\text{MSE})}{ds} \leq (C + \omega)\text{MSE} + \frac{1}{\omega} \int_S \left| \mathcal{F}(f_1) \right|^2 dX, \quad \text{MSE}(0) = \text{MSE}_0, \tag{24}
\]

where \( \omega = \omega(s) \) is an arbitrary positive function of time, \( \mathcal{F} \) denotes the \( N \)-dimensional Fokker-Planck operator defined by

\[
\mathcal{F}(\phi) = \frac{\partial \phi}{\partial s} - \text{div} \left( \frac{1}{2} \text{div} \left( G(X; \theta) \phi \right) - \alpha(X; \theta) \phi \right),
\]

and

\[
C = \max_{X \in S} \left| \text{div} \left( \frac{1}{2} \text{div} G(X; \theta) - \alpha(X; \theta) \right) \right|.
\]

**Proof** The proof of Proposition 2 is in the Appendix.

Proposition 2 establishes that there are two sources of error in approximating the mean square error of the transitional density function. The primary driver of error is the second term on the right hand side of inequality (24) which measures the extent to which the approximating density fails to satisfy the Fokker-Planck equation. The second source of error arises from the extent to which the approximating density differs initially from the true starting density. If the system is fully observed the initial condition is a delta function and both the approximating and true densities are identical initially and therefore the error is entirely driven by the misspecification in the density.

Taken at face value, Proposition 2 suggests that the MSE of the approximation grows exponentially. However, in this analysis \( \omega \) is an arbitrary positive function of \( s \) and the question of how best to use this freedom to sharpen inequality (24) is addressed in Proposition 3.

**Proposition 3**

Suppose that the non-negative function $\text{MSE}(s)$ satisfies the differential inequality
\[
\frac{d(\text{MSE})}{ds} \leq (C + \omega)\text{MSE}(s) + \frac{1}{\omega} \int_{S} |\mathcal{F}(f_1)|^2 dX, \quad \text{MSE}(0) = \text{MSE}_0,
\]
where $\omega = \omega(s)$ is an arbitrary positive function of time, then the root mean squared error in the approximation of the true transitional probability density function by the trial density satisfies
\[
\text{RMSE}(s) \leq \text{RMSE}_0 e^{Cs/2} + \int_{0}^{s} e^{C(s-u)/2} \left( \int_{S} |\mathcal{F}(f_1)|^2 dX \right)^{1/2} du.
\]

**Proof** The proof of Proposition 3 is in the Appendix.

Proposition 3 makes explicit the idea that if the approximating density satisfies the correct initial condition then the growth in the RMSE is driven entirely by the failure of the approximating density to satisfy the Fokker-Planck equation. If $f_1$ satisfies the Fokker-Planck equation, i.e. it is the true transitional density, then this inequality asserts that the $\text{RMSE}(s) \leq 0$. By construction, however, the $\text{RMSE}(s) \geq 0$ and therefore the $\text{RMSE}(s) \equiv 0$ in this case.

In terms of the efficacy of QML estimation, the most important consideration concerns the relationship between the moment conditions and the RMSE. Recall that the moment equations (11) are equivalent to the conditions
\[
\int_{S} X \mathcal{F}(f_1) dX = 0, \quad \int_{S} (X \otimes X) \mathcal{F}(f_1) dX = 0,
\]
which suggests that the quality of the fit of the approximating density $f_1$ to the true transitional density $f_0$ is related to unsatisfied moment conditions. This observation motivates Proposition 4.

**Proposition 4**
Suppose that the approximating density $f_1(X, s \mid X_t, \theta)$ satisfies exactly the first $m$ moment conditions with respect to the Fokker-Planck operator $\mathcal{F}$, that is, $K_1 = K_2 = \cdots = K_m = 0$ where
\[
K_j = \int_{S} X \otimes \cdots \otimes X \mathcal{F}(f_1) dX = 0,
\]
then the RMSE of the approximation may be expressed in terms of the $(m + 1)$ and higher order integrated moments of the approximating density $f_1$ with respect to the Fokker-Planck operator $\mathcal{F}$.

**Proof** The proof of Proposition 4 is in the Appendix.

Proposition 4 makes explicit how satisfying moment conditions for the approximating transitional density $f_1$ reduces the RMSE. In essence, the leading driver of error in the approximation arises from the first unsatisfied moment condition. In the case of the QML procedure for affine specifications of drift and diffusion, the approximating density matches exactly the first and second moments of
the true density and therefore the RMSE is driven by the extent to which the evolution of the third moment of the true density is misspecified by the approximating density. Whenever the initial conditions are properly specified for the approximating density, second and higher order moments are initially zero and therefore the RMSE of the QML procedure is small when \( \Delta \) is small.

5 Simulation Experiments

The performance of the QML estimation method for estimating the parameters of both affine and non-affine univariate diffusions and bivariate diffusions is now examined by means of simulation experiments.

5.1 Univariate

The CIR model

\[
dX = \lambda (\beta - X) \, ds + \sigma \sqrt{X} \, dW,
\]

and the CKLS model

\[
dX = \lambda (\beta - X) \, ds + \sigma X^{\gamma/2} \, dW,
\]

are used to generate 2000 samples of size 500 and 2000 respectively with true parameters \( \lambda = 0.20 \), \( \beta = 0.08 \), \( \sigma = 0.10 \) and, where necessary, \( \gamma = 0.75 \). The synthetic samples are generated with \( \Delta = 1/12 \) (so that the data may be interpreted as monthly data) using Milstein’s scheme with 1000 steps between observations to ensure accurate realizations of the process.

The bias and root mean square error (RMSE) of the estimates of the parameters of the CIR and CKLS processes over the 2000 samples are presented in Table 1. In the case of the CIR model for which an analytical expression for the transitional probability density function is available, exact maximum likelihood estimates are provided as a benchmark against which to the compare the QML estimates.\(^7\) For the CKLS model no closed-form expression for the transitional probability density function is available. In this instance the QML bias and RMSE of the QML estimates are reported for Gaussian quadratures of order 5 and 7.

Broadly speaking the simulation results conform to the well known patterns for estimating the parameters of univariate diffusions. For the CIR model, the speed of adjustment parameter, \( \lambda \), is by far the most difficult parameter to estimate. By comparison, the long-term mean, \( \beta \), and the volatility control parameter, \( \sigma \), are easier to estimate and are determined more accurately. For the CKLS model the pattern is similar with the levels effect parameter, \( \gamma \), relatively well resolved. The

\(^7\)All the computer code to implement the various methods was written in C and compiled with the Intel C++ Compiler.
Table 1: Bias and RMSE (in parentheses) of parameter estimates of the CIR model obtained by EML and QML. Also shown are the bias and RMSE of the CKLS model estimated by QML using Gaussian quadratures of order 5 and 7 respectively. Estimates are based on 2000 simulations for samples of size 500 and 2000 monthly observations ($\Delta = 1/12$).

<table>
<thead>
<tr>
<th></th>
<th>$T = 500$</th>
<th>$T = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda(0.20)$</td>
<td>$\beta(0.08)$</td>
</tr>
<tr>
<td>CIR (EML)</td>
<td>0.1016</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>(0.1318)</td>
<td>(0.0222)</td>
</tr>
<tr>
<td>CIR (QML)</td>
<td>0.1006</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>(0.1322)</td>
<td>(0.0226)</td>
</tr>
<tr>
<td>CKLS (O5)</td>
<td>0.0995</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>(0.1337)</td>
<td>(0.0368)</td>
</tr>
<tr>
<td>CKLS (O7)</td>
<td>0.0995</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>(0.1337)</td>
<td>(0.0125)</td>
</tr>
</tbody>
</table>

The accuracy of the QML procedure appears to be on a par with that of EML for these sample sizes and would stand comparison with any of the commonly-used methods for estimating the parameters of univariate stochastic differential equations (see for example, the surveys by Jensen and Poulsen (2002) and Hurn, Lindsay and Jeisman (2007).

### 5.2 Bivariate

Consider the bivariate Feller (1951) square root process $X = (X_1, X_2)$ satisfying the stochastic differential equations (SDEs)

$$dX = \begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 + k_{11}X_1 + k_{12}X_2 \\ \alpha_2 + k_{21}X_1 + k_{22}X_2 \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_1} \\ 0 \end{bmatrix} dW_1 + \begin{bmatrix} 0 \\ \sqrt{X_2} \end{bmatrix} dW_2,$$

(25)

where $dW_1$ and $dW_2$ are independent increments in the Wiener processes $W_1$ and $W_2$ and $\alpha_1$, $\alpha_2$, $k_{11}$, $k_{12}$, $k_{21}$ and $k_{22}$ are constant parameters. Let $\mu_1 = \mathbb{E}[X_1]$ and $\mu_2 = \mathbb{E}[X_2]$ where each expectation is taken with respect to the transitional density $f_0(X, s | X_t, \theta)$, and let $\sigma_{11}$, $\sigma_{12}$ and $\sigma_{22}$ denote the elements of the associated covariance matrix of $X$. For the model specified by equations (25), it is straightforward to show that the moment equations for the transitional density take the form

$$\frac{dM}{dt} = AM + B,$$

(26)
In particular, the stationary solution of equation (26) is straightforward to show that these properties are achieved with the restrictions \( k_1 < 0, k_2 < 0, k_{12} > 0, k_{11}k_2 - k_{12}k_2 > 0 \), and that in this case \( A \) has five real (and negative) eigenvalues. Although it is possible to construct a closed form solution to equation (26), these equations are most simply solved by numerical integration.

The nature of the solution of equations (26) depends on the properties of the eigenvalues of \( A \). Specifically, \( A \) has five real eigenvalues if \((k_{11} - k_{22})^2 + 4k_{12}k_2 > 0\), otherwise it has one real eigenvalue and two complex conjugate pairs of eigenvalues giving solutions with oscillatory behaviour. In particular, the stationary solution of equation (26) is \( \hat{M} = -A^{-1}B \) with components

\[
\begin{bmatrix}
\hat{\mu}_1 \\
\hat{\mu}_2
\end{bmatrix} = \begin{bmatrix}
k_{12}\alpha_2 - k_{22}\alpha_1 \\
k_{21}\alpha_1 - k_{11}\alpha_2
\end{bmatrix} \beta,
\begin{bmatrix}
\hat{\sigma}_{11} \\
\hat{\sigma}_{12} \\
\hat{\sigma}_{22}
\end{bmatrix} = \begin{bmatrix}
\frac{(\beta + k_{22}^2)\hat{\mu}_1 + k_{12}^2\hat{\mu}_2}{2(k_{11} + k_{22})} \\
\frac{k_{21}k_{22}\hat{\mu}_1 + k_{12}k_{11}\hat{\mu}_2}{2(k_{11} + k_{22})} \\
\frac{k_{21}\hat{\mu}_1 + (\beta + k_{11}^2)\hat{\mu}_1}{2(k_{11} + k_{22})}
\end{bmatrix}
\]

where \( \beta = k_{11}k_2 - k_{12}k_2 \). In a practical application based on equation (26), the parameters in (27) must be configured to ensure that \( X_1 \) and \( X_2 \) are restored to positive equilibrium states \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) and that the covariance matrix with entries \( \hat{\sigma}_{11}, \hat{\sigma}_{12} \) and \( \hat{\sigma}_{22} \) is positive definite. It is straightforward to show that these properties are achieved with the restrictions \( \alpha_1 > 0, \alpha_2 > 0, k_{11} < 0, k_{22} < 0, k_{12} > 0, k_{21} > 0 \) and \( k_{11}k_2 - k_{12}k_2 > 0 \), and that in this case \( A \) has five real (and negative) eigenvalues. Although it is possible to construct a closed form solution to equation (26), these equations are most simply solved by numerical integration.

Song (2009) provides parameter estimates for the bivariate Feller process (25) with parameters \((\alpha_1, \alpha_2, \kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}) = (0.56, 0.64, -0.7, 0.3, 0.4, -0.8)\) using exact maximum likelihood (EML) based on 1000 realisations in samples of 500 transitions and 1000 transitions. Table 2 compares bias and RMSE in the estimates of the parameters obtained by EML and QML.

The performance of QML and EML in estimating the parameters of the bivariate Feller process is broadly similar. For the parameters \( k_{11} \) and \( k_{22} \) the QML procedure delivers estimates with smaller bias and smaller RMSE in samples of 500 and 1000 transitions than EML. For the parameter \( k_{21} \) the situation is reversed. In the case of the parameter \( \alpha_1 \), QML delivers estimates with larger bias than EML but smaller RMSE in samples of 500 and 1000 transitions. For the parameter \( \alpha_2 \) the QML procedure delivers estimates with smaller RMSE than EML but the performance in bias is mixed whereas for the final parameter \( k_{12} \) the situation is exactly reversed; QML now generates estimates with smaller bias than EML but the situation regarding RMSE is unclear.
What this simulation evidence suggests therefore is QML performs with credit and does not suffer by comparison with EML as a method for parameter estimation in this bivariate setting for this particular model specification.

6 Unobservable State Variables

Unlike the one-dimensional problem where the state variable must be observed, in multi-dimensions the state variables naturally sub-divide into observable variables and latent variables. Probably the most important example in the financial econometrics literature is the stochastic volatility model, where financial returns are observed but the volatility of returns is unobserved. Consider an $N$-dimensional problem in which $K$ variables are observed and $(N - K)$ variables are latent. Let $V$ denote the observed variables and let $U$ denote the unobserved variables so that $X = [V^T, U^T]^T$ is the decomposition of the $N$-dimensional process governed by equation (1) into its observed and unobserved components.

The problem is now to estimate the parameters $\theta$ of equation (1) from $(T + 1)$ incomplete observations $V_0, \cdots, V_T$ of $X$ at the discrete times $(t_0, \cdots, t_T)$ where $V_t$ ($0 \leq t \leq T$) is the vector of dimension $K$ ($< N$) formed from the first $K$ entries of $X$ at time $t$.

The aim of the maximum likelihood procedure is to choose the value of $\theta$ so as to maximize the

<table>
<thead>
<tr>
<th>Parm</th>
<th>True Value</th>
<th>500 Transitions</th>
<th>1000 Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{11}$</td>
<td>(-0.70)</td>
<td>-0.119 (0.266)</td>
<td>-0.053 (0.170)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.342 (0.699)</td>
<td>-0.144 (0.481)</td>
</tr>
<tr>
<td>$k_{12}$</td>
<td>(0.30)</td>
<td>0.023 (0.199)</td>
<td>0.010 (0.132)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.077 (0.512)</td>
<td>0.027 (0.116)</td>
</tr>
<tr>
<td>$k_{21}$</td>
<td>(0.40)</td>
<td>0.017 (0.213)</td>
<td>0.008 (0.141)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001 (0.151)</td>
<td>0.001 (0.054)</td>
</tr>
<tr>
<td>$k_{22}$</td>
<td>(-0.80)</td>
<td>-0.110 (0.252)</td>
<td>-0.046 (0.155)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.310 (0.533)</td>
<td>-0.108 (0.350)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>(0.56)</td>
<td>0.102 (0.304)</td>
<td>0.041 (0.141)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.063 (0.316)</td>
<td>0.025 (0.193)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>(0.64)</td>
<td>0.111 (0.319)</td>
<td>0.048 (0.155)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.115 (0.391)</td>
<td>0.022 (0.188)</td>
</tr>
</tbody>
</table>

Table 2: Bias and RMSE (in parentheses) of parameter estimates of the Feller model (25) using QML are compared with those supplied by Song (2009) using EML. Estimates are based on 1000 simulations for samples of size 500 and 1000 monthly observations ($\Delta = 1/12$).
likelihood of observing $V_0, \cdots, V_T$ at the discrete times $(t_0, \cdots, t_T)$. Specifically, the conditional log-likelihood function of the observed sample, namely

$$\log L_T(\theta) = \sum_{t=1}^{T} \log f_1^{(m)}(V_t, \Delta | \bar{V}_{t-1}, \theta),$$  \hspace{1cm} (28)

is maximized (or typically the negative log-likelihood function is minimized) with respect to the parameters $\theta$, where $f_1^{(m)}(V_t, \Delta | \bar{V}_{t-1}, \theta)$ is the marginal density obtained by integrating the $(N-K)$ latent variables $U$ out of $f_1(X, \Delta | \bar{V}_{t-1}, \theta)$, and where $\bar{V}_{t-1} \equiv \{V_{t-1}, \cdots, V_0\}$ is the history of the observed variables up to and including that at time $t - 1$.\(^8\) Thus the computation of expression (28) requires a method for the construction of marginal transitional densities.

### 6.1 Marginal transitional density

Let the conditional covariance matrix of the $N$-dimensional process $X = [V^T, U^T]^T$, namely $\Sigma_1(\bar{V}_{t-1}, s, \theta)$, have block form

$$\Sigma_1 = \begin{bmatrix} L_V & 0 \\ P & L_U \end{bmatrix} \begin{bmatrix} L_V & 0 \\ P & L_U \end{bmatrix}^T = \begin{bmatrix} L_V L_V^T & L_V P^T \\ P L_V^T & PP^T + L_U L_U^T \end{bmatrix},$$  \hspace{1cm} (29)

where $L_V$ is a $K \times K$ lower triangular matrix, $P$ is an $(N-K) \times K$ full matrix and $L_U$ is an $(N-K) \times (N-K)$ lower triangular matrix. When expressed within this framework, the transitional density function $f_1(X, s | \bar{V}_{t-1}, \theta)$ in expression (6) takes the form

$$\frac{1}{(2\pi)^{N/2}} \frac{1}{|L_V||L_U|} \exp \left[ -\frac{1}{2}(V - \mu_V)^T L_V^{-T} L_V^{-1}(V - \mu_V) - \frac{1}{2}(Z - \mu_Z)^T L_U^{-T} L_U^{-1}(Z - \mu_Z) \right]$$  \hspace{1cm} (30)

where $Z = (U - P L_V^{-1}V)$. The functions $\mu_U = \mu_U(\bar{V}_{t-1}, s, \theta)$, $\mu_V = \mu_V(\bar{V}_{t-1}, s, \theta)$ and $\mu_Z = \mu_Z(\bar{V}_{t-1}, s, \theta)$ denote the respective conditional mean values of $U$, $V$ and $Z$ at time $s$, and thus $\mu_Z = (\mu_U - P L_V^{-1}\mu_V)$. It is transparent from expression (30) that the marginal density of $V$ at time $s$ is formed by integrating out the latent variables $U$ to obtain

$$f_1^{(m)}(V, s | \bar{V}_{t-1}, \theta) = \frac{1}{(2\pi)^{K/2}} \frac{1}{|L_V|} \exp \left[ -\frac{1}{2}(V - \mu_V)^T L_V^{-T} L_V^{-1}(V - \mu_V) \right],$$  \hspace{1cm} (31)

that is, it is the multivariate Gaussian probability density function with covariance matrix defined by the principal $K \times K$ minor of the covariance matrix $\Sigma_1$. Of course, this is an obvious result from a conceptual point of view. What is not obvious, however, is how to construct the conditional density of the latent variables once an observation, say $V_t$, becomes available after time $s = \Delta$ has passed.

---

\(^8\)The reason for the use of $\bar{V}_{t-1}$ as the conditioning information set in (28), as opposed to $V_{t-1}$ in isolation, is that in general $V_t$ will not inherit the Markov property from $X_t$. This is true of stock prices when volatility is an unobserved, persistent stochastic process.
6.2 Conditional transitional density

Given that \( X = [V^T, U^T]^T \) is distributed with probability density function \( f_1([V^T, U^T]^T, s | \vec{V}_{t-1}, \theta) \) and that \( V = V_t \) at time \( s = \Delta \), then the density of \( U \) conditioned on \( \vec{V}_t \) is, by definition,

\[
f_1^{(c)}(U, 0 | \vec{V}_t, \theta) = \frac{f_1([V^T, U^T]^T, \Delta | \vec{V}_{t-1}, \theta)}{f_1^{(m)}(V_t, \Delta | \vec{V}_{t-1}, \theta)}.
\]

The joint density of \( X = [V^T, U^T]^T \) is evaluated directly from equation (30), and the marginal density of \( V \) at \( V = V_t \) is evaluated from equation (31). Consequently, \( f_1^{(c)}(U, 0 | \vec{V}_t, \theta) \) is the ratio of two exponential functions, so that

\[
f_1^{(c)}(U, 0 | \vec{V}_t, \theta) = \frac{1}{(2\pi)^{(N-K)/2} |L_U|} \exp \left[ -\frac{1}{2} (Z - \mu_Z)^T L_U^{-1} L_U^{-1} (Z - \mu_Z) \right],
\]

where specifically

\[
Z = (U - PL_{V}^{-1}V_t), \quad \mu_Z = (\mu_U(\vec{V}_{t-1}, \Delta, \theta) - PL_{V}^{-1} \mu_V(\vec{V}_{t-1}, \Delta, \theta)).
\]

In conclusion, after observing \( V = V_t \) at \( t \) during the implementation of a maximum likelihood estimation procedure in the presence of latent variables, the initial conditions for equations (14) pertaining to the next transition are

\[
\begin{bmatrix}
\mu_U(\vec{V}_t, 0, \theta) \\
\mu_U(\vec{V}_t, 0, \theta)
\end{bmatrix}
= \begin{bmatrix}
V_t \\
\mu_Z
\end{bmatrix}, \\
\Sigma_1(\vec{V}_t, 0, \theta) = \begin{bmatrix}
0 & 0 \\
0 & L_UL_U^T
\end{bmatrix},
\]

where \( \mu_Z \) is given in (34).

7 Application: The Heston Model of Stochastic Volatility

The procedures described in this paper are now applied to the prototypical stochastic volatility model proposed by Heston (1993). A number of simple simulation experiments are presented that are based on those reported by A"it-Sahalia and Kimmel (2007), hereafter ASK. The model is then estimated using a similar dataset to that used by ASK.

7.1 The Heston Model

The stochastic volatility model of Heston (1993) posits that under the physical probability measure \( \mathbb{P} \), the stock price, \( S \), and variance, \( H \), evolve according to the stochastic differential equations,

\[
\begin{align*}
\frac{dS}{S} &= (r - \xi + \lambda(1 - \rho^2)H) dt + S \sqrt{H} (\sqrt{1 - \rho^2} dW_1 + \rho dW_2), \\
\frac{dH}{H} &= \kappa(\gamma - H) dt + \sigma \sqrt{H} dW_2,
\end{align*}
\]

(36)
where \( r \) is the risk-free rate, \( \xi \) is the dividend-price ratio, \( \lambda \) scales the equity premium, \( \kappa \) is the rate of mean reversion of volatility, \( \gamma \) is the long-run level of volatility, \( \sigma \) is the diffusion of volatility, \( \rho \) is the local correlation between returns and volatility and \( W = (W_1, W_2) \) is a two-dimensional orthogonal Brownian motion. With regard to the specification of the equity premium, we follow AKS who in the analogous situation propose a price of \( \lambda \sqrt{H} \sqrt{1 - \rho^2} \) per unit of \( W_1 \) risk, and leave \( W_2 \) risk unpriced. The choice of the latter owes to the fact that any additional parameters involved with the price of \( W_2 \) risk would be identifiable only with the use of observed option prices, and such a situation is not considered in this analysis.

When expressed in terms of returns, \( Y = \log S \), equations (36) have matrix representation

\[
\begin{bmatrix}
\frac{dY}{dH} \\
\end{bmatrix} = \begin{bmatrix}
\frac{dW_1}{dW_2}
\end{bmatrix},
\]

where \( \beta = \lambda (1 - \rho^2) - 1/2 \). The model for the process \( [Y, H]^T \) therefore has respective drift and covariance specifications

\[
\alpha(Y, H; \theta) = \begin{bmatrix}
\mu_Y(Y, H; \theta) \\
\mu_H(Y, H; \theta)
\end{bmatrix}, \quad \Gamma(Y, H; \theta) = \begin{bmatrix}
\sigma^2_H \\
\sigma_H \sigma_H^T
\end{bmatrix},
\]

Because these are affine functions of volatility, equations (14) provide exact solutions for the conditional mean and variance of the Heston diffusion. Specifically, the conditional mean at time \( t+s \) is

\[
\begin{bmatrix}
\mu_Y(s) \\
\mu_H(s)
\end{bmatrix} = \begin{bmatrix}
\mu_Y(0) + (r - \xi + \beta \gamma) \frac{\beta(\mu_H(0) - \gamma)}{\kappa} \left( 1 - e^{-\kappa s} \right) \\
\gamma + (\mu_H(0) - \gamma) e^{-\kappa s}
\end{bmatrix},
\]

where the initial condition \( \mu_Y(0) = Y_t \) as price is observable, where \( \mu_H(0) = H_t \) when variance is treated as observed, and where \( \mu_H(0) = \mathbb{E}[H_t | \mathcal{Y}_t] \) when variance is treated as unobserved, which is computed via the first of equations (35). The distinct entries of the conditional variance-covariance matrix \( \Sigma(s) \) at time \( t+s \) satisfy the differential equations

\[
\frac{d\Sigma_{YY}}{ds} = -\Sigma_{12} + \mu_H, \quad \frac{d\Sigma_{YH}}{ds} = -\kappa \Sigma_{12} - \frac{\rho \sigma}{2} \Sigma_{HH} + \rho \sigma \mu_H, \quad \frac{d\Sigma_{HH}}{ds} = -2 \kappa \Sigma_{HH} + \sigma^2 \mu_H.
\]

When price is observed but volatility is unobserved equations (40) are solved with the initial conditions \( \Sigma_{YY}(0) = \Sigma_{YH}(0) = 0 \) and \( \Sigma_{HH}(0) = \text{Var}[H_t | \mathcal{Y}_t] \), where this conditional variance is computed via the second of equations (35). The exact solution of (40) corresponding to these initial
If the state is fully observed, \( i.e. \) volatility is observed in addition to price, then solution (41) is likewise valid with \( \Sigma_{HH}(0) = 0 \). The QMLE procedure approximates the conditional distribution of \([Y_{t+1}, H_{t+1}]\) by the bivariate Gaussian density with mean value and covariance given respectively by

\[
\left[
\begin{array}{c}
\mu_Y(\Delta) \\
\mu_H(\Delta)
\end{array}
\right], \quad \left[
\begin{array}{cc}
\Sigma_{YY}(\Delta) & \Sigma_{YH}(\Delta) \\
\Sigma_{HY}(\Delta) & \Sigma_{HH}(\Delta)
\end{array}
\right].
\]

(42)

\[ If the state is fully observed, \( i.e. \) volatility is observed in addition to price, then solution (41) is likewise valid with \( \Sigma_{HH}(0) = 0 \). The QMLE procedure approximates the conditional distribution of \([Y_{t+1}, H_{t+1}]\) by the bivariate Gaussian density with mean value and covariance given respectively by

\[
\left[
\begin{array}{c}
\mu_Y(\Delta) \\
\mu_H(\Delta)
\end{array}
\right], \quad \left[
\begin{array}{cc}
\Sigma_{YY}(\Delta) & \Sigma_{YH}(\Delta) \\
\Sigma_{HY}(\Delta) & \Sigma_{HH}(\Delta)
\end{array}
\right].
\]

(42)

\[7.2 \text{ Simulation Results}\]

In order to examine the efficacy of QML for estimating the parameters of Heston’s model, data are simulated with \( \kappa = 3.00, \sigma = 0.25, \rho = -0.80, \gamma = 0.10 \) and \( \lambda = 4.00 \). The value of \((r - \xi)\), which represents the difference between the instantaneous risk-free interest rate and the instantaneous dividend yield of the stock, is fixed at 3%. To facilitate comparison, these parameters are similar to those used by ASK. The unconditional equity risk-premium underlying this model is \( \xi + \lambda(1 - \rho^2)\gamma = (\xi + 0.144) \) per annum.

The unconditioned results presented in Table 3 refer to a situation in which the simulated volatility is simply taken as observed and used in the parameter estimation. The conditioned results treat the volatility as unobserved and parameter estimation requires that the likelihood function be constructed from the marginal density of the observed stock price as described in Section 6. In this simulation experiment, it has been assumed that \( Y_0 \) is a fixed value and is uninformative with respect to \( H_0 \). Therefore, the conditional distribution of the initial value \( H_0 \) is approximated by a Gaussian density with mean, \( \gamma \), and variance \( \gamma\sigma^2/2\kappa \).

The results for the estimation of the unconditioned model are almost identical to those reported by ASK using their closed-form likelihood approximation. In the case where the state is completely observable, therefore, the degradation of the quality of parameter estimates due to using the Gaus-
The conditioned model uses precisely half of the information of the unconditioned model and a deterioration in performance is therefore inevitable. The mean point estimates of the model, however, hold up remarkably well. The loss of information is evident in the size of the standard errors which are significantly larger than those of the unconditional case.

An interesting exception to this general conclusion is the price of equity risk parameter, \( \lambda \), which in the 2000 transition conditioned case has a slightly smaller standard error than the corresponding unconditioned estimate. A tentative explanation for this result is that the sensitivity of the log-likelihood function to the parameter \( \lambda \) is increased in the absence of observable variance. From equation (39), it can be seen that the parameter \( \lambda \) (embedded linearly in \( \beta \)) is fundamental in scaling the mean volatility, \( \gamma \), which in turn governs the expected one-step ahead logarithm of the price. In the absence of observable variance the marginal distribution for the logarithm of price used in the construction of the log-likelihood function is driven by the difference between the observed log-price and its predicted value alone and essentially behaves as a quadratic function of \( \lambda \). When volatility is observed this key difference is but a component of a multivariate log-likelihood. Note that these comments apply mainly in a simulation environment where the model is known to be correct and there are a large number of data. In practice, however, the true model is never known and, as noted by ASK, the parameter \( \lambda \) is always likely to be problematic to estimate.

### Table 3: Bias and standard deviations (in parentheses) of parameter estimates for the Heston model with volatility treated as an observed variable using the QML procedure. The column headed ASK reproduces the results reported by Ait-Sahalia and Kimmel (2007). Results are reported for sample sizes of 500, 5000 and 10000 simulated daily observations (\( \Delta = 1/252 \)), with 1000 Monte Carlo replications between observations.

<table>
<thead>
<tr>
<th>Parm</th>
<th>500 Transitions</th>
<th>5000 Transitions</th>
<th>10000 Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASK</td>
<td>QML</td>
<td>ASK</td>
</tr>
<tr>
<td>( \kappa = 3.00 )</td>
<td>0.860 (1.570)</td>
<td>0.881 (1.542)</td>
<td>0.068 (0.380)</td>
</tr>
<tr>
<td>( \gamma = 0.10 )</td>
<td>0.001 (0.022)</td>
<td>-0.002 (0.016)</td>
<td>0.000 (0.006)</td>
</tr>
<tr>
<td>( \sigma = 0.25 )</td>
<td>0.000 (0.006)</td>
<td>0.000 (0.006)</td>
<td>0.000 (0.002)</td>
</tr>
<tr>
<td>( \rho = -0.8 )</td>
<td>-0.000 (0.013)</td>
<td>-0.000 (0.013)</td>
<td>-0.000 (0.004)</td>
</tr>
<tr>
<td>( \lambda = 4.00 )</td>
<td>0.920 (6.500)</td>
<td>1.696 (5.591)</td>
<td>0.070 (1.900)</td>
</tr>
</tbody>
</table>
Table 4: Bias and standard deviations (in parentheses) of parameter estimates for the Heston model using the QML procedure. The column headed QML* treats volatility as an unobserved variable and bases estimation on the marginal likelihood of log-prices. The column headed QML reproduces the results reported in Table 3 for comparative purposes. Results are reported for sample sizes of 2000, 5000 and 10000 simulated daily observations ($\Delta = 1/252$), with 1000 Monte Carlo replications between observations.

### 7.3 S&P500 Data

The Heston model is now estimated using data consisting of daily data on the S&P 500 Index, the Implied Volatility Index (VIX) and a risk-free interest rate in the form of the 1-month Treasury Bill return from Ibbotson and Associates, Inc.\footnote{The actual series used here is the risk-free market factor used by Ken French, downloadable from his website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/} Daily cash dividends of the S&P 500 are ignored in the estimation. The sample period is 2 January 1990 to 31 December 2008, which yield a total of 4971 observations. The S&P 500 Index and the VIX are plotted in Figure 1.

The VIX is constructed from European put and call option prices such that at any given time it represents the risk-neutral expectation of integrated variance averaged over the next 30 calendar days (or 22 trading days). Following the suggestion of ASK, the performance of a pair of proxies for unobserved volatility is explored, namely the direct proxy (the VIX itself) and what ASK call an integrated volatility proxy that is constructed by “unwinding” the averaging present in the VIX.\footnote{There are a number of other possible choices for proxies, including realised volatility measures (Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2001) or Black-Scholes implied volatilities.}

This integrated volatility proxy, $H_t$, is constructed from the VIX via the formula

$$H_t = \gamma + (\text{VIX}_t - \gamma) \frac{\kappa \Delta t}{1 - e^{-\kappa \Delta t}}. \quad (43)$$

<table>
<thead>
<tr>
<th>Parm</th>
<th>2000 Transitions</th>
<th>5000 Transitions</th>
<th>10000 Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QML*</td>
<td>QML</td>
<td>QML*</td>
</tr>
<tr>
<td>$\kappa = 3.00$</td>
<td>0.860 (1.570)</td>
<td>0.881 (1.542)</td>
<td>−0.126 (0.823)</td>
</tr>
<tr>
<td>$\gamma = 0.10$</td>
<td>0.001 (0.022)</td>
<td>−0.002 (0.016)</td>
<td>0.000 (0.007)</td>
</tr>
<tr>
<td>$\sigma = 0.25$</td>
<td>0.000 (0.006)</td>
<td>0.000 (0.006)</td>
<td>−0.002 (0.148)</td>
</tr>
<tr>
<td>$\rho = -0.8$</td>
<td>0.000 (0.013)</td>
<td>−0.000 (0.013)</td>
<td>0.020 (0.144)</td>
</tr>
<tr>
<td>$\lambda = 4.00$</td>
<td>0.920 (6.500)</td>
<td>1.696 (5.591)</td>
<td>−0.019 (1.729)</td>
</tr>
</tbody>
</table>
Figure 1: S&P 500 Index and the VIX Implied Volatility Index (annualized) for the period 2 January 1990 to 31 December 2008.

Note that implicit in this result is the assumption that the price of volatility risk is zero, from which it follows that the risk-neutral drift coefficients coincide with the physical drift coefficients. In the absence of such an assumption, it would be necessary to estimate the volatility risk-premium. It may be argued that this is not a particularly controversial assumption as estimating the volatility premium is notoriously difficult, and the existing literature presents conflicting results. For example, Pan (2002) finds that estimates of this parameter imply that the risk-neutral variance process is explosive, while Broadie, Chernov and Johannes (2007) find that the value of the parameter is not significantly different from zero.

QML estimation of the parameters of the Heston model may now be performed on the observable series of \( \{Y_t, VIX_t\}_{t=0}^T \). If the VIX itself is used for volatility the procedure is applied without any modification. If the integrated volatility proxy is used, the approximating likelihood function for the joint series \( \{Y_t, VIX_t\}_{t=0}^T \) is

\[
\log L_T(\theta) = \sum_{t=1}^T \log \left( f([Y_t, H_t]^T, \Delta | [Y_{t-1}, H_{t-1}]^T, \theta) \right) + \log \left| \frac{\partial H_t}{\partial VIX_t} \right| + \log \left( f(H_0; \theta) \right),
\]

where \( H_t \) is constructed from \( VIX_t \) by formula (43) and the Jacobian factor is given by

\[
\left| \frac{\partial H_t}{\partial VIX_t} \right| = \frac{\kappa \Delta t}{1 - e^{-\kappa \Delta t}}.
\]

23
The details of this change of variable method are given in ASK.

Table 5 reports the QML parameter estimates and QML standard errors based on the “sandwich” estimator for the Heston model using the direct volatility proxy. The model is estimated over two samples, one corresponding to that of ASK and the other using the entire span of the data. The models are also estimated for two different specifications of the drift term, \((r - \xi)\), namely one which uses the risk-free rate only, and one in which the term is fixed at 3%.

In general, the pattern of estimates for the period 1990-2003 is broadly very similar to that reported by ASK despite the fact that the two datasets are not identical, and the parameter estimates are basically unchanged for the alternative specification of drift. The correlation parameter \(\rho\) is found to be strongly negative \((-0.76)\), the volatility of volatility is almost identical \((0.48)\) and the long-term value of volatility returned by the two different estimation methods are comparable, being 21% per annum for the polynomial expansion and 20% for the QML method. The speed of mean reversion to this long-term level of volatility is faster for the polynomial expansion approach by comparison with the QML approach and the standard error is also larger for QML. For both estimation methods, there is substantial uncertainty in estimates of \(\lambda\). As argued by ASK, this is not surprising, given the relatively short sample period of the data. Overall, it appears that despite its simplicity, it is the contention here that the QML method provides a realistic alternative to estimating the parameters of stochastic volatility models when an implied volatility proxy is available.

The most noticeable difference when estimating the model over the extended data period is the significant fall in the estimated value of \(\kappa\). It is apparent from Figure 1 that the effect of the global financial crisis of 2007/8 had a marked effect on volatility. Indeed, it is doubtful that a model such as the Heston model presented here could have generated an increase in volatility of this magnitude. Consequently, the only way in which the model can compensate is by reducing the strength of reversion to the mean volatility. In all cases, the QML standard errors are comparable to those reported by ASK, given that slightly higher standard errors are to be expected.

\[\text{Data and C code for the results reported in this section are available at } \text{http://www.ncer.edu.au/data.}\]
Table 5: Estimated parameter values for the Heston stochastic volatility model using the S&P 500 Index and the VIX as a proxy for unobserved volatility.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Drift = Risk-free rate</th>
<th>Drift = 3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>κ</td>
<td>3.795</td>
<td>2.510</td>
</tr>
<tr>
<td></td>
<td>(1.133)</td>
<td>(1.826)</td>
</tr>
<tr>
<td>γ</td>
<td>0.039</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>σ</td>
<td>0.483</td>
<td>0.538</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>ρ</td>
<td>−0.765</td>
<td>−0.754</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>λ</td>
<td>6.403</td>
<td>4.946</td>
</tr>
<tr>
<td></td>
<td>(2.346)</td>
<td>(2.096)</td>
</tr>
</tbody>
</table>

Table 6: Estimated parameter values for the Heston stochastic volatility model using the S&P 500 Index and an integrated volatility proxy based on the VIX as described by Aït-Sahalia and Kimmel (2007).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Drift = Risk-free rate</th>
<th>Drift = 3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>κ</td>
<td>3.898</td>
<td>2.997</td>
</tr>
<tr>
<td></td>
<td>(1.138)</td>
<td>(1.821)</td>
</tr>
<tr>
<td>γ</td>
<td>0.046</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>σ</td>
<td>0.486</td>
<td>0.541</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>ρ</td>
<td>−0.764</td>
<td>−0.754</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>λ</td>
<td>4.263</td>
<td>2.851</td>
</tr>
<tr>
<td></td>
<td>(2.379)</td>
<td>(2.117)</td>
</tr>
</tbody>
</table>

The estimation of the Heston model is then repeated using the integrated volatility proxy in equation (43). Once again the fall in the value of κ is noticeable. Interestingly enough the estimated value of λ also falls quite dramatically. As the standard error on this parameter is of the order of the
parameter itself, it is not clear what significance should be attached to this change.

Finally, the model was estimated treating volatility as unobserved. This is clearly asking a great deal from the data as the full set has only 4971 observations and the results for the 5000 transition case in the simulation exercise of the model indicate that $\kappa$ and $\lambda$ are likely to be badly determined. The model was estimated over the full sample using the risk-free rate only for the drift term $(r - \xi)$. The parameter estimates recovered are:

\[
\begin{array}{cccccc}
\kappa & \gamma & \sigma & \rho & \lambda \\
6.309 & 0.027 & 0.356 & -0.533 & 3.241 \\
(1.400) & (0.002) & (0.093) & (0.150) & (2.020)
\end{array}
\]

Not surprisingly these results show some differences to the values obtained using proxy information. In particular, $\rho = -0.53$ is much lower than found previously and the estimate of long-term volatility, 16.5%, is below the unconditional mean value of the VIX of 21% over the period. Once again, the risk premium parameter is poorly determined. Despite these obvious differences in parameter estimates, given that this approach treats volatility as a truly unobservable state and therefore uses only half the data, a strong argument can be made that the results are credible. Even more pleasing is the fact that the QML standard errors for $\kappa$ and $\lambda$ are of the order of magnitude suggested by the simulation experiment for 5000 transitions. The conditioning approach, therefore, appears to offer a useful way of estimating the parameters of the Heston model in the absence of a suitable volatility proxy.

8 Concluding remarks

This paper has developed a quasi-maximum likelihood (QML) approach to estimating the parameters of stochastic differential equations. The essence of the procedure is to use a Gaussian density to approximate the unknown transitional density of the process and then to compute as accurately as possible the true mean and true variance of the unobserved transitional density function to pass to the approximating Gaussian distribution. If the stochastic differential equations have affine drift and diffusion functions, then the true mean and variance of the transitional density function may be obtained exactly and the resultant QML parameter estimates based on the Gaussian approximation are consistent. Although this result does not generalize to non-affine models, it is shown that the approximation error is determined by the lowest moment condition that is not satisfied. The main advantage of using QML is shown to be for multivariate diffusions where state variables are unobserved. It is demonstrated in a simulation experiment that the parameters of a classic stochastic volatility model may be easily estimated without much loss of efficiency when a volatility proxy is available. If volatility is treated as unobservable and only log-prices are used in the estimation,
then the conditioning method suggested here provides credible parameter estimates.
References


Appendix: Proofs

Proof of Proposition 1

Let \( f_1(X, \Delta \mid X_t, \theta) \) be a trial transitional probability density describing transitions of duration \( \Delta \) from state \( X_t \) to state \( X \) when the model parameters are \( \theta \), then the score function associated with the observations \( X_0, \cdots, X_T \) is

\[
S_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \log f_1(X_t, \Delta \mid X_{t-1}, \theta)}{\partial \theta}.
\]

(44)

When the approximating density \( f_1 \) is taken to be the multivariate Gaussian density

\[
f_1(X, \Delta \mid X_{t-1}, \theta) = \frac{1}{(2\pi)^{N/2} |\Sigma_1|^{1/2}} \exp \left[ -\frac{1}{2} (X - \mu_1)^T \Sigma_1^{-1} (X - \mu_1) \right],
\]

the contribution to the score function from the transition \( X_{t-1} \) to \( X_t \) is

\[
-\frac{1}{2} \frac{\partial |\Sigma_1(X_{t-1}, \Delta, \theta)|}{\partial \theta} + (X_t - \mu_1(X_{t-1}, \Delta, \theta))^T \Sigma_1(X_{t-1}, \Delta, \theta)^{-1} \frac{\partial \mu_1(X_{t-1}, \Delta, \theta)}{\partial \theta}
\]

\[
-\frac{1}{2} (X_t - \mu_1(X_{t-1}, \Delta, \theta))^T \frac{\partial \Sigma_1(X_{t-1}, \Delta, \theta)^{-1}}{\partial \theta} (X_t - \mu_1(X_{t-1}, \Delta, \theta)),
\]

(45)

(46)

where \( \mu_1 = \mu_1(X_{t-1}, \Delta, \theta) \) and \( \Sigma_1 = \Sigma_1(X_{t-1}, \Delta, \theta) \) are respectively the conditional mean and conditional covariance of transitions of duration \( \Delta \) from \( X_{t-1} \) to \( X_t \). Similarly let \( \mu_0 = \mu_0(X_{t-1}, \Delta, \theta_0) \) and \( \Sigma_0 = \Sigma_0(X_{t-1}, \Delta, \theta_0) \) be the true conditional mean and true conditional covariance of transitions of duration \( \Delta \) starting from \( X_{t-1} \). Given this, expression (46) may be usefully expanded into the form

\[
-\frac{1}{2} \frac{\partial |\Sigma_1|}{\partial \theta} - \frac{1}{2} (X_t - \mu_0)^T \Sigma_1^{-1} (X_t - \mu_0) - \frac{1}{2} (\mu_0 - \mu_1)^T \Sigma_1^{-1} (\mu_0 - \mu_1)
\]

\[
+ (X_t - \mu_0)^T \Sigma_0^{-1} \frac{\partial \mu_1}{\partial \theta} + (\mu_0 - \mu_1)^T \Sigma_1^{-1} \frac{\partial \mu_1}{\partial \theta} - (\mu_0 - \mu_1)^T \Sigma_0^{-1} (X_t - \mu_0),
\]

(47)

where \( \mu_0, \mu_1, \Sigma_0 \) and \( \Sigma_1 \) are functions of \( X_{t-1}, \Delta \) and \( \theta \) although this dependence has been suppressed for representational convenience. The contribution expressed in (47) is made to the score function with probability \( f_0(X_t, X_{t-1}, \Delta, \theta_0) \), the unconditional joint probability density function of observing the sequence \( (X_{t-1}, X_t) \). Consequently, the asymptotic form of the score function is

\[
S(\theta) = \mathbb{E}_{X_{t-1}} \left[ \mathbb{E}_{X_t} \left[ S_T(\theta) \mid X_{t-1}, \theta_0 \right] \right]
\]

\[
= \int_S \mathbb{E}_{X_t} \left[ S_T(\theta) \mid X_{t-1}, \theta_0 \right] f_0(X_{t-1}, \theta_0) dX_{t-1},
\]

(48)
where \( f_0(X_{t-1}, \theta_0) \) denotes the true marginal density of \( X_{t-1} \), namely
\[
f_0(X_{t-1}, \theta_0) = \int_S f_0(X_t, X_{t-1}, \Delta, \theta_0) \, dX_t.
\]
It follows immediately from expression (47) that
\[
\mathbb{E}_{X_t} [S_T(\theta) \mid X_{t-1}, \theta_0] = \mathbb{E}_{X_t} [(X_t - \mu_0)^T \mid X_{t-1}, \theta_0] \Sigma_1^{-1} \frac{\partial \mu_1}{\partial \theta} + \frac{1}{2} \frac{\partial |\Sigma_1|}{\partial \theta} + \mathbb{E}_{X_t} [(X_t - \mu_0)^T \mid X_{t-1}, \theta_0] (\mu_0 - \mu_1) + \frac{1}{2} \frac{\partial \Sigma_1^{-1}}{\partial \theta} \left( \mu_0 - \mu_1 \right)
\]
\[(49)\]
\[
- \frac{1}{2} \mathbb{E}_{X_t} [(X_t - \mu_0)^T \Sigma_1^{-1} \frac{\partial \Sigma_1^{-1}}{\partial \theta} (X_t - \mu_0) \mid X_{t-1}, \theta_0]
\]
\[
- (\mu_0 - \mu_1)^T \mathbb{E}_{X_t} [(X_t - \mu_0) \mid X_{t-1}, \theta_0].
\]
The true conditional mean \( \mu_0(X_{t-1}, \Delta, \theta_0) \) and true conditional covariance \( \Sigma_0(X_{t-1}, \Delta, \theta_0) \) of \( X_t \) are defined by the probability limits
\[
\mu_0(X_{t-1}, \Delta, \theta_0) = \mathbb{E} \left[ X_t \mid X_{t-1}, \theta_0 \right] = \operatorname{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} X_t \mid X_{t-1},
\]
\[
\Sigma_0(X_{t-1}, \Delta, \theta_0) = \mathbb{E} \left[ (X_t - \mu_0(X_{t-1}, \Delta, \theta_0))^2 \mid X_{t-1}, \theta_0 \right]
\]
\[(50)\]
\[
= \operatorname{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (X_t - \mu_0(X_{t-1}, \Delta, \theta_0)) (X_t - \mu_0(X_{t-1}, \Delta, \theta_0))^T \mid X_{t-1}.
\]
On taking account of these definitions, equation (49) simplifies to obtain
\[
\mathbb{E}_{X_t} [S_T(\theta) \mid X_{t-1}, \theta_0] = -\frac{1}{2} \frac{\partial |\Sigma_1|}{\partial \theta} - \frac{1}{2} \text{Trace} \left[ \Sigma_0 \frac{\partial \Sigma_1^{-1}}{\partial \theta} \right] + (\mu_0 - \mu_1)^T \left[ \Sigma_1^{-1} \frac{\partial \mu_1}{\partial \theta} - \frac{1}{2} \frac{\partial \Sigma_1^{-1}}{\partial \theta} (\mu_0 - \mu_1) \right].
\]
\[(51)\]
Further progress requires some basic properties of matrices and determinants. Given a nonsingular square matrix \( A \), it can be proved that \( A^{-1} \) satisfies the identities
\[
\frac{\partial A}{\partial \theta} = -A \frac{\partial A^{-1}}{\partial \theta} A, \quad \frac{\partial |A|}{\partial A} = |A| A^{-T}.
\]
\[(52)\]
Thus
\[
\frac{1}{|\Sigma_1|} \frac{\partial |\Sigma_1|}{\partial \theta} = \frac{1}{|\Sigma_1|} \text{Trace} \left[ \Sigma_1 \frac{\partial \Sigma_1}{\partial \theta} \right] = -\text{Trace} \left[ \left( \Sigma_1 \frac{\partial \Sigma_1^{-1}}{\partial \theta} \Sigma_1 \right) \Sigma_1^{-T} \right].
\]
However \( \Sigma_1 \) is a symmetric matrix and therefore
\[
\frac{1}{|\Sigma_1|} \frac{\partial |\Sigma_1|}{\partial \theta} = -\text{Trace} \left[ \Sigma_1 \frac{\partial \Sigma_1^{-1}}{\partial \theta} \right].
\]
In conclusion, equation (51) becomes

\[ \mathbb{E}_X [S_T(\theta) \mid X_{t-1}, \theta_0] = \frac{1}{2} \text{Trace} \left[ (\Sigma_1 - \Sigma_0) \frac{\partial \Sigma_1^{-1}}{\partial \theta} \right] + (\mu_0 - \mu_1)^T \left[ \Sigma_1^{-1} \frac{\partial \mu_1}{\partial \theta} - \frac{1}{2} \Sigma_1^{-1} (\mu_0 - \mu_1) \right], \] (53)

and the final asymptotic expression for the score function of the approximating transitional probability density function is

\[ S(\theta) = \frac{1}{2} \int_S \text{Trace} \left[ (\Sigma_1 - \Sigma_0) \frac{\partial \Sigma_1^{-1}}{\partial \theta} \right] f_0(X_{t-1}, \theta_0) dX_{t-1} \]

\[ + \int_S (\mu_0 - \mu_1)^T \left[ \Sigma_1^{-1} \frac{\partial \mu_1}{\partial \theta} - \frac{1}{2} \Sigma_1^{-1} (\mu_0 - \mu_1) \right] f_0(X_{t-1}, \theta_0) dX_{t-1}. \] (54)

In the presence of affine drift and diffusion specifications, the conditional mean and the conditional covariance of the true transitional process are identified for any legitimate choice of \( \theta \). The choices \( \mu_1(X_t, \Delta, \theta) = \mu_0(X_t, \Delta, \theta) \) and \( \Sigma_1(X_t, \Delta, \theta) = \Sigma_0(X_t, \Delta, \theta) \) therefore ensure that the approximating score function based on the multivariate Gaussian transitional density satisfies \( S(\theta_{QML}) \rightarrow 0 \) as \( \theta_{QML} \rightarrow \theta_0 \). Consequently using a multivariate Gaussian transitional probability density function with conditional mean \( \mu_1(X_t, \Delta, \theta) = \mu_0(X_t, \Delta, \theta) \) and conditional covariance \( \Sigma_1(X_t, \Delta, \theta) = \Sigma_0(X_t, \Delta, \theta) \) will provide consistent estimates of the parameters \( \theta_0 \) despite the fact that the true transitional density is misspecified.

**Proof of Proposition 2**

The argument begins by noting that

\[ \frac{d(\text{MSE})}{ds} = 2 \int_S (f_0 - f_1) \left( \frac{\partial f_0}{\partial s} - \frac{\partial f_1}{\partial s} \right) dX \]

\[ = 2 \int_S (f_0 - f_1) \sum_{j=1}^N \frac{\partial}{\partial X_j} \left( \frac{1}{2} \sum_{k=1}^N \frac{\partial (G_{jk} f_0)}{\partial X_k} - \alpha_j f_0 \right) dX - 2 \int_S (f_0 - f_1) \frac{\partial f_1}{\partial s} dX, \]

where \( G(X; \theta) = [G_{jk}] \) and \( \alpha(X; \theta) = [\alpha_j] \). The divergence theorem is applied to the first integral on the right hand side of the previous equation to obtain

\[ \frac{d(\text{MSE})}{ds} = -2 \int \sum_{j=1}^N \left( \frac{1}{2} \sum_{k=1}^N \frac{\partial (G_{jk} f_0)}{\partial X_k} - \alpha_j f_0 \right) \frac{\partial (f_0 - f_1)}{\partial X_j} dX - 2 \int_S (f_0 - f_1) \frac{\partial f_1}{\partial s} dX \]

\[ = -2 \int \sum_{j=1}^N \frac{\partial (f_0 - f_1)}{\partial X} \left( \frac{1}{2} \sum_{k=1}^N \frac{\partial G_{jk}}{\partial X_k} - \alpha_j \right) f_0 dX \]

\[ - \int \sum_{j=1}^N \sum_{k=1}^N G_{jk} \frac{\partial (f_0 - f_1)}{\partial X_j} \frac{\partial f_0}{\partial X_k} dX - 2 \int_S (f_0 - f_1) \frac{\partial f_1}{\partial s} dX. \] (55)
The first and second integrals on the right hand side of equation (55) are analysed in turn. The first integral is expanded into the form

\[- \int_S \sum_{j=1}^N \partial (f_0 - f_1) \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} - \alpha_j \right) dX - \int_S \sum_{j=1}^N \partial (f_0 - f_1) \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} - \alpha_j \right) f_1 dX,\]

which, after some further manipulation, becomes

\[\frac{1}{2} \int_S (f_0 - f_1)^2 \sum_{j=1}^N \frac{\partial}{\partial X_j} \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} - \alpha_j \right) dX + \int_S (f_0 - f_1) \sum_{j=1}^N \frac{\partial}{\partial X_j} \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} f_1 - \alpha_j f_1 \right) dX.\] (56)

The second integral is expanded into the form

\[- \int_S \sum_{j=1}^N \sum_{k=1}^N G_{jk} \frac{\partial (f_0 - f_1)}{\partial X_j} \frac{\partial (f_0 - f_1)}{\partial X_k} dX - \int_S \sum_{j=1}^N \sum_{k=1}^N G_{jk} \frac{\partial (f_0 - f_1)}{\partial X_j} \frac{\partial f_1}{\partial X_k} dX,\]

which, after some further manipulation, becomes

\[- \int_S \sum_{j=1}^N \sum_{k=1}^N G_{jk} \frac{\partial (f_0 - f_1)}{\partial X_j} \frac{\partial (f_0 - f_1)}{\partial X_k} dX + \int_S (f_0 - f_1) \sum_{j=1}^N \sum_{k=1}^N \frac{\partial}{\partial X_j} \left( G_{jk} \frac{\partial f_1}{\partial X_k} \right) dX.\] (57)

Results (56) and (57) are incorporated into equation (55) to get, after some regrouping of terms, that

\[\frac{d(MSE)}{ds} = \int_S (f_0 - f_1)^2 \sum_{j=1}^N \frac{\partial}{\partial X_j} \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} - \alpha_j \right) dX - 2 \int_S (f_0 - f_1) F(f_1) dX \]

\[- \int_S \sum_{j=1}^N \sum_{k=1}^N G_{jk} \frac{\partial (f_0 - f_1)}{\partial X_j} \frac{\partial (f_0 - f_1)}{\partial X_k} dX,\] (58)

Because the diffusion \(G\) is a positive definite\(^{12}\) array, then

\[- \int_S \sum_{j=1}^N \sum_{k=1}^N G_{jk} \frac{\partial (f_0 - f_1)}{\partial X_j} \frac{\partial (f_0 - f_1)}{\partial X_k} dX \leq 0.\] (59)

with equality if and only if the trial density \(f_1\) is the true transitional density. The properties of \(G(X; \theta)\) and \(\alpha(X; \theta)\) included in the statement of the theorem and basic properties of integration lead to the inequality

\[\int_S (f_0 - f_1)^2 \sum_{j=1}^N \frac{\partial}{\partial X_j} \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} - \alpha_j \right) dX \]

\[\leq \int_S (f_0 - f_1)^2 \sum_{j=1}^N \frac{\partial}{\partial X_j} \left( \frac{1}{2} \sum_{k=1}^N \partial G_{jk} - \alpha_j \right) dX \leq C \int_S (f_0 - f_1)^2 = C \times MSE(s).\] (60)

\(^{12}\)If per chance the sample space \(S\) is finite then the right hand side of inequality (59) (which is of Poincaré type) can be sharpened to \(-\lambda E\) where \(\lambda\) is a positive constant determined jointly by the smallest eigenvalue of the symmetric array \(G = [G_{jk}]\) and the smallest eigenvalue of the region \(S\).
When results (59) and (60) are incorporated into equation (58) the outcome is the inequality
\[
\frac{d(MSE)}{ds} \leq C \times MSE(s) - 2 \int_S (f_0 - f_1) F_1 dX .
\] (61)

The final step in the proof of Proposition 2 is to note that if \( \omega(s) \) is an arbitrary positive function of \( s \) then the application of the identity
\[
\omega a^2 + \frac{b^2}{\omega} = -2ab + \left( a\sqrt{\omega} + \frac{b}{\sqrt{\omega}} \right)^2 \geq -2ab
\]
to inequality (61) with \( a = (f_0 - f_1) \) and \( b = F_1 \) yields the final inequality
\[
\frac{d(MSE)}{ds} \leq (C + \omega) \times MSE(s) + \frac{1}{\omega} \int_S |\mathcal{L}(f_1)|^2 dX .
\] (62)

The initial condition measures the extent to which the initial value of the approximating density matches that of the true transitional density. If both probability densities share the same initial condition then MSE(0) = 0.

**Proof of Proposition 3**

The proof begins by noting that the arbitrariness of \( \omega(s) \) can be used to sharpen the inequality
\[
\frac{d(MSE)}{ds} \leq (C + \omega)MSE + \frac{1}{\omega} \int_S |\mathcal{L}(f_1)|^2 dX ,
\]
by minimizing its right hand side to obtain the inequality
\[
\frac{d(MSE)}{ds} \leq C \times MSE + 2\sqrt{MSE} \left( \int_S |\mathcal{L}(f_1)|^2 dX \right)^{1/2} , \quad MSE(0) = MSE_0 .
\]

This equation is a linear ordinary differential inequality for RMSE(s) = \( \sqrt{MSE(s)} \) with general solution
\[
RMSE(s) \leq RMSE_0 e^{Cs/2} + \int_0^s e^{-C(s-u)/2} \left( \int_S |\mathcal{L}(f_1)|^2 dX \right)^{1/2} du .
\]

For example, if \( f_1 \) satisfies the Fokker-Planck equation of the SDE (1) but for misspecified initial conditions then the misspecification error simply grows exponentially. When particularized to the usual situation in which the approximating transitional density and the true transitional density share the same initial condition, then MSE_0 = 0 and
\[
RMSE(s) \leq \int_0^s e^{C(s-u)/2} \left( \int_S |\mathcal{L}(f_1)|^2 dX \right)^{1/2} du .
\]
Proof of Proposition 4
It is required to show that the value of
\[ \int_S |\mathcal{F}(f_1)|^2 dX \]
depends only on moments of the trial density of order \((m + 1)\) and above with respect to the Kolmogorov operator \(\mathcal{F}\). The essence of the proof is the observation that \(\mathcal{F}(f_1)\) has a convergent Taylor series in \(S\) of the form
\[ \mathcal{F}(f_1) = \sum_{k=0}^{\infty} C_k \times \underbrace{X \times \cdots \times X}_{k \text{ times}}, \]
where \(C_k\) is typically a tensorial array formed from all the possible \(k\)-th order derivatives of \(\mathcal{F}(f_1)\) with respect to \(X\). Thus
\[ \int_S |\mathcal{F}(f_1)|^2 dX = \int_S \mathcal{F}(f_1) \left( \sum_{k=0}^{\infty} C_k \times \underbrace{X \times \cdots \times X}_{k \text{ times}} \right) dX \]
\[ = \sum_{k=0}^{\infty} \frac{C_k}{k!} \times \int_S \underbrace{X \times \cdots \times X}_{k \text{ times}} \mathcal{F}(f_1) dX. \]
However, all moments of \(f_1(X, s | X_t, \theta)\) with respect to the Kolmogorov operator \(\mathcal{F}\) are zero for \(k \leq m\) and therefore
\[ \int_S |\mathcal{F}(f_1)|^2 dX = \sum_{k=m+1}^{\infty} \frac{C_k}{k!} \times \int_S \underbrace{X \times \cdots \times X}_{k \text{ times}} \mathcal{F}(f_1) dX = \sum_{k=m+1}^{\infty} \frac{C_k K_k}{k!}, \]
thereby establishing the proposition. \[ \blacksquare \]
List of NCER Working Papers

No. 64  [Download full text]  
Ralf Becker and Adam Clements  
Volatility and the role of order book structure?

No. 63  [Download full text]  
Adrian Pagan  
Can Turkish Recessions be Predicted?

No. 62  [Download full text]  
Lionel Page and Katie Page  
Evidence of referees’ national favouritism in rugby

No. 61  [Download full text]  
Nicholas King, P Dorian Owen and Rick Audas  
Playoff Uncertainty, Match Uncertainty and Attendance at Australian National Rugby League Matches

No. 60  [Download full text]  
Ralf Becker, Adam Clements and Robert O’Neill  
A Cholesky-MIDAS model for predicting stock portfolio volatility

No. 59  [Download full text]  
P Dorian Owen  
Measuring Parity in Sports Leagues with Draws: Further Comments

No. 58  [Download full text]  
Don Harding  
Applying shape and phase restrictions in generalized dynamic categorical models of the business cycle

No. 57  [Download full text]  
Renee Fry and Adrian Pagan  
Sign Restrictions in Structural Vector Autoregressions: A Critical Review

No. 56  [Download full text]  
Mardi Dungey and Lyudmyla Hvozdyk  
Cojumping: Evidence from the US Treasury Bond and Futures Markets

No. 55  [Download full text]  
Martin G. Kocher, Marc V. Lenz and Matthias Sutter  
Psychological pressure in competitive environments: Evidence from a randomized natural experiment: Comment

No. 54  [Download full text]  
Adam Clements and Annastiina Silvennoinen  
Estimation of a volatility model and portfolio allocation
<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>The Credit Channel and Monetary Transmission in Brazil and Chile: A Structured VAR Approach</td>
<td>Luis Catão and Adrian Pagan</td>
</tr>
<tr>
<td>52</td>
<td>Testing the Profitability of Technical Analysis as a Portfolio Selection Strategy</td>
<td>Vlad Pavlov and Stan Hurn</td>
</tr>
<tr>
<td>51</td>
<td>Substitution Between Managers and Subordinates: Evidence from British Football</td>
<td>Sue Bridgewater, Lawrence M. Kahn and Amanda H. Goodall</td>
</tr>
<tr>
<td>50</td>
<td>Structural Macro-Econometric Modelling in a Policy Environment</td>
<td>Martin Fukac and Adrian Pagan</td>
</tr>
<tr>
<td>49</td>
<td>Detecting Common Dynamics in Transitory Components</td>
<td>Tim M Christensen, Stan Hurn and Adrian Pagan</td>
</tr>
<tr>
<td>48</td>
<td>Inter-market Arbitrage in Sports Betting</td>
<td>Egon Franck, Erwin Verbeek and Stephan Nüesch</td>
</tr>
<tr>
<td>47</td>
<td>Relational Good at Work! Crime and Sport Participation in Italy. Evidence from Panel Data Regional Analysis over the Period 1997-2003.</td>
<td>Raul Caruso</td>
</tr>
<tr>
<td>46</td>
<td>The Influence of Social Pressure and Nationality on Individual Decisions: Evidence from the Behaviour of Referees</td>
<td>Peter Dawson and Stephen Dobson</td>
</tr>
<tr>
<td>45</td>
<td>Forecast performance of implied volatility and the impact of the volatility risk premium</td>
<td>Ralf Becker, Adam Clements and Christopher Coleman-Fenn</td>
</tr>
<tr>
<td>44</td>
<td>On the economic benefit of utility based estimation of a volatility model</td>
<td>Adam Clements and Annastiina Silvennoinen</td>
</tr>
<tr>
<td>43</td>
<td>A nonparametric approach to forecasting realized volatility</td>
<td>Adam Clements and Ralf Becker</td>
</tr>
<tr>
<td>42</td>
<td>The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition</td>
<td>Uwe Dulleck, Rudolf Kerschbamer and Matthias Sutter</td>
</tr>
</tbody>
</table>
No. 41  [Download full text]
Adam Clements, Mark Doolan, Stan Hurn and Ralf Becker
On the efficacy of techniques for evaluating multivariate volatility forecasts

No. 40  [Download full text]
Lawrence M. Kahn
The Economics of Discrimination: Evidence from Basketball

No. 39  [Download full text]
Don Harding and Adrian Pagan
An Econometric Analysis of Some Models for Constructed Binary Time Series

No. 38  [Download full text]
Richard Dennis
Timeless Perspective Policymaking: When is Discretion Superior?

No. 37  [Download full text]
Paul Frijters, Amy Y.C. Liu and Xin Meng
Are optimistic expectations keeping the Chinese happy?

No. 36  [Download full text]
Benno Torgler, Markus Schaffner, Bruno S. Frey, Sascha L. Schmidt and Uwe Dulleck
Inequality Aversion and Performance in and on the Field

No. 35  [Download full text]
T M Christensen, A. S. Hurn and K A Lindsay
Discrete time-series models when counts are unobservable

No. 34  [Download full text]
Adam Clements, A S Hurn and K A Lindsay
Developing analytical distributions for temperature indices for the purposes of pricing temperature-based weather derivatives

No. 33  [Download full text]
Adam Clements, A S Hurn and K A Lindsay
Estimating the Payoffs of Temperature-based Weather Derivatives

No. 32  [Download full text]
T M Christensen, A S Hurn and K A Lindsay
The Devil is in the Detail: Hints for Practical Optimisation

No. 31  [Download full text]
Uwe Dulleck, Franz Hackl, Bernhard Weiss and Rudolf Winter-Ebmer
Buying Online: Sequential Decision Making by Shopbot Visitors

No. 30  [Download full text]
Richard Dennis
Model Uncertainty and Monetary Policy
No. 29 (Download full text)
Richard Dennis
The Frequency of Price Adjustment and New Keynesian Business Cycle Dynamics

No. 28 (Download full text)
Paul Frijters and Aydogan Ulker
Robustness in Health Research: Do differences in health measures, techniques, and time frame matter?

No. 27 (Download full text)
Paul Frijters, David W. Johnston, Manisha Shah and Michael A. Shields

No. 26 (Download full text)
Paul Frijters and Tony Beatton
The mystery of the U-shaped relationship between happiness and age.

No. 25 (Download full text)
T M Christensen, A S Hurn and K A Lindsay
It never rains but it pours: Modelling the persistence of spikes in electricity prices

No. 24 (Download full text)
Ralf Becker, Adam Clements and Andrew McClelland
The Jump component of S&P 500 volatility and the VIX index

No. 23 (Download full text)
A. S. Hurn and V. Pavlov
Momentum in Australian Stock Returns: An Update

No. 22 (Download full text)
Mardi Dungey, George Milunovich and Susan Thorp
Unobservable Shocks as Carriers of Contagion: A Dynamic Analysis Using Identified Structural GARCH

No. 21 (Download full text) (forthcoming)
Mardi Dungey and Adrian Pagan
Extending an SVAR Model of the Australian Economy

No. 20 (Download full text)
Benno Torgler, Nemanja Antic and Uwe Dulleck
Mirror, Mirror on the Wall, who is the Happiest of Them All?

No. 19 (Download full text)
Justina AV Fischer and Benno Torgler
Social Capital And Relative Income Concerns: Evidence From 26 Countries

No. 18 (Download full text)
Ralf Becker and Adam Clements
Forecasting stock market volatility conditional on macroeconomic conditions.
No. 17  [Download full text]  
Ralf Becker and Adam Clements  
Are combination forecasts of S&P 500 volatility statistically superior?

No. 16  [Download full text]  
Uwe Dulleck and Neil Foster  
Imported Equipment, Human Capital and Economic Growth in Developing Countries

No. 15  [Download full text]  
Ralf Becker, Adam Clements and James Curchin  
Does implied volatility reflect a wider information set than econometric forecasts?

No. 14  [Download full text]  
Renee Fry and Adrian Pagan  
Some Issues in Using Sign Restrictions for Identifying Structural VARs

No. 13  [Download full text]  
Adrian Pagan  
Weak Instruments: A Guide to the Literature

No. 12  [Download full text]  
Ronald G. Cummings, Jorge Martinez-Vazquez, Michael McKee and Benno Torgler  
Effects of Tax Morale on Tax Compliance: Experimental and Survey Evidence

No. 11  [Download full text]  
Benno Torgler, Sascha L. Schmidt and Bruno S. Frey  
The Power of Positional Concerns: A Panel Analysis

No. 10  [Download full text]  
Ralf Becker, Stan Hurn and Vlad Pavlov  
Modelling Spikes in Electricity Prices

No. 9  [Download full text]  
A. Hurn, J. Jeisman and K. Lindsay  
Teaching an Old Dog New Tricks: Improved Estimation of the Parameters of Stochastic Differential Equations by Numerical Solution of the Fokker-Planck Equation

No. 8  [Download full text]  
Stan Hurn and Ralf Becker  
Testing for nonlinearity in mean in the presence of heteroskedasticity.

No. 7  [Download full text]  (published)  
Adrian Pagan and Hashem Pesaran  
On Econometric Analysis of Structural Systems with Permanent and Transitory Shocks and Exogenous Variables.

No. 6  [Download full text]  (published)  
Martin Fukac and Adrian Pagan  
Limited Information Estimation and Evaluation of DSGE Models.
No. 5  (Download full text)
Andrew E. Clark, Paul Frijters and Michael A. Shields
Income and Happiness: Evidence, Explanations and Economic Implications.

No. 4  (Download full text)
Louis J. Maccini and Adrian Pagan
Inventories, Fluctuations and Business Cycles.

No. 3  (Download full text)
Adam Clements, Stan Hurn and Scott White

No. 2  (Download full text)
Stan Hurn, J.Jeisman and K.A. Lindsay

No. 1  (Download full text)
Adrian Pagan and Don Harding
The Econometric Analysis of Constructed Binary Time Series.