# A quasi-likelihood approach to estimating parameters in diffusion-type processes

## C. C. HEYDE

### Abstract

  Estimation of parameters in diffusion models is usually handled by maximum likelihood and involves the calculation of a Radon–Nikodym derivative. This methodology is often not available when minor changes are made to the model. However, these complications can usually be avoided and results obtained under more general conditions using quasi-likelihood methods. The basic ideas are explained in this paper and are illustrated through discussion of the Cox–Ingersoll–Ross model and a modification of the Langevin model.

COX–INGERSOLL–ROSS MODEL; LANGEVIN MODEL; INFERENCE; PARAMETER ESTIMATION; DIFFUSION PROCESSES

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60J60

## 1. Introduction

The standard method of estimation for parameters in the drift coefficient of a diffusion process involves calculation of a likelihood ratio (Radon–Nikodym derivative) and thence the maximum likelihood estimator(s). This is less than straightforward for more complicated models, and indeed it is often not available at all because of the non-existence of the Radon–Nikodym derivative. New methods of quasi-likelihood, however, allow estimators to be obtained straightforwardly under very general conditions. They can deal, in particular, with the situation in which the Brownian motion in a diffusion is replaced by a general square-integrable martingale. The approach, which is based on selection of an optimal estimating function from within a specified class of such functions, involves assumptions on only the first two conditional moments of the underlying process. Nevertheless, the quasi-likelihood estimators will ordinarily be true maximum likelihood estimators in a context where the Radon–Nikodym derivative is available. Furthermore, they will generally be consistent, asymptotically normally distributed, and can be used to construct minimum-size asymptotic confidence zones for the unknown parameters among estimators coming from the specified class. The purpose of this paper is to illustrate the value of the methodology through a general discussion and application to the

Cox–Ingersoll–Ross model for interest rates and to a modification of the Langevin model for dynamical systems.

Since the general theory of quasi-likelihood is not widely known we shall begin with a sketch of the basic principles. For a detailed discussion of the general framework see Godambe and Heyde (1987). Various amplifications are provided in Heyde (1988) and Sørensen (1990). For a particularly accessible introduction see Heyde (1989).

It should be noted that the term 'quasi-likelihood' also appears in the literature with a rather different interpretation from that given here (e.g. Wedderburn (1974), McCullagh and Nelder (1989)). A discussion of the relationship between these interpretations is provided in Lin and Heyde (1993).

## 2. General principles

The general framework is as follows. We are given a sample $\{Z_t, t \in T\}$, say, $T$ being discrete or continuous, taking values in $r$-dimensional Euclidean space, and the set of probability measures $\{P_\theta\}$ for $\{Z_t\}$ is a union of families of models, each being indexed by a characteristic $\theta$ belonging to an open subset of $p$-dimensional Euclidean space. The object is the efficient estimation of $\theta$ in situations where a likelihood function may not be available.

Attention is focused on the class $\mathscr{G}$ of zero-mean square-integrable estimating functions $G_T(\theta) = G_T(\{Z_t, t \in T\}, \theta)$ which are vectors of dimension $p$ for which $EG_T(\theta) = 0$ for each $P_\theta$ and for which the $p$-dimensional matrices $E\dot{G}_T(\theta) = (E\partial G_{T,i}(\theta)/\partial\theta_j)$ and $EG_T(\theta)G_T'(\theta)$ are non-singular, the prime denoting transpose. The expectations are always with respect to $P_\theta$. Within $\mathscr{G}$, estimators $\hat{\theta}_T$ are found by solving the estimating equation $G_T(\hat{\theta}_T) = 0$.

Quasi-likelihood theory is focused on suitably chosen subsets of $\mathscr{G}$ and involves choice of the estimating function $G$ to maximize, in the partial order of non-negative definite (n.n.d.) matrices, the information criterion

$$\mathscr{E}(G_T) = (E\dot{G}_T)'(EG_TG_T')^{-1}E\dot{G}_T$$

which is a natural generalization of Fisher information (see e.g. Godambe and Heyde (1987)). (Here and below we omit the argument $\theta$ for simplicity.)

*Definition.* Suppose that $G_T^* \in \mathscr{H} \subseteq \mathscr{G}$. If

$$\mathscr{E}(G^*) - \mathscr{E}(G_T)$$

is n.n.d. for all $G_T \in \mathscr{H}$ we say that $G_T^*$ is a quasi-score estimating function within $\mathscr{H}$.

The estimator $\theta_T^*$ obtained from $G_T^*(\theta_T^*) = 0$, which is termed a *quasi-likelihood estimator*, has, under broad conditions, certain minimum variance and minimum

size asymptotic confidence zone properties for $\theta$, at least within the class $\mathscr{H}$. Indeed, the basic properties are those of the maximum likelihood estimator but restricted to the class $\mathscr{H}$.

The theory does not require a parametric setting, let alone the existence of a likelihood score function $U_T$ (derivative of the log-likelihood with respect to $\theta$), but if $U_T$ exists a judicious choice of $\mathscr{H}$ will usually produce it as the quasi-score estimating function. If $U_T \in \mathscr{H}$, as can ordinarily be arranged in exponential family problems, then $U_T$ is the quasi-score estimating function within $\mathscr{H}$ and can easily be calculated. Godambe (1960) showed that $U_T$ is the quasi-score estimating function within $\mathscr{G}$.

It is not usually practicable to find a quasi-score estimating function directly from the use of the definition. However, the following proposition (Heyde (1988)) provides for ready derivation.

*Proposition* 1. *Let* $\mathscr{H} \subset \mathscr{G}$. *Then* $G_T^* \in \mathscr{H}$ *is a quasi-score estimating function within* $\mathscr{H}$ *if*

$$(1) \qquad (E\dot{G}_T)^{-1} EG_T G_T^{*'} = C_T$$

*for all* $G_T \in \mathscr{H}$, *where* $C_T$ *is a fixed matrix. Conversely, if* $\mathscr{H}$ *is convex and* $G_T^*$ *is a quasi-score estimating function then* (1) *holds.*

## 3. The application

The models which we shall consider in this paper can all be written in the semimartingale form

$$(2) \qquad dX_t = dA_t(\theta) + dM_t(\theta)$$

where the finite variation process $\{A_t\}$ can be interpreted as the *signal* and the local martingale $\{M_t\}$ can be interpreted as the *noise*. The local martingale has a natural role to play in inference as it represents the residual after fitting of the signal. Then, the quasi-score estimating function based on the family of local martingale estimating functions

$$\mathscr{H} = \left\{ \int_0^T a_t(\theta) dM_t(\theta), \{a_t\} \text{ non-anticipating} \right\}$$

is easily seen from Proposition 1 to be

$$(3) \qquad \int_0^T (E(d\dot{M}_t(\theta)|\mathscr{F}_{t-}))'(d\langle M(\theta)\rangle_t)^+ dM_t(\theta)$$

where $\{\mathscr{F}_t\}$ is a filtration of past-history $\sigma$-fields, $\langle M(\theta)\rangle_t$ is the quadratic characteristic and the $+$ denotes the Moore–Penrose generalized inverse which is, for a matrix $A$, the unique matrix $A^+$ possessing the properties $AA^+A = A$, $A^+AA^+ = A^+$, $A^+A = AA^+$.

Now (3) can be rewritten as

$$\int_0^T (E(d\dot{A}_t(\theta)|\mathscr{F}_{t-}))'(d\langle M(\theta)\rangle_t)^+(dA_t(\theta) - dX_t)$$

from which it is clear that the quasi-score estimating function is unaffected if the local martingale noise $\{M_t(\theta)\}$ is replaced by another whose quadratic characteristic is the same. The precise distributional form of the noise does not need to be known. In the commonly met situation where $M_t(\theta) = \sigma W_t$ with $\sigma > 0$ and $W_t$ being standard Brownian motion, the results are robust to the extent that $\{M_t(\theta)\}$ could be replaced by any local martingale $\{Z_t(\theta)\}$, for example one with independent increments, for which $\langle Z\rangle_t = \sigma^2 t$ without changing the estimators.

In the particular case of a diffusion process the components on the right-hand side of the representation (2) can be written as

$$(4) \qquad\qquad dA_t(\theta) = a(t, X_t, \theta)dt, \qquad dM_t(\theta) = b^{1/2}(t, X_t)dW_t$$

where $a$ and $b$ are known vector and matrix functions respectively and $\{W_t\}$ is standard Brownian motion. Then, an appropriate Radon–Nikodym derivative of the measure induced by the process $\{X_t, 0 \le t \le T\}$ with parameter $\theta$ with respect to the corresponding measure for parameter $\theta_0$ can be calculated and is given by

$$(5) \qquad\qquad \exp\left\{\int_0^T C(t, X_t)dX_t - \int_0^T D(t, X_t)dt\right\}$$

where

$$b(t, x)C(t, x) = a(t, x, \theta) - a(t, x, \theta_0)$$
$$D(t, x) = (a(t, x, \theta_0))'C(t, x) + \tfrac{1}{2}(C(t, x))'b(t, x)C(t, x)$$

(e.g. Basawa and Prakasa Rao (1980), p. 219).

From (5) it is easily checked that the likelihood score function (derivative of the logarithm of the Radon–Nikodym derivative with respect to $\theta$) is given by (3). This means that the quasi-likelihood estimator is the maximum likelihood estimator for the model (4). However, as indicated above, the quasi-likelihood estimator is available much more generally.

The explanation for the quasi-score corresponding to the likelihood score for the model (4) is not hard to discern. A likelihood score is a martingale under modest regularity conditions and all square integrable martingales living on the same probability space as the Brownian motion in the noise term of the model (4) can be described as stochastic integrals with respect to the Brownian motion (see e.g. Theorem 5.17 of Lipster and Shiryaev (1977)). The likelihood score will be one such martingale and will therefore be included in the relevant family $\mathscr{H}$ over which optimization takes place and it solves the optimization problem.

As a concrete illustration of the methodology we shall discuss the stochastic differential equation

$$(6) \qquad dX_t = \alpha(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t$$

where $X_0 > 0$, $\alpha > 0$, $\beta > 0$, $\sigma > 0$. This form was proposed by Cox, Ingersoll and Ross (1985) as a model for interest rates and it has been widely used in finance.

In considering the model (6) we shall be concerned with the estimation of $\theta = (\alpha, \beta)'$. The parameter $\sigma$ can be regarded as known whenever Brownian motion and continuous sampling are involved. Indeed, $\sigma$ can be calculated with probability 1 on the basis of knowledge of a path of the process on any finite time interval. This follows from the definitions of the quadratic variation process and stochastic integrals with respect to Brownian motion (e.g. Rogers and Williams (1987), Chapter IV, Section 4) from which one obtains that, writing $t_i^{(n)} = \min(T, 2^{-n}i)$,

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} (X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}})^2 = \sigma^2 \int_0^T X_t dt \text{ a.s.}$$

and

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} X_{t_i^{(n)}} (t_{i+1}^{(n)} - t_i^{(n)}) = \int_0^T X_t dt \text{ a.s.}$$

For the model (6) the Radon–Nikodym derivative of the measure based on $(\alpha, \beta)'$ with respect to that based on $(\alpha_0, \beta_0)'$ is easily seen from (5) to be

$$\exp\left\{ \sigma^{-2} \int_0^T X_t^{-1}[\alpha(\beta - X_t) - \alpha_0(\beta_0 - X_t)]dX_t \right.$$
$$\left. - \frac{1}{2}\sigma^{-2} \int_0^T X_t^{-1}[\alpha^2(\beta - X_t)^2 - \alpha_0^2(\beta_0 - X_t)^2]dt \right\}$$

and differentiating the logarithm of this likelihood ratio with respect to $\theta = (\alpha, \beta)'$ gives the likelihood score

$$(7) \qquad U_T = \sigma^{-2} \int_0^T \binom{\beta - X_t}{\alpha} X_t^{-1/2} dW_t$$

which is also the quasi-score given by (2).

If we modify the model to the form

$$dX_t = \alpha(\beta - X_t)dt + \sigma(\alpha, \gamma)\sqrt{X_t}dW_t$$

where $\sigma(\alpha, \gamma)$ reflects a possibly rate-dependent noise, then the likelihood ratio does not exist in general. Indeed, when $\sigma(\alpha_1, \lambda) \neq \sigma(\alpha_2, \lambda)$ the supports of the distributions of the two processes are disjoint. The quasi-likelihood methodology, however, is unaffected by this change. The quasi-score estimating function

continues to be given by (7) and the asymptotic properties of the estimators are also unaffected.

From (7), the maximum likelihood/quasi-likelihood estimators $\hat{\alpha}_T$, $\hat{\beta}_T$ are given by

$$\int_0^T (\hat{\beta}_T - X_t)X_T^{-1}[dX_t - \hat{\alpha}_T(\hat{\beta}_T - X_t)dt] = 0$$

$$\int_0^T X_t^{-1}[dX_t - \hat{\alpha}_T(\hat{\beta}_T - X_t)dt] = 0$$

and putting

$$I_T = \int_0^T X_t^{-1}dX_t, \qquad J_T = \int_0^T X_t^{-1}dt, \qquad K_T = \int_0^T X_t dt,$$

we find that

$$\hat{\alpha}_T = (I_T T - J_T(X_T - X_0))/(J_T K_T - T^2)$$

$$\hat{\beta}_T = (I_T K_T - T(X_T - X_0))/(I_T T - J_T(X_T - X_0)).$$

For the model with $2\alpha\beta \geq \sigma^2$ there is a strictly positive stationary ergodic solution to (6) at $T \to \infty$ whose distribution has gamma density $\Gamma(2\alpha\beta/\sigma^2, 2\alpha/\sigma^2)$ (see e.g. Kloeden and Platen (1992), p. 38). Suppose $X_\infty$ has this density; then

(8) $$EX_\infty = \beta, \qquad EX_\infty^{-1} = 2\alpha/(2\alpha\beta - \sigma^2).$$

Using the ergodic theorem we obtain

(9) $$T^{-1}J_T = T^{-1}\int_0^T X_t^{-1}dt \xrightarrow{\text{a.s.}} EX_\infty^{-1},$$

(10) $$T^{-1}K_T = T^{-1}\int_0^T X_t dt \xrightarrow{\text{a.s.}} EX_\infty$$

and, since

$$I_T = \int_0^T X_t^{-1}dX_t = \log X_0^{-1}X_T + \tfrac{1}{2}\sigma^2 J_T$$

using Itô's formula,

$$T^{-1}I_T \xrightarrow{\text{a.s.}} \tfrac{1}{2}\sigma^2 EX_\infty^{-1}$$

as $T \to \infty$. These results readily give the strong consistency of the estimators $\hat{\alpha}_T$, $\hat{\beta}_T$. Asymptotic normality of $T^{1/2}(\hat{\alpha}_T - \alpha, \hat{\beta}_T - \beta)'$ can be obtained by applying Theorem 2.1, p. 405 of Basawa and Prakasa Rao (1980).

All these results continue to hold under substantially weakened conditions on the noise component in the model. For example, they hold if $\{W_t\}$ is replaced by a

square integrable martingale with stationary independent increments $\{Z_t\}$ for which $\langle Z \rangle_t \equiv t$. The details involve straightforward applications of the martingale strong law and central limit theorem and are omitted.

The formulation via (2) and (3) has to be used with care for problems with multiple sources of variation. Suppose for example that the Langevin stochastic differential equation (e.g. Kloeden and Platen (1992), pp. 104–105) is augmented with jumps coming from a Poisson process and becomes

$$(11) \qquad dX_t = \theta X_t dt + dW_t + dN_t,$$

$N_t$ being a Poisson process with intensity $\lambda$. Then, the process may be written in semi-martingale form as

$$dX_t = (\theta X_t + \lambda)dt + dM_t$$

where $\{M_t\}$ is a martingale given by

$$(12) \qquad M_t = W_t + N_t - \lambda t.$$

Using (11) and (12) in (3), the quasi-score estimating function based on noise $\{M_t\}$ is

$$\int_0^T (X_{t-} 1)' dM_t,$$

leading to the estimating equations

$$\int_0^T X_{t-} dX_t = \hat{\theta}_T \int_0^T X_t^2 dt + \hat{\lambda}_T \int_0^T X_t dt$$

$$X_T = \hat{\theta}_T \int_0^T X_t^2 dt + \hat{\lambda}_T T.$$

These, however, are the maximum likelihood estimating equations for the model

$$dX_t = (\theta X_t + \lambda)dt + dW_t,$$

i.e. a version of (11) in which $N_t$ has been replaced by its compensator $\lambda t$. In this model the entire stochastic fluctuation is described by the Brownian process and this is only realistic if $\lambda \ll 1$.

This problem, first noted by Sørensen (1990), can be circumvented and the true maximum likelihood estimators for $\theta$, $\lambda$ obtained if we treat the sources of variation separately. We replace (11) by the equations

$$dX_t^c = \theta X_t dt + dW_t, \qquad X_t^c = X_t - N_t$$
$$dX_t - dX_t^c = dN_t = \lambda dt + (dN_t - \lambda t)$$

which focus on the continuous and discrete components respectively. Then, using separate quasi-score estimating functions based on the martingales $\{W_t\}$ and

$\{N_t - \lambda t\}$ respectively, we obtain estimating equations

$$\int_0^T X_{t-} dX_t^c = \tilde{\theta}_T \int_0^T X_t^2 dt$$

$$N_T = \tilde{\lambda}_T T$$

leading to maximum likelihood estimators $\tilde{\theta}_T$, $\tilde{\lambda}_T$.

Models of the above kind are quite common and the general message is to identify relevant (local) martingales which focus on the individual sources of variation. It is then possible to obtain quasi-score estimating functions for each, and to combine them, provided the appropriate sample information is available.

## References

BASAWA, I. V. AND PRAKASA RAO, B. L. S. (1980) *Statistical Inference for Stochastic Processes*. Academic Press, London.

COX, J. S., INGERSOLL, J. E. AND ROSS, S. A. (1985) A theory of the term structure of interest rates. *Econometrica* **53**, 363–384.

GODAMBE, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208–1211.

GODAMBE, V. P. AND HEYDE, C. C. (1987) Quasi-likelihood and optimal estimation. *Int. Statist. Rev.* **55**, 231–244.

HEYDE, C. C. (1988) Fixed sample and asymptotic optimality for classes of estimating functions. *Contemp. Math.* **80**, 241–247.

HEYDE, C. C. (1989) Quasi-likelihood and optimality for estimating functions: some current unifying themes. *Bull. Int. Statist. Inst.* **53**(1), 19–29.

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

LIN, Y.-X. AND HEYDE, C. C. (1993) Optimal estimating functions and Wedderburn's quasi-likelihood. *Commun. Statist. Theory Meth.* **22**, 2341–2350.

LIPSTER, R. S. AND SHIRYAEV, A. N. (1977) *Statistics of Random Processes I. General Theory*. Springer-Verlag, New York.

McCULLAGH, P. AND NELDER, J. A. (1989) *Generalized Linear Models*, 2nd edn. Chapman and Hall, London.

ROGERS, L. C. G. AND WILLIAMS, D. (1987) *Diffusions, Markov Processes and Martingales. Vol. 2 Itô Calculus*. Wiley, Chichester.

SØRENSEN, M. (1990) On quasi-likelihood for semimartingales. *Stoch. Proc. Appl.* **25**, 331–346.

WEDDERBURN, R. W. M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**, 439–447.