Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

October 2021

Part VIII
SV: QMLE and the Kalman Filter

In this chapter we consider QML-based estimation of the log-SV model. Due to the fact that the volatility process $\{\sigma_t\}_{t=1}^T$ is unobserved, one cannot write down the log-likelihood function explicitly. Instead, we consider a log-linearization of the model, and use a filtering method, the so-called Kalman Filter, to compute a Gaussian quasi-log-likelihood function.

VIII.1 The log-Normal SV Model

Recall that the log-SV model for log-returns y_t is given by,

$$x_t = \sigma_t z_t, \quad t = 1, ..., T \tag{VIII.1}$$

with $\{z_t\}_{t=1}^T$ *i.i.d.* $\mathrm{N}(0,1)$ and

$$\log \sigma_t = (1-\phi)\alpha + \phi \log \sigma_{t-1} + \kappa_t \tag{VIII.2}$$

with $\{\kappa_t\}_{t=1}^T$ *i.i.d.* $\mathrm{N}(0,\sigma_\kappa^2)$, and independent of $\{z_t\}_{t=1}^T$. Moreover, $\phi \in (-1,1)$ and

$$\sigma_\kappa^2 := \beta^2 \left(1-\phi^2\right),$$

such that the three parameters to be estimated are given by,

$$\theta = (\alpha, \beta, \phi)' \in \mathbb{R} \times (0, \infty) \times (-1, 1).$$

Recall that, since $|\phi| < 1$, the stationary version of $\log \sigma_t$ is unconditionally normal with mean and variance given, respectively by $E \log \sigma_t = \alpha$, and $V \log \sigma_t = \beta^2$.

VIII.1.1 The Likelihood function

With observations $\{x_t\}_{t=1}^T$ from the log-SV model, the log-likelihood function is

$$\ell_T(\theta) = \log L_T(x_1, ..., x_T \; ; \; \theta) = \log f_\theta(x_1) + \sum_{t=2}^T \log f_\theta(x_t | x_{t-1}, ..., x_1),$$

with f_θ denoting a generic density function parametrized by θ. However, even in the simple case of the standard SV model in (VIII.1), $\ell_T(\theta)$ cannot be computed - let alone stated in any closed-form - since $f_\theta(x_1)$ and $f_\theta(x_t | x_{t-1}, ..., x_1)$ are unknown. Likewise, if one used the representation in terms of the unobserved σ_t,

$$\begin{aligned} L_T(x_1, ..., x_T \; ; \; \theta) &= \int L_T(x_1, ..., x_T, \sigma_1, ..., \sigma_T \; ; \; \theta) \, d(\sigma_1, ..., \sigma_T) \\ &= \int [\prod_{t=2}^T f_\theta(x_t | \sigma_t) f_\theta(\sigma_t | \sigma_{t-1})] f_\theta(x_1 | \sigma_1) f_\theta(\sigma_1) \, d(\sigma_1, ..., \sigma_T) \end{aligned}$$

the likelihood value cannot be computed, since the state space for (the Markov chain) $\log \sigma_t$ is $\mathbb{R}$, making the integration infeasible.

Note that in the case where σ_t had a finite valued (2 or more) state space, we demonstrated in Part VII that a filtering algorithm could be used to obtain the MLE of θ. With an infinite state space as here, various alternative routes have been followed to overcome the problem of finding the MLE. Some are simulation-based and approximate the likelihood function by various simulation schemes, see below for a quick introduction.

Some compute moments of x_t and apply the so called method of moments to obtain non-MLE estimators and study their properties, see Andersen and Sørensen (1996) and Kim et al. (1998) for discussion of some of the problems with method of moments.

Here we discuss in detail the Quasi-MLE (QMLE) approach where the state space form (SSF) is used to define a quasi-likelihood function which can be recursively computed by the Kalman filter. Thus the QMLE of θ, $\hat{\theta}_T$, is not the MLE as we shall see, due to simplifying assumptions (basically those of Gaussanity). However, in practice often the QMLE is used. This is despite the fact that it can have poor small sample properties.

The discussion will be structured as follows. After a brief introduction to simulation-based estimation, an introduction to the state space model is given. Next, MLE is discussed for linear Gaussian state space models based on the Kalman filter. Finally, the QMLE is discussed for the standard log-SV model.

VIII.1.1.1 The idea of simulated likelihood - a detour

The idea behind this can be – very – briefly described as follows:

Suppose that we wish to compute the density (and hence the likelihood) function for the first observed variable x_1 with σ_0 fixed. Now, by definition,

$$f(x_1|\sigma_0) = \int f(x_1, \sigma_1|\sigma_0)\, d\sigma_1 = \int f(x_1|\sigma_1) f(\sigma_1|\sigma_0)\, d\sigma_1,$$

where $f(x_1|\sigma_1)$ is the Gaussian $\mathrm{N}(0, \sigma_1^2)$ density, while $f(\sigma_1|\sigma_0)$ is the density of σ_1 given σ_0. As the distribution of $\log\sigma_1$ conditional on $\log\sigma_0$ is known, so is $f(\sigma_1|\sigma_0)$, such that the density $f(\sigma_1|\sigma_0)$ can be used to simulate draws of σ_1 from (with σ_0 fixed). Let $i = 1, ..., N$ be simulations of σ_1 which are denoted $\sigma_1(i)$, then one can approximate $f(x_1|\sigma_0)$ by,

$$f(x_1|\sigma_0) = \int f(x_1|\sigma_1) f(\sigma_1|\sigma_0)\, d\sigma_1 \simeq f^{\mathrm{a}}(x_1|\sigma_0) := \frac{1}{N}\sum_{i=1}^{N} f(x_1|\sigma_1(i)).$$

For the next x_2 conditional on x_1 and σ_0, analogously,

$$f(x_2|x_1, \sigma_0) = \int f(x_2|\sigma_2) f(\sigma_2|x_1, \sigma_0)\, d\sigma_2 \simeq f^{\mathrm{a}}(x_2|x_1, \sigma_0) = \frac{1}{N}\sum_{i=1}^{N} f(x_2|\sigma_2(i)).$$

However, here we need to simulate, or sample, $\sigma_2(i)$ conditional on (x_1, σ_0), which is not straightforward (in particular as we do not observe σ_1). Rewriting gives,

$$f(\sigma_2|x_1, \sigma_0) = \frac{f(\sigma_2|\sigma_1) f(\sigma_1|\sigma_0) f(x_1|\sigma_1)}{f(x_1|\sigma_0)} = \frac{K(\sigma_2, \sigma_1, \sigma_0, x_1)}{f(x_1|\sigma_0)}$$

Thus the problem is how to sample $\sigma_2(i)$ from a density expressed in terms of $K(\cdot)$ and $f(x_1|\sigma_0)$. The three densities in $K(\cdot)$ are known, but the latter we do not know so little can be done.

However, we can compute an approximate value $f^a(x_1|\sigma_0)$ as above for the $f(x_1|\sigma_0)$ term, and next we may use the trick to insert some "sampling density $\phi(\sigma_2)$" of our own choice and write:

$$\begin{aligned} f(x_2|x_1, \sigma_0) &= \int f(x_2|\sigma_2) f(\sigma_2|x_1, \sigma_0)\, d\sigma_2 = \int f(x_2|\sigma_2) \frac{f(\sigma_2|x_1\sigma_0)}{\phi(\sigma_2)} \phi(\sigma_2)\, d\sigma_2 \\ &\simeq f^a(x_2|x_1, \sigma_0) = \frac{1}{N}\sum f(y_2|\sigma_2(i)) \left(\frac{f(\sigma_2(i)|x_1, \sigma_0)}{\phi(\sigma_2(i))}\right) \end{aligned}$$

where $\sigma_2(i)$ is now drawn from $\phi(\cdot)$, and

$$f(\sigma_2(i)|x_1, \sigma_0) = \frac{f(\sigma_2(i)|\sigma_1(i)) f(\sigma_1(i)|\sigma_0) f(x_1|\sigma_1(i))}{f^a(x_1|\sigma_0)}.$$

Clearly, the choice of $\phi(\cdot)$ is important, and one should aim at choosing it "close" to $f(\sigma_2(i)|x_1,\sigma_0)$.

For $t \geq 2$ the equivalent problems occur, and it should be clear that computing the likelihood this way is non-trivial.

VIII.2 The State Space Model

We consider here the linear Gaussian state space model formulated in terms of two sets of equations. The first set of equations states the dynamics of the observed variable $Z_t \in \mathbb{R}^p$, in terms of the unobserved variable, $X_t \in \mathbb{R}^q$, and an innovation, w_t. The second set of equations states the dynamics of the unobserved (state) variable X_t, $X_t \in \mathbb{R}^q$, in terms of its own past and an innovation v_t. Specifically, for $t = 1, 2, ..., T$,

$$Z_t = AX_t + w_t, \tag{VIII.3}$$
$$X_t = \Phi X_{t-1} + v_t, \tag{VIII.4}$$

with some initial value X_0. The innovation w_t is p-dimensional and the innovation v_t q-dimensional with $(w_t, v_t)'$ *i.i.d.* $\mathrm{N}_{p+q}(0, \Omega)$, with covariance,

$$\Omega = \begin{pmatrix} \Omega_w & 0 \\ 0 & \Omega_v \end{pmatrix}.$$

Finally, A is a $(p \times q)$-dimensional matrix, and Φ $(q \times q)$ dimensional.

Equations (VIII.3) and (VIII.4) constitute the state space form (SSF), where equation (VIII.3) is referred to as the observation equation, and equation (VIII.4) is referred to as the state equation, specifying that the unobserved variable X_t as a (vector) autoregression of order one.

In terms of the initialization of the system, one may keep X_0 fixed, or one may consider X_0 as drawn from the $\mathrm{N}_q(\mu_x, \Omega_x)$ distribution, independently of $\{(w_t, v_t)'\}_{t=1}^T$. This we will return to when discussing the likelihood-based estimation. Let us consider first the log-SV model written in SSF:

Example VIII.1 (The log-SV model) *The log-Normal SV model can not be written in the SSF directly. However, an approximate SSF exists, which can be seen by considering the following transformation of the model. Define the observed variable Z_t as*

$$Z_t = \log|x_t|,$$

and set $w_t = \log|z_t| - \mu_z$, $\mu_z = E\log|z_t|$. With z_t $N(0,1)$, then $\mu_z = E\log|z_t| = -0.63518...$ and $\Omega_w = V(\log|z_t|) = \pi^2/8$. Hence w_t is indeed i.i.d. $(0, \Omega_w)$ but not Gaussian.

Moreover, with this notation, and $p = 1$, $q = 2$, one can write the transformed log-Normal SV model as,

$$Z_t = (1,1)\, X_t + w_t$$
$$X_t = \begin{pmatrix} \log \sigma_t - \alpha \\ \alpha + \mu_z \end{pmatrix} = \begin{pmatrix} \phi & 0 \\ 0 & 1 \end{pmatrix} X_{t-1} + \begin{pmatrix} \kappa_t \\ 0 \end{pmatrix},$$

or

$$Z_t = AX_t + w_t$$
$$X_t = \Phi X_{t-1} + v_t$$

where $A = (1,1)$, $\Phi =$diag$(\phi, 1)$ and $v_t = (\kappa_t, 0)'$.

In particular, the SV model is not a Gaussian state space model as w_t is not Gaussian, while v_t indeed is (singular) Gaussian with

$$\Omega_v = \begin{pmatrix} \beta^2 (1 - \phi^2) & 0 \\ 0 & 0 \end{pmatrix}.$$

The fact that w_t is not Gaussian is ignored in the estimation of the parameters $\theta = (\alpha, \phi, \beta^2)'$, which explains the use of the terminology Quasi-MLE (QMLE).

Note that the SSF is not unique in this case, and more generally, often different state space representations exist for the same model(s).

VIII.2.1 State Space Model Likelihood

For the state space model in (VIII.3) and (VIII.4), the log-likelihood in terms of the observations $\{Z_t\}_{t=1}^T$ is given by,

$$l_T(\theta) = \log L_T(Z_1, ..., Z_T\,;\theta) = \log f_\theta(Z_1) + \sum_{t=2}^{T} \log f_\theta(Z_t | Z_{t-1}, ..., Z_1), \tag{VIII.5}$$

where for $f_\theta(Z_1)$ the initial distribution of X_0 is used as by definition $Z_1 = AX_1 + w_1$, and $X_1 = \Phi X_0 + v_1$.

Furthermore, under the assumption of normality of $(w_t, v_t)'$ and X_0 (and hence X_1), we can see that the process $(X_1, Z_1, X_2, Z_2, ..., X_T, Z_T)$ is multivariate Gaussian. In particular, this means that $(Z_1, ..., Z_T)$ is Gaussian and the conditional distribution $Z_t | Z_{t-1}, ..., Z_1$ as well.

Hence for a given value of the parameter θ, the likelihood in (VIII.5) is completely specified by the conditional mean and variance,

$$E(Z_t | Z_{t-1}, ..., Z_1) \quad \text{and } V(Z_t | Z_{t-1}, ..., Z_1), \tag{VIII.6}$$

in addition to the initial distribution of Z_1, as noted above. The thing to emphasize here is that $\sum_{t=2}^{T} \log f_\theta (Z_t|Z_t, ..., Z_1)$ can be calculated from $E(X_t|Z_{t-1}, ..., Z_1)$ and $V(X_t|Z_{t-1}, ..., Z_1)$ due to Gaussianity. Moreover, the conditional mean and variance can be recursively calculated by the Kalman filter, as explained below.

VIII.2.2 The Kalman Filter

The Kalman filter is a recursive algorithm from which the conditional moments in (VIII.6) can be obtained. Define for $0 \leq s \leq t$, the conditional moments:

$$X_{t|s} = E(X_t|Z_s, ..., Z_1) \quad \text{and} \quad \Omega_{t|s} = V(X_t|Z_s, ..., Z_1) \tag{VIII.7}$$

We may set $X_{0|0} = \mu_x$ and $\Omega_{0|0} = \Omega_x$ corresponding to $X_0 = N(\mu_x, \Omega_x)$.

Straightforward calculations as demonstrated below, see also Shumway and Stoffer (2000) and Hamilton (1994), give the following lemma:

Lemma VIII.1 (Kalman) *Consider the Gaussian state space model given by (VIII.3) and (VIII.4). With initial conditions $X_{0|0} = \mu_x$ and $\Omega_{0|0} = \Omega_x$, the conditional moments in (VIII.7) for $t = 1, 2, ..., T$ are given by:*

$$X_{t|t-1} = \Phi X_{t-1|t-1}, \tag{VIII.8}$$
$$\Omega_{t|t-1} = \Phi \Omega_{t-1|t-1} \Phi' + \Omega_v, \tag{VIII.9}$$

with

$$X_{t|t} = X_{t|t-1} + K_t \left(Z_t - A X_{t|t-1} \right), \tag{VIII.10}$$
$$\Omega_{t|t} = (I - K_t A) \Omega_{t|t-1}, \tag{VIII.11}$$

where

$$K_t \equiv \Omega_{t|t-1} A' \left(A \Omega_{t|t-1} A' + \Omega_w \right)^{-1}. \tag{VIII.12}$$

Proof of Lemma VIII.1: To see (VIII.8), set $Z_{1:t-1} = (Z_1, ..., Z_{t-1})$ and observe that

$$E(X_t|Z_{1:t-1}) = E(\Phi X_{t-1} + v_t|Z_{1:t-1}) = \Phi E(X_{t-1}|Z_{1:t-1}) = \Phi X_{t-1|t-1}$$

as $E(v_t|\mathbb{Z}_{t-1}) = 0$ by definition. Likewise, (VIII.9) holds as

$$V(X_t|Z_{1:t-1}) = \Phi \Omega_{t-1|t-1} \Phi' + \Omega_v.$$

Next, using obvious notation, recall from well-known results on the Gaussian distribution, that if

$$\left(\begin{array}{c} X \\ Y \end{array} \middle| Z \right) \overset{d}{=} \mathrm{N}(\left(\begin{array}{c} \mu_{x|z} \\ \mu_{y|z} \end{array} \right), \left(\begin{array}{cc} \Omega_{xx|z} & \Omega_{xy|z} \\ \Omega_{yx|z} & \Omega_{yy|z} \end{array} \right)),$$

then

$$\begin{aligned} E\left(X|Y,Z\right) &= \mu_{x|z} + \Omega_{xy|z}\Omega_{yy|z}^{-1}\left(Y - \mu_{y|z}\right) \quad \text{and} \\ V\left(X|Y,Z\right) &= \Omega_{xx|z} - \Omega_{xy|z}\Omega_{yy|z}^{-1}\Omega_{yx|z}. \end{aligned}$$

Set now $X = X_t$, $Y = Z_t$ and $Z = \mathbb{Z}_{t-1} = (Z_{t-1}, ..., Z_1)$, then

$$\begin{aligned} X_{t|t} &= E\left(X_t|Z_t, Z_{1:t-1}\right) \\ &= E\left(X_t|Z_{1:t-1}\right) + \mathrm{Cov}\left(X_t, Z_t|Z_{1:t-1}\right)\mathrm{V}\left(Z_t|Z_{1:t-1}\right)^{-1}\left(Z_t - E\left(Z_t|Z_{1:t-1}\right)\right). \end{aligned} \tag{VIII.13}$$

Next, $E\left(X_t|Z_{1:t-1}\right) = X_{t|t-1}$,

$$\mathrm{Cov}\left(X_t, Z_t|Z_{1:t-1}\right) = \mathrm{Cov}\left(X_t, AX_t + w_t|Z_{1:t-1}\right) = \Omega_{t|t-1}A', \tag{VIII.14}$$

$$\mathrm{V}\left(Z_t|Z_{1:t-1}\right) = \mathrm{V}\left(AX_t + w_t|Z_{1:t-1}\right) = A\Omega_{t|t-1}A' + \Omega_w, \tag{VIII.15}$$

and finally,

$$E\left(Z_t|Z_{1:t-1}\right) = AX_{t|t-1}. \tag{VIII.16}$$

Inserting (VIII.14)-(VIII.16) in (VIII.13) gives directly (VIII.10). Likewise for (VIII.11), where

$$\begin{aligned} \Omega_{t|t} &= \mathrm{V}\left(X_t|Z_t, Z_{1:t-1}\right) \\ &= V\left(X_t|Z_{1:t-1}\right) - \mathrm{Cov}\left(X_t, Z_t|Z_{1:t-1}\right)\mathrm{V}\left(Z_t|Z_{1:t-1}\right)^{-1}\mathrm{Cov}\left(Z_t, X_t|Z_{1:t-1}\right). \end{aligned} \tag{VIII.17}$$

□

VIII.2.3 Computing the likelihood

The likelihood value for a given parameter value θ can be computed by using Lemma VIII.1 and (VIII.16) as well as (VIII.15). First note that, we have directly that,

$$\mathrm{V}\left(Z_t|Z_{1:t-1}\right) = A\Omega_{t|t-1}A' + \Omega_w \qquad \text{and} \qquad E\left(Z_t|Z_{1:t-1}\right) = AX_{t|t-1}.$$

This means that for $t = 1, 2, ..., T$, we can write the likelihood in terms of the Gaussian innovations,

$$\epsilon_t := Z_t - AX_{t|t-1}, \tag{VIII.18}$$

with conditional mean zero, and conditional variance,

$$\Sigma_t = A\Omega_{t|t-1}A' + \Omega_w. \tag{VIII.19}$$

Hence the Gaussian likelihood function equals,

$$l_T(\theta) = \log L_T(Z_1, ..., Z_T\ ;\ \theta) = \log f_\theta(Z_1) + \sum_{t=2}^{T} \log f_\theta(Z_t | Z_t, ..., Z_1) \tag{VIII.20}$$

$$= -\frac{pT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\Sigma_t| + \epsilon_t'\Sigma_t^{-1}\epsilon_t\right) \tag{VIII.21}$$

As to the initial conditions for $X_{0|0}$ and $\Omega_{0|0}$, we use as mentioned, $X_{0|0} := \mu_x$ and $\Omega_{0|0} := \Omega_x$, corresponding to X_0 chosen as $\mathrm{N}(\mu_x, \Omega_x)$. Hence, the parameters μ_x and Ω_x may be fixed or considered as additional parameters to be found in the QML algorithm maximizing $l_T(\theta)$.

Example VIII.2 *For the log-SV model, since $|\phi| < 1$, $\log\sigma_t - \alpha$ has a stationary representation with*

$$E(\log\sigma_t - \alpha) = 0 \text{ and } V(\log\sigma_t - \alpha) = \beta^2.$$

Hence, X_0 can be given the initial distribution, $N(\mu_x, \Omega_x)$ with

$$\mu_x = (0, \alpha + \mu_z)' \quad \text{and} \quad \Omega_x = diag\left(\beta^2, 0\right).$$

VIII.2.4 Asymptotic theory

Estimating the parameters in the state space model in (VIII.3)-(VIII.4) using the described Kalman-Filter gives the MLE $\hat{\theta}_T$ if $(w_t, v_t)'$ are *i.i.d.* jointly Gaussian as stated, while $\hat{\theta}_T$ is the QMLE if this assumption does not apply. From Watson (1989), see also Ruiz (1994) and Dunsmuir (1979), it holds that $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is asymptotically Gaussian under the assumption that (the observations) Z_t are stationary and weakly mixing, and provided $(w_t, v_t)'$ are martingale differences with finitie fourth order moments. Note that this in particular requires that Φ has eigenvalues smaller than one in absolute value, see also Shumway and Stoffer (2000).

References

Andersen, T. & B. Sørensen (1996), 'GMM estimation of a stochastic volatility model: a Monte Carlo study', Journal of Business and Economic Statistics 14, 328-52.

Dunsmuir, W. (1979), A Central Limit Theorem for parameter estimation in stationary time series and its applications to models for a signal observed with noise, Annals of Statistics, 7:490-506.

Kim, S., N. Shephard & S. Chib (1998), 'Stochastic volatility: likelihood inference and comparison with ARCH models', Review of Economic Studies 65, 361-393.

Koopman, S.J., N. Shephard & J.A. Doornik (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', Econometrics Journal 2, 113-166.

Ruiz, E. (1994), Quasi Maximum Likelihood Estimation of Stochastic Volatility Models, Journal of Econometrics, 63:289–306

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility, in D.R. Cox, David V. Hinkley and Ole E. Barndorff-Nielsen (eds.), Time Series Models in Econometrics, Finance and Other Fields, Chapman & Hall London.

Shumway, R.H. & D.S. Stoffer (2000), Time Series Analysis and its Applications, Springer-Verlag New York.

Taylor, S.J. (2005), Asset Prices Dynamics, Volatility, and Prediction, Princeton University Press. 6

Watson, M.W. (1989), Recursive Solution Methods for Dynamic Linear Rational Expectations Models, Journal of Econometrics, 41:65-89