

On the Nonlinear Estimation of GARCH Models Using an Extended Kalman Filter

Sebastián Ossandón and Natalia Bahamonde

Abstract—A new mathematical representation, based on a discrete-time nonlinear state space formulation, is presented to characterize a Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) model. Nonlinear parameter estimation and nonlinear state estimation, for this state space model, using an Extended Kalman Filter (EKF) are described. Finally some numerical results, which make evident the effectiveness and relevance of the proposed nonlinear estimation are given.

Index Terms—GARCH models; Discrete-time nonlinear state space model; Nonlinear parameter estimation; Nonlinear state estimation; Extended Kalman Filter.

I. INTRODUCTION

During the last to decades GARCH type modeling, Engle [5] and Bollerslev [2], has been an extremely active area of research. These models are often used in financial econometrics literature because their properties are close to the observed properties of empirical financial data. Also these properties can capture various stylized facts. These models have gained popularity, specially due to their easy applicability and flexibility, allowing simple extensions that better fit the empirical financial data.

For the class of GARCH models, the most commonly used estimation procedure has been the Quasi Maximum Likelihood (QMLE) approach. Weiss [14] was the first to study the asymptotic properties of QMLE in GARCH models. The asymptotic properties of the QMLE for classical GARCH models have been extensively studied; see, for recent references, Berkes et al [1], Francq and Zakoian [6], Hall and Yao [8].

Alternatively, other estimation procedures are available based on the Autoregressive Moving Average Model (ARMA) representation of the squared GARCH process. This idea was taken by Giraitis and Robinson [7] who studied the Whittle estimator of parametric Auto Regressive Conditional Heteroskedasticity (∞) (ARCH(∞)) models, which involve the GARCH(p, q) case. Recently Kristensen and Linton [11] have proposed the use of the Yule Walker estimator for the GARCH(1,1) model. In Bose and Mukherjee [3] the asymptotic properties of two-stage least-squares estimator of the parameters of ARCH models is investigated, which has a closed-form expression and is computationally easy to

obtain. Simulation results show that for small samples size, this estimator has a better performance than the QMLE.

On the other hand, space state models are a flexible family of models which fits for the modelling of many scenarios. These models, which were introduced in Kalman [9] and Kalman and Bucy [10], are frequently constructed and applied by modern stochastic controllers. In Durbin and Koopman [4], state space models was applied to time series analysis treatment.

This work proposes a novel estimation procedure for nonlinear time series models based on the EKF. It is shown that for GARCH processes, it is possible to have a novel state space formulation and an efficient approach, based on the EKF, in order to obtain an estimation for the parameters and predictions for the states. The EKF proposed for GARCH models is derived from a discrete-time nonlinear state space formulation of the studied model. This method is adequate to obtain initial conditions for a maximum-likelihood iteration, or to provide the final estimation of the states and the parameters when maximum-likelihood is considered inadequate or costly.

The structure of the paper is as follows. Section II defines the basic notation and the nonlinear problem. In Section III the EKF methodology is derived. In Section IV, nonlinear estimation algorithm is presented and Sections V their performance are evaluated in finite samples. Some concluding remarks are given in Section VI.

II. THE DISCRETE-TIME NONLINEAR PROBLEM

Let us consider the following discrete-time nonlinear state space mathematical model:

$$\begin{cases} \mathbf{x}(k) = \mathbf{f}(\mathbf{x}(k-1), \mathbf{u}(k-1), \boldsymbol{\theta}) + \boldsymbol{\sigma}(\mathbf{u}(k-1), \boldsymbol{\theta}) \mathbf{w}(k), \\ \mathbf{y}(k) = \mathbf{h}(\mathbf{x}(k), \mathbf{u}(k), \boldsymbol{\theta}) + \boldsymbol{\nu}(k), \end{cases} \quad (1)$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state unknown vector, $\mathbf{u}(k) \in \mathbb{R}^r$ is the input known vector, $\mathbf{y}(k) \in \mathbb{R}^m$ is the noisy observation vector or output vector of the stochastic process, $\mathbf{w}(k) \in \mathbb{R}^n$ and $\boldsymbol{\nu}(k) \in \mathbb{R}^m$ are, respectively, the process noise (due, mainly, to disturbances and modelling inaccuracies of the process) and the measurement noise (due, mainly, to sensor inaccuracy). Moreover $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector that is generally unknown, $\mathbf{f}(\cdot) \in \mathbb{R}^n$, $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$ and $\mathbf{h}(\cdot) \in \mathbb{R}^m$ are nonlinear functions that characterize the stochastic system.

Manuscript received Mars 23, 2011; revised April 06, 2011. Natalia Bahamonde was supported in part by FONDECYT Grant 3080009.

S. Ossandón is with the Institute of Mathematics, Pontificia Universidad Católica de Valparaíso, Blanco Viel 596, Cerro Barón, Valparaíso, Chile. E-mail: sebastian.ossandon@ucv.cl.

N. Bahamonde is with the Institute of Statistics, Pontificia Universidad Católica de Valparaíso, Blanco Viel 596, Cerro Barón, Valparaíso, Chile. E-mail: natalia.bahamonde@ucv.cl.

With respect to the noises of the process, we assume the following assumptions:

- The vector $\mathbf{w}(k)$ is assumed to be Gaussian, zero-mean $\mathbb{E}(\mathbf{w}(k)) = 0$ and white noise with covariance matrix $\mathbb{E}(\mathbf{w}(k) \cdot \mathbf{w}(j)^T) = \mathbf{Q} \cdot \delta(k - j)$.
- The vector $\mathbf{v}(k)$ is assumed to be Gaussian, zero-mean $\mathbb{E}(\mathbf{v}(k)) = 0$ and white noise with covariance matrix $\mathbb{E}(\mathbf{v}(k) \cdot \mathbf{v}(j)^T) = \mathbf{R} \cdot \delta(k - j)$.

Where $\delta(k - j) =$ identity matrix when $k = j$, otherwise, $\delta(k - j) =$ zero matrix.

A. The nonlinear state space formulation of the GARCH model used

The GARCH model (GARCH(1,1)), that we will use throughout this work, is characterized by the following discrete-time equations:

$$\begin{aligned} x(k) &= \sigma(k)\varepsilon(k), & (2) \\ \sigma^2(k) &= \alpha_0 + \alpha_1 x^2(k-1) + \beta_1 \sigma^2(k-1), & (3) \end{aligned}$$

where $x(k)$ and $\sigma(k) > 0$ are, respectively the return and the volatility, in the discrete-time $k \in \mathbb{Z}$, associated to a financial process, and $(\varepsilon(k))_{k \in \mathbb{Z}}$ is a i.i.d. Gaussian sequence, with $\mathbb{E}(\varepsilon(k)) = 0$, $\mathbb{E}(\varepsilon(k) \cdot \varepsilon(j)) = Q\delta(k - j)$, and parameters $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$. Moreover $x(0)$ is independent of sequence $(\varepsilon(k))_{k > 0}$.

The only state space representation of equations (2) and (3) is,

$$\begin{bmatrix} X_1(k) \\ X_2(k) \end{bmatrix} = \begin{bmatrix} f_1(X_1(k-1), X_2(k-1), \boldsymbol{\theta}) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \varepsilon(k) \quad (4)$$

$$Y(k) = x(k) = X_2(k)\sqrt{X_1(k)}, \quad (5)$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1)$, $f_1(X_1(k), X_2(k), \boldsymbol{\theta}) = \alpha_0 + \alpha_1 X_2^2(k)X_1(k) + \beta_1 X_1(k)$, $X_1(k) = \sigma^2(k)$ and $X_2(k) = x(k)/\sigma(k)$. Let us notice the obvious nonlinearity of this state space representation, due to the nonlinearity of the process and observation equations.

III. THE EXTENDED KALMAN FILTER

The discrete-time EKF generalizes, for a discrete-time nonlinear stochastic process, the standard Kalman Filter (KF) used in discrete-time linear stochastic process. This extension is based on a successive linearization of the nonlinear state space model proposed for the stochastic process under study (see Wan and Nelson [12] and Wan et al [13]).

The functions \mathbf{f} and \mathbf{h} (see equation (1)) are used to compute the predicted state and the predicted measurement from the previous estimate state. The following equation shows the computation of the predicted state from the previous estimate:

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{f}(\hat{\mathbf{x}}(k-1|k-1), \mathbf{u}(k-1), k, \boldsymbol{\theta}). \quad (6)$$

To compute the predicted estimate covariance a matrix \mathbf{A} of partial derivatives (the Jacobian matrix) is previously computed. This matrix is evaluated, with the predicted states, at each discrete timestep and used in the Kalman filter equations. In other words, \mathbf{A} is a linearized version of the nonlinear function \mathbf{f} around the current estimate.

$$\mathbf{P}(k|k-1) = \mathbf{A}(k-1)\mathbf{P}(k-1|k-1)\mathbf{A}^T(k-1) + \mathbf{Q}. \quad (7)$$

After making the prediction stage, we need to update the equations. So we have the residual measure innovation

$$\tilde{\mathbf{y}}(k) = \mathbf{y}(k) - \mathbf{h}(\hat{\mathbf{x}}(k|k-1), \mathbf{u}(k), k, \boldsymbol{\theta}) \quad (8)$$

and the conditional covariance innovation

$$\mathbf{S}(k|k-1) = \mathbf{C}(k)\mathbf{P}(k|k-1)\mathbf{C}^T(k) + \mathbf{R}, \quad (9)$$

where \mathbf{C} is a linearized version of the nonlinear function \mathbf{h} around the current estimate.

The Kalman gain is given by

$$\mathbf{K}(k) = \mathbf{P}(k|k-1)\mathbf{C}^T(k)\mathbf{S}^{-1}(k|k-1), \quad (10)$$

and the corresponding updates by

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\tilde{\mathbf{y}}(k) \quad (11)$$

and

$$\mathbf{P}(k|k) = (\mathbf{I} - \mathbf{K}(k)\mathbf{C}(k))\mathbf{P}(k|k-1). \quad (12)$$

The state transition and observation matrices (the linearized versions of \mathbf{f} and \mathbf{h}) are defined, respectively, by

$$\mathbf{A}(k-1) = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}(k-1|k-1), \mathbf{u}(k-1)} \quad (13)$$

and

$$\mathbf{C}(k) = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}(k|k-1)} \quad (14)$$

IV. NONLINEAR ESTIMATION

A. Nonlinear parameter estimation

Given a discrete-time nonlinear stochastic system, as the presented in equation (1), the maximum likelihood estimation technique can be used to find the unknown parameters $\boldsymbol{\theta}$, of the model, from data of the state and output observations. In other words, given a sequence of measurement or observations $\mathcal{Y}_N = [\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(k), \dots, \mathbf{y}(N)]$, the likelihood function is given by the following joint probability density function:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = p(\mathcal{Y}_N | \boldsymbol{\theta}), \quad (15)$$

or equivalently:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = p(\mathbf{y}(0) | \boldsymbol{\theta}) \prod_{k=1}^N p(\mathbf{y}(k) | \mathcal{Y}_{k-1}, \boldsymbol{\theta}). \quad (16)$$

Since the dynamics of the stochastic system presented in equation (1) depends of Gaussian, white noise processes, it seems reasonable to assume that under certain regularity conditions, the probability density functions $p(\mathbf{y}(k)|\mathcal{Y}_{k-1}, \boldsymbol{\theta})$ can be approximated by functions of Gaussian probability densities. Therefore we can rewrite equation (16) as follows:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = \frac{p(\mathbf{y}(0)|\boldsymbol{\theta})}{(2\pi)^{m/2}} \prod_{k=1}^N \frac{g(k)}{(\det(\mathbf{S}(k|k-1)))^{1/2}}, \quad (17)$$

where $g(k) = \exp\{-0.5\tilde{\mathbf{y}}^\top(k)\mathbf{S}^{-1}(k|k-1)\tilde{\mathbf{y}}(k)\}$, $\tilde{\mathbf{y}}(k)$ is the residual measure innovation defined in equation (8), $\hat{\mathbf{y}}(k|k-1) = \mathbb{E}(\mathbf{y}(k)|\mathcal{Y}_{k-1}, \boldsymbol{\theta})$ is the conditional mean of $\mathbf{y}(k)$ given $\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(k-1)$ and $\boldsymbol{\theta}$, and finally $\mathbf{S}(k|k-1)$ is the conditional covariance innovation, defined in equation (9), given $\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(k-1)$ and $\boldsymbol{\theta}$.

Conditioning on $\mathbf{y}(0)$, and considering the function:

$$l(\boldsymbol{\theta}) = -\ln(L(\boldsymbol{\theta}; \mathcal{Y}_k|\mathbf{y}(0))), \quad (18)$$

the maximum likelihood estimator of $\boldsymbol{\theta}$ can be obtained solving the following nonlinear optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} l(\boldsymbol{\theta}). \quad (19)$$

Let us remark that for a fixed $\boldsymbol{\theta}$, the values of $\tilde{\mathbf{y}}(k)$ and $\mathbf{S}(k|k-1)$, at each discrete timestep, are obtained from the Kalman filter equations, described in section III, and subsequently used in the construction of the likelihood function. Therefore the success of the optimization of likelihood function depends strictly on the behavior of the EKF designed.

B. Nonlinear state estimation

Once estimated the parameters of the nonlinear stochastic process, such as described in the previous subsection, using the Extended Kalman Filter, the goal is to calculate from the observations an estimation of the state of the nonlinear system. For this calculation, we use again Kalman filter equations described in section III (see particularly equation (11)).

V. NUMERICAL RESULTS

In this section a numerical example is presented in order to make evident the effectiveness and relevance of the proposed nonlinear estimation method.

A GARCH(1,1) model, with $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1) = (1, 0.3, 0.5)$ and noise process covariance $\mathbf{Q} = 0.1$, is considered. The sample size used is 1000. Similar results are obtained with different initial values.

Figure V shows the nonlinear parameter estimation of GARCH(1,1) model described below. It can be seen the fast convergence of the parameter evolution (see Table 1 for details). For the parameter estimation, it is assumed that the

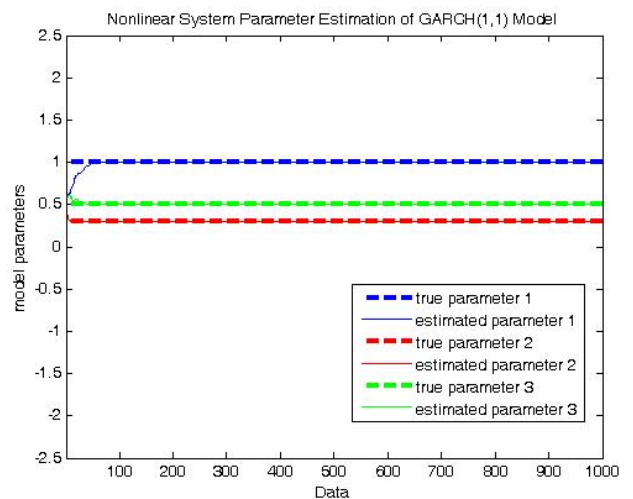


Figure 1. Nonlinear Parameter Estimation

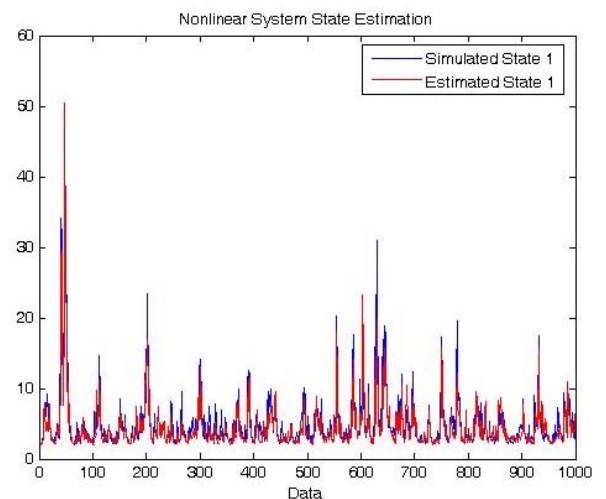


Figure 2. Nonlinear State 1 Estimation

states of the model is always known. As seen in Figures 2 and 3, the simulated states are very close to the estimated states (using the EKF technique), being the Mean Squared Errors 3.2484 and 0.2332 respectively.

Table I
MEAN AND MEAN SQUARE ERROR (MSE) OF THE NONLINEAR
PARAMETER ESTIMATION

	α_0	α_1	β_1
MEAN	0.9817	0.3049	0.5084
MSE	0.0062469	0.00086165	0.0013252

VI. CONCLUSIONS

This work presents for the first time a state space representation for GARCH family of time series models. Moreover,

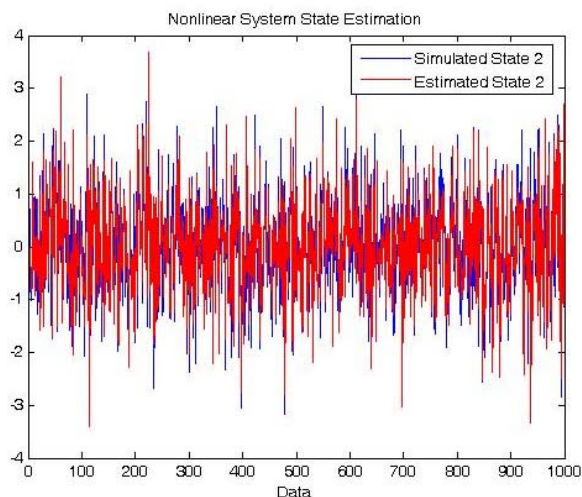


Figure 3. Nonlinear State 2 Estimation

an efficient numerical method, for nonlinear estimation of GARCH processes, is presented. This procedure is implemented using a EKF technique. The numerical results demonstrate the effectiveness of state representation, and shows that it's appropriate when the objective is to estimate the parameters and the state of a nonlinear system.

ACKNOWLEDGMENT

The authors would like to thank S. Eyheramendy and C. Reyes for many fruitful discussions related to this work.

REFERENCES

- [1] I. Berkes, L. Horváth and P. Kokoszka, "GARCH processes: structure and estimation," *Bernoulli*, vol. 9, no. 2, pp. 201-227, 2003.
- [2] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no 3, pp. 307-327, 1986.
- [3] A. Bose and K. Mukherjee, "Estimating the ARCH parameters by solving linear equations," *J. Time Ser. Anal.*, vol. 24, no. 2, pp. 127-136, 2003.
- [4] J. Durbin and S.J. Koopman, "Time Series Analysis by State Space Methods," *Oxford: Oxford University Press*, 2001.
- [5] R. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987-1007, 1982.
- [6] C. Francq and J-M. Zakoian, "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli* vol. 10, pp. 605-637, 2004.
- [7] L. Giraitis and P. M. Robinson, "Whittle estimation of arch models," *Econometric Theory*, vol. 17, no. 3, pp. 608-631, 2001.
- [8] P. Hall and Q. Yao, "Inference in ARCH and GARCH models with heavy-tailed errors," *Econometrica* 71, 2853-2877, 2003.
- [9] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Trans ASME J. Basic Eng.*, 82, 35-45, 1960.
- [10] R.E. Kalman and R.S. Bucy, "New results in filtering and prediction theory," *Trans. ASME J. Basic Eng.*, 83, 95-108, 1961.
- [11] D. Kristensen and O. Linton, "A closed-form estimator for the Garch (1, 1) model," *Econometric Theory*, vol. 22, no. 2, pp. 323-337, 2006.
- [12] E. A. Wan and A. T. Nelson. "Neural dual extended Kalman filtering: applications in speech enhancement and monaural blind signal separation," *In Proc. Neural Networks for Signal Processing Workshop. IEEE*, 1997.

- [13] E. A. Wan, R. van der Merwe, and A. T. Nelson. "Dual Estimation and the Unscented Transformation," *In S. Solla, T. Leen, and K.-R. Muller; editors, Advances in Neural Information Processing Systems, 12*, pp. 666-672. MIT Press, 2000.
- [14] A. Weiss, "Asymptotics theory for Arch models: estimation and testing," *Econometric Theory*, vol. 2, pp. 107-131, 1986.