GARCH Models Estimation with Missing Observations using State Space Representation

Natalia Bahamonde* Sebastián Ossandón†

Abstract
A mathematical formulation, based on a discrete time nonlinear state space formulation, is presented to characterize Generalized AutoRegresive Conditional Heteroskedasticity (GARCH) models. In order to improve the parameter and state estimation techniques in GARCH models, a novel estimation procedure for nonlinear time series model with missing observations, based on an Extended Kalman Filter (EKF) approach, is described and successfully evaluated herein. Finally, through a comparison analysis between our proposed nonlinear estimation method and a Quasi Maximum Likelihood Estimation (QMLE) technique based on different methods of imputation, some numerical results with real data, which make evident the effectiveness and relevance of the proposed nonlinear estimation technique are given.

Key Words: GARCH models, Missing observations, Nonlinear state space model, Nonlinear estimation, Extended Kalman Filter.

1. Introduction

Most of the work on time series assume that the observations are consecutive and equally spaced. In practice, however, in real time series data set it is not unusual to find a large number of missing observations.
Missing values may follow a variety of pattern, however, two schemas are of special interest:

- The Bernoulli pattern in which each measurement has a fixed probability of being missing, and the missed are independent.
- The periodic pattern in which one or a few blocks of missing values are repeated periodically (e.g. holiday or calendar effects). The time series is sampled in groups of A consecutive data separated by B missed observations.

Periodically or randomly missing data appear, either by nature, either because aberrant data were detected and thus eliminated. The following examples of incomplete data seem to occur frequently in practice: speech in presence of interference, fading communication channel, astronomical measurements (available only during the night), informations resulting from signals sonar and radar (passive electronic intelligence) or radar studies of the moon surface.
Missing data are nonignorable in the context of time series analysis. When one or more observations are missing it may be necessary to estimate the model and also to obtain estimates of the

*Instituto de Estadística, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile
†Instituto de Matemáticas, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

missing values. By including estimates of missing values, a better understanding of the nature of the data is possible with more accurate forecasting.

If we consider a time series studies with varying proportions of missing data, different strategies to estimate these missing values it is necessary to use. Little and Rubin (2002) classify missing data mechanisms into three types: missing completely at random (MCAR), where the probability of missingness does not depend on observed or unobserved data measures; missing at random (MAR), where the probability of missingness depends only on observed data measures; and nonignorable missing data (NI), where the probability of missingness depends on the unobserved data measures.

The process of estimating missing values in time series data for univariate data is complex. The more frequently strategy to the treatment of missing observations is to replace the unobserved data for another value (e.g. mean, median or last observed data), which is know as imputation or numerical interpolation methods. However we note that all the treatments of missing observation in time series based on the standard imputation techniques will lead to biased estimates.

In practice, most of the literature above time series with missing observations is concerned with linear processes with normal innovations. In addition, these perturbations are usually regarded as strict white noise. This assumption is very restrictive; this characteristic implies only linear models with homoskedastic conditional variances.

As far as we know, the first study that extended the sample autocorrelation function to the case of missing observations is due to Parzen (1963). Their study formulated that the values of the observed series at unequally spaced times can be represented as an amplitude modulated time series $Y_k = C_k X_k$ where $(C_k)_{k\in\mathbb{Z}}$ represents the censoring process, with $C_k = 0$ when X_k is missing at time k and $C_k = 1$ in otherwise. The asymptotic properties of this modified autocorrelation function, under various assumptions on the noise $(\varepsilon_k)_{k\in\mathbb{Z}}$, were investigated by Dunsmuir and Robinson (1981). More recently, Yajima and Nishino (1999) compare three estimators of the autocorrelation function for a stationary process with missing observations. The first estimator is the sample autocorrelation function extended to the case with censored data proposed originally by Parzen (1963). The others estimators are extensions of this first estimator. The authors derive asymptotic distribution for both short memory and long memory models for the three estimators of the autocorrelation function with missing observations. They impose the same assumptions on the innovations $(\varepsilon_k)_{k\in\mathbb{Z}}$ as those in Dunsmuir and Robinson (1981).

Autoregressive conditionally heteroscedastic (ARCH) type modeling, introduced by Engle (1982), are often used in finance because their properties are close to the observed properties on empirical financial data such as heavy tails, volatility clustering, white noise behavior or autocorrelation of the squared series. Financial time series often exhibit that the conditional variance can change over time, namely heteroskedasticity. The ARCH family of model is a class of nonlinear time series models including the GARCH (General ARCH) process, introduced by Bollerslev (1986) and their many variations and extensions. The GARCH(p,q) model with normal error is

$$\begin{cases} r_k = \sigma_k \varepsilon_k, \\ \sigma_k^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{k-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{k-j}^2, \end{cases} \tag{1}$$

where r_k and σ_k ($> 0, \forall k$) 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 $\alpha_0 > 0$, $\alpha_i \geq 0, p \geq i \geq 1$, and $\beta_j \geq 0, q \geq j \geq 1$ and $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Moreover, let us consider r_0 independent of sequence $\{\varepsilon_k\}_{k>0}$.
Given that r_k follows Gaussian distribution conditional on past history, the log-likelihood function of the GARCH model conditional on initial values is

$$\begin{aligned} -\ln L &= \ell(\boldsymbol{\theta}) \\ &= \frac{N}{2}\log(2\pi) + \frac{1}{2}\sum_{k=1}^{N}\log\sigma_k^2 + \frac{1}{2}\sum_{k=1}^{N}\frac{\varepsilon_k}{\sigma_k^2}. \end{aligned} \tag{2}$$

The conditional log-likelihood function is used in practice since the unconditional distribution of the initial values is not known in a closed form expression. This method can not be easily implemented, require numerical optimization procedures and the choice of initial values.
For the class of GARCH models with complete data, the most commonly used estimation procedure has been the QMLE approach. Weiss (1986) 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. (2003), Francq and Zakoïan (2004), Hall and Yao (2003). 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 (2001) who studied the Whittle estimator of parametric AutoRegresive Conditional Heteroskedasticity of (∞) order (ARCH(∞)) models, which involve the GARCH(p, q) case. Recently Kristensen and Linton (2006) have proposed the use of the Yule Walker estimator for the GARCH(1,1) model. In Bose and Mukherjee (2003) 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. Simulations show that for small sample sizes, this estimator has a better performance than the QMLE.
As discussed by many authors, the most common problem with this estimation procedure is the fact that the formula of the log-likelihood function requires calculating unobservable values from the observed sample $r_0, \ldots, r_N$. A common technique for solving this problem is to choose initial values, hoping that the initial values dependence vanishes for large values of N. Once the log-likelihood function is initialized, it can be maximized using numerical optimization algorithms. GARCH models use a Quasi-Newton optimizer to find the maximum likelihood estimates of the conditionally normal model. The first max(p, q) values are assumed to be fixed. The optimizer uses a hessian approximation computed from a method for solving nonlinear optimization update. The gradient is either computed analytically or using a numerical approximation. Also there exist some evidence that the Gaussian QMLE, using imputation methods, does not work too well in the case of GARCH processes with missing observations, because in a missing data scenario it is not possible to apply directly QMLE approach, therefore we consider a QMLE based on a complete data set, obtained after filling the missing observations by some imputation procedures.

It is well known that in linear time series analysis state space formulation (see Durbin and Koopman (2001)) and the associated Kalman Filter (KF), introduced by Kalman (see Kalman(1960) and Kalman and Bucy (1961)), provide convenient tools to handle missing observations in the measurement equation. However, the applicability of this approach to GARCH processes with missing observations is not immediately apparent. Let us remark that the dynamics of a GARCH process is nonlinear, and consequently the associated state space formulation will follow the same type of behavior. Standard KF algorithm can not be applied to nonlinear state space formulation of GARCH models, because this algorithm only works when the processes have a linear behavior. The original idea of applying a KF to GARCH models was proposed by Harvey et al. (1992). In their implementation, they consider only a GARCH linear equation and the state space representation of volatilities is given by an ARMA model (see Wan et al. (2000) and Galka et al. (2004)).
The goal of this research is to develop a novel nonlinear estimation procedure, based on an EKF approach, for GARCH models considering missing observations. The EKF technique proposed is derived from a nonlinear state space formulation of the discrete time GARCH equation (see equation (1)). This method is adequate to obtain initial conditions for a maximum-likelihood iteration, or to provide the final estimation of the parameters and the states when maximum-likelihood is considered inadequate or costly.
The structure of the paper is as follows. Section 2 introduces the nonlinear state space formulation for the GARCH(1, 1) model. The EKF methodology and the nonlinear estimation algorithm are presented in Section 3 and 4 respectively. In order to evaluate the numerical performance of the proposed methodology, a comparison between our proposed nonlinear estimation method and QMLE based on different methods of imputation is carried out in Section 5. An analysis between estimated states and simulated states (see equation (3)), using a EKF approach, is also detailed herein. Finally some concluding remarks are given in Section 6.

2. Nonlinear state space formulation for GARCH(1,1)

A possible state space representation of equation (1) when $p = q = 1$ is,

$$\begin{cases} \boldsymbol{X}_k := \begin{bmatrix} X_k^{(1)} \\ X_k^{(2)} \end{bmatrix} = \boldsymbol{f}(\boldsymbol{X}_{k-1}, \boldsymbol{\theta}) + \mathcal{B} \cdot \boldsymbol{w}_k, \\ \boldsymbol{Y}_k := r_k = X_k^{(2)} \sqrt{X_k^{(1)}}, \end{cases} \tag{3}$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1) \in \mathbb{R}^3$, $\mathcal{B} = (0, 1)^\top \in \mathbb{R}^2$, $X_k^{(1)} = \sigma_k^2 \in \mathbb{R}\ (> 0, \forall k)$, $X_k^{(2)} = r_k/\sigma_k \in \mathbb{R}$, $\boldsymbol{w}_k = \varepsilon_k$ and

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{X}_{k-1}, \boldsymbol{\theta}) := \begin{bmatrix} \alpha_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix} X_{k-1}^{(1)} (X_{k-1}^{(2)})^2 + \\ + \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix} X_{k-1}^{(1)}. \end{aligned} \tag{4}$$

For details see Ossandón and Bahamonde (2011).
A state space representation for the ARCH(1) model can be obtained from the state space formulation of GARCH(1, 1), replacing $\boldsymbol{f}(\boldsymbol{X}_{k-1}, \boldsymbol{\theta})$ in this formulation by

$$\boldsymbol{f}(\boldsymbol{X}_{k-1}, \boldsymbol{\theta}) := \begin{bmatrix} \alpha_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix} X_{k-1}^{(1)} (X_{k-1}^{(2)})^2. \tag{5}$$

It is important to remark that rest of equations characterizing ARCH(1) state space representation remains invariant in state space formulation for the GARCH(1, 1) model.

3. The Extended Kalman Filter

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

$$\begin{cases} \boldsymbol{X}_k = \boldsymbol{f}(\boldsymbol{X}_{k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{\theta}) + \boldsymbol{\sigma}(\boldsymbol{u}_{k-1}, \boldsymbol{\theta}) \cdot \boldsymbol{w}_k, \\ \boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) + \boldsymbol{\nu}_k, \end{cases} \tag{6}$$

where $\boldsymbol{X}_k \in \mathbb{R}^n$ is the state unknown vector, $\boldsymbol{u}_k \in \mathbb{R}^r$ is the input known vector, $\boldsymbol{Y}_k \in \mathbb{R}^m$ is the noisy observation vector or output vector of the stochastic process, $\boldsymbol{w}_k \in \mathbb{R}^n$ 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 $\boldsymbol{\theta} \in \mathbb{R}^l$ is the parameter vector that is generally unknown, $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$, $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n\times n}$ and $\boldsymbol{h}(\cdot) \in \mathbb{R}^m$ are nonlinear functions that characterize the stochastic system.

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

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

Where $\boldsymbol{\delta}_{k-j} =$ identity matrix when $k = j$, otherwise, $\boldsymbol{\delta}_{k-j} =$ zero matrix.

The EKF generalizes, for a discrete time nonlinear stochastic process, the standard 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 (1997) and Wan et al. (2000)).
Let $\mathcal{Y}_N = [\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_k, ..., \boldsymbol{y}_N]$ be a known sequence of measurement or observations. The functions $\boldsymbol{f}$ and $\boldsymbol{h}$ (see equation (6)) 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:

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{f}(\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{\theta}). \tag{7}$$

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 KF equations. In other words, $\mathbf{A}$ is a linearized version of the nonlinear function $\boldsymbol{f}$ around the current estimate.

$$\mathbf{P}_{k|k-1} = \mathbf{A}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{A}_{k-1}^{\top} + \mathbf{Q}. \tag{8}$$

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

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{y}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{u}_k, \boldsymbol{\theta}) \tag{9}$$

and the conditional covariance innovation

$$\mathbf{S}_{k|k-1} = \mathbf{C}_k\mathbf{P}_{k|k-1}\mathbf{C}_k^{\top} + \mathbf{R}, \tag{10}$$

where $\mathbf{C}$ is a linearized version of the nonlinear function $\boldsymbol{h}$ around the current estimate.
The Kalman gain is given by

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{C}_k^{\top}\mathbf{S}_{k|k-1}^{-1}, \tag{11}$$

and the corresponding updates by

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \mathbf{K}_k\tilde{\boldsymbol{y}}_k \tag{12}$$

and

$$\mathbf{P}_{k|k} = (I - \mathbf{K}_k\mathbf{C}_k)\mathbf{P}_{k|k-1}. \tag{13}$$

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

$$\mathbf{A}_{k-1} = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{X}}\right|_{\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{u}_{k-1}} \tag{14}$$

and

$$\mathbf{C}_k = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{X}}\right|_{\hat{\boldsymbol{x}}_{k|k-1}} \tag{15}$$

4. Nonlinear estimation with missing observations

4.1 Nonlinear parameter estimation

Given a discrete time nonlinear stochastic system, as the presented in equation (6), 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 $\mathcal{Y}_N$, the likelihood function is given by the following joint probability density function:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = p(\mathcal{Y}_N|\boldsymbol{\theta}), \tag{16}$$

or equivalently:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = p(\boldsymbol{y}_0|\boldsymbol{\theta}) \prod_{k=1}^{N} p(\boldsymbol{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}). \tag{17}$$

Since the dynamics of the stochastic system presented in equation (6) depends of Gaussian, white noise processes, it seems reasonable to assume that under certain regularity conditions, the probability density functions can be approximated by functions of Gaussian probability densities. Therefore we can rewrite equation (17) as follows:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = \frac{p(\boldsymbol{y}_0|\boldsymbol{\theta})}{(2\pi)^{m/2}} \prod_{k=1}^{N} \frac{g(k)}{\det(\mathbf{S}_{k|k-1})^{1/2}}, \tag{18}$$

where $g(k) = \exp\{-0.5\tilde{\boldsymbol{y}}_k^{\top} \cdot \mathbf{S}_{k|k-1}^{-1} \cdot \tilde{\boldsymbol{y}}_k\}$, $\tilde{\boldsymbol{y}}_k$ is the residual measure innovation defined in equation (9), $\widehat{\boldsymbol{y}}_{k|k-1} = \mathbb{E}(\boldsymbol{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta})$ is the conditional mean of $\boldsymbol{y}_k$ given $\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_{k-1}$ and $\boldsymbol{\theta}$, and finally $\mathbf{S}_{k|k-1}$ is the conditional covariance innovation, defined in equation (10), given $\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_{k-1}$ and $\boldsymbol{\theta}$.
Conditioning on $\boldsymbol{y}_0$, and considering the function:

$$\ell(\boldsymbol{\theta}) = -\ln(L(\boldsymbol{\theta}; \mathcal{Y}_k|\boldsymbol{y}_0)), \tag{19}$$

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

$$\widehat{\boldsymbol{\theta}} = arg \min_{\boldsymbol{\theta}}(\ell(\boldsymbol{\theta})). \tag{20}$$

State space formulation and EKF provide a powerful tool for the analysis of data in the context of maximum likelihood estimation. Let us remark that for a fixed $\boldsymbol{\theta}$, the values of $\tilde{\boldsymbol{y}}_k$ and $\mathbf{S}_{k|k-1}$, at each discrete timestep, can be obtained from the Kalman filter equations, described in section 3, and subsequently used in the construction of the log-likelihood function. In this context, the success of the optimization of log-likelihood function depends strictly on the behavior of the EKF designed.

Under the assumption that there is a large number of missing values in the data, we use a Kalman approach, as proposed by Brockwell and Davis (1991) for ARIMA processes, in the context of nonlinear parameter estimation. This procedure calculates maximum likelihood estimates by means of a nonlinear space state formulation of models and then computing the Gaussian log-likelihood function using Kalman recursive equations.
In a first time, we give the evaluation of the Gaussian log-likelihood function bases on $\mathcal{Y}_r = \left[\boldsymbol{y}_{i_0}, \boldsymbol{y}_{i_1}, \ldots, \boldsymbol{y}_{i_r}\right]$, where $\mathcal{I}_r = [i_0, i_1, \ldots, i_r]$ are positive integers such that $0 \leq i_0 < i_1 < \cdots < i_r \leq N$. This allows for observation of the process at irregular intervals, or equivalently for the possibility that $(N-r)$ observations are missing from the sequence $\mathcal{Y}_N$.
To deal with possibly irregularly spaced observations or data with missing values, we introduce a new series $\{\boldsymbol{Y}_k^*\}_k$, related to the state $\{\boldsymbol{X}_k\}_k$ by the modified observation equation (see equation (6)):

$$\boldsymbol{Y}_k^* = \boldsymbol{h}^*(\boldsymbol{X}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) + \boldsymbol{\nu}_k^*, \qquad k = 0, 1, \ldots, \tag{21}$$

where

$$\boldsymbol{h}^*(\boldsymbol{X}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) = \begin{cases} \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) & \text{if } k \in \mathcal{I}_r, \\ \mathbf{0} & \text{otherwise}, \end{cases}$$

$$\boldsymbol{\nu}_k^* = \begin{cases} \boldsymbol{\nu}_k & \text{if } k \in \mathcal{I}_r, \\ \boldsymbol{\eta}_k & \text{otherwise}, \end{cases}$$

and $\boldsymbol{\eta}_k$ is a i.i.d. Gaussian, zero-mean and white noise sequence with covariance matrix $\mathbb{E}(\boldsymbol{\eta}_k \cdot \boldsymbol{\eta}_s^T) = \mathbf{I}_{w \times w} \cdot \boldsymbol{\delta}_{k-s}$ (w is the dimension of noise $\boldsymbol{\nu}_k$ and $\boldsymbol{\eta}_k$ at each discrete timestep k). Moreover $\boldsymbol{\eta}_s \perp \boldsymbol{X}_0$, $\boldsymbol{\eta}_s \perp \boldsymbol{w}_k$, and $\boldsymbol{\eta}_s \perp \boldsymbol{\nu}_k$ $(s, k = 0, \pm 1, \ldots)$.
Let us consider the joint Gaussian probability density function $L_1(\boldsymbol{\theta}; \mathcal{Y}_r)$ and the related Gaussian log-likelihood function $\ell_1(\boldsymbol{\theta})$ based on the measured values $\mathcal{Y}_r$. From these measured values, let us define the sequence $\mathcal{Y}_N^* = [\boldsymbol{y}_0^*, \boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_N^*]$ as follows:

$$\boldsymbol{y}_k^* = \begin{cases} \boldsymbol{y}_k & \text{if } k \in \mathcal{I}_r, \\ \mathbf{0} & \text{otherwise}. \end{cases} \tag{22}$$

Let $L_2(\boldsymbol{\theta}; \mathcal{Y}_N^*)$ be the joint Gaussian probability density function based on the values $\mathcal{Y}_N^*$ and let $\ell_2(\boldsymbol{\theta})$ be the related log-likelihood function.
So it is clear that $L_1(\boldsymbol{\theta}; \mathcal{Y}_r)$ and $L_2(\boldsymbol{\theta}; \mathcal{Y}_N^*)$ are related by:

$$L_1(\boldsymbol{\theta}; \mathcal{Y}_r) = (2\pi)^{(N-r)w/2} L_2(\boldsymbol{\theta}; \mathcal{Y}_N^*), \tag{23}$$

and consequently $\ell_1(\boldsymbol{\theta})$ and $\ell_2(\boldsymbol{\theta})$ are related by:

$$\ell_1(\boldsymbol{\theta}) = \frac{\ln(2\pi)(N-r)w}{2} + \ell_2(\boldsymbol{\theta}). \tag{24}$$

Then we can now compute the required $L_1(\boldsymbol{\theta}; \mathcal{Y}_r)$ and the corresponding $\ell_1(\boldsymbol{\theta})$ of the realized sequence $\mathcal{Y}_r$, using the Kalman approach described above and applying equations (23) and (24).

4.2 Nonlinear state estimation

Once estimated the parameters of the nonlinear stochastic process with missing observations, such as described in the previous subsection, using the EKF, the next goal is to calculate from the data values an estimation of the state of the nonlinear system. For this calculation, we use again EKF equations described in section 3 (see particularly equations (7) and (12)).

5. Numerical results: a real time series example

In this section a real time series example is presented in order to make evident the effectiveness and relevance of the proposed nonlinear estimation method. We consider daily real return observations, using the Chile's IPSA stock index, an index composed of the 40 most heavily traded stocks, from 1994 through 2004. These time series contains approximately 10% of missing observations on the data considered. It is important to remark that numerical examples were made using simulated data with different percentages (5%, 10% and 15%) of missing data, obtaining, in all of cases, better mean square errors with our methodology. In this article, these examples were not included, and we only work with the real data from the IPSA.

Let $r_k = \ln P_k - \ln P_{k-1}$ be the log return in the discrete time k, computed from the associated IPSA stock price index P_k in the discrete time k. It should be noted that we will work directly with daily log return time series (r_k) for IPSA stock and not with the related price index time series.

Let us notice that for the treatment of missing data using QMLE technique is necessary to replace the unobserved data by other values (which is know as imputation methods). Different imputation methods can be applied to obtain a complete data set. In our real time series example presented in this section, three imputation methods are used to replace the missing data: QMLE-Mean (replaces unobserved data by mean of the observed values); QMLE-NonNA (when each missing value is deleted and the observed data are bound together); QMLE-Last (replaces unobserved data by last observed data).

Let us consider a GARCH(1,1) model, with initial random parameters $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1)^\top$. The GARCH(1,1) model is the most popular empirical specification because it performs well in a wide range of applications. Also let us assume a initial noise process covariance $\mathbf{Q} = \mathbf{I}_{1\times1}$. The data set used, to adjust this model, consists of 2000 daily observations on the IPSA Price Index considering missing observations.

As seen in Figures 1, the simulated states, using equation (3) with the estimated parameters given in Table 1 using a EKF technique, are very close to the estimated states, being the mean squared errors 5.1270×10^{-4} and 2.5106×10^{-4} respectively. Also let us notice that a new noise process estimated covariance $\widehat{\mathbf{Q}} = 0.0111\mathbf{I}_{1\times1}$, which best fits the data set of the numerical example proposed when the parameters are obtained using a EKF approach, is obtained from the recursive equations described in section 3.

Finally in Figure 2, a comparison between simulated log return and real log return, using EKF and using QMLE with different imputation methods for the unobserved data, is presented. To measure the performance of these methods, we computed the Mean Square Errors (MSE) of parameters estimates. Clearly, as shown in Table 2, we can see that the return MSE (simulated return vs. real return) is lower when the parameters are estimated using the EKF methodology proposed in this work. For these reasons, the proposed nonlinear estimation method presents better performances than the classical imputation techniques considered here. Also our methodology is readily extended to other nonlinear time series models and non-stationary and asymmetric cases.

6. Discussion

This article introduces a new efficient numerical method, based on an EKF approach, for parameter and state estimation of GARCH processes in a nonlinear state space formulation. The framework that we propose it is valuable if the processes have unobserved values. The strategy of estimation associated with this representation allows computationally efficient parameter estimation.

Other state space representations of GARCH processes has been proposed in the literature. To the best of our knowledge, there are two scenarios for state space representation proposed for GARCH models: after replacing or approximating some variables in equation (1).

An estimation technique for GARCH models is suggested by Chou et al. (1992) and Harvey et al. (1992). On the other hand, the Kalman filter approach is employed by Chou and Wu (2008) for non-GARCH models applied in competition with GARCH for predicting the conditional beta in the capital asset pricing model (CAPM). Other result related to state space representation of GARCH model can be found in Bougerol and Picard (1986) where the considered a multivariate stochastic difference equations to give a condition for existence of a strictly stationary solution for GARCH models.

Let us notice that quasi maximum likelihood methods are very accurate by estimating GARCH parameters in a complete time series. In these cases, the EKF involve a more complex procedure for the estimation and the gain is not realy significant. On the other hand, it is well known that time series with missing values presents a serious problem to conventional estimation methodologies such as QMLE. Studies show that, for linear time series models, the standard KF approach can be easily modified in order to obtain an efficient method to deal with missing observations. A natural extension, for nonlinear time series models, has been proposed in this work to treat problems with missing values. The numerical results presented herein demonstrate the effectiveness of this methodology, and show that it is more appropriate than other QMLE-imputation methods used in practice (see Table 2).

In conclusion, the methodology proposed in this article is an innovative and effective way to solve the problem of estimation of parameters in GARCH processes with missing observations.

Future work will seek to exploit the nonlinear state space models, presented in this article, in order to develop more optimized nonlinear techniques for parameter estimation, state estimation and observation estimation. In addition, new procedures for prediction of observations in several discrete time steps will be developed. Finally, we should mention that due to the versatility of

Table 1: Parameter Estimation using different methods

	EKF	$QMLE - Mean$	$QMLE - NonNA$	$QMLE - Last$
α_0	0.9992	3.8101×10^{-6}	4.2709×10^{-6}	4.9328×10^{-6}
α_1	0.1	0.1249	0.13904	0.1589
β_1	0.2006	0.8466	0.8327	0.8081

Table 2: Mean Square Error using different methods

	EKF	$QMLE - Mean$	$QMLE - NonNA$	$QMLE - Last$
MSE	3.1365×10^{-4}	0.0028	4.8087×10^{-4}	4.8119×10^{-4}

state-space models, important properties of GARCH models, such as stability, controllability and observability, can be studied from a new perspective in future research.

Acknowledgements

This research was supported by the projects DII 037.397/2012 and DII 037.395/2012, Dirección de Investigación e Innovación, Pontificia Universidad Católica de Valparaíso. Also the authors would like to thank to Camilo Reyes, and Wilfredo Palma for many fruitful discussions about this work.

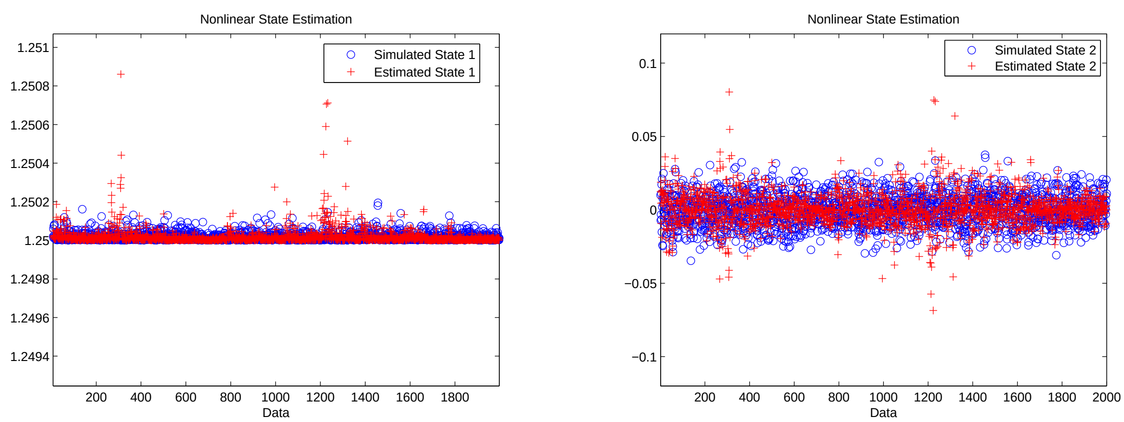

Figure 1: Nonlinear State Estimation: Simulated State vs. Estimated State

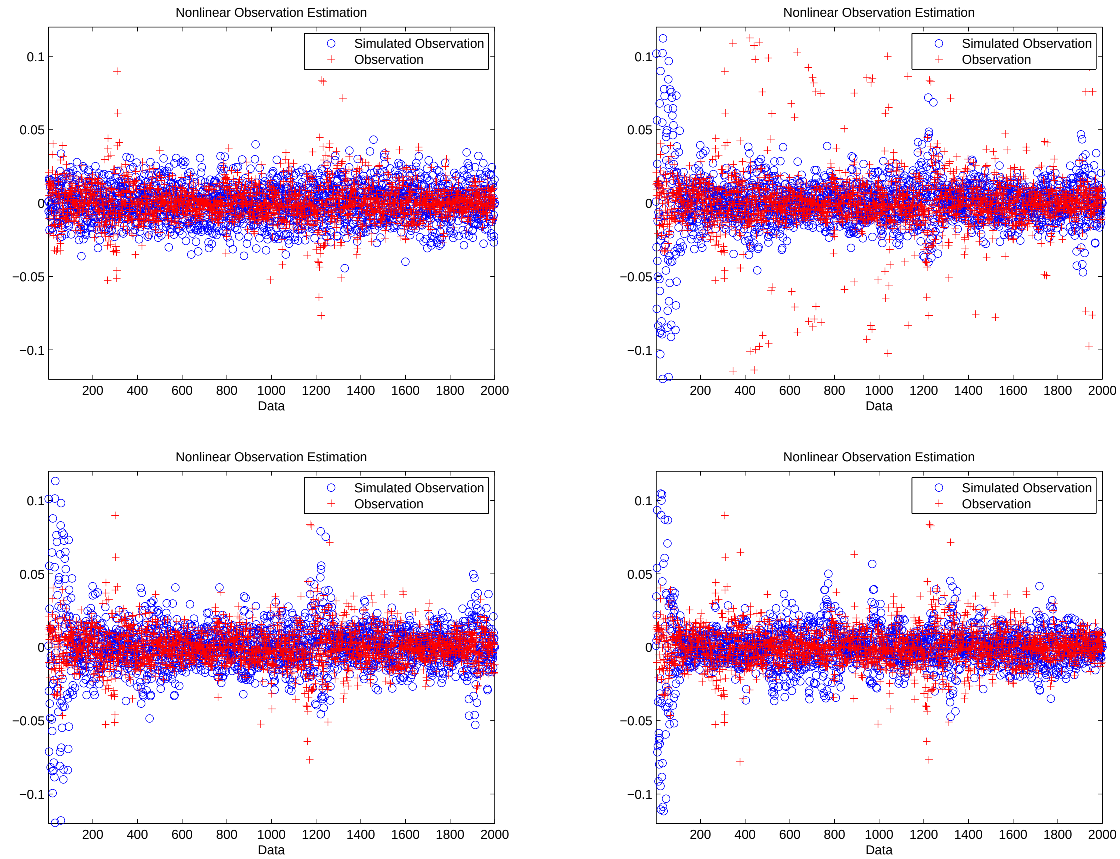

Figure 2: Simulated Observation vs. Real Observation under different methods: EKF method (top, left), QMLE-Mean method (top, right), QMLE-NonNA method (bottom, left) and QMLE-Last method (bottom, right).

REFERENCES

Berkes, I., Horváth, L. and Kokoszka, P. (2003), GARCH processes: structure and estimation, Bernoulli 9(2), 201-227.

Bose A. and Mukherjee, K. (2003), Estimating the ARCH parameters by solving linear equations, J. Time Ser. Anal. 24(2), 127-136.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31(3), 307-327.

Bougerol, P. and Picard, N. (1992), Strict stationary of generalized autoregressive processes, The annals of Probability 20(4), 1714-1730.

Brockwell P. and Davis R. (1991), Time series: theory and methods, second edition, Springer Series in Statistics .

Chou, R., Engle, R. and Kane, A. (1992), Measuring Risk Aversion from Excess Returns on a Stock Index, Journal of Econometrics, 52, 201-224.

Choudhry, T. and Wu, H. (2008), Forecasting Ability of GARCH vs Kalman Filter Method: Evidence from Daily UK Time-Varying Beta, J. Forecast. 27, 670-689.

Dunsmuir, W. and Robinson, P. M. (1981). Asymptotic theory for time series containing missing and amplitude modulated observations. *Sankhyā Ser. A*, 43(3):260–281.

Durbin, J. and Koopman, S.J. (2001), Time series analysis by state space methods. Oxford Statistical Science Series, 24, Oxford University Press, Oxford.

Engle, R. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50(4), 987-1007.

Francq, C. and Zakoïan, J.M. (2004), Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, Bernoulli 10(4), 605-637.

Galka, A., Yamashita, O. and Ozaki, T. (2004), GARCH modelling of covariance in dynamical estimation of inverse solutions, Physics letters, Section A, 333, 261-268.

Giraitis, L. and Robinson, P.M. (2001), Whittle estimation of arch models, Econometric Theory 17(3), 608-631.

Hall, P. and Yao, Q. (2003), Inference in ARCH and GARCH models with heavy-tailed errors, Econometrica, 71(1), 285-317.

Harvey, A.C., Ruiz, E. and Sentana, E. (1992), Inference in ARCH and GARCH models with heavy-tailed errors, Journal of Econometrics 52, 129-157.

Kalman, R. (1960), A new approach to linear filtering and prediction problems, Trans. ASME J. Basic Eng. 82, 35-45.

Kalman, R. and Bucy, R. (1961), New results in filtering and prediction theory, Trans. ASME J. Basic Eng., 83, 95-108.

Kristensen, D. and Linton, O. (2006), A closed-form estimator for the Garch $(1,1)$ model, Econometric Theory 22(2), 323-337.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley Series in Probability and Statistics. Wiley-Interscience John Wiley & Sons, Hoboken, NJ, second edition.

Ossandón, S. and Bahamonde, N. (2011), On the Nonlinear Estimation of GARCH Models Using an Extended Kalman Filter, Lectures Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2011, WCE 2011, 6-8 July, 2011, London, U.K., pp 148-151.

Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhyā Ser. A*, 25:383–392.

Yajima, Y. and Nishino, H. (1999). Estimation of the autocorrelation function of a stationary time series with missing observations. *Sankhyā Ser. A*, 61(2):189–207.

Wan, E.A. and Nelson, A.T. (1997), Neural dual extended kalman filtering: Applications in speech enhancement and monaural blind signal separation, In Proc. of IEEE Workshop on Neural Networks and Signal Processing, 466-475.

Wan, E.A., Merwe, R.V.D. and Nelson, A.T. (2000), Dual estimation and the unscented transformation, In Neural Information Processing Systems, MIT Press, 666-672.

Weiss, A. (1986), Asymptotics theory for Arch models: estimation and testing, Econometric Theory, 2(1), 107-131.

Wong, K. F. K., Galka, A., Yamashita, O. and Ozaki, T. (2006), Modelling non-stationary variance in EEG time series by state space GARCH model, Comput. Biol. Med., 36, 1327-1335.