A PRACTICAL PROCEDURE FOR ESTIMATION OF LINEAR MODELS VIA ASYMPTOTIC QUASI-LIKELIHOOD

RICCARDO BIONDINI, YAN-XIA LIN AND SIFA MVOI
School of Mathematics and Applied Statistics, University of Wollongong, Australia

Abstract. This paper is concerned with the application of an asymptotic quasi-likelihood practical procedure to estimate the unknown parameters in linear stochastic models of the form $y_t = f_t(\theta) + M_t(\theta)$ $(t = 1, 2, .., T)$, where f_t is a linear predictable process of θ and M_t is an error term such that $E(M_t|\mathcal{F}_{t-1}) = 0$ and $E(M_t^2|\mathcal{F}_{t-1}) < \infty$ and $\mathcal{F}_t$ is a σ-field generated from $\{y_s\}_{s \leq t}$. For this model, to estimate the parameter $\theta \in \Theta$, the ordinary least squares method is usually inappropriate (if there is only one observable path of $\{y_t\}$ and if $E(M_t^2|\mathcal{F}_{t-1})$ is not a constant) and the maximum likelihood method either does not exist or is mathematically intractable. If the finite dimensional distribution of M_t is unknown, to obtain a good estimate of θ an appropriate predictable process g_t should be determined. In this paper, criteria for determining g_t are introduced which, if satisfied, provide more accurate estimates of the parameters via the asymptotic quasi-likelihood method.

Keywords: asymptotic quasi-likelihood, linear models, Hurst parameter

1. Introduction

Linear models are very popular in practical situations. Examples of such applications may be found in Weisberg (1985) and references therein. In a similar manner we also limit our attention to linear models in this paper. In particular we concentrate on the following model:

$$y_t = f_t(\theta) + M_t(\theta), \qquad t \leq T,$$

where $f_t(\theta)$ is a linear function of θ and M_t is random error. Here $\{y_t\}$ could be a sequence of independent observations from a population or a sequence of observations from a stochastic process (in this case, we call it one realisation of the stochastic process). To estimate the parameter θ in this model, the ordinary least squares (OLS) method, in general, can provide a very good estimate subject to the $\{M_t\}$ being mutually uncorrelated and the variances of the $\{M_t\}$ being equal. However, the last condition does not always hold in reality. To reduce the effect from unequal variances of random error, a weighted least squares method is needed and proper weights need to be allocated. How to determine the weights becomes interesting. When $\{y_t\}$ is a sequence of independent observations from a population, we can sometimes estimate the weights via groups of observations. In each group of observations, all of the observations are associated with approximately the same value of $f_t(\theta)$ (see Weisberg, 1985). However, when $\{y_t\}$ is a path of realisation of a

process, it seems that it is not possible to use the same method to obtain an estimate of the weights. To cope with this problem a practical inference procedure, called the asymptotic quasi-likelihood method, is derived by Lin (1995). The asymptotic quasi-likelihood method given by Lin (1995) is different from that of Heyde and Gay (1989) although they share the same name. Lin (1995) also proved the asymptotic quasi-likelihood method is asymptotically optimal. Mvoi, Lin and Biondini (1998) prove the consistency of the asymptotic quasi-likelihood estimate for linear models. This paper will only focus on some techniques in applying the practical asymptotic procedure. For the theoretical discussion on the properties of the asymptotic quasi-likelihood method see Lin (1995) and Mvoi, Lin and Biondini (1998).

In Section 2, we can see in order to succeed in our application of the asymptotic quasi-likelihood method a predictable process, denoted by g_t, needs to be determined, which is used to help adjust the estimation procedure. In practice, for given $\{y_t\}$, several predictable processes can be determined. The choice of g_t, however, will affect the accuracy of the estimate of θ. In this paper, we will focus on the important question: how do we choose a proper g_t for a given data set. We show how the different choice of g_t affects inference results. To simplify our discussion, in this paper all g_t's are given in autoregressive form and determined by the Box-Jenkins method. Instead of focusing on the form of g_t we are interested in whether or not g_t satisfies certain criteria. Based on our experience, several criteria are presented in Section 3 to help choose a proper g_t. In Section 4, the criteria are applied to three models. In Section 5 one hundred simulations of each of these first two models are performed in order to assess their qualities.

2. The Asymptotic Quasi-likelihood Method

Assume that the observed process $\{y_t\}$ satisfies the model

$$y_t = f_t(\theta) + M_t(\theta), \tag{1}$$

where $t = 1, 2, .., T$, f_t is a predictable process, θ is an unknown parameter from an open parameter space Θ, $\mathcal{F}_t$ denotes a standard filtration generated from $\{y_s\}_{s \leq t}$, $\mathcal{F}_0 = \bigcap_{t=1}^{T} \mathcal{F}_t$ and M_t is an error process such that

$$E(M_t|\mathcal{F}_{t-1}) = 0, \tag{2}$$
$$E(M_t^2|\mathcal{F}_{t-1}) < \infty.$$

Equation (2) implies that M_i and M_j are uncorrelated, $i \neq j$. The case where the M_i are mutually independent with mean 0 is a special case of (2).

According to the quasi-likelihood method (see Godambe and Heyde, 1987) a quasi-score estimating function can be determined based on (1) and has the expression

$$G_T^*(\theta) = \sum_{t=1}^{T} \frac{\dot{f}_t(\theta)}{E(M_t^2|\mathcal{F}_{t-1})} M_t, \tag{3}$$

where $\dot{f}_t$ represents the derivative of f_t with respect to the unknown parameter θ.

The quasi-likelihood estimate of θ is obtained by solving the quasi-score normal equation $G_T^*(\theta) = 0$. When $f_t(\theta)$ is a linear function of θ, the quasi-likelihood method always provides a good estimate of θ subject to $E(M_t^2|\mathcal{F}_{t-1})$ being known. However, it is very difficult to accurately determine $E(M_t^2|\mathcal{F}_{t-1})$ and thus the expression for the quasi-score estimating function may not be determined. Therefore, a possible approach of the asymptotic quasi-likelihood method for linear models was discussed by Lin (1995) and Mvoi, Lin and Biondini (1998) and an inference procedure was given. The procedure is as follows; we accept the true model is that given by (1). If, for given y_t, we can determine a predictable process g_t such that $E(y_t^2 - g_t|\mathcal{F}_{t-1})$ is small enough for all t, then

$$
\begin{aligned}
E(M_t^2|\mathcal{F}_{t-1}) &= E((y_t - f_t)^2|\mathcal{F}_{t-1}) \\
&= E((y_t^2 - f_t^2)|\mathcal{F}_{t-1}) \\
&\approx g_t - f_t^2(\theta),
\end{aligned}
$$

and the asymptotic quasi-score estimating function

$$
\tilde{G}_T^*(\theta) = \sum_{t=1}^{T} \frac{\dot{f}_t(\theta)M_t}{g_t - f_t^2(\theta)}
$$

is obtained. The solution of the asymptotic quasi-score normal equation $\tilde{G}_T^*(\theta) = 0$ is called the asymptotic quasi-likelihood estimate. In practice, the asymptotic quasi-score normal equation that has been used is of the following form

$$
\tilde{G}_T^*(\theta) = \sum_{t=1}^{T} \frac{\dot{f}_t(\theta)M_t}{|g_t - f_t^2(\theta)|} = 0.
$$

Since $g_t - f_t^2(\theta)$ is used to estimate $E(M_t^2|\mathcal{F}_{t-1})$, and $E(M_t^2|\mathcal{F}_{t-1})$ is always positive, it is reasonable to use $|g_t - f_t^2(\theta)|$ rather than $g_t - f_t^2(\theta)$ in the quasi-score normal equation. A detailed discussion of this issue appears in the next section. For the new form of the quasi-score normal equation, Mvoi, Lin and Biondini (1998) have proved that, when $f_t(\theta)$ is a linear function of θ, under certain conditions, the asymptotic quasi-likelihood estimate is a good estimate of the true parameter. Also, this estimate is consistent as sample size is increasing.

To show how the asymptotic quasi-likelihood method can apply to real-life data, the following example is presented.

Example 1: This example utilises data obtained from a physics experiment given in Weisberg (1985, pp. 83-87). The experiment involves aiming a beam a, having various values of incident momentum p_a^{lab} which are measured in the laboratory frame of reference, at a target containing protons and results in the emission of other particles. The quantity measured y is the scattering cross-section of a particular particle. A quantity of more basic significance than p_a^{lab} is s, the square of the total energy in the centre-of-mass frame of reference system. The quantity s, under the conditions of the experiment is given by

$$
s = 2m_p p_a^{lab}.
$$

where s is measured in $(GeV)^2$, where $1GeV = 1 \times 10^9$ electron volts is the energy that an elementary particle reaches on being accelerated by an electric potential of one billion volts. The momentum p_a^{lab} and the mass m_p are measured in GeV, and $m_p = 0.938GeV$ for a proton.

Theoretical physicists believe that, under certain conditions (satisfied by this experiment), the cross-section y is given by the model

$$y = \theta_0 + \theta_1 s^{-1/2} + \text{relatively small terms.}$$

Table 1 summarises the results of the experiment. At each p_a^{lab}, a very large number of particles N_a was used so that the variance of the observed y values could be accurately estimated from theoretical considerations. The square root of these variances are given in the fourth column of Table 1. The best approach to estimate θ_0 and θ_1 is by using the weighted least squares (WLS) method subject to the estimated standard variances being known. Here we apply the asymptotic quasi-likelihood method to only one single path of realisation based on $g_t = 29705 + 856104(s^{-1/2})^2$ (the graphs of g_t and y_t^2 are shown in Figure 1). Although Weisberg does not state whether or not the data is ordered by time a look at the data will show that a strong linear relationship appears to exist between $s^{-1/2}$ and the estimated standard deviation and thus a time-ordered sequence may also appear plausible.

Table 1. Data for the Example 1.

p_a^{lab} GeV/c	$s^{-1/2}$ GeV/c^{-1}	y (μb)	Estimated Stand. Dev.
4	0.345	367	17
6	0.287	311	9
8	0.251	295	9
10	0.225	268	7
12	0.207	253	7
15	0.186	239	6
20	0.161	220	6
30	0.132	213	6
75	0.084	193	5
150	0.060	192	5

The estimates obtained via the asymptotic quasi-likelihood method are compared with those via the WLS and OLS methods and are given in Table 2. The asymptotic quasi-likelihood estimates are closer to the WLS estimates than the OLS estimates. This indicates that in the availability of only one realisation of data and lacking any knowledge regarding the nature of the error, the asymptotic quasi-likelihood method provides the best estimates.

The choice of g_t will affect the accuracy of the estimates of the unknown parameters. To obtain the g_t from given information of y_t^2, we have found that, in practice, the autoregressive technique discussed by Box et al (1994) is a simple method to use. One of the main reasons is that g_t obtained from autoregression is in a form of predictable function which we would like. Therefore, in the following, we are only concerned with those g_t which are obtained via autoregressive methods.
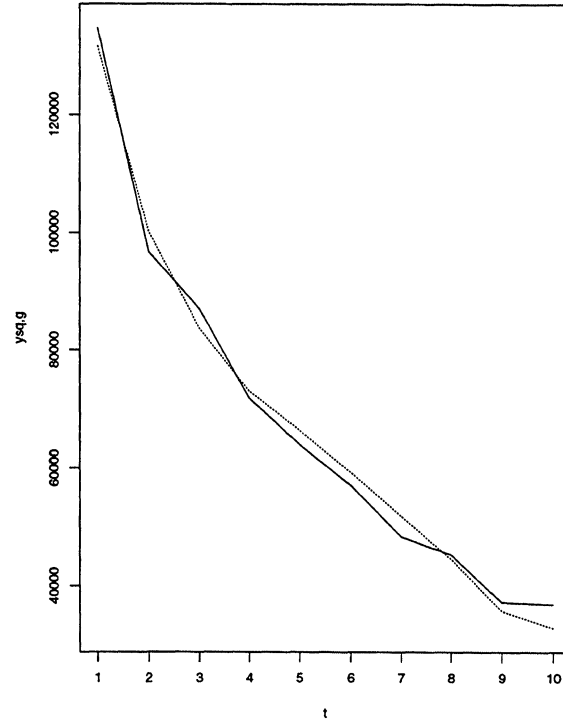
Figure 1. y_t^2 (hard line) and one possible g_t (dotted line) for Example 1.

Table 2. Weighted least squares, ordinary least squares and asymptotic quasi-likelihood estimates for Example 1.

Method	$\hat{\theta}_0$	$\hat{\theta}_1$
WLS	148.47	530.84
OLS	135.00	619.71
AQLM	148.93	537.46

3. Criteria in Selection of g_t

In this section we will list the criteria in selecting a predictable process g_t to approach y_t^2 and then provide the logic to each point listed. The criteria discussed here will be applied to three simulated models in the next section.

In practice, we found that the criteria in selecting a predictable process which yields accurate asymptotic quasi-likelihood estimates can be summarised as follows;

- Examine the time series plot of g_t and y_t^2, g_t should be chosen such that it is close to y_t^2.

- Examine the stationarity of ϵ_t (where $\epsilon_t = y_t^2 - g_t$), g_t should be chosen such that ϵ_t for that particular g_t is stationary. The correlation between g_t and ϵ_t should not be very large.

As mentioned previously, it would be expected that the better the g_t approaches y_t^2, the more accurate the asymptotic quasi-likelihood estimates of the parameters are likely to be. This is because of the relationship between these two quantities, i.e.

$$
\begin{aligned}
E(M_t^2|\mathcal{F}_{t-1}) &= E(y_t^2|\mathcal{F}_{t-1}) - f_t^2(\theta) \\
&\approx g_t - f_t^2(\theta).
\end{aligned}
$$

If g_t is close to y_t^2, then the predictable process g_t will be close to $E(y_t^2|\mathcal{F}_{t-1})$. Therefore $g_t - f_t^2(\theta)$ will be close to $E(M_t^2|\mathcal{F}_{t-1})$. In practice it is better to use $|g_t - f_t^2|$ to approach $E(M_t^2|\mathcal{F}_{t-1})$ since $E(M_t^2|\mathcal{F}_{t-1})$ is always positive. If, however, we cannot find a g_t such that $g_t - f_t^2(\theta)$ is close to $E(M_t^2|\mathcal{F}_{t-1})$ we may be able to find a g_t such that this g_t "mimics" y_t^2. By "mimics" we mean that the change in successive observations of y_t^2 should be subsequently accounted for by g_t. Therefore, even if the graph of g_t does not accurately approach y_t^2 it should model the pattern of the y_t^2, in these cases a positive constant may be added (or subtracted) to improve both the graphical approximation of y_t^2 by g_t and thus the corresponding asymptotic quasi-likelihood estimates.

In examining the adequacy of the predictable process g_t an analysis of the residuals ϵ_t should be carried out. Analysis of the autocorrelation and partial autocorrelation functions of the residuals is of utmost importance. It is preferred that minimal correlation between g_t and ϵ_t exists. Once the model is fitted, the residuals should resemble a white noise process, i.e. $\{\epsilon_t\}$ should be a sequence of uncorrelated random variables with constant mean (0 in this case) and constant variance and the autocorrelation and partial autocorrelation functions of $\{\epsilon_t\}$ should ideally be identically equal to 0.

In the process of obtaining the AQL estimate, as we mentioned before, the asymptotic quasi-score normal equation

$$
\tilde{G}_T^*(\theta) = \sum_{t=1}^{T} \frac{\dot{f}_t(\theta) M_t}{|g_t - f_t^2(\theta)|} = 0
$$

is considered. The two-stage estimation procedure is applied to $\tilde{G}_T^*(\theta) = 0$. The algorithm for the procedure is as follows:

(i) Start with preliminary estimates of $\hat{\theta}$. These initial values are determined via the method of ordinary least squares.

(ii) Substitute the $\hat{\theta}$ into the $g_t - f_t^2(\theta)$ in the asymptotic quasi-score normal equation and solve this equation for θ;

i.e. $\displaystyle \tilde{G}_T^*(\theta) = \sum_{t=1}^{T} \frac{\dot{f}_t(\theta)}{\left| g_t - f_t^2(\hat{\theta}) \right|} M_t = 0.$

(iii) Let the solutions obtained in step (ii) be the updated estimates of $\hat{\theta}$.

(iv) Repeat steps (ii) and (iii) until the sequence of estimates is convergent. However, sometimes due to $\left| g_t - f_t^2(\hat{\theta}) \right|$ being too close to zero the iteration may not converge quickly, we add a positive constant c to $\left| g_t - f_t^2(\hat{\theta}) \right|$. A discussion about this technique is given below.

According to the proof given by Mvoi, Lin and Biondini (1998), based on the AQL procedure, a good estimate of θ will be obtained if

$$\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2|}$$

is bounded. Therefore, there is a $k_1 > 0$ and $k_2 > 0$ such that

$$k_2 \leq \frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2|} \leq k_1.$$

The smaller the difference between k_1 and k_2, the quicker the iteration will converge. A problem arises when $|g_t - f_t^2|$ is too close to zero. To ensure the iteration converges in such circumstances, $|g_t - f_t^2|$ is replaced by $|g_t - f_t^2| + c$, where c is a suitable positive constant. Therefore, if a good estimate is expected to be obtained, the ratio

$$\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c} \tag{4}$$

should be bounded between two finite values for each t, c denoting a suitable positive constant. This constant must not be too large, further investigation is needed to determine the appropriate value of c. We may rewrite the above ratio in the following form;

$$
\begin{aligned}
\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c} &= \frac{E(y_t^2|\mathcal{F}_{t-1}) - f_t^2}{|g_t - f_t^2| + c} \\
&= \frac{g_t - f_t^2 + c}{|g_t - f_t^2| + c} + \frac{E(\epsilon_t|\mathcal{F}_{t-1}) - c}{|g_t - f_t^2| + c}.
\end{aligned} \tag{5}
$$

It is easily seen that if $\frac{E(\epsilon_t|\mathcal{F}_{t-1}) - c}{|g_t - f_t^2| + c}$ is bounded then $\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c}$ is also bounded because the quantity $\frac{g_t - f_t^2 + c}{|g_t - f_t^2| + c}$ is bounded between -1 and 1, i.e. $-1 \leq \frac{g_t - f_t^2 + c}{|g_t - f_t^2| + c} \leq 1$.

If $\left|g_t - f_t^2\right|$ is not negligible and $E(\epsilon_t|\mathcal{F}_{t-1})$ is not much larger than $\left|g_t - f_t^2\right|$, c is not needed for the ratio (4) to become bounded. If $\left|g_t - f_t^2\right|$ is negligible and $E(\epsilon_t|\mathcal{F}_{t-1}) >> \left|g_t - f_t^2\right|$, then from (5), a proper c can make

$$\frac{E(\epsilon_t|\mathcal{F}_{t-1}) - c}{|g_t - f_t^2| + c}$$

bounded, so the ratio (4) becomes bounded. The closer the ratio is to 1, the quicker the convergence of the asymptotic quasi-likelihood estimates and the more accurate these estimates will be (for theoretical details see Mvoi, Lin and Biondini, 1998). The resulting asymptotic quasi-likelihood estimates are found to be also very accurate providing the quasi-likelihood estimates are accurate. Even though in practical situations this ratio is unavailable, by the inclusion of a suitable positive constant c in the denominator in (4) we can assure that the ratio (4) for each t is much smaller than it would be if this constant was not taken into consideration. Since the ratio is no longer large this ensures convergence. It must be noted that this constant must not be too large as $\left|g_t - f_t^2\right| + c$ will be dominated by the value of c and the value of the asymptotic quasi-likelihood estimates will be closer to the ordinary least squares estimates.

As noted before, when the two-stage estimation procedure is applied, a suitable positive constant c may need to be added to $\left|g_t - f_t^2(\theta)\right|$. In this paper we suggest that the initial value of c be 0 and increase this value to 0.01 if the asymptotic quasi-likelihood estimates do not appear to converge. The value of c continues to increase by increments of 0.01 until the asymptotic quasi-likelihood estimates converge. It is important to note that c must be increased only to the point of immediate convergence, a choice of c which is significantly greater than that necessary for convergence might lead to the asymptotic quasi-likelihood estimates becoming closer to those obtained via the method of ordinary least squares and thus compromise the effectiveness of the procedure outlined here.

It does not appear that the sample mean of ϵ_t would help determine whether or not g_t adequately approaches y_t^2, the reason being that the sample mean for such a statistic will be very close to zero if the predictable process overestimates and then underestimates the true series y_t^2, the positive values of ϵ_t might "balance" out the negative values of ϵ_t thus producing a low mean value of ϵ_t.

4. Application of Criteria

If the quantity $E(M_t^2|\mathcal{F}_{t-1})$ is known, the quasi-likelihood method can be used to estimate $\hat{\theta}$ (see (3)). In practice this quantity is unknown and thus must be estimated. If $E(M_t^2|\mathcal{F}_{t-1})$ is non-constant, ordinary least squares is not a suitable method for estimating the unknown parameters. In this section we will firstly apply the procedure to three examples. It is realised that it is better to apply the technique to real-life data than to artificial data. However the comparison between the estimated parameter and the true parameter value becomes very difficult when

such data is used since the true parameter values are unknown. In the next section we also list the estimates of the parameters via the asymptotic quasi-likelihood method over one hundred simulations and the standard errors of the corresponding estimates and compare them to the estimates obtained via the ordinary least squares method.

As an application of the asymptotic quasi-likelihood method we will apply the procedure in this section to estimate the parameter d in a fractional ARIMA process X_t. A standard procedure used for estimating d is to use Hurst's R/S-statistic and to estimate the Hurst parameter H. The estimate of d can be obtained via the estimate of H since $H = d + 1/2$.

The estimation procedure is briefly described as follows; for a given set of observations $(X_k : k = 1, 2, ..., n)$ with sample mean $\bar{X}(n)$ and sample variance $S^2(n)$, the modified rescaled adjusted range or the modified R/S-statistic (see Lo, 1991) is given by

$$R(n)/S(n) = 1/S(n)[\max(0, W_1, W_2, ..., W_n) - \min(0, W_1, W_2, ..., W_n)],$$

with $W_k = (X_1 + X_2 + ... + X_k) - k\bar{X}(n)$, $k = 1, 2, ..., n$ and $S(n)$ being the sample standard deviation of $X_1, X_2, ..., X_n$.

In our application of R/S analysis the window length n is doubled each time the R/S-statistic is calculated. Hurst (1965) found that many naturally occurring time-series appear to be well represented by the relation

$$E[R(n)/S(n)] = bn^H, \qquad \text{as } n \to \infty,$$

with Hurst parameter $0 < H < 1$, and b is a finite positive constant that does not depend on n.

A straight line is plotted in the log-log scale:

$$log[R(n)/S(n)] = c + H log(n) + \epsilon(n) \tag{6}$$

and the estimate of H is obtained via the ordinary least squares method.

We note the similarity between (5) and (1). The standard method for estimating H is based on the assumption that the $\{\epsilon(n)\}$ are independently and identically distributed. However, this assumption is not necessarily true based on the R/S estimation procedure mentioned above. Therefore it would be expected that the ordinary least squares method may not provide a good estimate of H. In this paper, instead of applying the method of ordinary least squares we apply the asymptotic quasi-likelihood method to estimate H.

The quantity $log[R(n)/S(n)]$ is then plotted against $log(n)$. This plot is known as the rescaled adjusted range plot (also called the pox diagram of R/S). A typical rescaled adjusted range plot commences with a transient region representing the short-range dependence structure in the sample (in this transient region the R/S-statistic grows faster than $n^{0.5}$ for small n than it does for relatively larger n). That is why we must examine the accuracy of the estimates when different initial window sizes are considered. With smaller initial window sizes the power of R/S analysis may be severely compromised. This behaviour, however, eventually settles

down and fluctuates in a straight "street" with a certain asymptotic slope. For a fractional Gaussian process this graph should have slope H as n increases.

In this section we apply the asymptotic quasi-likelihood method to (5) for data simulated from a fractional ARIMA(2,0.3,0) model and we also examine the affect of selecting different initial window sizes. Also we apply the asymptotic quasi-likelihood method to stochastic processes represented by model (1).

For a comparison of the accuracy of both the asymptotic quasi-likelihood method, the quasi-likelihood method and the method of ordinary least squares we introduce a quantity S defined as

$$S = \sqrt{(\theta_1 - \hat{\theta}_1)^2 + ... + (\theta_p - \hat{\theta}_p)^2}$$

where $\theta = (\theta_1, ..., \theta_p)^T$ is an unknown parameter and $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_p)$ is an estimate of θ. Since this value is determined by the true value of the parameters, it is a quantity which is unable to be calculated in practice.

To demonstrate the criteria listed in Section 3 and show how these criteria are to be applied, three simulated data sets are analysed in this section.

Example 2: Data is generated from the following model:

$$y_t = 0.3 + 0.5y_{t-1} + M_t, \quad t \geq 2,$$

where $M_t = N_t - 0.5(y_{t-1}^2 + y_{t-2}^2)$ and N_t has the Poisson distribution with rate $0.5(y_{t-1}^2 + y_{t-2}^2)$. Therefore M_t is a martingale difference.

For this random sample of 240 data values four different predictable processes g_t are analysed, the first g_t is that based on the ARIMA(1,0,0) model (denoted by g_{1t}), the second based on the ARIMA(1,1,0) model (denoted by g_{2t}), the third g_t analysed is that based on the ARIMA(2,0,0) model (denoted by g_{3t}) and the last is based on the ARIMA(2,1,0) model (denoted by g_{4t}). The predictable processes are listed below:

$$g_{1t} = 6.743 + 0.193y_{t-1}^2,$$

$$g_{2t} = 0.517y_{t-1}^2 + 0.483y_{t-2}^2,$$

$$g_{3t} = 5.864 + 0.168y_{t-1}^2 + 0.131y_{t-2}^2,$$

$$g_{4t} = 0.340y_{t-1}^2 + 0.291y_{t-2}^2 + 0.368y_{t-3}^2.$$

These predictable processes are analysed because the coefficients of each g_t are all significant. The predictable processes g_{2t} and g_{4t} are examined because it appears as though y_t^2 is non-stationary from the time-series plots, we know this is true from our generation of the data. The plots of the autocorrelation and partial autocorrelation functions of y_t^2 however do not reveal any significant non-stationarity. The second and fourth predictable processes g_{2t} and g_{4t} (see Figure 3) are found to graphically approach the quantity y_t^2 better than when the two other predictable processes are chosen. The first and third predictable processes also appear to be good if the constant is not taken into consideration but rather only the y_{t-1}^2 term. If this is done, however, the predictable process g_t will be poor at approaching y_t^2 at large values of y_t^2.
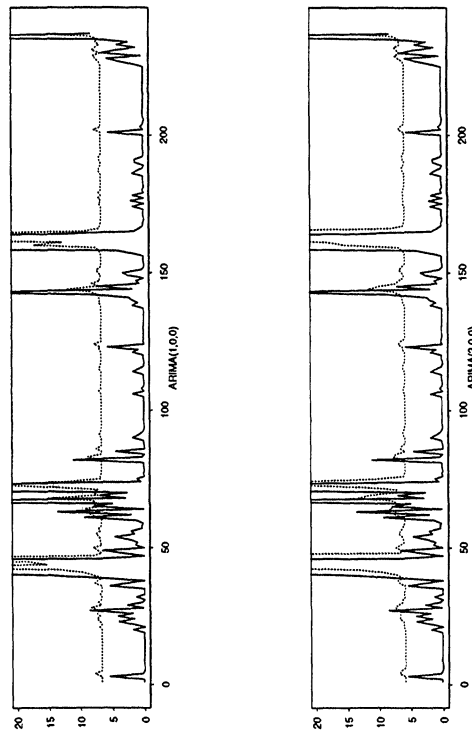
Figure 2. y_t^2 (hard line) and two possible g_t's (dotted line) for Example 2.

From Figure 2 the ARIMA(1,0,0) model is not good at approaching y_t^2 if the constant term is taken into account. Similarly it can be seen that the ARIMA(2,0,0) model is also not good at approaching y_t^2. The $\{\epsilon_t\}$ are stationary for each of the four g_t's. From Table 3 it is seen that the most accurate asymptotic quasi-likelihood estimates occur when the second predictable process is used. The S-values for the asymptotic quasi-likelihood estimates when each predictable process is considered are 0.111, 0.009, 0.080 and 0.092 respectively. As we shall see in the next section the ARIMA(1,1,0) model (i.e. g_{2t}) is always better at graphically approaching y_t^2 and in the vast majority of simulations provides a much better approximation of the unknown parameters than a predictable process which does not consider possible non-stationarity. In this simulation the S-value associated with the estimates obtained when the ARIMA(1,1,0) process is used is much smaller than the S-value obtained when each of the other predictable processes are used.
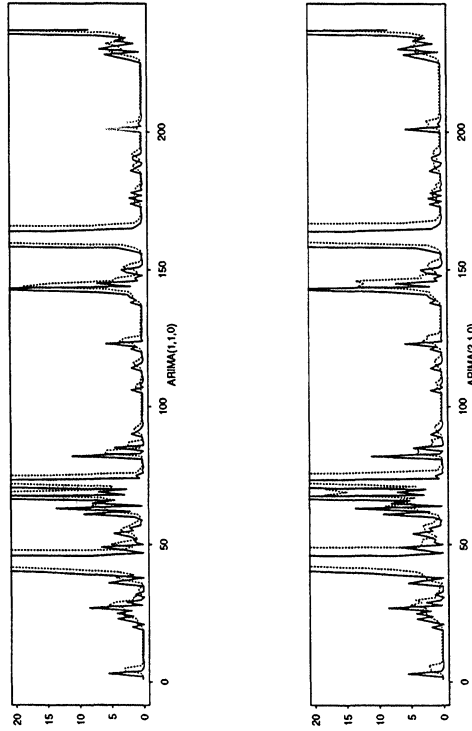
Figure 3. y_t^2 (hard line) and another two possible g_t's (dotted line) for Example 2.

The S-value surprisingly is much smaller than when the quasi-likelihood method is invoked.

Table 3. Ordinary least squares, quasi-likelihood and asymptotic quasi-likelihood estimates (for four possible predictable processes) for Example 2.

Method	$\hat{\theta}_0$	$\hat{\theta}_1$	S-value
OLS	0.204	0.445	0.111
AQLM(g_1)	0.210	0.435	0.111
AQLM(g_2)	0.292	0.504	0.009
AQLM(g_3)	0.315	0.421	0.080
AQLM(g_4)	0.378	0.451	0.092
QLM	0.330	0.598	0.103

Example 3: Data is generated from the following model:

$$y_t = 0.2 + 0.6y_{t-1} + 0.8x_t + M_t, \quad t \geq 2,$$

where M_t, given $\{y_s\}_{s \leq t}$, is generated from the normal distribution with mean 0 and variance $0.3(y_{t-1}^2 + y_{t-2}^2)$, i.e. $E(M_t|\mathcal{F}_{t-1}) = 0$ and $E(M_t^2|\mathcal{F}_{t-1}) = 0.3(y_{t-1}^2 + y_{t-2}^2)$. x_t is a standard normal random variable and is independent of $\{y_s\}_{s \leq t}$.

For this random sample of 240 data values two different predictable processes are analysed, the first g_t is based on the ARIMA(1,0,0) model (denoted by g_{1t}) and the second is based on the ARIMA(1,1,0) model (denoted by g_{2t}). The predictable processes are listed below;

$$g_{1t} = 2.059 + 0.539y_{t-1}^2,$$
$$g_{2t} = 0.603y_{t-1}^2 + 0.397y_{t-2}^2.$$

These predictable processes are analysed because the coefficients of each g_t are all significant. The predictable process g_{2t} is examined because it appears as though y_t^2 is non-stationary from the time-series plots. The second predictable process g_{2t} is found to graphically approach the quantity y_t^2 better than if the other predictable process (g_{1t}) is chosen.

We wish to compare both models with the first order autoregressive component, that with the differencing term (g_{1t}) and that without the differencing term (g_{2t}). From Figure 4 the latter predictable process is found to approach the quantity y_t^2 better than if the former predictable process is chosen. The $\{\epsilon_t\}$ are stationary when either g_{1t} and g_{2t} are selected.

The asymptotic quasi-likelihood estimates when each of the two g_t's are chosen are shown in Table 4. However, the estimates when g_t is based on the ARIMA(1,0,0) model result in an S-value of 0.234 whereas the S-value, when the second predictable process is selected, is equal to 0.054. The estimates of each parameter can be seen to be much more accurate when the second predictable process is used. The first predictable process in this particular simulation even produces an S-value which is greater than that obtained via the method of ordinary least squares.

Table 4. Ordinary least squares, quasi-likelihood and asymptotic quasi-likelihood estimates (for two possible predictable processes) for Example 3.

Method	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	S-value
OLS	0.274	0.552	0.677	0.152
AQLM(g_1)	0.348	0.530	0.629	0.234
AQLM(g_2)	0.200	0.590	0.747	0.054
QLM	0.219	0.657	0.796	0.060

The ordinary least squares method yields estimates of 0.274, 0.552 and 0.678 respectively which results in an S-value of 0.152. On the other hand the estimates via the quasi-likelihood method are 0.219, 0.657 and 0.796 respectively and the resultant S-value is 0.060. The asymptotic quasi-likelihood estimates when g_{2t} is chosen as the predictable process are very similar to the quasi-likelihood estimates.
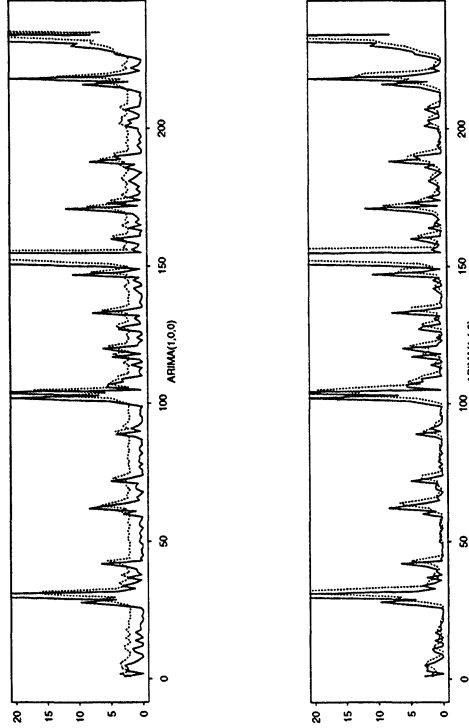
Figure 4. y_t^2 (hard line) and two possible g_t's (dotted line) for Example 3.

Example 4: Our analysis now turns to the simulating of data from the fractional autoregressive integrated process or fractional ARIMA(2,0.3,0) process. The ARIMA model is specified as

$$(1 - 0.2B - 0.6B^2)(1 - B)^{0.3}Y_t = h_t,$$

where B denotes the backward shift operator.

The process is stationary and we want to estimate the value of H, which from the selection of d we know to be 0.8 (since $H = d + 1/2$). The data was analysed using the R/S-statistic and the estimate of the Hurst parameter is obtained via both the method of ordinary least squares and the asymptotic quasi-likelihood method.

There are 10,000 values generated for this simulation. By applying R/S analysis to the data we transform the 10,000 data values to twelve data points. Three g_t's are determined based on this sample of twelve data points. The predictable

processes are listed below;

$$g_{1t} = -0.743 - 0.312y_{t-1}^2 + 0.850(logn_t)^2,$$
$$g_{2t} = -1.426 - 0.477y_{t-1}^2 - 0.427y_{t-2}^2 + 1.172(logn_t)^2,$$
$$g_{3t} = -1.714 - 0.536y_{t-1}^2 - 0.486y_{t-2}^2 - 0.094y_{t-3}^2 + 1.282(logn_t)^2.$$

From Figure 5 it can be seen that g_{1t} is not as good at approaching y_t^2 as the other two predictable processes. It can be seen that there is very little difference between g_{2t} and g_{3t}. Turning our attention to the $\{\epsilon_t\}$, they appear to be stationary for each of the three g_t's. From Table 5 it is seen that the most accurate asymptotic quasi-likelihood estimate occurs when the third predictable process is used (0.835) followed by g_{2t} (0.870). When the first predictable process is chosen the resulting estimate is 0.975. The ordinary least squares estimate is 0.928 (when the initial window size is 4).
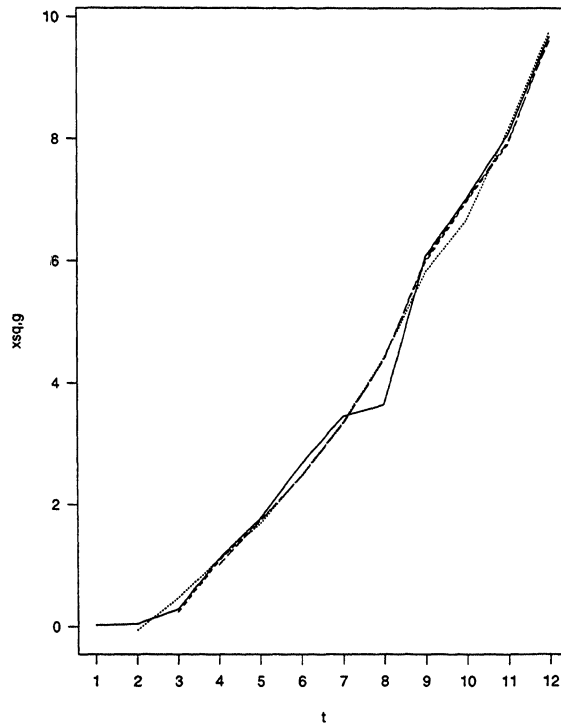

Figure 5. y_t^2 (hard line) and three possible g_t's (dotted lines) for Example 4. g_{1t} starts at the first lag, g_{2t} starts at the second lag and g_{3t} starts at the third lag.

We now wish to reduce any bias which may exist from the initial transient region of the R/S plot (as mentioned previously). We would like to do this as the emphasis is not on the short-range dependence structure but on the nature of long range dependence in the sample. Mandelbrot and Taqqu (1979) suggest $n \approx 10$ as the minimal lag to use. The question is at what initial window size will the resulting estimates be more accurate? We examined the results when the initial window size is 4, 8 and 16 respectively. It seems as though there is minimal improvement in the estimates in this particular example. When the initial window sizes are 8 and 16 respectively the ordinary least squares estimate of H becomes 0.938 and 0.903 respectively. This finding also illustrates the power of the asymptotic quasi-likelihood method; resulting estimates will not be unduly affected even with small initial window lengths. The method works well notwithstanding the observed values in the transient region of the sample.

Table 5. Ordinary least squares (for three different initial window sizes) and asymptotic quasi-likelihood estimates (for three possible predictable processes) for Example 4.

Method	$\hat{H}$
OLS(4)	0.928
OLS(8)	0.938
OLS(16)	0.903
AQLM(g_1)	0.975
AQLM(g_2)	0.870
AQLM(g_3)	0.834

5. Simulations

In this section one hundred simulations are performed via each of the first two models mentioned in Section 4 (i.e. Examples 2 and 3). We now wish to examine the results.

The ordinary least squares and asymptotic quasi-likelihood estimates for the simulations based on the model in Example 2 are included in Table 6. From the table it can be seen that the ordinary least squares estimates are very inaccurate for both parameters. The mean value of the estimates of $\hat{\theta}_0$ and $\hat{\theta}_1$ over one hundred simulations via the asymptotic quasi-likelihood method (when the second predictable process is used to approach y_t^2) are very accurate (0.296 and 0.493 for the with standard errors of 0.004 and 0.009 respectively. In the method of ordinary least squares, however, the mean of the estimates of the two unknown parameters are

0.329 and 0.441 respectively. The corresponding standard errors of these estimates are 0.012 and 0.013 respectively.

The asymptotic quasi-likelihood method based on the remaining predictable processes also yields accurate estimates. In each and every simulation g_t approaches y_t^2 extremely well when we consider the predictable processes with a differencing term $(g_{2t}$ and $g_{4t})$. These predictable processes also provide the most accurate estimates.

Table 6. Ordinary least squares and asymptotic quasi-likelihood estimates (with standard errors in brackets) for one hundred simulations of Example 2.

Method	mean$\hat{\theta}_0$(st. error $\hat{\theta}_0$)	mean$\hat{\theta}_1$(st. error $\hat{\theta}_1$)	$\bar{S}$-value
LS	0.329(0.012)	0.441(0.013)	0.019
AQLM(g_1)	0.318(0.006)	0.462(0.011)	0.013
AQLM(g_2)	0.296(0.004)	0.493(0.009)	0.009
AQLM(g_3)	0.317(0.006)	0.466(0.011)	0.013
AQLM(g_4)	0.306(0.005)	0.486(0.011)	0.012

It is obvious from these results that the method of ordinary least squares is very inaccurate and since the quasi-likelihood method is unable to be applied the asymptotic quasi-likelihood method is very effective providing an appropriate g_t is chosen.

$\bar{S}$ is defined as the average of the S-values from the one hundred simulations, the S-value being previously defined. The $\bar{S}$-value for the estimates via the ordinary least squares method is 0.019. The corresponding $\bar{S}$-value for each of the four predictable processes are 0.013, 0.009, 0.013 and 0.012 respectively. The second predictable process, g_{2t}, always approaches y_t^2 better than any other possible g_t. The $\bar{S}$-value associated with g_{2t} is over two times smaller than that obtained via ordinary least squares. The other three predictable processes also yield smaller $\bar{S}$-values than those obtained via the traditional ordinary least squares method.

The ordinary least squares and asymptotic quasi-likelihood estimates for the simulations based on the model in the Example 3 are included in Table 7. From the table it can be seen that the ordinary least squares estimates are not as accurate as the asymptotic quasi-likelihood estimates when taking the mean of the one hundred sets of estimates. The estimates of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ when the method of ordinary least squares is invoked are 0.211, 0.541 and 0.844 respectively. The mean values of the estimates of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ over one hundred simulations via the asymptotic quasi-likelihood method when the predictable process g_{2t} is chosen are 0.221, 0.584 and 0.797 respectively. The asymptotic quasi-likelihood estimates based on the predictable process g_{1t} also yield more accurate estimates than the method of ordinary least squares. The mean of the estimates of the parameters when any predictable process is chosen is within 0.02 of the true values of the unknown parameters but when ordinary least squares is invoked only the mean of the estimates of $\hat{\theta}_0$ lie within 0.04 of the true value of the parameters.

The graphs of y_t^2 and g_t are useful. By observing these graphs we can obtain information as to which predictable process will produce the most accurate estimates

Table 7. Ordinary least squares and asymptotic quasi-likelihood estimates (with standard errors in brackets) for one hundred simulations of Example 3.

Method	mean$\hat{\theta}_0$(st. error$\hat{\theta}_0$)	mean$\hat{\theta}_1$(st. error$\hat{\theta}_1$)	mean$\hat{\theta}_2$(st. error$\hat{\theta}_2$)	$\bar{S}$-value
LS	0.211(0.010)	0.541(0.009)	0.844(0.010)	0.018
AQLM(g_1)	0.205(0.010)	0.581(0.011)	0.781(0.012)	0.019
AQLM(g_2)	0.191(0.007)	0.591(0.009)	0.808(0.010)	0.015

of the unknown parameters. If g_t approaches y_t^2 very well then the asymptotic quasi-likelihood estimates will be much more accurate than if this approximation was not as good. How accurate these estimates will be will depend on the accuracy of the corresponding quasi-likelihood estimates. Obviously, if these quasi-likelihood estimates are not precise the corresponding asymptotic quasi-likelihood estimates will, more than likely, be not as precise as they would be if the quasi-likelihood estimates were accurate.

For the higher order models the approximation of y_t^2 by g_t was not as good as it was for the lower order models. This could well be due to the fact that the y_t is generated to be dependent of y_{t-1} and not dependent upon the observation of y at higher order lags. However, lower order processes will be more respondent to outliers than higher order processes and therefore place all weight upon the previous observation of y_t^2. Higher order models will tend to spread the weight upon previous values (e.g. the ARIMA(1,1,0) model will consider only the y_{t-1}^2 and y_{t-2}^2 terms whereas an ARIMA(2,1,0) process will consider the y_{t-1}^2, y_{t-2}^2 and y_{t-3}^2 terms). Higher order autoregressive processes will therefore produce a "smoothing" effect on the predictable process, thus it is preferred that a lower order g_t be chosen. Taking into account that y_t^2 may possibly be non-stationary is important when selecting g_t. The graphs of g_t and y_t^2 seem to be very similar when the g_t is based on the ARIMA(1,1,0) model. However, even if the g_t is selected by not taking into consideration any possible non-stationary the estimates via asymptotic quasi-likelihood will usually be better than the ordinary least squares estimates.

6. Conclusions

The method of ordinary least squares does not yield accurate estimates of the unknown parameters when the residuals are not uncorrelated and/or do not have equal variance. The quasi-likelihood estimates are accurate when the form of $E(M_t^2|\mathcal{F}_{t-1})$ is known, which in practical situations is not true. In Example 2 the asymptotic quasi-likelihood method is much better than the method of ordinary least squares. It appears that if the residuals do not have a common variance the method of ordinary least squares will be very inaccurate indeed. In Example 3 the ordinary least squares estimates are once again inaccurate (though not as inaccurate as they were in Example 2 where the errors are observations from the Poisson distribution). The estimates obtained via the asymptotic quasi-likelihood method will be much more accurate than the ordinary least squares estimates for

this particular model. The accuracy of the estimates can be improved by applying the asymptotic quasi-likelihood procedure. From our practical procedure since the asymptotic quasi-likelihood method takes account of the effect from errors, in general, this method will always produce a better estimate of the parameters than the method of ordinary least squares if the errors are not independently and identically distributed.

The quality of improvement by applying the asymptotic quasi-likelihood method can also be maximised by careful selection of a predictable process g_t. If this g_t approaches the quantity y_t^2 very well (we must also check ϵ_t as well) then the asymptotic quasi-likelihood estimates will be more accurate. That is why this selection is very critical. We can improve on the ordinary least squares estimates in most cases but to maximise this improvement we must choose the best possible g_t.

Furthermore, when a positive constant c is added the asymptotic quasi-likelihood estimates will converge immediately as the ratio $\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c}$ (which is unable to be calculated in practice) is closer to 1. The positive constant c is of utmost importance as it leads to convergent asymptotic quasi-likelihood estimates. By doing so we ensure the ratio is bounded even if we are unable to calculate it due to the form of $E(M_t^2|\mathcal{F}_{t-1})$ being unknown.

References

1. G.E.P. Box, G.M. Jenkins and G.C. Reinsel. Time Series Analysis, Forecasting and Control, 3rd Edition, Prentice-Hall Inc., Sydney, 1994.
2. V.P. Godambe and C.C. Heyde. Quasi-likelihood and Optimal Estimation, International Statistical Review, 55, 231-244, 1987.
3. C.C. Heyde and R. Gay. On Asymptotic Quasi-likelihood Estimation, Stochastic Processes Applications, 31, 223-236, 1989.
4. H.E. Hurst, R.P. Black and Y.M. Simaika. Long Term Storage: An Experimental Study, Constable, London, 1965.
5. Y-X. Lin. On Asymptotic Quasi-score Functions and its Applications, Australian National University Scientific Research Report, SRR 024-95, 1995.
6. A.W. Lo. Long-Term Memory In Stock Market Prices, Econometrica, 59(5), 1279-1313, 1991.
7. B.B. Mandelbrot and M.S. Taqqu. Robust R/S Analysis of Long-Run Serial Correlation, Proceedings of the 42nd Session of the International Statistical Institute Manila, Bulletin of the International Statistical Institute, 48(2), 69-104, 1979.
8. S. Mvoi, Y-X. Lin and R.F. Biondini. Consistency of the Asymptotic Quasi-likelihood Estimate on Linear Models, Biometrical Journal, 40, 57-78, 1998.
9. S. Weisberg. Applied Linear Regression, 2nd Edition, Wiley Inc., New York, 1985.