2011

Initial values in estimation procedures for State Space Models (SSMs)

Raed Alzghool
Al-Balqa Applied University

Yan-Xia Lin
University of Wollongong, yanxia@uow.edu.au

Publication Details
Initial values in estimation procedures for State Space Models (SSMs)

Abstract
In this paper, we will focus on State Space Models (SSMs), especially the stochastic volatility model, and look for a standard approach for assigning initial values in the Quasi-Likelihood (QL) and Asymptotic Quasi-Likelihood (AQL) estimation procedures.

Keywords
Initial, values, estimation, procedures, for, State, Space, Models, SSMs

Disciplines
Physical Sciences and Mathematics

Publication Details
Initial Values in Estimation Procedures for State Space Models (SSMs)

Raed Alzghool and Yan-Xia Lin

Abstract—In this paper, we will focus on State Space Models (SSMs), especially the stochastic volatility model, and look for a standard approach for assigning initial values in the Quasi-Likelihood (QL) and Asymptotic Quasi-Likelihood (AQL) estimation procedures.

Index Terms—State Space Models (SSMs), Quasi-Likelihood (QL), Asymptotic Quasi-Likelihood (AQL), Kalman filter, Non-linear and/or Non-Gaussian SSMs.

I. INTRODUCTION

The class of state space models (SSM) provides a flexible framework for describing a wide range of time series in a variety of disciplines. For extensive discussion on SSM and their applications see Harvey [16] and Durbin and Koopman [13]. A state-space model can be written as

$$y_t = f_1(\alpha_t, \theta) + h_1(y_{t-1}, \theta)\varepsilon_t, \ t = 1, 2, \ldots, T$$

where \(y_1, \ldots, y_T\) represent the time series of observations; \(\theta\) is an unknown parameter that needs to be estimated; \(f_1(.)\) is a known function of state variable \(\alpha_t\) and \(\theta\); and \(\{\varepsilon_t\}\) are uncorrelated disturbances with \(E_{t-1}(\varepsilon_t) = 0, \text{Var}_{t-1}(\varepsilon_t) = \sigma^2_{\varepsilon}\); in which \(E_{t-1}, \text{and Var}_{t-1}\) denote conditional mean and conditional variance associated with past information updated to time \(t-1\) respectively. State variables \(\alpha_1, \ldots, \alpha_T\) are unobserved and satisfy the following model

$$\alpha_t = f_2(\alpha_{t-1}, \theta) + h_2(\alpha_{t-1}, \theta)\eta_t, \ t = 1, 2, \ldots, T,$$

where \(f_2(.)\) is a function of past state variables and \(\theta\); \(\{\eta_t\}\) are uncorrelated disturbances with \(E_{t-1}(\eta_t) = 0, \text{Var}_{t-1}(\eta_t) = \sigma^2_{\eta}\); \(h_1(.)\) and \(h_2(.)\) are unknown functions.

One special application that we will consider in detail is the Stochastic Volatility Model (SVM), a frequently used model for returns of financial assets. Applications, together with estimation for SVM, can be found in Jacquier, et al [45]; in which \(\eta_t\) might not available. The AQL approach provides an alternative method of parameter estimation when unknown form of heteroscedasticity is present.

The estimation procedure for SSMs consists of two parts. The first part is, given observations \(\{y_1, \ldots, y_T\}\), to estimate state variables \(\alpha_t\). The second part is to combine the information of \(\{y_t\}\) and \(\{\alpha_t\}\) to estimate unknown parameter \(\theta\) in the model. The Kalman filter and the smoother methods are widely used to estimate an unobservable series, state variables, in SSMs (Anderson and Moore [7], Harvey [17]). In summary, the QL and AQL estimating procedures discussed in Alzghool and Lin (12),[3], [5], Alzghool [4], and Alzghool, et al [6], consist of the following steps:

(i) Assign initial values to \(\alpha_0, \theta_0\) and \(\Sigma_0 = I\).

(ii) Obtain the QL/AQL estimates \(\hat{\alpha}_t\) of \(\alpha_t\) for \(t = 1, 2, \ldots, T\).

(iii) For the AQL estimation procedure, obtain \(\hat{\Sigma}_{t,n}\) by using the kernel method.

(iv) Obtain the QL/AQL estimate \(\hat{\theta}\) of \(\theta\).

(v) Steps (ii), (iii) and (iv) will be alternatively repeated until estimates converge.

The final estimation results for SSMs might be jointly affected by the initial values \(\alpha_0\) and \(\theta_0\) which initially assigned to the underlying model during the inference procedure. In this paper, following two issues are investigated.

(1) How sensitive are the final estimates to the initial values assigned to the state variable \(\alpha_0\) and \(\theta_0\)?

(2) If the estimation results are sensitive to the choice of the initial values, what should initial value of the state variable \(\alpha_0\) be and how is the final estimate of \(\theta\) determined?

This paper is structured as follows. In Section II, the sensitivity of the QL and AQL estimation procedures to the initial values assigned to state variable \(\alpha_0\) is investigated...
via simulation studies. In Section III, a new suggestion for choosing the initialisation of the state variable \( \alpha_0 \) is given. In Section IV, the impact of the starting values of system parameters \( \theta \) in the estimation results is investigated via simulation studies. In Section V, a standard procedure to improve the grid search procedure for obtaining a better estimation of \( \theta \) is established. In Section VI applications of the QL and AQL methods to real data modelled by SSMs are given. In Section VII, a conclusion is provided.

II. EFFECT OF INITIALISATION OF \( \alpha_0 \)

The impact of the initial value of the state variable \( \alpha_0 \) on the final inference result is illustrated via simulation studies in this section. Simulation study based on stochastic volatility model (SVM) is presented below.

A. Stochastic Volatility Models (SVM)

Consider the stochastic volatility model,

\[
\ln(y_t^2) = \alpha_t + \ln \xi_t^2, \quad t = 1, 2, \ldots, T,
\]

\[
\alpha_t = \gamma + \phi\alpha_{t-1} + \eta_t, \quad t = 1, 2, \ldots, T,
\]

where both \( \xi_t \) and \( \eta_t \) are i.i.d. r.v.'s; \( \eta_0 \) has mean 0 and variance \( \sigma^2_\eta \).

In order to show how the initial value \( \alpha_0 \) effects the final estimation in the SVM when the QL and AQL approaches are applied, we carried out a simulation study on SVM Model defined by (3) and (4). The simulation was conducted as follows. First, 1,000 independent samples of size 500 are generated from (3) and (4) based on a true parameter \( \theta = (\gamma, \phi) \), where \( \eta_t \sim N(0, \sigma^2_\eta) \), \( \xi_t \sim N(0, 1) \), and the initial value for \( \alpha_0 \) in the true model is \( \alpha_0 = 0 \). Once \( \{y_t\} \) and \( \{\alpha_t\} \) are generated, pretend that \( \{\alpha_t\} \) is unobserved and \( \gamma \), and \( \phi \) are unknown. Then apply the QL and AQL estimation procedures to \( \{y_t\} \) only to obtain the estimate of \( \alpha_t \), \( \gamma \), and \( \phi \). Different parameter settings for \( (\gamma, \phi, \sigma^2_\eta) \) are considered in the simulation. The mean and root mean squared errors for \( \hat{\gamma} \) and \( \hat{\phi} \) based on 1,000 independent samples are calculated.

Let \( \hat{\alpha}_0 \) be the initial value used in the inference procedure. In Table I, different values of \( \hat{\alpha}_0 \), mean and root mean squared errors for \( \hat{\gamma} \), and \( \hat{\phi} \) given by the QL and AQL methods are reported.

We can see from Table I that the RMSE of QL and AQL estimates are increased when \( \hat{\alpha}_0 \) is chosen farther from the true value \( \alpha_0 \). Since the increase in the RMSE for QL is less than for AQL, this indicates that the QL approach is less sensitive to the initial value of state variable than the AQL approach.

III. DETERMINATION OF \( \hat{\alpha}_0 \)

Consider the univariate time series \( y_t \) satisfying

\[
y_t = \alpha_t + \epsilon_t, \quad t = 1, 2, \ldots, T
\]

\[
\alpha_t = \alpha_{t-1} + \eta_t, \quad t = 1, 2, \ldots, T
\]

where \( \epsilon_t \sim N(0, \sigma^2_\epsilon) \), \( \eta_t \sim N(0, \sigma^2_\eta) \), and \( \alpha_0 \sim N(0, P_0) \). \( \{\epsilon_t\} \) and \( \{\eta_t\} \) are two independent Gaussian white noise series. The initial value \( \alpha_0 \) is independent of \( \{\epsilon_t\} \) and \( \{\eta_t\} \) for \( t > 0 \). In literature, \( \alpha_t \) is referred to as the trend of the series, which is not directly observable, and \( y_t \) is observable. The model is called a local level model in Durbin and Koopman ([13], Chapter 2), which is a simple case of the structural time series model of Harvey [17].

When nothing is known about the initial value \( \alpha_0 \), the initialisation of \( \alpha_0 \) is usually given by a diffuse prior approach that fixes \( \alpha_0 \) at an arbitrary value and let \( P_0 \rightarrow \infty \) (Zivot et al. [30], Durbin and Koopman [13], Harvey [16]). However, some researchers consider that the diffuse approach is not realistic because they regard that the assumption of infinite variance is unnatural, given that all observed time series have finite values. From this point of view an alternative approach is suggested, which assumes that \( \alpha_0 \) is an unknown constant and needs to be estimated from the data. In Harvey [18], it is suggested that the initial value of \( \alpha_0 \) can be taken as \( y_1 \). This is the same value as that obtained by assuming that \( \alpha_0 \) is diffuse. More details about the initialisation of the Kalman filter under the normality assumption for SSM are provided in Durbin and Koopman ([13], Chapter 5 and references therein). Several other suggestions on initialisation for the state variable in SSM under normality assumption are given in a recent survey by Casals and Sotoca [9]. They derived an exact expression for the conditional mean and variance of the initial state of SSM.

In this paper, we follow the QL method to derive a simple method for determining \( \alpha_0 \) without assigning any probability distribution to \( \alpha_0 \).

Consider the following state-space model:

\[
y_t = f(\alpha_t, \theta) + \epsilon_t, \quad t = 1, 2, \ldots, T,
\]

\[
\alpha_t = g(\alpha_{t-1}, \theta) + \eta_t, \quad t = 1, 2, \ldots, T.
\]

For \( t = 1 \), we have

\[
y_1 = f(\alpha_1, \theta) + \epsilon_1,
\]

\[
\alpha_1 = g(\alpha_0, \theta) + \eta_1.
\]
In models (9), and (10), $\alpha_1$, $\alpha_0$, $\epsilon_1$, and $\eta_1$ are unobserved. Assume $\theta$ is known or determined by empirical knowledge.

The rule used to determine $\hat{\alpha}_0$ should meet the condition that given observation $y_1$, $\hat{\alpha}_0$ is able to ensure that $f(\hat{\alpha}_1, \theta)$ is an optimal estimation of $E(y_1)$.

From (9), consider
\[
\epsilon_1 = y_1 - f_1(\alpha_1, \theta)
\]

Let $\alpha_1$ be an unknown parameter and consider estimating function space
\[
G_T^{(1)} = \{a_1(y_1 - f_1(\alpha_1, \theta)) \mid \alpha_1 \in R\}
\]

A standardised optimal estimating function in $G_T^{(1)}$ is
\[
G_{(1)}^*(\alpha_1) = -E(\frac{\partial f}{\partial \alpha_1})[Var(\epsilon_1)]^{-1}(y_1 - f(\alpha_1, \theta)).
\]

If $E(\frac{\partial f}{\partial \alpha_1}) \neq 0$, and $f^{-1}$ exists, the optimal estimator of $\alpha_1$ will be given by $G_{(1)}^*(\alpha_1) = 0$, that is,
\[
\hat{\alpha}_1 = f^{-1}(y_1, \theta).
\]

Using (10), consider
\[
\eta_1 = \alpha_1 - g(\alpha_0, \theta).
\]

Let $\alpha_0$ be an unknown parameter and consider estimating function space
\[
G_T^{(0)} = \{a_0(\alpha_1 - g(\alpha_0, \theta)) \mid \alpha_0 \in R\}
\]

A standardised optimal estimating function in $G_T^{(0)}$ is
\[
G_{(0)}^*(\alpha_0) = -E(\frac{\partial g}{\partial \alpha_0})[Var(\epsilon_1)]^{-1}(\alpha_1 - f(\alpha_0, \theta)).
\]

If $E(\frac{\partial g}{\partial \alpha_0}) \neq 0$, and $g^{-1}$ exists, the optimal estimator of $\alpha_0$ will be given by $G_{(0)}^*(\alpha_0) = 0$, that is,
\[
\hat{\alpha}_0 = g^{-1}(\alpha_1, \theta).
\]

Therefore, we make the following suggestion for determining the initial state $\hat{\alpha}_0$ in inference process.

**Suggestion:** For a SSM
\[
y_t = f(\alpha_t, \theta) + \epsilon_t, \quad t = 1, 2, \ldots, T
\]
\[
\alpha_t = g(\alpha_{t-1}, \theta) + \eta_t, \quad t = 1, 2, \ldots, T.
\]

If $E(\frac{\partial g}{\partial \alpha_0}) \neq 0$, $E(\frac{\partial g}{\partial \alpha_0}) \neq 0$, $f^{-1}$ and $g^{-1}$ exist, the optimal decision on $\hat{\alpha}_0$ is
\[
\hat{\alpha}_0 = g^{-1}(f^{-1}(y_1)).
\]

For convenience, denote this $\hat{\alpha}_0$ as $\hat{\alpha}_0^*$. As an example for (5) and (6), the optimal value for $\hat{\alpha}_0$ is $y_1$, which is the same as the one given under diffuse conditions.

In the following, we apply the Suggestion to stochastic volatility model, and use simulation to investigate whether the Suggestion is practicable or not.

### A. Stochastic Volatility Model

Consider stochastic volatility process defined by (3) and (4), i.e.
\[
\ln(y_t^2) = \alpha_t + \ln(\xi_t^2), \quad t = 1, 2, \ldots, T.
\]
\[
\alpha_t = \gamma + \phi \alpha_{t-1} + \eta_t, \quad t = 1, 2, \ldots, T,
\]

where both $\xi_t$ and $\eta_t$ are i.i.d. r.v.'s; $\eta_t$ has mean 0 and variance $\sigma^2_\eta$, $\phi \neq 0$.

Let
\[
\epsilon_t = \ln(\xi_t^2) - E(\ln(\xi_t^2))
\]

Using (3) and (4), it follows that
\[
\epsilon_t = \ln(y_t^2) - \alpha_t - E(\ln(\xi_t^2)) = \ln(y_t^2) - f(\alpha_t, \theta),
\]

and
\[
\eta_t = \alpha_t - (\gamma + \phi \alpha_t) = \alpha_t - g(\alpha_0, \theta),
\]

where $\theta = (\gamma, \phi)^T$, $f(\alpha_1, \theta) = \alpha_1 + E(\ln(\xi_1^2))$, and $g(\alpha_0, \theta) = \gamma + \phi \alpha_0$.

Since $E(\frac{\partial f}{\partial \alpha_1}) = 1 \neq 0$, $E(\frac{\partial g}{\partial \alpha_0}) = \phi \neq 0$, and $f^{-1}$, $g^{-1}$ exist, therefore,
\[
\hat{\alpha}_0 = g^{-1}(f^{-1}(y_1)) = \frac{\ln(y_t^2) - E(\ln(\xi_t^2)) - \gamma}{\phi}.
\]

If $\xi_t$ has standard normal distribution, then $E(\ln(\xi_1^2)) = -1.2704$ and $Var(\ln(\xi_1^2)) = \pi^2/2$ (see Abramowitz and Stegun [1], p. 943). Then, substituting in (14)
\[
\hat{\alpha}_0 = g^{-1}(f^{-1}(y_1)) = \frac{\ln(y_t^2) + 1.2704 - \gamma}{\phi}.
\]

To show how the optimal initial value $\hat{\alpha}_0^*$ effects the final estimation when the QL and AQL approaches are applied, we carried out a simulation study on SVM model defined by (3) and (4). We camper the estimation of $(\gamma, \phi)$ given by the true $\alpha_0$ and $\hat{\alpha}_0^*$. Results are presented by Table II.

Table II shows that, compared to results in Table I, the estimation given by $\hat{\alpha}_0^*$ are close related to those given by the true $\alpha_0 = 0$. 

---

**TABLE II**

QL and AQL estimates based on 1,000 replication. The root mean square error of each estimate is reported below that estimate. $\hat{\alpha}_0^*$ is different from sample to sample. ($T = 500$).

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_\eta = 0.675$</th>
<th>$\sigma_\eta = 0.260$</th>
<th>$\sigma_\eta = 0.061$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma \phi$</td>
<td>$\gamma \phi$</td>
<td>$\gamma \phi$</td>
</tr>
<tr>
<td>$\alpha_0 = 0$</td>
<td>0.163 ± 0.024</td>
<td>0.243 ± 0.051</td>
<td>0.402 ± 0.060</td>
</tr>
<tr>
<td>AQL</td>
<td>0.163 ± 0.024</td>
<td>0.243 ± 0.051</td>
<td>0.402 ± 0.060</td>
</tr>
<tr>
<td>QL</td>
<td>0.104 ± 0.034</td>
<td>0.082 ± 0.019</td>
<td>0.071 ± 0.014</td>
</tr>
</tbody>
</table>
IV. THE STARTING VALUES FOR SYSTEM PARAMETER $\theta_0$

In this section, we consider the starting value for system parameters $\theta_0$. As described in literature, the outputs of non-linear inference procedures rely strongly on the appropriate value of the initial parameter $\theta_0$. It is usually suggested that $\theta_0$ should be chosen from a close neighbourhood of its true value (Zivot et al. [30]). Since the true value of $\theta_0$ is unknown, it is an issue how to identify the close neighbourhood of $\theta_0$.

The impact of the starting values of system parameters $\theta_0$ is illustrated via simulation studies below.

A. Stochastic Volatility Models

Consider SVM as given in (3) and (4) where $\eta_t \sim N(0, 0.675^2)$, $\xi_t \sim N(0, 1)$, and the initial value for $\alpha_0$ in the true model is given by $\alpha_0 = 0$. In this example, the state space model is involved with the parameter $\theta = (\gamma, \phi)$. Let $\theta = (-0.368, 0.95)$, a sequence of observations $y_1, \ldots, y_{1000}$ from the state space model were generated. Then we pretend $\theta$ is unknown. Consider a two-dimensional range $(-0.868, 0.132; 0.80, 0.99)$ for $\theta = (\gamma, \phi)$, which covers the true parameter $(-0.368, 0.95)$. Then we apply a two-dimensional grid search to $(-0.868, 0.132; 0.80, 0.99)$ with increment of 0.01. For each starting value of $\theta$ from the grid area, we apply the QL and AQL estimating procedures to the realisation $y_1, \ldots, y_{1000}$ and obtain the QL and AQL estimation of $\theta$ where $\hat{\alpha}_0 = \alpha_0$ are used. In Figure 1 - 4, we show the histograms of QL and AQL estimation of $\gamma$ and $\phi$ based on 2000 different starting values.

Like others estimation procedures described in literature, the QL and AQL estimations of $\theta$ rely strongly on the value of the initial parameter $\theta_0$.

We note an interesting phenomenon in the histograms illustrated in Figures 1 - 4. The true value of a parameter is not always allocated in the low frequency area. Obviously, the size of the low frequency area relies on the nature of the true model. This suggests that, although it is not appropriate to quantitatively identify an optimal estimation on system parameters utilising the information provided by a histogram diagram indirectly through the grid search approach, it is possible to narrow down and obtain a potential range covering the true value of parameters in underlying model by using the information provided by the histogram diagrams.
V. Determination of the Estimation of the System Parameter $\theta$

In their survey article, Zivot et al. [30] suggested choosing a starting value $\theta_0$ close to the true value of $\theta$. The estimation of $\theta$ using a Monte Carlo approximation for count data given by Kuk [23] is only good when the initial value of $\theta$ is assigned around the true value of $\theta$. Other approaches to decide $\theta_0$ are also suggested in literature. For example, Durbin and Koopman [14] numerically maximised the approximate likelihood for non-Gaussian SSMs to obtain the starting value for $\theta_0$; Sandmann and Koopman [27] used a two-dimensional grid search procedure which searches for an appropriate starting value for $\theta_0$ across the surface of a Gaussian log-likelihood function; Geweke and Tanizaki [15] and Tanizaki and Mariano ([28], [29]) used a simple grid search for $\theta_0$ where the expected log-likelihood function is maximised.

The ML method is a popular method for estimating the parameters of SSMs. The ML method works if the probability structure of the underlying state space system is known. In practice, it is not realistic to assume that the system’s probability structure is known. Then, the maximum likelihood method becomes impracticable. Therefore, searching the starting values in the histogram as a potential region to get the QL or AQL estimation of the parameter. It is sensible to obtain the estimate $\hat{\theta}$ based on maximising the log-likelihood function cannot be applied. Without knowledge of the log-likelihood, a distribution-free procedure can be considered. It is implemented by a grid search over a feasible region of the parameter space, and the parameter estimation will be the one giving the minimum residual sum of squares (RSS) (see Coakley et al. [11] and Naik-nimbalkar and Rajarshi [24]).

In this paper, we adapt grid search procedure but with some improvements. It is sensible to obtain the estimate of $\theta$ by utilising a grid search, and the residual sum of squares. However, if the grid search area is relatively large, the smallest sum of residuals might not lead to the best estimation of $\theta$. One example can be found from the simulation study discussed below. To improve the outcomes of the grid search procedure and sum of residuals, we need to reduce the area of the grid search into a reasonable size.

We suggest the following steps in determining the estimation of $\theta$ for SSMs: (in the following, we used a two-dimensional parameter as an example.)

Step 1. First determine a reasonable range. Based on experience, this range should cover the true parameter $\theta$. For example, for PM and SVM, decide a two-dimensional area $[a;b; c;d]$, covering the true parameter $\theta$.

Step 2. Following the two-dimensional grid search procedure, we assign $\theta_0$ with a different starting value, and obtain the QL or AQL estimation of the parameter.

Step 3. Draw the histogram of the QL or AQL estimates obtained from step 2.

Step 4. Consider the region with the highest frequency estimation values in the histogram as a potential region to cover the true value of the parameter. Obviously this potential region tends to be smaller than the range in Step 1.

Step 5. Let $y_t(\theta)$ be the predicted value of $y_t$ based on the observation equation. Find $\theta$, which minimises $RSS_y(\theta) = \sum_{t=1}^{T}(y_t - y_t(\theta))^2$ in the potential region.

The above steps used to determine the estimate of $\theta$ for SSMs are illustrated by the following example.

Example: Consider SVM as given in (22) and (23), where $\eta_t \sim N(0,0.675^2)$, $\xi_t \sim N(0,1)$, and the initial value for $\alpha_0$ in the true model is given by $\alpha_0 = 0$. In this example, the state space model is involved with the parameter $\theta = (\gamma, \phi)$. Let $\theta = (-0.368, 0.95)$, a sequence of observations $y_1, \cdots, y_{1000}$ and $\alpha_1, \cdots, \alpha_{1000}$ from the SVM were generated. Then we pretend $\{\alpha_t\}$ and $\theta$ are unknown.

Step 1. Consider a two-dimensional range $(-0.868,0.122; 0.80,0.99)$ for $\theta = (\gamma, \phi)$, which covers the true parameter $(-0.368,0.95)$.

Step 2. Apply a two-dimensional grid search to $(-0.868,0.122; 0.80,0.99)$ with increments of 0.01. For each starting value of $\theta$ from the grid area, we apply the QL/AQL estimating procedures and obtain the QL/AQL estimate of $\theta$.

Step 3. In Figures 1-4, we show the histograms of the QL and AQL estimates of $\gamma$ and $\phi$, based on 2,000 different starting values.

Step 4. From the histograms of the QL estimates of $\gamma$ and $\phi$ given in Figures 1 and 2, the potential region for parameter $\theta = (\gamma, \phi)$ is chosen as $[-0.36, -0.12; 0.91, 0.95]$. By using the histogram of the AQL estimates of $\gamma$ and $\phi$ given in Figures 3 and 4, the potential region for parameter $\theta = (\gamma, \phi)$ is chosen as $[-1.0,-0.30; 0.80,0.95]$.

Step 5. Find the estimate of $\gamma$ and $\phi$ by minimising the residual sum of squares (RSS$_y(\theta)$) in the potential region and give the QL estimate of $\theta = (-0.32,0.95)$, and the AQL estimate of $\theta = (-0.31,0.94)$.

In Table III, the QL and AQL denote the estimation of $\theta$, which gives the smallest RSS$_y(\theta)$ based on the region given in Step 1, and the QL* and AQL* denote the estimates of $\theta$, which gives the smallest RSS$_y$ based on the potential region determined in Step 4. We can see from Table III, that the estimate of $\theta$ has improved in all cases after using the potential region determined by the information provided by histogram diagram. The above examples indicate that using the potential region is able to significantly improve the performance of RSS$_y$.

VI. Real Data Application

In this section, we consider log returns of Pound/Dollar exchange rates. The data are the daily observation of weekdays’ closing pound to dollar exchange rates $x_t$ from 1/10/81 to 28/6/85 and have been taken from the site:www//staff.feweb.vu.nl/koopman/sv/. This data set has

<table>
<thead>
<tr>
<th>$\sigma_1$</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.30</td>
</tr>
<tr>
<td>0.95</td>
<td>323.53</td>
</tr>
<tr>
<td>0.45</td>
<td>660.62</td>
</tr>
<tr>
<td>0.31</td>
<td>457.65</td>
</tr>
<tr>
<td>0.32</td>
<td>725.03</td>
</tr>
</tbody>
</table>
been studied and analysed by Harvey et al. [20], Davis and Rodriguez-Yam [12]; Rodriguez-Yam [26]; Durbin and Koopman [13] and Alzghool and Lin [2].

Let \( y_t = \log(x_t/x_{t-1}), t = 1, 2, \ldots, 945 \). To model \( y_t \), we adopt the same SVM used by Davis and Rodriguez-Yam [12].

\[
y_t = \sigma_t \xi_t = e^{\alpha_t/2} \xi_t, \quad t = 1, 2, \ldots, 945,
\]

\[
\alpha_t = \gamma + \phi \alpha_{t-1} + \eta_t, \quad t = 1, 2, \ldots, T,
\]

where both \( \xi_t \) and \( \eta_t \) are i.i.d. r.v.'s; \( \eta_t \) has mean 0 and variance \( \sigma^2_\eta \). Therefore,

\[
\ln(\sigma^2_t) = \alpha_t + \ln \xi^2_t, \quad t = 1, 2, \ldots, T.
\]

If \( \xi_t \) were standard normal, then \( E(\ln \xi^2_t) = -1.2704 \) and \( Var(\ln \xi^2_t) = \pi^2/2 \) (see Abramowitz and Stegun [1], p. 943).

Let \( \epsilon_t = \ln \xi^2_t + 1.2704 \), and \( \delta_t = (\epsilon_t, \eta_t)' \).

We apply the QL method to the data model under the assumption that the conditional covariance matrix is known as follows:

\[
Var_{t-1}(\delta_t) = \Sigma_t = \begin{pmatrix} \frac{\pi^2}{2} & 0 \\ 0 & \sigma^2_\eta \end{pmatrix}.
\]

The AQL method is applied to the data by assuming no knowledge of the conditional covariance matrix. In the QL approach, \( \sigma_\eta \) will estimate from the residuals, but in AQL approach it is estimated by the Kernel estimator.

Following steps are for obtaining the estimate of \( \theta = (\phi, \gamma) \) for the Pound/ Dollar exchange rate data:

Step 1. Decide a grid search area, based on previous studies: (-0.813,0.177; 0.80,0.99).

Step 2. Apply a two-dimensional grid search to \((-0.813,0.177; 0.80,0.99)\) with increases of 0.01. For each starting value of \( \theta \) from the grid area, we apply the QL/AQL estimating procedures and obtain the QL/AQL estimate of \( \theta \).

Step 3. In Figures 5-8, we show the histograms of the QL and AQL estimates of \( \gamma \) and \( \phi \), based on 2,000 different starting values.

Step 4. From the histograms of the QL estimates of \( \gamma \) and \( \phi \) given in Figures 5 and 6, the potential region for parameter \( (\gamma, \phi) \) is chosen as (-0.17,-0.04; 0.86,0.95). By using the histogram of the AQL estimates of \( \gamma \) and \( \phi \) given in Figures 7 and 8, the potential region for parameter \( (\gamma, \phi) \) is chosen as (-0.45,0.1; 0.825,0.99).

Step 5. Find the estimate of \( \gamma \) and \( \phi \) by minimising the residual sum of squares \( RSS_\theta(\theta) \) in the potential region and the QL estimate of \( \theta \) is (-0.048,0.949), and the AQL estimate of \( \theta \) is (-0.082,0.971).

Table IV shows estimations of \( \theta = (\phi, \gamma) \) obtained by different methods. AQL denotes the asymptotic quasi-likelihood estimate, QL the estimate obtained by quasi-likelihood approach, AL the estimate obtained by maximising the approximate likelihood proposed by Davis and Rodriguez-Yam [12] and MCL the estimate obtained by maximising the estimate of the likelihood proposed by Durbin and Koopman [14]. AL and MCL outputs are taken from Rodriguez-Yam [26].

In Table IV, the estimate of \( \gamma \) and \( \phi \) by QL, AL and MCL are close to each other. These three methods are carried out under the same assumption where \( \xi_t \) and \( \eta_t \) are independent. This might indicate that the performance of QL, AL and MCL will be similar. However, the AQL estimates are slightly different from those of QL, as well as the estimates of AL and MCL.

The estimates of AQL and QL are obtained based on different model settings. The main difference between their models is that one assumes that \( \text{cov}(\eta_t, \xi_t) = 0 \) and the other
does not. To understand which model setting is appropriate, we require checking whether we can accept $cov(\eta_t, \xi_t) = 0$. We consider $\eta_t$ and $\xi_t$ given by QL and find that $\eta_t$ and $\xi_t$ are highly correlated with $r = 0.91$ and significant at level 0.01. So, the assumption of $\eta_t$ and $\xi_t$ uncorrelated is not valid. Therefore, it is not appropriate to apply the QL method to the data. Thus, we rather accept the estimations given by the AQL method than those given by the QL method.

VII. Conclusion

In this paper, we investigated the sensitivity of the QL and AQL estimation procedures to initial values assigned to state variable $\alpha_0$ and $\theta_0$ via simulation studies. A suggestion on choosing the initial value of state variable $\alpha_0$, without knowing the system’s probability structure has been given. Simulation studies indicate that it is relatively reliable to follow the suggestion in determining the initialisation of the state variable $\alpha_0$ during inference procedure. Apart from the impact of $\alpha_0$, the QL and AQL estimates of $\theta$ also sensitive to the value of the starting parameter $\theta_0$. In literature, it always suggests that $\theta_0$ has to be chosen from a neighbourhood close to the true value of $\theta$. But, it does not mention how to determine the close neighbourhood given the location of the true $\theta$ is unknown. In this paper, we established a standard procedure for determining the "close neighbourhood" and the estimation of $\theta$ in terms of minimizing $RSS_\theta$.

REFERENCES