Article

Parsimonious Heterogeneous ARCH Models for High Frequency Modeling

Juan Carlos Ruilova [1] and **Pedro Alberto Morettin** [2,*]

[1] Itaú Bank, São Paulo 04344-902, Brazil; juan.teran@itau.com.br
[2] Department of Statistics, University of São Paulo, São Paulo 05508-090, Brazil
[*] Correspondence: pam@ime.usp.br

check for
updates

Abstract: In this work we study a variant of the GARCH model when we consider the arrival of heterogeneous information in high-frequency data. This model is known as HARCH(n). We modify the HARCH(n) model when taking into consideration some market components that we consider important to the modeling process. This model, called parsimonious HARCH(m,p), takes into account the heterogeneous information present in the financial market and the long memory of volatility. Some theoretical properties of this model are studied. We used maximum likelihood and Griddy-Gibbs sampling to estimate the parameters of the proposed model and apply it to model the Euro-Dollar exchange rate series.

Keywords: GARCH model; HARCH model; PHARCH model; Griddy-Gibs; Euro-Dollar

1. Introduction

High frequency data are those measured in small time intervals. This kind of data is important to study the micro structure of financial markets and also because their use is becoming feasible due to the increase of computational power and data storage.

Perhaps the most popular model used to estimate the volatility in a financial time series is the GARCH(1,1) model; see Engle (1982), Bollerslev (1986):

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\,(0,1)\ , \\
\sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2,
\end{aligned}
\tag{1}
$$

with $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$.

When we use high frequency data in conjunction with GARCH models, these need to be modified to incorporate the financial market micro structure. For example, we need to incorporate heterogeneous characteristics that appear when there are many traders working in a financial market trading with different time horizons.

The HARCH(n) model was introduced by Müller et al. (1997) to try to solve this problem. In fact, this model incorporates heterogeneous characteristics of high frequency financial time series and it is given by

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= c_0 + \sum_{j=1}^{n} c_j \left(\sum_{i=1}^{j} r_{t-i} \right)^2,
\end{aligned}
\tag{2}
$$

where $c_0 > 0, c_n > 0, c_j \geq 0\ \forall j = 1, \ldots, n-1$ and ε_t are identically and independent distributed (i.i.d.) random variables with zero expectation and unit variance.

However, this model has a high computational cost to fit when compared with GARCH models, due to the long memory of volatility, so the number of parameters to be estimated is usually large.

We propose a new model known as the parsimonious heterogeneous autoregressive conditional heteroscedastic model, in short-form PHARCH, as an extension of the HARCH model. Specifically, we call a PHARCH(m,p), with aggregations of different sizes $a_1, \ldots, a_m$, where m is the number of the market components, the model given by

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= C_0 + C_1 \left(r_{t-1} + \ldots + r_{t-a_1} \right)^2 + \ldots + \\
&\quad + C_m \left(r_{t-1} + \ldots + r_{t-a_m} \right)^2 + b_1 \sigma_{t-1}^2 + \ldots + b_p \sigma_{t-p}^2,
\end{aligned}
\tag{3}
$$

where $\varepsilon_t \sim$ i.i.d. $(0,1)$, $C_0 > 0$, $C_j \geq 0$, $\forall j = 1, \ldots, m-1$, $C_m > 0$, $b_j \geq 0$, $j = 1, \ldots, p$.

HARCH models are important because they take account the natural behavior of the traders in the market. However they have some problems, mainly because they need to include several aggregations, so the number of parameters to estimate is large, because of the large memory feature of financial time series. Parsimonious HARCH includes only the most important aggregations in its structure, which makes the model more realistic. We can see some simulations in Figure 1, where the characteristics of clustering and volatility are better represented in PHARCH processes than in ARCH or HARCH processes.
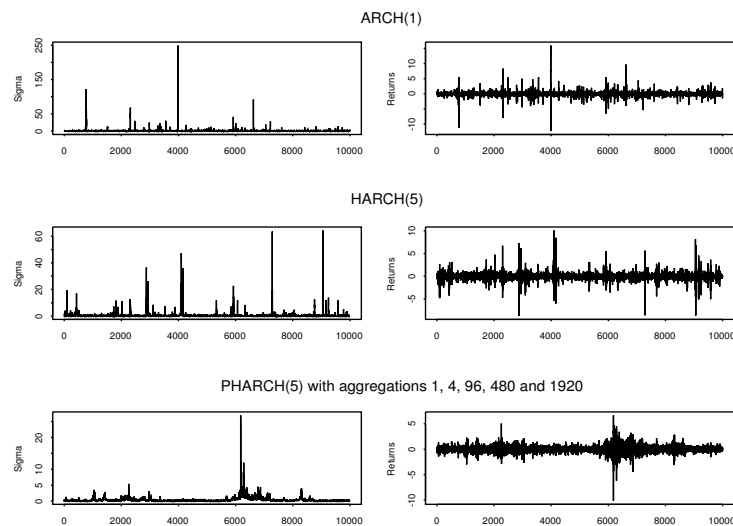

Figure 1. Simulations of ARCH, HARCH and PHARCH processes.

The organization of the paper is as follows. In Section 2 we provide some background information on Markov chains and give the necessary and sufficient conditions for the PHARCH model to be stationary. In Section 3 we obtain forecasts for the proposed model, and in Section 4 we introduce the data that will be used for illustrative purposes. The actual application is given in Section 5, and we close the paper with some conclusions in Section 6.

2. Background

In this section we provide briefly some background on Markov chains and results on stationarity of PHARCH models.

2.1. Markov Chains

Suppose that $\mathbf{X} = \{X_n, n \in \mathbb{Z}^+\}$, $\mathbb{Z}^+ := \{0, 1, 2, \ldots\}$ are random variables defined over $(\Omega, \mathcal{F}, \mathcal{B}(\Omega))$, and assume that $\mathbf{X}$ is a Markov chain with transition probability $P(x, A), x \in \Omega, A \subset \Omega$. Then we have the following definitions:

1. A function $f : \Omega \to \mathbb{R}$ is called the smallest semi-continuous function if $\liminf_{y \to x} f(y) \geq f(x)$, $x \in \Omega$. If $P(\cdot, A)$ is the smallest semi-continuous function for any open set $A \in \mathcal{B}(\Omega)$, we say that (the chain) **X** is a *weak Feller chain*.

2. A chain **X** is called φ-irreducible if there exists a measure φ on $\mathcal{B}(\Omega)$ such that, for all x, whenever $\varphi(A) > 0$, we have,

$$U(x, A) = \sum_{n=1}^{\infty} P^n(x, A) > 0.$$

3. The measure ψ is called maximal with respect to φ, and we write $\psi > \varphi$, if $\psi(A) = 0$ implies $\varphi(A) = 0$, for all $A \in \mathcal{B}(\Omega)$. If **X** is φ-irreducible, then there exists a probability measure ψ, maximal, such that **X** is ψ-irreducible.

4. Let $d = \{d(n)\}$ a distribution or a probability measure on $\mathbb{Z}^+$, and consider the Markov chain $\mathbf{X}_d$ with transition kernel

$$K_d(x, A) := \sum_{n=0}^{\infty} P^n(x, A) d(n).$$

If there exits a transition kernel T satisfying

$$K_d(x, A) \geq T(x, A), \quad x \in \Omega, \ A \in \mathcal{B}(\Omega),$$

then T is called the continuous component of K_d.

5. If **X** is a Markov chain for which there exits a (sample) distribution d such that K_d has a continuous component T, with $T(x, \Omega) > 0$, $\forall x$, then **X** is called a *T-chain*.

6. A measure π over $\mathcal{B}(\Omega)$, σ-finite, with the property

$$\pi(A) = \int_{\Omega} \pi(dx) P(x, A), \ A \in \mathcal{B}(\Omega),$$

is called an *invariant measure*.

The following two lemmas will be useful. See Meyn and Tweedie (1996) for the proofs and further details. We denote by $I_A(\cdot)$ the indicator function of A.

Lemma 1. *Suppose that X is a weak Feller chain. Let $C \in \mathcal{B}(\Omega)$ be a compact set and V a positive function. If*

$$V(X_n) - E[V(X_{n+1}) | X_n] \geq 0, X_n \in C^c,$$

then there exists an invariant measure, finite on compact sets of Ω.

Lemma 2. *Suppose that X is a weak Feller chain. Let $C \in \mathcal{B}(\Omega)$ be a compact set, and V a positive function that is finite at some $x_0 \in \Omega$. If*

$$V(X_n) - E[V(X_{n+1}) | X_n] \geq 1 - bI_C(X_n), X_n \in \Omega,$$

with b a constant, $b < \infty$, then there exists an invariant probability measure π on $\mathcal{B}(\Omega)$.

Lemma 3. *Suppose that X is a ψ-irreducible aperiodic chain. Then the following conditions are equivalent:*

1. *There exists a function $f : \Omega \to [1, \infty)$, a set $C \in \mathcal{B}(\Omega)$, a constant $b < \infty$ and a function $V : \Omega \to [0, \infty)$, such that*

$$V(X_n) - E[V(X_{n+1}) | X_n] \geq f(X_n) - bI_C(X_n), X_n \in \Omega.$$

2.   *The chain is positive recurrent with invariant probability measure π, and $\pi(f) < \infty$.*

2.2. Stationarity of PHARCH(m,p) Models

We first give a necessary condition for Model (3) to be stationary. We know that $E\left(r_{t-i}r_{t-j}\right) = 0, \forall i \neq j$, and if r_t is stationary, we must have

$$E\left(r_t^2\right) = E\left(\sigma_t^2\right) = C_0 + E\left(r_t^2\right) \sum_{i=1}^{m} a_i C_i + E\left(\sigma_t^2\right) \sum_{i=1}^{p} b_i,$$

so

$$E\left[r_t^2\right] = \frac{C_0}{1 - \left(\sum_{i=1}^{m} a_i C_i + \sum_{i=1}^{p} b_i\right)}.$$

Therefore,

$$\sum_{i=1}^{m} a_i C_i + \sum_{i=1}^{p} b_i < 1. \tag{4}$$

To prove a sufficient condition it will be necessary to represent the PHARCH(m,p) as a Markov process. We use the definitions given in the previous section, so the process

$$X_t = \left(r_{t-1}, \ldots, r_{t-a_m+1}, \sigma_t, \ldots, \sigma_{t-p+1}\right), \tag{5}$$

whose elements follow Equation (3), is also a *T*-chain.

The proofs of the following results are based on Dacorogna et al. (1996), and they are given in the Appendix A.

Proposition 1. *The Markov Chain X_t that represents a PHARCH(m,p) process is a T-chain.*

Proposition 2. *The Markov Chain that represents a PHARCH(m,p) process is recurrent with an invariant probability measure (stationary distribution), and its second moments are finite if the condition given in (4) is satisfied.*

Note that if $\varepsilon_t \sim t(v)$ (a Student's t distribution with v degrees of freedom) in (3), then the necessary and sufficient condition becomes

$$\sum_{i=1}^{m} \frac{v}{v-2} a_i C_i + \sum_{i=1}^{p} b_i < 1, \quad \text{for } v > 2.$$

3. Forecasting

In this section we make some considerations about forecasting and validation of the proposed model. Usually two data bases are used for testing tha forecasting ability of a model: one (in-sample), used for estimation, and the other (out-of-sample) used for comparing forecasts with true values. There is an extra complication in the case of volatility models: there is no unique definition of volatility. Andersen and Bollerslev (1998) show that if wrong estimates of volatility are used, evaluation of forecasting accuracy is compromised. We could use the realized volatility as a basis for comparison, or use some trading system.

We could, for example, have a model for hourly returns and use the realized volatility computed from 15 min returns for comparisons. In general, we can compute $v_{h,t} = \sum_{i=1}^{a_h} r_{t-i}^2$, where a_h is the aggregation factor (4, in the case of 15 min returns). Then use some measure based on $s_h = \tilde{v}_{h,t} - v_{h,t}$, for example, mean squared error, where $\tilde{v}_{h,t}$ is the volatility predicted by the proposed model. See Taylor and Xu (1997), for example.

Now consider Model (3). The forecast of volatility at origin t and horizon ℓ is given by

$$\hat{\sigma}_t^2(l) = E(\sigma_{t+l}^2|X_t)$$
$$= E(C_0 + C_1 \left(r_{t+l-1} + \ldots + r_{t+l-a_1}\right)^2 + \ldots +$$
$$+ C_m \left(r_{t+l-1} + \ldots + r_{t+l-a_m}\right)^2 + b_1\sigma_{t+l-1}^2 + \ldots + b_p\sigma_{t+l-p}^2|X_t),$$

where $X_t = (r_t, \sigma_t, r_{t-1}, \sigma_{t-1}, \ldots)$, for $l = 1, 2, \ldots$

Since $a_0 = 1 < a_1 < a_2 < \ldots < a_m < \infty$, then we have three cases:

(i)   If $l = 1$,

$$
\begin{aligned}
\hat{\sigma}_t^2(l) &= E(C_0 + C_1 \left(r_{t+l-1} + \ldots + r_{t+l-a_1}\right)^2 + \ldots + \\
&\quad + C_s \left(r_{t+l-1} + \ldots + r_{t+l-a_s}\right)^2 + \ldots + \\
&\quad + C_m \left(r_{t+l-1} + \ldots + r_{t+l-a_m}\right)^2 + \\
&\quad + b_1\sigma_{t+l-1}^2 + \ldots + b_p\sigma_{t+l-p}^2/X_t) \\
&= C_0 + C_1 \left(r_t + \ldots + r_{t+1-a_1}\right)^2 + \ldots + \\
&\quad + C_s \left(r_t + \ldots + r_{t+1-a_s}\right)^2 + \ldots + \\
&\quad + C_m \left(r_t + \ldots + r_{t+1-a_m}\right)^2 + b_1\sigma_t^2 + \ldots + b_p\sigma_{t+1-p}^2.
\end{aligned}
$$

(ii)   If l is such that $a_{s-1} < l < a_s, s = 1, 2, \ldots, m$, then we have,

$$
\begin{aligned}
\hat{\sigma}_t^2(l) &= E(C_0 + C_1 \left(r_{t+l-1} + \ldots + r_{t+l-a_1}\right)^2 + \ldots + \\
&\quad + C_s \left(r_{t+l-1} + \ldots + r_{t+l-a_s}\right)^2 + \ldots + \\
&\quad + C_m \left(r_{t+l-1} + \ldots + r_{t+l-a_m}\right)^2 + \\
&\quad + b_1\sigma_{t+l-1}^2 + \ldots + b_p\sigma_{t+l-p}^2/X_t) \\
&= E(C_0 + \sum_{i=1}^{s-1} C_i \left(\sum_{j=1}^{a_i} r_{t+l-j}\right)^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=1}^{l-1} r_{t+l-j}\right)^2 + \sum_{i=s}^{m} C_i \left(\sum_{j=l}^{a_i} r_{t+l-j}\right)^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=1}^{l-1} r_{t+l-j}\right)\left(\sum_{j=l}^{a_i} r_{t+l-j}\right) + \\
&\quad + b_1\sigma_{t+l-1}^2 + \ldots + b_p\sigma_{t+l-p}^2/X_t) \\
&= E(C_0 + \sum_{i=1}^{s-1} C_i \left(\sum_{j=1}^{a_i} \sigma_{t+l-j}\varepsilon_{t+l-j}\right)^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=1}^{l-1} \sigma_{t+l-j}\varepsilon_{t+l-j}\right)^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=l}^{a_i} r_{t+l-j}\right)^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=1}^{l-1} \sigma_{t+l-j}\varepsilon_{t+l-j}\right)\left(\sum_{j=l}^{a_i} r_{t+l-j}\right) + \\
&\quad + b_1\sigma_{t+l-1}^2 + \ldots + b_p\sigma_{t+l-p}^2/X_t),
\end{aligned}
$$

and given the independence of ε_t and $E(\varepsilon_t) = 0$, we have $E\left(r_{t-i}r_{t-j}\right) = 0, \forall i \neq j$; hence,

$$
\begin{aligned}
\hat{\sigma}_t^2(l) &= E(C_0 + \sum_{i=1}^{s-1} C_i \sum_{j=1}^{a_i} \sigma_{t+l-j}^2 + \sum_{i=s}^{m} C_i \sum_{j=1}^{l-1} \sigma_{t+l-j}^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=l}^{a_i} r_{t+l-j}\right)^2 + b_1\sigma_{t+l-1}^2 + \ldots + b_p\sigma_{t+l-p}^2/X_t) \\
&= C_0 + \sum_{i=1}^{s-1} C_i \sum_{j=1}^{a_i} \hat{\sigma}_{t+l-j}^2 + \sum_{i=s}^{m} C_i \sum_{j=1}^{l-1} \hat{\sigma}_{t+l-j}^2 + \\
&\quad + \sum_{i=s}^{m} C_i \left(\sum_{j=l}^{a_i} r_{t+l-j}\right)^2 + b_1\tilde{\sigma}_{t+l-1}^2 + \ldots + b_p\tilde{\sigma}_{t+l-p}^2,
\end{aligned}
$$

where, for $i = 1, \ldots, p$, we have that $\tilde{\sigma}_{t+l-i}^2 = \begin{cases} \sigma_{t+l-i}^2, & i \geq l \\ \hat{\sigma}_{t+l-i}^2, & i < l \end{cases}$

(iii)   If l is such that $l > a_m, s = 1, 2, \ldots, m$, then it follows

$$\hat{\sigma}_t^2(l) = E\left(C_0 + \sum_{i=1}^{m} C_i \sum_{j=1}^{a_i} \sigma_{t+l-j}^2 + b_1 \sigma_{t+l-1}^2 + \ldots + b_p \sigma_{t+l-p}^2 / X_t\right)$$

$$= C_0 + \sum_{i=1}^{m} C_i \sum_{j=1}^{a_i} \hat{\sigma}_{t+l-j}^2 + b_1 \tilde{\sigma}_{t+l-1}^2 + \ldots + b_p \tilde{\sigma}_{t+l-p}^2,$$

where for $i = 1, \ldots, p$, we have $\tilde{\sigma}_{t+l-i}^2 = \begin{cases} \sigma_{t+l-i}^2, & i \geq l \\ \hat{\sigma}_{t+l-i}^2, & i < l \end{cases}$

4. High Frequency Data

In this section we further elaborate on high frequency data and introduce the series that will be analyzed later. High frequency data are very important in the financial environment, mainly because there exist large movements in short intervals of time. This aspect represents an interesting opportunity for trading. Furthermore, it is well known that volatilities in different frequencies have significant cross-correlation. We can even say that coarse volatility predicts fine volatility better than the inverse, as shown in Dacorogna et al. (2001).

As an example, take the tick by tick foreign exchange (FX) time series Euro-Dollar, from January First 1999 to December 31, 2002. Returns are calculated using bid and ask prices, as

$$r_t = \ln\left(\left(p_t^{bid} + p_t^{ask}\right)/2\right) - \ln\left(\left(p_{t-1}^{bid} + p_{t-1}^{ask}\right)/2\right). \tag{6}$$

We discard Saturdays and Sundays, and we replace holidays with the means of the last ten observations of the returns for each respective hour and day. After cleaning the data (see Dacorogna et al. (2001), for details) we will consider equally spaced returns, with sampling interval $\Delta t = 15$ min. This seems to be adequate, as many studies indicate.

Figure 2 shows Euro-Dollar returns calculated as above. The length of this time series is 95,317. The figure shows that the absolute returns present a seasonal pattern. This is due to the fact that physical time does not follow, necessarily, the same pattern as the business time. This is a typical behavior of a financial time series and we will use a seasonal adjustment procedure similar to that of Martens et al. (2002). However, we will use absolute returns instead of squared returns; that is, we will compute the seasonal pattern as

$$S_{d,s,h} = \frac{1}{s} \sum_{j=1}^{s} |(r_{d,j,h}|, \tag{7}$$

where $r_{d,s,h}$ is the return in the weekday d, week s and hour h, and s is the number of weeks from the beginning of the series. Therefore, $S_{d,N_s,h}$ is the rolling window mean of the absolute returns with the beginning fixed.

In Figure 3 we have the autocorrelation function of these returns and of squared returns. The seasonality pattern is no longer present.

FX data has some distinct characteristics, mainly because they are produced twenty four hours a day, seven days a week. In particular, Euro-Dollar is the most liquid FX in the world. However, there are periods where the activity is greater or smaller, causing seasonal patterns to occur, as seen above.

Let us analyze some facts about these returns that we will denote simply by r_t. We can see in Figure 4 the histogram fitted with a non-parametric density kernel estimate, using unbiased cross-validation method to estimate the bandwidth. It shows fat tails and high kurtosis, namely, 121, while its skewness coefficient is -0.079, showing almost symmetry. A normality test (Jarque-Bera) rejects the hypothesis that these returns are normal.
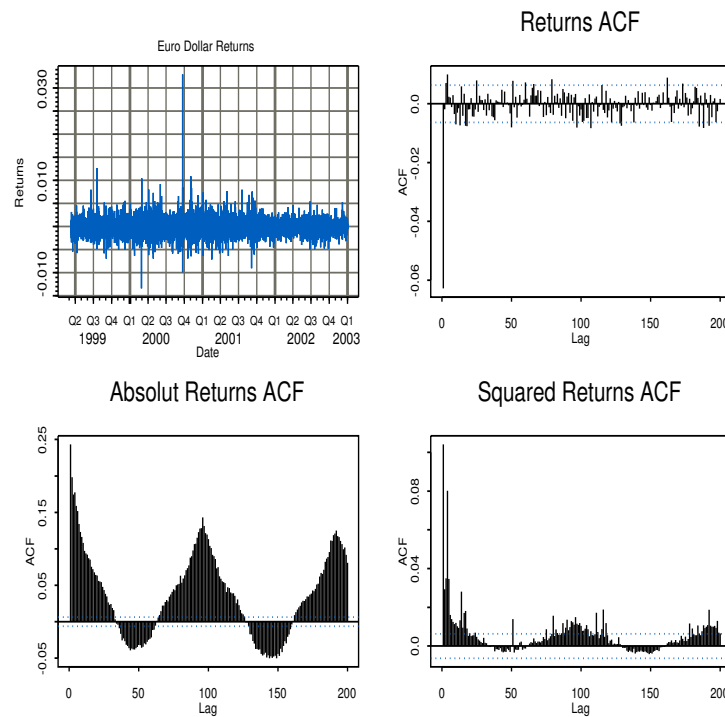
Figure 2. Euro-Dollar returns: acf of returns, acf of absolute returns and acf of squared returns.

The seasonally adjusted returns are then given by

$$\tilde{r}_t = \tilde{r}_{d,s,h} = \frac{r_{d,s,h}}{S_{d,s,h}}. \tag{8}$$

We may assume for example that the errors of a GARCH model fitted to these returns follow a Student's *t* distribution or a generalized error distribution, which represents better the fat tails of the distribution.

Often the optimization of the likelihood function can be a very difficult task, due mainly to the flat behavior of likelihood function, as can be seen in Zumbach (2000). Bayesian methods are an alternative, and in the next section we will use the Griddy-Gibbs sampling to estimate the parameters of a PHARCH model.

Figure 4 also shows that the Euro-Dollar series has some clusters of volatility. This is a typical behavior of financial time series. A problem is that we do not know how many clusters there are and what their sizes are. The reason for this is that the information arriving is different for each sampling frequency.

We can look these clusters as *market components* and they depend on the heterogeneity of the market. These market components are considered in our PHARCH model, as seen in Equation (3). Differently from GARCH-type models, PHARCH models have a variance equation with returns over intervals of different sizes. Therefore PHARCH models take into account the sign of the returns and not only their absolute value as GARCH models do. Two subsequent returns with similar sizes in the same direction will cause a higher impact on the variance than two subsequent returns with similar sizes but opposite signs.

Now we need to determine the number and the size of the market components for the Euro-Dollar FX series. Ruilova (2007) proposed some technical rules to determine these market components, and Dacorogna et al. (2001) proposed some empirical rules.

To help us to determine if the component sizes chosen are correct we can use the *impact of the component*.

We define the impact I_i of the ith component as,

$$I_i = a_i C_i, \forall i. \tag{9}$$

Note that the stationary condition to PHARCH(m) models can be written in terms of these impacts; namely,

$$\sum_{i=1}^{m} I_i < 1.$$

We also notice that if we consider the Student's t distribution with ν degrees of freedom, the impact should be defined as

$$I_i = \frac{v}{v-2} a_i C_i, \quad \forall i \geq 1. \tag{10}$$

As remarked above, the number of components in a financial series can vary depending how the returns are being traded in this market. That is, liquid series can have a structure with more components than a non-liquid series.

5. Application

Due to the complexity of the proposed model, the likelihod function may be flat in the neighbourhood of the maxima, so the optimization procedure using traditional procedures may fail. An alternative is to use Bayesian methods. Some references on the use of Bayesian procedures for the family of ARCH processes are Geweke (1989), Kleibergen and Dijk (1993), Geweke (1994) and Bauwens and Lubrano (1998).
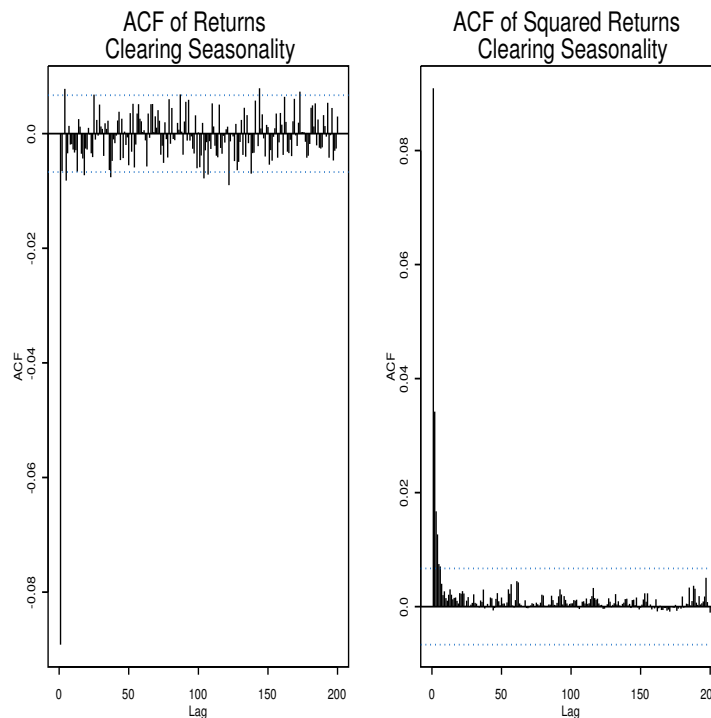

Figure 3. Autocorrelation functions of the Euro-Dollar returns and squared returns after seasonal adjustment.
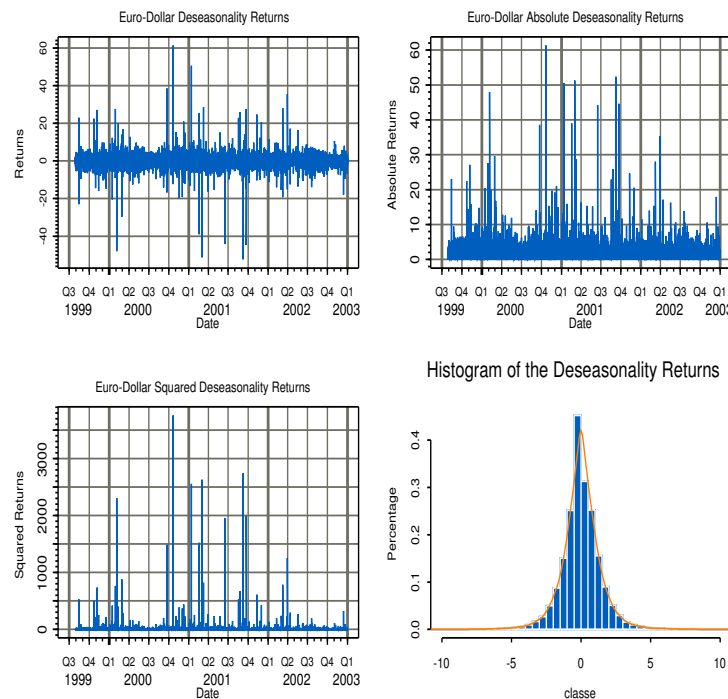
Figure 4. Euro-Dollar returns, absolute returns, squared returns and histogram after taking off seasonal pattern.

It is well known that when the analytical expressions of the full conditional distributions are known we can use Gibbs sampling. However, if the conditional distributions are not known, we need to modify the algorithm or to use another algorithm, such as the Metropolis-Hastings one. Another alternative to solve this is to use the Griddy-Gibbs sampler of Ritter and Tanner (1992).

Griddy-Gibbs sampling can be used when the joint conditional distribution of at least one parameter does not have a distribution form known but it has an analytical expression that can be evaluated on a grid of points. For that, we evaluate the analytical expressions of the joint conditional distribution function, and by numerical integration we can generate random variables of this distribution; see Davis and Rabinowitz (1975).

A problem that appears when we use Griddy-Gibbs sampling is to determine a window and the number of points where we will evaluate numerically the desired function. An inadequate determination of this grid of points could cause errors in the parameters estimation. In general it seems suitable to have 50 points in the grid for a good evaluation.

We will use a technique to reduce the variance that is to compute the conditional mean

$$\sum_{n=1}^{N} E\left(\theta_i | \theta_1^n, \ldots, \theta_{i-1}^n, \theta_{i+1}^n, \ldots, \theta_n^n, y\right) / N,$$

instead of $\sum_{n=1}^{N} \theta_i^n / N$ to estimate $E\left(\theta_i | \theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n, y\right)$. Here θ_i^n denotes the value of the parameter θ_i at iteration n.

An important fact is that aggregate returns have, generally, a magnitude greater than non-aggregate returns, so the components with larger aggregations have smaller values. For this reason, we can use the impacts defined above to study the contribution of each component to the model.

In order to establish the number of components in the PHARCH model, we will use information of the financial market behavior based on the behavior of the traders. So, we consider five components, as seen in the Table 1, corresponding to information arriving at the market at the rate of 15 min, 1 h, 1 day, 1 week and 1 month.

This means that we need to estimate the parameters of a PHARCH(5) process with aggregations 1, 4, 96, 480 and 1920 as follows.

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= C_0 + C_1 r_{t-1}^2 + C_2 \left(r_{t-1} + \ldots + r_{t-4} \right)^2 + \\
&\quad + C_3 \left(r_{t-1} + \ldots + r_{t-96} \right)^2 + C_4 \left(r_{t-1} + \ldots + r_{t-480} \right)^2 + \\
&\quad + C_5 \left(r_{t-1} + \ldots + r_{t-1920} \right)^2
\end{aligned}
\tag{11}
$$

where $C_j \geq 0$, $j = 1, \ldots, 5$, $C_0 > 0$, and $\varepsilon_t \sim t(0, 1, v)$.

The number of parameters to estimate is seven because we considered $\varepsilon_t \sim t(0, 1, v)$, $v > 2$. We use an autoregressive processes to filter the data and to take into account the information given by the acf function of the returns shown in the Figure 3.

We consider non-informative priors, that is, uniform distributions on the parametric space, as follows: $C_0, C_1, C_2, C_3, C_4, C_5 \sim U(0, 1)$ and $v \sim U(3, ct)$, where U denotes the uniform distribution and ct is a large number; in particular, we used $ct = 50$.

Estimates using maximum likelihood (ML) are shown in Table 2, and the corresponding impacts are shown in Figure 5. The optimizer used to evaluate the impacts was simulated annealing; see Belisle (1992). Several problems were faced in the process of using ML because in some situations the optimizer did not converge. Sometimes we can solve this problem, using initial values near to optimum. But this may not be a normal situation in real cases. So the need for alternative procedures.

As we can see, the impact of the components decreases for larger aggregations. This is a natural result because intraday traders are those who dominate the market. Another fact is that the weekly component has a similar impact to the monthly component, meaning that both have similar weight contributions to predicting volatility. The results show that an impact can be significant even when the parameter is small. The above estimates will serve as a comparison with Griddy-Gibbs estimates.

In the Griddy–Gibbs sampling we use a moving window: we define a new window in each iteration as a function of the mean, mode and standard deviation.

Table 1. Component description of PHARCH process for Euro-Dollar.

Component	Aggregations of Euro-Dollar	Range of Time Intervals	Description
1	1	15 min	Short term, intraday traders, market makers.
2	4	1 h	Intraday traders with few transactions per day.
3	96	1 day	Daily traders.
4	480	1 week	Medium term traders.
5	1920	1 month	Long term traders, investors, derivative traders.

Table 2. Parameter estimation by maximum likelihood using simulate annealing optimizer.

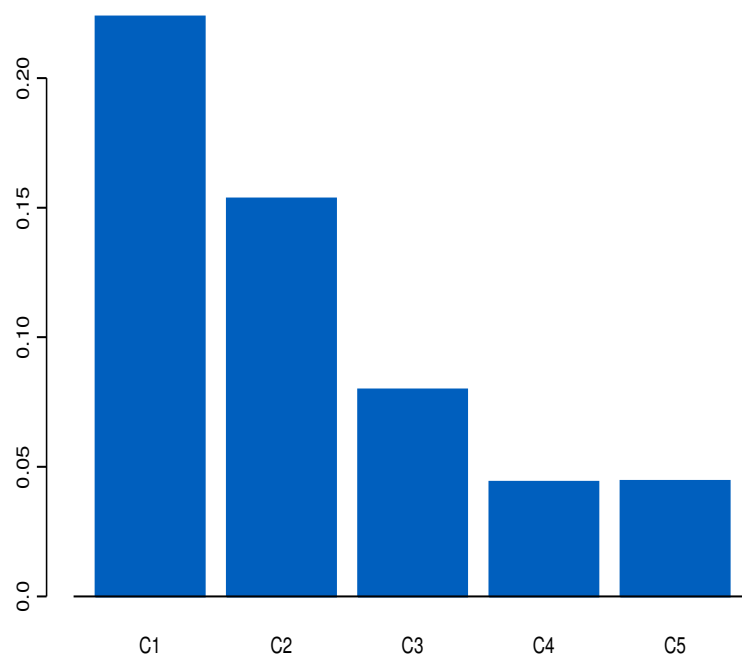| Parameter	Estimates
C_0	0.529
C_1	0.101
C_2	0.0173
C_3	0.000374
C_4	0.0000415
C_5	0.0000105
ν	3.638

Figure 5. Impact of the components estimated by maximum likelihood.

Table 3 shows the results of the estimation of a parsimonious HARCH(5) model for the Euro-Dollar, using this criterion of selection for a moving window for each parameter, where the conditional density will be computed. The number of points where this density was evaluated was 50.

We used the non conditional and conditional mean in each step of the iteration to calculate the estimate parameters. As expected, conditional method was faster than non-conditional, but the difference was very small.

We see that the values are practically the same by both methods (maximum likelihood and Griddy-Gibbs sampling).

In Figure 6 we can see the convergence of the parameters using Griddy-Gibbs for each iteration step. We can see the fast convergence of the parameters.
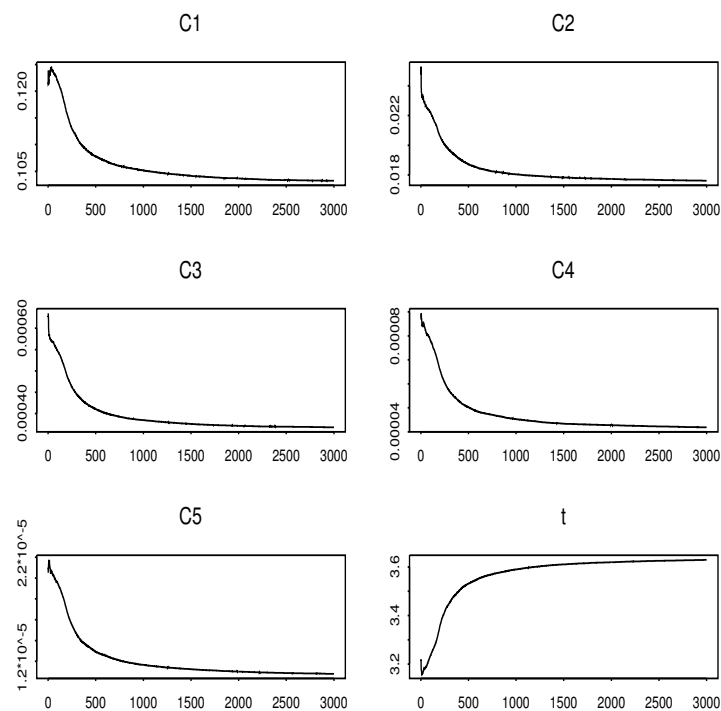
Figure 6. Convergence of the parameters using Griddy-Gibbs sampler.

Now, we compare HARCH modeling with GARCH modeling. In Figures 7–9 we present a residual analysis after the fitting of a GARCH model. In Figures 10–12 we have the corresponding graphs for the PHARCH(5) fitting. We see a slightly better fit of the PHARCH model. If we use the prediction mean squared error (PMSE) as a criterion for comparison, we obtain the values 15.58 and 15.20, for GARCH and PHARCH modeling, respectively, using the standardized residuals and 1000 values for the prediction period.

Table 3. Estimated parameters for the PHARCH(5)model with aggregations 1, 4, 96, 480 and 1920 for the Euro-Dollar series, using Griddy-Gibbs sampling.

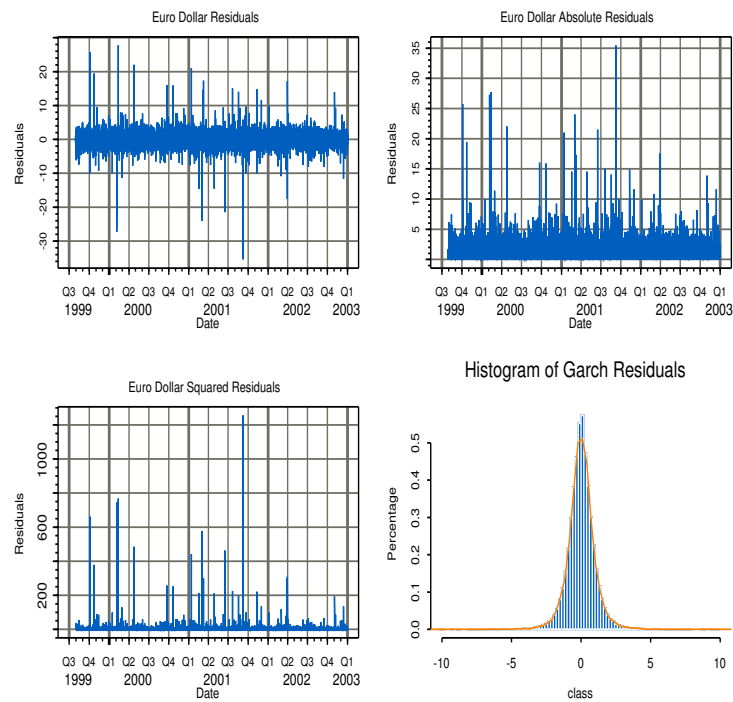| Parameter	Non-Conditional	Conditional	Std. Dev.
C_0	0.532	0.532	0.00471
C_1	0.102	0.102	0.00321
C_2	0.0174	0.0174	0.000709
C_3	0.00359	0.000360	0.0000205
C_4	0.0000402	0.0000403	0.00000421
C_5	0.0000104	0.0000103	0.00000105
ν	3.649	3.649	0.034

Figure 7. Euro-Dollar residuals, absolute residuals, squared residuals and histogram of the residuals after fitting a GARCH process.
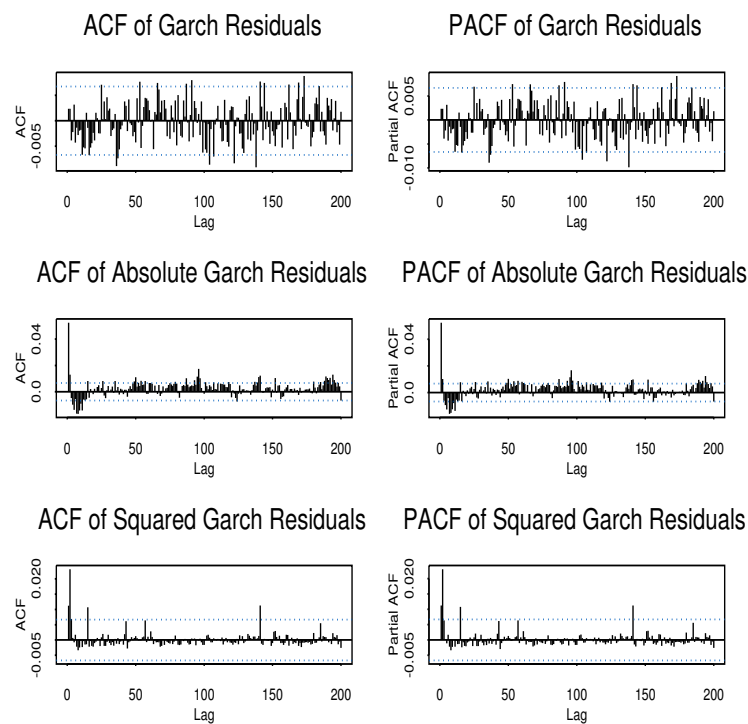

Figure 8. Autocorrelation and partial autocorrelation functions of the residuals, absolute residuals and squared residuals after GARCH fitting.
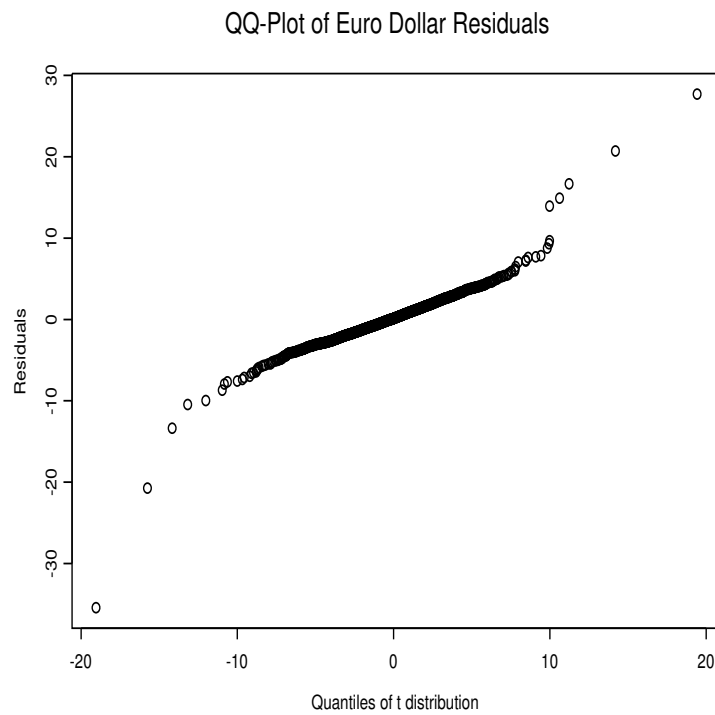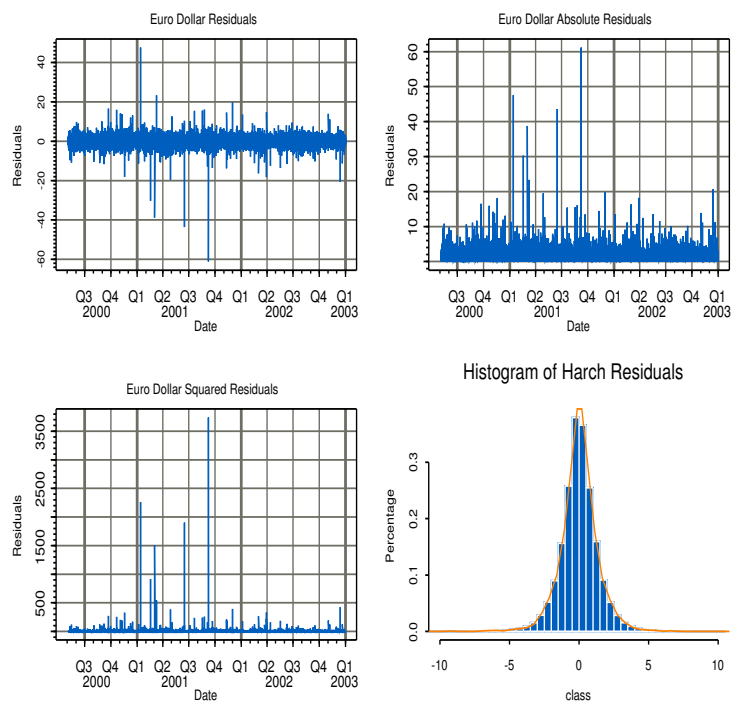
Figure 9. QQ-Plot of GARCH Residuals.

Figure 10. Euro-Dollar residuals, absolute residuals, squared residuals and histogram of the residuals after fitting a PHARCH process.
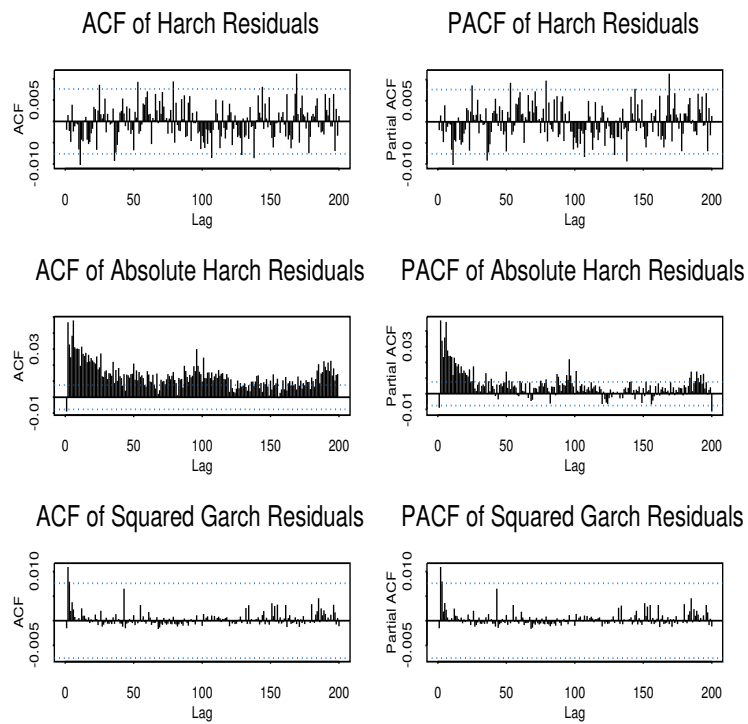
Figure 11. Autocorrelation and partial autocorrelation functions of the residuals, absolute residuals and squared residuals after PHARCH fitting.
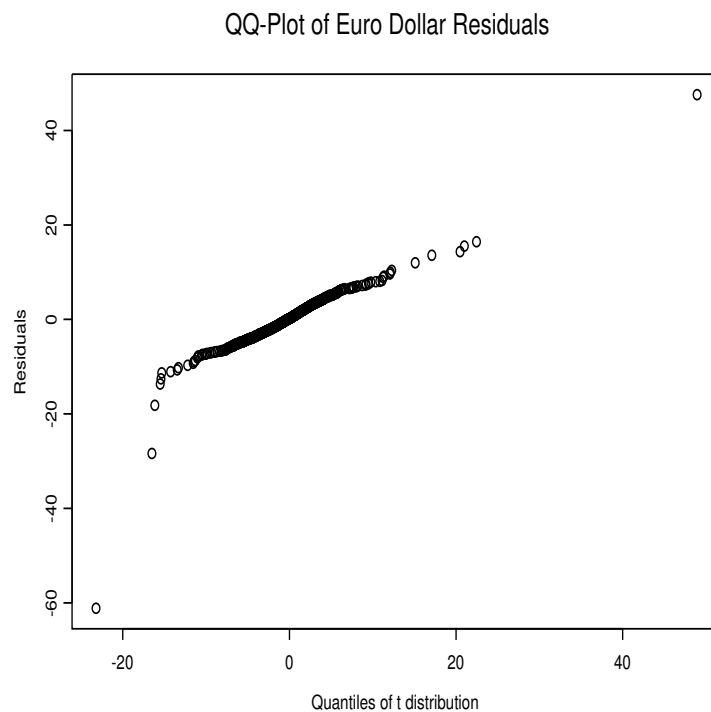

Figure 12. QQ-plot of PHARCH residuals.

6. Conclusions

PHARCH models are good models for the analysis of high frequency data, since the financial market agents behave differently, incorporating heterogeneous information to the

market microstructure. Nevertheless, their use still depends on solving some issues, mainly computational ones.

One big challenge in the analysis of high frequency data is dealing with large amounts of observations, and complex models bring computational difficulties, even with the recent technological breakthroughs in computing technology. Therefore, the first issue here is to develop techniques that help us to improve the computational algorithms. Maximum likelihood estimation may collapse, as we have described earlier. Techniques such as genetic algorithms and neural networks are viable optimization alternatives.

Another possibility is to use Bayesian techniques, such as the Griddy-Gibbs samples that we have used. The disadvantage of the Griddy-Gibbs sampler lies in its high computational load. From another viewpoint, more sophisticated volatility models might be developed, taking into account the arrival of information, for example, a stochastic volatility model or stochastic duration model; or we could adapt existing models such as CHARN (conditional heteroskedasticity nonlinear autoregressive) models to heterogeneity of information characteristics. Finally, extensions similar to those proposed to GARCH models could be studied for HARCH models.

A feature of the HARCH models is that the market components are chosen in a subjective way. In the analysis of the Euro-Dollar series, we considered five components, with different aggregations. A different number of components could be proposed, depending of the degree of information one has. This is clearly a matter for further studies.

One last remark is that the performance of the different estimation methods should be evaluated. This evaluation could be done using prediction capabilities, for example. Other possibility is calculating some measure of risk. Volatility models are often established with the purpose of computing the VaR (value at risk) or other risk measure or for establishing trading strategies. In this context, an evaluation of the performance of the proposed model and several estimation procedures should be interesting. A comparison of returns of different trading systems that use a proposed model will be of fundamental importance. Further details on these aspects can be found in Acar and Satchell (2002), Dunis et al. (2003), Ghysels and Jasiak (1994) and Park and Irwin (2005).

Other models for high frequency data use the realized volatility as a basis, instead of models such as ours and models of the ARCH family, which assume that volatility is a letent variable. Among the former, we mention the autoregressive fractionally integrated moving average (ARFIMA) models, the heterogeneous autoregressive model of realized volatilidade (HAR–RV) of Corsi (2009) and the mixed data sampling regression (MIDAS) proposed by Ghysels Santa-Clara and Valkanov (2002). A comparison of the PHARCH models with HAR and MIDAS would be useful, but due to the length of the present paper, this will be the object of future research.

Author Contributions: The authors study some theoretical properties of the PHARCH models and illustrate the theory with an application to real data. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Proposition 1. Let X_t following (3) and (5). Assume that the innovations distribution has a non-zero absolutely continuous component, with a positive density on a Borel set with a non-empty interior. Examples of this are the normal and Student's t distribution.

Then X_t can be written as $X_t = H(X_{t-1}, \varepsilon_t)$, where H is a non-linear continuous function for each ε_t fixed. Then, using the *Continuous Mapping Theorem* we obtain the weak convergence of X_t, namely, the conditional distribution X_t given $X_{t-1} = y_k$ converges to the conditional distribution of X_t given $X_{t-1} = y$ if $y_k \to y$. So, the Markov Chain X_t that represents the PHARCH(m,p) process is also a T-chain. □

Proof of Proposition 2. Define the function V as:

$$V\left(X_t\right) := \sum_{i=1}^{a_m-1} \alpha_i r_{t-i}^2 + 2 \sum_{i=1}^{a_m-2} \sum_{j=i+1}^{a_m-1} \beta_j r_{t-i} r_{t-j} + \sum_{i=0}^{p-1} \gamma_i \sigma_{t-i}^2.$$

A simple algebraic computation gives

$$\mathbb{E}\left[\sum_{i=1}^{a_m-1} \alpha_i r_{t+1-i}^2 \,|\, X_t\right] = \alpha_1 \sigma_t^2 + \sum_{i=1}^{a_m-2} \alpha_{i+1} r_{t-i}^2,$$

$$\mathbb{E}\left[2 \sum_{i=1}^{a_m-2} \sum_{j=i+1}^{a_m-1} \beta_j r_{t+1-i} r_{t+1-j} \,|\, X_t\right] = 2 \sum_{i=1}^{a_m-3} \sum_{j=i+1}^{a_m-2} \beta_{j+1} r_{t-i} r_{t-j},$$

and using (4) we have,

$$\mathbb{E}\left[\sum_{i=0}^{p-1} \gamma_i \sigma_{t+1-i}^2 \,|\, X_t\right] = \gamma_0 \left[C_0 + \sum_{i=1}^{m} C_i \left[\sigma_t^2 + \left(\sum_{j=1}^{a_i-1} r_{t-j}\right)^2\right] + \sum_{i=1}^{p} b_i \sigma_{t+1-i}^2\right] +$$

$$+ \sum_{i=1}^{p-1} \gamma_i \sigma_{t+1-i}^2.$$

Therefore, taking $\alpha_{a_m} = \beta_{a_m} = \gamma_p = 0$ and $a_0 = 1$ and grouping we have

$$V\left(X_t\right) - \mathbb{E}\left[V\left(X_{t+1}\right) \,|\, X_t\right] = \sum_{k=1}^{m} \sum_{j=a_{k-1}}^{a_k-1} \left(\alpha_j - \alpha_{j+1} - \gamma_0 \sum_{i=k}^{m} C_i\right) r_{t-j}^2 +$$

$$+ \sum_{l=1}^{m} \sum_{j=1}^{a_l-2} \sum_{k=a_{l-1}+j}^{a_l-1} \left(\beta_k - \beta_{k+1} - \gamma_0 \sum_{i=l}^{m} C_i\right) r_{t-j} r_{t-k} +$$

$$+ \sum_{i=1}^{p-1} \left(\gamma_i - \gamma_{i+1} - \gamma_0 b_{i+1}\right) \sigma_{t-i}^2 +$$

$$+ \left(\gamma_0 - \gamma_1 - \gamma_0 b_1 - \alpha_1 - \gamma_0 \sum_{i=1}^{m} C_i\right) \sigma_t^2 - \gamma_0 C_0.$$

We choose $k \in \mathbb{Z}^+$, $a_{l-1} < k < a_l$, $l \in \{1, \dots, m\}$, and $\beta_k = \beta_{k+1} + \sum_{i=l}^{m} C_i$.

If we take $\alpha_j > \beta_j$, for all j, then we have $V\left(X_t\right) \geq 0$, and we can take $\gamma_0 = 1$, so

$$V(X_t) - \mathbb{E}\left[V\left(X_{t+1}\right) \,|\, X_t\right] = \sum_{k=1}^{m} \sum_{j=a_{k-1}}^{a_k-1} \left(\alpha_j - \alpha_{j+1} - \sum_{i=k}^{m} C_i\right) r_{t-j}^2$$

$$+ \sum_{i=1}^{p-1} \left(\gamma_i - \gamma_{i+1} - b_{i+1}\right) \sigma_{t-i}^2 \qquad \text{(A1)}$$

$$+ \left(1 - \gamma_1 - b_1 - \alpha_1 - \sum_{i=1}^{m} C_i\right) \sigma_t^2 - C_0.$$

Similarly, we can choose $k \in \mathbb{Z}^+$, $a_{l-1} < k < a_l$, $l \in \{1, \dots, m\}$, $\alpha_k > \alpha_{k+1} + \sum_{i=l}^{m} C_i$ and $\gamma_i > \gamma_{i+1} + b_{i+1}$.

If $\sum_{i=1}^{m} a_i C_i + \sum_{i=1}^{p} b_i < 1$, then we have for the expressions in Equation (A1):

$\xi_i = \left(\alpha_i - \alpha_{i+1} - \sum_{j=k}^{m} C_j\right) > 0$, $i = 1, \dots, a_m - 1$ and k is chosen such that $i \in (a_{k-1}, a_k)$;

$\xi_{a_m-1+i} = \left(\gamma_i - \gamma_{i+1} - b_{i+1}\right) > 0$, $i = 1, \dots, p-1$; and

$$\xi_{a_m+p-1} = \left(1 - \gamma_1 - b_1 - \alpha_1 - \sum_{i=1}^{m} C_i\right) > 0.$$

So, $V(X_t) - \mathbb{E}[V(X_{t+1})|X_t]$ can be as large as we want if $X_t \in C^c$.

Then, using Lemma 1 we have that there exists an invariant measure, finite on compact sets of Ω.

Choosing $\varpi_{\min} = \min(\xi_i)$, for $i = 1, \ldots, a_m + p - 1$, we have,

$$V(X_t) - \mathbb{E}[V(X_{t+1})|X_t] \geq \varpi_{\min} \sum_{i=1}^{a_m-1} r_{t-j}^2 + \varpi_{\min} \sum_{i=1}^{p} \sigma_{t-i}^2 - C_0$$

$$\geq 1 - (C_0 + 1) I_{B\left(0, \sqrt{\frac{C_0+1}{\varpi_{\min}}}\right)},$$

where $B(c, r)$ is the ball with center c and radius r.

Therefore, the Markov chain $X_t = (r_{t-1}, \ldots, r_{t-a_m+1}, \sigma_t, \ldots, \sigma_{t-p+1})$ that represents the PHARCH(m,p) process is recurrent, with an invariant probability measure (stationary distribution).

Now, if we consider $f(x_1, \ldots, x_{a_m+p-1}) = x_1^2 + \ldots + x_{a_m+p-1}^2$, then,

$$V(X_t) - \mathbb{E}[V(X_{t+1})|X_t] \geq \frac{\varpi_{\min}}{2} f(r_{t-1}, \ldots, r_{t-a_m+1}, \sigma_t, \ldots, \sigma_{t-p+1}) +$$

$$+ C_0 1_{B\left(0, \sqrt{\frac{2C_0}{\varpi_{\min}}}\right)}.$$

We conclude, using Lemmas 2 and 3, that the process X_t is a T chain having a stationary distribution with finite second order moments. $\square$

References

Acar, Emmanuel, and Stephen Satchell. 2002. *Adavanced Trading Rules*. Oxford: Butterworth Heinemann.

Andersen, Torben G., and Tim Bollerslev. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39: 885–905. [CrossRef]

Bauwens, Luc, and Michel Lubrano. 1998. Bayesian inference on GARCH models using the gibbs sampler. *Econometrics Journal* 1: 23–46. [CrossRef]

Bélisle, Claude J. 1992. Convergence theorems for a class of simulated annealing algorithms on R^d. *Journal of Applied Probability* 29: 885–95. [CrossRef]

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Corsi, Fulvio. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7: 174–96 [CrossRef]

Dacorogna, Michel M., Ramazan Gençay, Ulrich Muller, Richard B. Olsen, and Olivier V. Pictet. 2001. *An Introduction to High-Frequency Finance*. Amsterdam: Elsevier.

Dacorogna, Michel M., Ulrich A. Müller, Paul Embrechts, and Gennady Samorodnitsky. 1996. How heavy are the tails of a stationary HARCH(k) process? A study of the moments. In *Stochastic Processes and Related Topics*. Boston: Birkäuser.

Davis, Philip J., and Philip Rabinowitz. 1975. *Methods Of Numerical Integration*. New York: Academic Press.

Dunis, Christian L., Jason Laws, and Patrick Naïm. 2003. *Applied Quantitative Methods For Trading And Investment*. Chichester: John Wiley & Sons Ltd.

Engle, Robert F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of u. k. inflation. *Econometrica* 50: 987–1008. [CrossRef]

Geweke, John. 1989. Exact predictive densities in linear models with ARCH disturbances. *Journal of Econometrics* 40: 63–86. [CrossRef]

Geweke, John. 1994. *Bayesian Comparison of Econometric Models*. Technical Report Working Paper 532, Research Department, Federal Reserve Bank of Minneapolis, Minneapolis, MN, USA.

Ghysels, Eric, and Joanna Jasiak. 1994. *Stochastic Volatility and Time Deformation: An Apllication to Trading Volume and Leverage Effects*. Technical report. Santa Fé: Western Finance Association Meeting.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2002. *The MIDAS Touch: Mixed Data Sampling Regression Models*. Working paper, UNC and UCLA, Chapel Hill, NC, USA.

Kleibergen, Frank, and Herman K. Van Dijk. 1993. Non-stationarity in GARCH models: A bayesian analysis. *Journal of Applied Econometrics* 8 (Suppl. 1): 41–61.

Martens, Martin, Yuan-Chen Chang, and Stephen J. Taylor. 2002. A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research* 25: 283–99. [CrossRef]

Meyn, Sean P., and Richard L. Tweedie. 1996. *Markov Chains and Stochastic Stability*. Heidelberg: Springer.

Müller, Ulricn, Michel M. Dacorogna, Rakhal D. Davé, Richard B. Olsen, Olivier V. Pictet, and Jacob E. von Weizsäcker. 1997. Volatilities of different time resolutions—Analyzing the dynamics of market components. *Journal of Empirical Finance* 4: 213–89. [CrossRef]

Park, Cheol-Ho, and Scott H. Irwin. 2005. *The Profitability of Technical Trading Rules in US Futures Markets: A Data Snooping Free Test*. Technical report, AgMAS Project Research Report. Urbana-Champaign: University of Illinois.

Ritter, Christian, and Martin A. Tanner. 1992. Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler. *Journal of the American Statistical Association* 87: 861–68. [CrossRef]

Ruilova, Juan Carlos. 2007. Modelos Arch Heterogêneos E Aplicações À Análise De Dados De Alta Freqüência. Ph.D. thesis, Institute of Mathematic and Statistics, University of São Paulo, São Paulo, Brazil.

Taylor, Stephen J., and Xinzhong Xu. 1997. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4: 317–40. [CrossRef]

Zumbach, Gilles. 2000. The pitfalls in fitting GARCH(1,1) processes. In *Advances in Quantitative Asset Management*. Technical report. Boston: Springer.