UNIVERSIDAD NACIONAL DE COLOMBIA

A transfer function model for volatilities between water inflows and spot prices for Colombian electricity market

National University of Colombia
Sciences Faculty, Department of Statistics
Medellín, Colombia
2015

A transfer function model for volatilities between water inflows and spot prices for Colombian electricity market

David Andrés Galeano González

Thesis proposed as a prerequisite to obtain the title of:
Master of Science - Statistics

Director:
MSc. Norman Diego Giraldo Gómez.
Associated Professor.

National University of Colombia
Sciences Faculty, Department of Statistics
Medellín, Colombia
2015

'Keep an open mind but not so open that your brain falls out'
Richard Feynman

Acknowledgement

To my beloved wife, my mother and all my family for their patience and understanding in this difficult but fruitful path.

Also, I want to thank Professor Norman Giraldo for his guidance and assistance throughout the process of elaboration of this work.

Resumen

La canasta de generación eléctrica en Colombia es predominantemente hidroeléctrica. Fenómenos que pudieran generar hidrologías extremas como El Niño o La Niña causan nerviosismo entre las empresas generadoras de electricidad y por lo tanto, aumenta la volatilidad del precio de bolsa de energía eléctrica.

En este trabajo nosotros introducimos un modelo de función de transferencia entre las volatilidades de los Aportes Hidrológicos y el Precio de la Energía con base en modelos del tipo SARFIMA-GARCH. El modelo para el Precio de Energía incorpora estacionalidad y memoria larga. El modelo de transferencia confirma que el régimen hidrológico influencia el precio y la parte GARCH del modelo de Precios puede incorporar la volatilidad de los Aportes Hidrológicos como una variable exógena. Es importante hacer notar que aunque la volatilidad del precio de bolsa está influenciada por otras variables, aquí solo se analizará la influencia de los aportes hidrológicos.

Palabras clave:
Rezagos autoregresivos, Volatilidad, Contagio, Función de transferencia,ADL-Koyck.

Abstract

The electricity generation mix in Colombia is predominantly hydroelectric. Phenomena that could generate extreme hydrology as El Niño or La Niña cause nervousness among electricity generators and therefore, the Electricity Spot Price increases the volatility.

In this paper we propose a transfer function model between the volatilities of Water Inflows and Energy Price based on models of SARFIMA-GARCH type. The model for Energy Prices incorporates seasonality and long memory. The transfer model confirms that the hydrology regimen influences the price and the GARCH part of the Price model can incorporate the volatility of Water Inflows as an exogenous variable. It is important to note that, although the spot price volatility is influenced by other variables, we only analyze the influence of water inflows.

Keywords:
Autoregressive lags, Volatility, Contagion, Transfer Function, ADL-Koyck

Table of Contents

List of Figures

List of Tables

1. Introduction

The electricity generation in Colombia is predominantly hydroelectric and therefore, in the electricity market there exists always an uncertainty about the future behavior of hydrology because the reservoirs are limited and Water Inflows have a stochastic behavior. This uncertainty implies a change in the strategies of commercial operation among the electricity generating companies, especially during periods (or expected) of El Niño and La Niña phenomena. Throughout Colombian history it has been observed that there is a close relationship between the rainfall variations and the occurrence of El Niño or La Niña phenomena.

The hydrological uncertainty causes that the generation companies review the different strategies depending on the current and future availability of resources, however, the future availability of resources will always have a degree of uncertainty which causes some "nervousness" about what is optimal trading strategy.

When we speak of El Niño or La Niña phenomena, this does not necessarily imply drought flows or heavy rains in Colombia, respectively. It has been seen that there is a close correlation between the occurrence of these phenomena and the respective hydrological event in most of the country (see Puertas and Carvajal [36]). Additionally, it has been observed that the simple expectation of the occurrence of El Niño or La Niña, changes (up or down) the electricity price (See Quinero et al [37]).

It is important to clarify that the purpose of this thesis is not to find the correlation between the occurrence of the phenomena of La Niña or El Niño and the hydrology behavior in Colombia - unlike Poveda et al. [35], [34]. **The purpose of this work is to find a transfer model between the Energy Spot Price volatility[1] and Water Inflows volatility, and thereby analyze the response due to the occurrence of phenomena such as those.**

A model for volatility transfer between Water Inflow level and Energy Spot Market Price allows a better estimation of risk from extreme hydrological phenomena, considering that the hydrological behavior has a strong stochastic component.

[1] In this work is very important to distinguish the difference between volatility and variance. Volatility is a statistical measure of dispersion around the average of any random variable such as market parameters etc. whereas, the variance measures how far a set of numbers is spread out.

To propose the statistical model to solve the problem (section 4.1), we are inspired by methods for measuring contagion among financial markets. The model of contagion among financial markets refers to the way that a particular event in a financial market is transferred to other financial market.

In Section 1.1 we make a brief description of some useful technical terms of electricity market, then in Chapter 2 we show the state of the art of main concepts for the development of this work. In Chapter 3 we explain some useful statistical concepts and definitions. In this chapter we explain the definitions and models of time series, volatility, transfer function and other terms that will be useful in Chapter 4.

In Chapter 4 we explain the statistical model that we use to solve the proposed problem and in Section 4.2 we show the results of of applying it. Finally, we give some recommendations and conclusions in Chapter 5. Due to the great importance of the GARCH model for this work, in Appendix A.3 we describe how to model this kind of processes in R software.

1.1. Some Useful Definitions for the Electricity Market

Here are some basic definitions related to the electricity market.

1.1.1. ONI (Ocean Niño Index) and Water Inflows

The term Water Inflows refers to the sum of the amount of electricity that could be generated from water flowing down the rivers to the power plants and dams.

The ONI is an index produced by the NOAA (National Oceanic and Atmospheric Administration) which is a three-month moving average of sea temperature anomalies in the Niño 3.4 region (5N-5S, 120-170W) from ERSST (Extended Reconstructed Sea Surface Temperature).

According to Ramírez and Jaramillo [38], there is a statistically significant linear relationship ($P < 0{,}01$ and $P < 0{,}05$) between ONI and rains in Colombia from December to February and from June to September (months historically considered to have low hydrology). They showed that in the central Andean region of Colombia exists a greater correlation between rainfall and surface temperature of the Pacific Ocean, represented by ONI.

1.1.2. Spot Price of Electricity in Colombia

The Electricity Spot Price (also called Marginal Price) is defined as the sum of the Maximum Bid Price (MPO for its acronym in Spanish) for power plants participating in the spot market, in other words, the Spot Price is the price offered by the last plant to provide energy to the demand every hour.

In the Colombian Electricity Market, there is a minimum price for energy supplied from the agents, that is calculated by XM [2] as the sum of the following terms:

- Real Equivalent Cost of Energy (CERE): this is the payment for Reliability Charge
- Contributions: Law 99 of 1993 (Environmental Law)
- Secondary Service or Frequency Control (AGC)
- Contribution for Non-Interconnected Zones - FAZNI.

According to Franco et al. [10], the efficient operation of the electricity market implies that all generating companies must take the best operational decisions with the best information available; therefore the electricity companies must have a thorough knowledge about the dynamics of the electricity price and the mechanisms that determine their evolution.

[2]Manager Wholesale Energy Market in Colombia

2. State of the Art

2.1. Models for Energy Markets

We briefly review some contributions to the modeling of the electricity price, which we consider are of relevance for the present work. Koopman et at [40], analyze daily data from the Nord Pool electricity market from January 4, 1993 to April 10, 2005, and introduce a seasonal periodic autoregressive fractionally integrated moving average model, of the form

$$\Phi_P(L^s)(1-L^s)^{D_{j(t)}}(\Phi_p(L)P_t+S_t)=\Theta_Q(L^S)\epsilon_t, \quad \text{Where } \epsilon_t = \text{White Noise} \sim (0,\sigma^2)$$

They use a period of $s = 7$ days and argue that the series for each day needs a particular order of seasonal fractional differentiation $D_{j(t)}, j = 1, 2, ..., 7$, each of which is a periodic function of t of period $s = 7$. The autoregressive $\Phi_p(L)$ term is a periodic autoregressive model, i.e. its coefficients are periodic functions of period s as well as the term S_t which contains a linear term of exogenous periodic variables.

In Gil and Maya [13], the authors introduce a model for the volatility of Electricity Prices in the Colombian market. They apply an EGARCH type for the residuals. In contrast, we found that a GARCH model also provides a satisfactory fit to volatilities from a type of SARFIMA model. In Castaño and Sierra [6] the authors provide a deep analysis of the monthly time series of Electricity Prices in the Colombian market. They found that the monthly prices are not integrated I(1) but did not consider the seasonal component neither the fractional integration feature of prices, considered in the present contribution. In section 2.1.1 we will come back to discuss these and other studies to modeling electricity price in Colombia.

2.1.1. Other Models Related to the Colombian Electricity Market

Time series models are widely used to analyze the behavior of the Spot Price of energy and how it responds to exogenous events, such as El Niño and La Niña.

Relative to the Energy Spot Price series, many articles have been written that attempt to describe mathematically this series and to try to understand their behavior over time. Castaño and Sierra [6] presented evidence that the Colombian Spot Price is a stationary process that changes around several levels. According to the statistical tests, they rejected the existence of a unit root in the monthly electricity prices series. However, they suggest that the price evolves as a stationary process around different levels and there is no linear deterministic trend. Then, they proposed the following model.

$$X_t = (\text{Time Series Level Shift})_t + \frac{1}{1-\phi_1 L - \phi_3 L^3} e_t, \quad \text{Where } e_t \sim i.i.d(0,\sigma^2) \tag{2-1}$$

Moreover, Botero and Cano [3] found an increase in volatility due to the latent risk of intervention by Colombian energy regulator for the Spot Price. The authors claim that a special modeling is required for periods of a market intervention by the regulator.

For Energy Spot Price Series they adjusted an Autoregressive Model of order 2, known as $AR(2)$ (see equation (2-2)). However, due to multiple interventions and extreme hydrological phenomena in some periods, the residuals model was modeled in a particular way. That is, they did an analysis for shorter periods of time where there was some kind of anomaly.

$$z_t = \phi_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \epsilon_t, \quad \text{Where } \epsilon_t = \text{White Noise} \sim (0,\sigma^2) \tag{2-2}$$

Other authors, such as Maya and Gil [13] have also developed models of volatility of the Energy Spot Price in Colombia in order to manage the risk. They found that the Spot Price of energy can be modeled as a mean reversion, an autoregressive process of the previous day and with seasonal patterns.

They considered $P_t = f(t) + X_t$ where P_t is the Energy Spot Price, $f(t)$ represents a deterministic function of time and X_t represents a stochastic diffusion process, such that $dX_t = -kX_t dt + \sigma dZ$ where dZ is a Brownian motion and $k \in \mathbb{R}$. Additionally, the deterministic component is represented by $f(t) = \alpha + \beta D_t + \sum_{i=1}^{12} \beta_i M_{it}$ where D is a dummy variable that is associated with a weekend or workday, and M is associated for the month i.

Regarding to the Spot Price volatility, they found that the best model is a $EGARCH(1,1)$.

$$\begin{aligned} \epsilon_t = e_t \sigma_t, &\qquad e_t \sim i.i.d(0,\sigma^2) \\ log(\sigma_t^2) = \omega + \theta_1 log(\sigma_{t-1}^2) &+ \delta_1 \left| \frac{e_{t-1}}{\sigma_{t-1}} \right| \end{aligned} \tag{2-3}$$

Where ω, θ and δ are coefficients, and σ_t^2 is the conditional variance.

Recently, Uribe and Trespalacios [45] contrasted different stochastic models for the Spot Price in Colombia. They proposed a model for the Spot Price P_t that has two components: a deterministic component $f(t)$ and a stochastic component x_t.

$$P_t = f(t) + x_t \tag{2-4}$$

Where, $f(t)$ is a model with linear trend and deterministic seasonality. The stochastic component $x(t)$ of the equation (2-4) is represented by $AR(p)$ or $MA(q)$ or $ARMA(p,q)$ models.

$$f(t) = c + \beta_0 t + \sum_{m=1}^{11} \beta_m D_m + \beta_{\text{niño}} D_{\text{strongniño}} \tag{2-5}$$

where, c is a constant, β_0 represents a price level increase (in kWh), β_m represents the coefficient of each month m, D_m is a categorical variable that takes the value of 1 if the month is m and 0 otherwise, $D_{\text{strongniño}}$ represents the occurrence or nonoccurrence of El Niño, $\beta_{\text{niño}}$ is the expected price level increase.

Uribe and Trespalacios [45] found that the instantaneous variance of monthly average price changes due to the occurrence of El Niño.

An interesting alternative to simulate the electrical market in the short term was proposed by Franco et al. [10]. In their work, they combine a climate model to simulate periods of El Niño and La Niña with a model to simulate the deal price of generators.

2.2. Models of Contagion

The aim of the present work is to apply the idea of financial contagion between stock markets to the electricity market. Our proposal is to show that there is a transmission of volatility between the Water Inflows and the Energy Spot Prices.

The volatility is a non-observable time series and it can be estimated by several methods (see Soofi [41], Chapter 5). The concept of volatility may be interpreted as the "temperature" or "nervousness" measurement of a stock market; then we want to model this level of "nervousness" introduced in the electricity market by variations (or expected change) in Water Inflows, measured by their volatility.

Recent works regarding contagion models have been focused on increasing financial returns correlation between markets. Forbes et al. [9], discuss the current ambiguity and disagreement in the use of the term "contagion". They say that this term can be defined as a significant increase in cross-market linkages after the occurrence of an event.

Pesaran et al, [33] argue that the financial contagion is complex to estimate econometrically. They say that the selection of the crisis period introduces a bias in the selection of the sample, this may cause that the estimate of the correlation is not appropriate.

Currently, there is no consensus in the literature about what is a contagion between markets, however, all definitions converge to make correlation analysis. However, the contagion transfer analysis is useful in improving our understanding of several mechanism that have the potential to destabilize economies and produce financial crises, as Gray [15] page 304, says. This concept can be extrapolated to the energy sector and its basic components.

Similarly, Forbes et al. [9] argue that the strong links between the globalized economies do not necessarily reflect a contagion, but rather to a cooperative and coordinated market movement in all periods and therefore are always “interdependent” and not that spread in certain periods, as Peckham [32] says.

2.2.1. Auto-regressive Distributed Lag (ADL) Models to Measure Contagion

The ADL models are widely used to model the values of a dependent variable and it is based on the current and lagged values of one or several explanatory variables. In Section 3.3 of this document ADL models are formally presented.

Jung et al. [21] investigated about the transmission of volatility between stock markets in Hong Kong, Europe and The United States. They proposed a heterogeneous ADL model in which, when there was a sudden strong increase in volatility, this phenomenon is transferred to other markets.

However, for issues related to energy markets, the application of ADL has been of recent use. This is the case of Zachmann and Hirschhausen [49] who tested the following hypothesis: “an increase in the prices of carbon emissions have a greater impact on wholesale energy prices than the decrease in the marginal cost of generation in the German energy market”. This is a hypothesis which is known for many years in the energy sector, however, only recently a geometric ADL model (See Section 3.3.1) was used to evaluate it .

3. Framework

We provide a brief description of the principal concepts and theories used in the development of this work.

3.1. Some Time Series Models

Models for time series data can have many forms and represent different stochastic processes. Throughout this work, we will refer to ARMA, SARMA, ARIMA, SARIMA, ARFIMA and SARFIMA models. According to Holand and Lund [18], in sections 3.1.1, 3.1.2 and 3.1.3 we will make a formal definition of these concepts. Additionally, in Appendix A.2 we explain the conceptual bases to model time series in R software.

3.1.1. ARMA and SARMA Models

The Autoregressive Moving Average (ARMA) process refers to a model which provide a parsimonious description of a stationary stochastic process in terms of two polynomials: auto-regressive and moving average.

A SARMA model is an ARMA model, but with a seasonal component S. In the definitions 3.1 and 3.2 we explain these concepts.

Definition 3.1. If p, q are non-negative integers, then an $ARMA(p, q)$ process is defined as $(X_n, n \in \mathbb{Z})$ with zero mean, solution of the recursive equation

$$\Phi_p(L)X_n = \Theta_q(L)\epsilon_n \qquad (3\text{-}1)$$

Where $\epsilon_n \sim$ White Noise$(0, \sigma^2)$, $\Phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\Theta_q(L) = 1 - \theta_1 L - ... - \theta_q L^q$. The roots of $\Phi_p(L)$ and $\Theta_q(L)$ lie outside the unit circle in the complex plane and have not common roots.

Definition 3.2. A process $(X_n, n \in \mathbb{Z})$ is $SARMA(p, q)(p_s, q_s)$, if it satisfies the following recursive equation:

$$\Phi_p(L)\Phi_{P_s}(L^s)X_n = \Theta_q(L)\Theta_{Q_s}(L^s)\epsilon_n \qquad (3\text{-}2)$$

Where $\epsilon_n \sim$ White Noise$(0, \sigma^2)$, S is the period of the process X_n (S is a positive integer, for example S= 4, 12, ...), $\Phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$, $\Theta_q(L) = 1 - \theta_1 L - ... - \theta_q L^q$,

$\Phi_{P_s}(L) = 1 - \phi_1 L - ... - \phi_{P_s} L^{P_s}$ and $\Theta_{Q_s}(L) = 1 - \theta_1 L - ... - \theta_{Q_s} L^{Q_s}$. The roots of $\Phi_p(L)$, $\Theta_q(L)$, $\Phi_{P_s}(L)$ and $\Theta_{Q_s}(L)$ lie outside the unit circle in the complex plane and have not common roots between $\Phi_p(L)$ and $\Theta_q(L)$ and between $\Phi_{P_s}(L)$ and $\Theta_{Q_s}(L)$.

3.1.2. ARIMA and SARIMA Models

The Autoregressive Integrated Moving Average (ARIMA) model is a generalization of an ARMA model (see section 3.1.1). The ARIMA models are applied in some cases where data show evidence of non-stationarity or when a differentiation can be applied to remove the homogeneous non-stationarity. A SARIMA model is an ARIMA model, but with a seasonal component.

Definition 3.3. If p, d, q are non-negative integers, then an $ARIMA(p, d, q)$ process is defined as $(X_n, n \in \mathbb{Z})$ with zero mean, solution of the recursive equation

$$\Phi_p(L)(1 - L)^d X_n = \Theta_q(L)\epsilon_n \tag{3-3}$$

Where $\epsilon_n \sim$ White Noise$(0, \sigma^2)$, $\Phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\Theta_q(L) = 1 - \theta_1 L - ... - \theta_q L^q$. d is the number of nonseasonal differences needed for stationarity. The roots of $\Phi_p(L)$ and $\Theta_q(L)$ lie outside the unit circle in the complex plane and have not common roots.

Definition 3.4. If p, d, q, P, D, Q, s are non-negative integers, then a $SARIMA(p, d, q)(P, D, Q)[s]$ process is defined as $(X_n, n \in \mathbb{Z})$ with zero mean, solution of the recursive equation

$$\Phi_{P_s}(L)\Phi_p(L)(1 - L)^d(1 - L^s)^D X_n = \Theta_q(L)\Theta_{Q_s}(L)\epsilon_n \tag{3-4}$$

Where $\epsilon_n \sim$ White Noise$(0, \sigma^2)$, S is the period of the process X_n (S is a positive integer, for example S= 4, 12, ...), $\Phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$, $\Theta_q(L) = 1 - \theta_1 L - ... - \theta_q L^q$, $\Phi_{P_s}(L) = 1 - \phi_1 L - ... - \phi_{P_s} L^{P_s}$ and $\Theta_{Q_s}(L) = 1 - \theta_1 L - ... - \theta_{Q_s} L^{Q_s}$. d and D are the number of nonseasonal differences needed for stationarity in the ordinary and seasonal components, respectively.

The roots of $\Phi_p(L)$, $\Theta_q(L)$, $\Phi_{P_s}(L)$ and $\Theta_{Q_s}(L)$ lie outside the unit circle in the complex plane and have not common roots between $\Phi_p(L)$ and $\Theta_q(L)$ and between $\Phi_{P_s}(L)$ and $\Theta_{Q_s}(L)$.

3.1.3. ARFIMA and SARFIMA Models

These models are a generalization of ARIMA models (see Section 3.1.2) by allowing non-integer values of the differencing parameter. The ARFIMA and SARFIMA models are useful in modeling time series with long memory. That is, the ACF decay more slowly than an exponential decay.

Definition 3.5. If p, q are non-negative integers and $d \in \mathbb{R}$ is the fractional differencing parameter. Then an $ARFIMA(p, d, q)$ process is defined as $(X_n, n \in \mathbb{Z})$ with zero mean, solution of the recursive equation

$$\Phi_p(L)(1 - L)^d X_n = \Theta_q(L)\epsilon_n \tag{3-5}$$

Where $\epsilon_n \sim$ White Noise$(0, \sigma^2)$, $\Phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\Theta_q(L) = 1 - \theta_1 L - ... - \theta_q L^q$. The roots of $\Phi_p(L)$ and $\Theta_q(L)$ lie outside the unit circle in the complex plane and have not common roots. Also, in case of stationary series:

- If $d \in (0{,}0; 0{,}5)$ the process exhibits the property of long memory
- If $d \in (-0{,}5; 0{,}0)$ the process exhibits the property of intermediate memory
- If $d = 0$ the process exhibits the property of short memory
- If $d \geq 0{,}5$ the ARFIMA process is non-stationary although for $d \in [0{,}5; 1{,}0)$ it is level reverting in the sense that there is no long run impact of an innovation on the value of the process, as Lopes et al. [28] says.
- If $d \geq 1$ the level-reversion property no longer holds
- If $d \leq -0{,}5$ the ARFIMA process is non invertible.

Definition 3.6. If p, q, P, D, Q, s are non-negative integers and $d \in \mathbb{R}$ is the degree or parameter of differencing. Then a $SARFIMA(p, d, q)(P, D, Q)[s]$ process is defined as $(X_n, n \in \mathbb{Z})$ with zero mean, solution of the recursive equation

$$\Phi_{P_s}(L)\Phi_p(L)(1 - L)^d(1 - L)^D X_n = \Theta_q(L)\Theta_{Q_s}(L)\epsilon_n \tag{3-6}$$

Where $\epsilon_n \sim$ White Noise$(0, \sigma^2)$, S is the period of the process X_n (S is a positive integer, for example S= 4, 12, ...), $\Phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$, $\Theta_q(L) = 1 - \theta_1 L - ... - \theta_q L^q$, $\Phi_{P_s}(L) = 1 - \phi_1 L - ... - \phi_{P_s} L^{P_s}$ and $\Theta_{Q_s}(L) = 1 - \theta_1 L - ... - \theta_{Q_s} L^{Q_s}$.The roots of $\Phi_p(L)$, $\Theta_q(L)$, $\Phi_{P_s}(L)$ and $\Theta_{Q_s}(L)$ lie outside the unit circle in the complex plane and have not common roots between $\Phi_p(L)$ and $\Theta_q(L)$ and between $\Phi_{P_s}(L)$ and $\Theta_{Q_s}(L)$.

The properties of d remain similar to those described in the definition 3.5.

3.2. Some Volatility Models

The volatility concept underlies the financial mathematics field and it refers to the measurement of changes in frequency and intensity of series. It is used as a measure of "nervousness"that affects an asset because of particular events. It is important to clarify that the volatility does not measure changes in the time series, it measures the degree of dispersion of series. In other words, volatility is equivalent to the standard deviation of the returns of an asset (daily, monthly, yearly).

The volatility σ_T for a time horizon T, is expressed as $\sigma_T = \sigma_x \sqrt{T}$ Where σ_x is volatility in the horizon of a unit.

3.2.1. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Models

Based on Francq and Zakoian [11] Chapter 2, the GARCH models assume the volatility of the current error term to be a function of the actual sizes of the previous time periods' error term. Let ϵ_t denote the error terms (return residuals, with respect to a mean process). These ϵ_t are split into a stochastic piece e_t and a time-dependent standard deviation σ_t, so that

$$\epsilon_t = \sigma_t e_t, \quad e_t \sim \text{i.i.d}(0, 1) \tag{3-7}$$

Where e_t and σ_t are two independent stationary processes, this ensures that the ϵ_t series has zero marginal mean because $E[\epsilon_t] = E[\sigma_t]E[e_t] = 0$. Similarly, it is shown that the conditional mean is zero because $E[\epsilon_t|\epsilon_{t-1}] = E[\sigma_t|\epsilon_{t-1}]E[e_t|\epsilon_{t-1}] = 0$.

Regarding the variance, as we assumed that ϵ_t is a stationary process, then, the marginal variance (σ^2) is constant. Therefore, $E[\epsilon_t^2] = E[\sigma_t^2]E[e_t^2] = E[\sigma_t^2] = \sigma^2$. However, the process ϵ_t has a not constant conditional variance: $Var[\epsilon_t^2|\epsilon_{t-1}] = E[\sigma_t^2|\epsilon_{t-1}]E[e_t^2] = \sigma_t^2$. Therefore, σ_t has a dynamical structure, with the value t according to previous values of t. That is, $\sigma_t = f(\sigma_{t-1}, \sigma_{t-2}, ...)$.

Making a correlation analysis, it is important to remember the independence of e_t and σ_t and therefore ϵ_t lacks of autocorrelation. Regarding the autocovariance it is noted that these are zero. In fact, $E[\epsilon_t \epsilon_{t-k}] = E[\sigma_t e_t \sigma_{t-k}] = E[e_t]E[\sigma_t \sigma_{t-k}] = 0$.

Definition 3.7. (See Francq and Zakoian [11] Chapter 2) We say that a process $(\epsilon_n, n \in \mathbb{Z})$ is $GARCH(p, q)$ if its two conditional moments exist and satisfy:

1. $E(\epsilon_t|\epsilon_u, u < t) = 0, t \in \mathbb{Z}$

2. There exist constants $\omega, \alpha_i, i = 1, 2, ..., q$ and $\beta_j, j = 1, 2, ..., p$ such that

$$\sigma_t^2 = Var(\epsilon_t|\epsilon_u, u < t) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z} \tag{3-8}$$

proposition 3.1. *According to Lindner [26], A necessary and sufficient condition for a process $\epsilon_n \sim GARCH(p,q)$ defined by the equations (3-8) be strictly stationary with $\mathbb{E}(\epsilon_n^2) < \infty$ is*

$$\sum_{j=1}^{p} \beta_i + \sum_{i=1}^{q} \alpha_i < 1 \tag{3-9}$$

If additionally it holds that $Max(1, \sqrt{\mathbb{E}(\epsilon_n^4)}) \sum_{i=1}^{q} \frac{\alpha_i}{(1-\sum_{j=1}^{p}\beta_j)} < 1$ then $\mathbb{E}(\epsilon_n^4) < \infty$

In (3-8), if $p = 0$ then it is known as $ARCH(q)$ model. According to Bera and Higgins [2], when a high-order $ARCH(q)$ model is obtained, it is not easy to work with. Therefore, you should try to fit a $GARCH(p,q)$ model.

FIGARCH Model

The acronym $FIGARCH(p,d,q)$ refers to a model Fractionally Integrated Generalized Autoregressive Conditional Heteroscedastic. This kind of model is mainly used for describing the observed persistence and long memory in the volatility of a time series, as Tayefi and Ramanathan, says [42].

Definition 3.8. We say that a process $(\epsilon_n^2, n \in \mathbb{Z})$ is FIGARCH if ϵ_n^2 can be defined as $ARMA(\text{máx}(p,q),q)$ process with white noise $\eta_t = \epsilon_n^2 - \sigma_t^2$,

If $\epsilon_n^2 = \sigma_n e_n$, Such that:

$$\left(1 - \sum_{j=1}^{q} \alpha_j L^j - \sum_{i=1}^{p} \beta_i L^i\right)(1-L)^d \epsilon_n^2 = \beta_0 + \left(1 - \sum_{j=1}^{q} \alpha_j L^j\right)\eta_t. \tag{3-10}$$

Where $e_t \sim$ i.i.d, $\beta_0 > 0, \beta_i \geq 0, 1 \leq i \leq p$ y $\alpha_j \geq 0, 1 \leq j \leq q$ and d is a fraction $0 < d < 1$.

For a FIGARCH model to be covariance stationary and invertible the parameter d must be defined as $0 < d < 0{,}5$. In the FIGARCH model, d must not be less than zero because of the non-negativity conditions imposed on the conditional variance equation.

3.3. Review of Some Infinite Distributed Lags Models

A Distributed Lag Model is a time series model in which a regression equation is used to predict current values of a dependent variable. In this section we follow closely Judge et al [20], Chapter 10.

This kind of models are based on the current and lagged values of an explanatory variable, x_t. Thus, $y_t = (\nu_0 + \nu_1 L + \nu_2 L^2 + ...)x_t + \epsilon_t = \nu(L)x_t + \epsilon_t$, where y_t is the dependent variable,

x_t is the explanatory variable and ϵ_t is a zero mean stationary stochastic process. As this model have an infinite number of parameters, then we will assume a more parsimonious model:

$$y_t = \frac{\gamma(L)}{\phi(L)}x_t + \frac{\alpha(L)}{\theta(L)}\epsilon_t \quad , \quad \epsilon_t = \text{white noise} \sim N(0, \sigma^2_{\epsilon_t}) \tag{3-11}$$

Where $\gamma(L) = 1 - \gamma_1 L - \gamma_2 L^2 - ... - \gamma^{s'}_{s'L}$, $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \gamma_{r'} L^{r'}$, $\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - ... - \alpha_{q'} L^{q'}$, $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_{p'} L^{p'}$ and r', s', p', q' are non-negative integers. The roots of $\gamma(L)$, $\phi(L)$, $\alpha(L)$ and $\theta(L)$ lie outside the unit circle in the complex plane and have not common roots between $\gamma(L)$ and $\phi(L)$ and between $\alpha(L)$ and $\theta(L)$.

According to Judge et al [20], (page 378, Table 10.1 "Infinite Distributed Lag Models"), some types of Autoregressive Distributed Lag (ADL) are:

- Rational Lag whith ARMA Errors
- Gamma Lag
- Exponential Lag

In sections 3.3.1, 3.3.2 and 3.3.3, we explain each of these ADL types.

3.3.1. Rational Lag With ARMA errors

According to Judge et al. this kind of models can be specified using the transfer functions methodology of Box and Jenkins [5].

Due to the historical and practical significance of Box and Jenkins identification methodology, we will make a full statement of the method for fitting the parameters of the model (3-11). There exists other methods to find the parameters such as The Approximation of Lag Structure or The Common Factor Analysis, however, in this section we will discuss only the methodology of Box and Jenkins.

Let, x_t and y_t be stationary series. Additionally, consider that x_t and y_t are related by a linear filter $v(L)$, so that, as explained Wei [46], page 322, the equation (3-12) is called *the transfer function model*.

$$\begin{aligned} y_t &= v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + ... + n_t \\ &= v(L) x_t + n_t \end{aligned} \tag{3-12}$$

where, n_t is the noise of the system, which it is independent of the input time series x_t. $v(L) = \sum_{j=0}^{\infty} v_j L^j$ is the transfer function filter of (3-12). The coefficients (v_j) are called the impulse response function. When x_t is assumed as an ARMA process then the equation

(3-12) is knows as the ARMAX model.

In relation to v_j:

- If $\sum |v_j| < \infty$ then the transfer function model is stable, that is to say, a bounded input always produces a bounded output
- If $v_j = 0$ for $j < 0$ the transfer function model is causal. Thus, the system does not respond to input series until they have been applied to the system. That is to say, the present output is affected only by the current and past input values x_t. *From now on, we will assume causal models.*

Additionally, for the equation (3-12) we assume that x_t follows an ARMA process $\phi_x(L)x_t = \theta_x(L)\alpha_t$, then $\alpha_t = \frac{\phi_x(L)}{\theta_x(L)}x_t$ where α_t is white noise, also called prewhitened input series. Moreover, y_t is defined by the expression $\theta_x(L)\beta_t = \phi_x(L)y_t$, then $\beta_t = \frac{\phi_x(L)}{\theta_x(L)}y_t$.

The purposes of Box and Jenkins methodology is to identify and estimate the impulse response function $v(L)$ and the noise n_t based in the input series x_t and the output series y_t.

It is common to represent the impulse response function $v(L)$ as a rational function

$$v(L) = \sum_0^\infty v_j L^j = \frac{\omega(L)L^b}{\delta(L)} \tag{3-13}$$

Where b is a delay parameter representing the actual time lag that elapses before the impulse of the input variable produces an effect on the output variable, $\omega(L) = \omega_0 - \omega_1 L - ... - \omega_s L^s$ and $\delta(L) = 1 - \delta_1 L - ... - \delta_r L^r$. For a stable system the roots of $\omega(L)$ and $\delta(L)$ lie outside the unit circle in the complex plane and have not common roots as indicated by Wei [46], page 325.

Rewriting the equation (3-13) and expanding it, we get the equation (3-14)

$$[1 - \delta_1 L - ... - \delta_r L^r][v_0 + v_1 + v_2 L^2 + ...] = [\omega_0 - \omega_1 L - ... - \omega_s L^s]L^b \tag{3-14}$$

Thus, we have that,

- b is determined by $v_j = 0$ for $j < b$ and $v_b \neq 0$
- r is determined by the pattern of the impulse response weights through pattern of the autocorrelation function
- After found b and r values, we can found s easily using that $v_j = 0$ for $j > b + s$ if $r \neq 0$. The value of s is found by checking where the pattern of decay for impulse response weight begins.

According to Wei [46], pages 324 and 325, some transfer functions may exhibit the following behavior:

1. The transfer function contains a finite number of impulse response weights. For this kind of transfer function $r = 0$ and the impulse respond weights start in $v_b = \omega_0$ and finish at $v_{b+s} = -\omega_s$. *(See: typical impulse weights of (b,0,0)(b,0,1)(b,0,2) in table **3-1**)*

2. The impulse response weights decay exponentially from the value of s. *(See: typical impulse weights of (b,1,0)(b,1,1)(b,1,2) in table **3-1**)*

3. The impulse response weights shows an exponential decay or damped sinusoid wave. The shape depends on the roots of the polynomial $\delta(L) = 1 - \delta_1 L - \delta_2 L^2 - ... = 0$. If the roots are real, then the impulse response weights follows an exponential decay. In contrast if the roots are complex the impulse response weights follows a damped sine wave. *(See: typical impulse weights of (b,2,0)(b,2,1)(b,2,2) in table **3-1**)*

The Cross Correlation Function (CCF)

To understand the definition of the CCF it is necessary to remember the concepts related to autocovariance and cross-autocorrelation. Here we follow closely Wei [46], Chapter 14.

The cross-covariance between two processes $(X_t, Y_t, t \in \mathbb{Z})$, is defined as the function

$$\gamma_{x,y}(k) = Cov(X_t, Y_{t+k}) = \mathbb{E}((X_t - \mu_x)(Y_{t+k} - \mu_y)) \text{ for } k = 0, \pm 1, \pm 2, ... \tag{3-15}$$

such that,

$$\gamma_{x,y}(k) = \gamma_{y,x}(-k). \tag{3-16}$$

An estimator for $\gamma_{x,y}(k)$ is defined as

$$\widehat{\gamma}_{x,y}(k) = \begin{cases} \dfrac{1}{T}\displaystyle\sum_{t=1}^{T-k}(X_t - \bar{X})(Y_{t+k} - \bar{Y}), & k = 0, 1, 2, \dots \\ \dfrac{1}{T}\displaystyle\sum_{t=1}^{T+k}(X_t - \bar{X})(Y_{t-k} - \bar{Y}), & k = 0, -1, -2, \dots \end{cases} \tag{3-17}$$

The CCF measures the strenght and direction of correlation between two stochastic processes and it is defined as

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \tag{3-18}$$

If the input series is white noise and $\rho_x(k) = 0$ for $k \neq 0$, it can be shown that the equation (3-18) can be written as

$$v_k = \frac{\sigma_x}{\sigma_y}\rho_{xy} \tag{3-19}$$

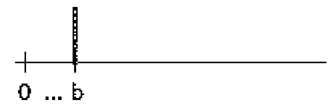
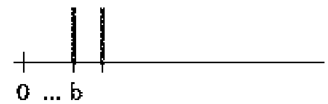
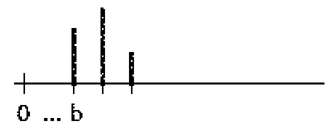
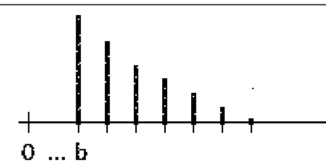
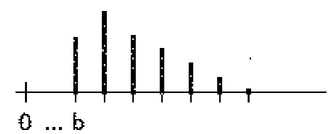
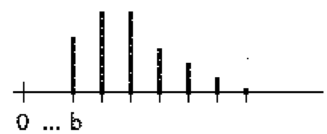
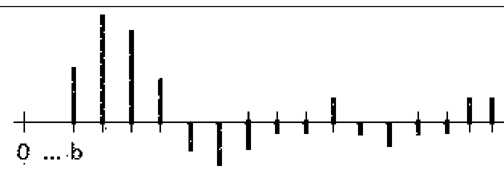
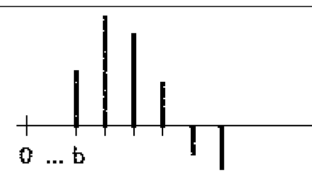
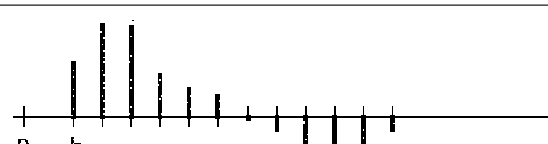
(b,r,s)	Transfer function	Typical impulse weights
(b,0,0)	$v(L)x_t = \omega_0 x_{t-b}$	
(b,0,1)	$v(L)x_t = (\omega_0 - \omega_1 L)x_{t-b}$	
(b,0,2)	$v(L)x_t = (\omega_0 - \omega_1 L - \omega_2 L^2)x_{t-b}$	
(b,1,0)	$v(L)x_t = \frac{\omega_0}{1-\delta_1 L} x_{t-b}$	
(b,1,1)	$v(L)x_t = \frac{\omega_0-\omega_1 L}{1-\delta_1 L} x_{t-b}$	
(b,1,2)	$v(L)x_t = \frac{\omega_0-\omega_1 L-\omega_2 L^2}{1-\delta_1 L} x_{t-b}$	
(b,2,0)	$v(L)x_t = \frac{\omega_0}{1-\delta_1 L-\delta_2 L^2} x_{t-b}$	
(b,2,1)	$v(L)x_t = \frac{\omega_0-\omega_1 L}{1-\delta_1 L-\delta_2 L^2} x_{t-b}$	
(b,2,2)	$v(L)x_t = \frac{\omega_0-\omega_1 L-\omega_2 L^2}{1-\delta_1 L-\delta_2 L^2} x_{t-b}$	

Table 3-1.: Typical impulse weights according to the transfer function type.

Thus, the impulse response function v_k is directly proportional to the cross-correlation function $\rho_{xy}(k)$. It is important to say that the CCF is defined only when x_t and y_t are jointly stationary bivariate process.

It can be shown that, from the definition of α_t and β_t, the equation (3-19) can be written as

$$v_k = \frac{\sigma_\beta}{\sigma_\alpha}\rho_{\alpha\beta}(k) \tag{3-20}$$

Construction of Transfer Function Models

To build a transfer function model it is necessary to define the cross-correlation function, according to the equation (3-18). To test if certain values of the $\rho_{xy}(k)$ are zero, we compare the CCF with their standard error.

In this work, it is used the prewhitening methodology proposed by Box and Jenkins [4], Chapters 10 and 11, and which is clearly explained by Wei, [46], Chapter 14, and by Hipel and McLeod [17], Chapter 16. To identify the Transfer Function Model the first step is to prewhite the input series, that is to say, if $\phi_x(L)x_t = \theta_x(L)\alpha_t$ then, $\alpha_t = \frac{\phi_x(L)}{\theta_x(L)}x_t$ where α_t is a white noise series with zero mean and variance σ_α^2.

After prewhittening the series, it is calculated the filtered ouput series (β_t) to transformate the ouput series y_t, such that, $\beta_t = \frac{\phi_x(L)}{\theta_x(L)}y_t$.

Then, it is calculated the prewhiten input series and the filtered output series, then we proceed to calculate the sample CCF ($\hat{\rho}_{\alpha\beta}(k)$) between α_t and β_t to estimate ν_k.

$$\hat{\nu}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}\hat{\rho}_{\alpha\beta}(k)$$

To identify the parameters b, r and s of the equation (3-21), we should compare whith the known theoretical patterns of ν_k, which are shown in table **3-1**

$$\hat{\nu}(L) = \frac{\hat{\omega}(L)}{\hat{\delta}(L)}L^b = \frac{\omega_0 - \omega_1 L - ... - \omega_s L^s}{1 - \delta_1 L - ... - \delta_r L^r}L^b \tag{3-21}$$

Once we obtain the preliminary transfer function, we can calculate the estimated noise series:

$$\hat{n}_t = y_t - \frac{\hat{\omega}(L)}{\hat{\delta}(L)}L^b x_t \tag{3-22}$$

Combining the equations (3-21) and (3-22), we obtain the equation

$$y_t = \frac{\omega(L)}{\delta(L)}x_{t-b} + n_t, \quad n_t = \frac{\theta(L)}{\phi(L)}\epsilon_t, \quad \epsilon_t \sim \text{ White Noise}(0, \sigma_{a_t}^2) \tag{3-23}$$

Note that the equations (3-23) contains a finite number of parameters. In Appendix A.1 we explain how to implement in R Software the Box and Jenkins methodology.

Geometric Lag Model

A special case of Rational Lag Model with ARMA errors, is the Geometric Lag Model, also known as ADL Koyck model.

Definition 3.9. The ADL-Koyck is defined as a transfer model between two stationary process in covariance. $(X_n), (Y_n), n \in \mathbb{Z}$ is given by

$$Y_n = \mu + \alpha \sum_{j=0}^{\infty} \lambda^j X_{n-j} + Z_n, \quad \text{where } Z_n \sim ARMA(p,q) \tag{3-24}$$

Let, (3-25) a model that describes y_t in terms of x_t^* which is defined as an unobservable expected value.

$$y_t = \alpha_0 + \alpha_1 x_t^* + e_t \tag{3-25}$$

If the expectation of x_t^* is revised in proportion to the error associated with the previous level of expectation, it can be demostrated that

$$x_t^* = \sum_{i=0}^{\infty} \lambda(1-\lambda)^i x_{t-1-i} \tag{3-26}$$

where $0 < \lambda < 1$ is the coefficient of expectation and x_{t-1} is the actual observable value. Note that x_t^* describes a decay geometrically lag form as function of all past values. If we substitute (3-26) in (3-25), then:

$$y_t = \alpha_0 + \alpha_1 \sum_{i=0}^{\infty} \lambda(1-\lambda)^i x_{t-1-i} + \epsilon_t \tag{3-27}$$

Koyck [25] shows that (3-27) can be rewritten as

$$\begin{aligned} y_t =& \alpha_0 + \alpha_1 \sum_{i=0}^{\infty} \lambda^i x_{t-i} + \epsilon_t \\ =& \alpha_0 + \frac{\alpha_1}{1-\lambda L} x_t + \epsilon_t \\ =& \alpha_0(1-\lambda) + \alpha_1 x_t + \lambda y_{t-1} + \epsilon_t - \lambda \epsilon_{t-1} \end{aligned} \tag{3-28}$$

3.3.2. Gamma Lag

The Gamma Distributed Lag is a model porposed by Tsurumi [44]. It is described by the equation (3-29).

$$y_t = \alpha \sum_{i=1}^{\infty} i^{s-1} e^{-i} x_{t-i} + \epsilon_t, \quad \epsilon_t \sim \text{White Noise}(0, \sigma^2) \tag{3-29}$$

If we truncate the sum of 3-29 at m, such that m is sufficiently large, then it is true that:

$$y_t = \alpha \sum_{i=1}^{\infty} i^{s-1} e^{-i} x_{t-i} + \epsilon_t = \alpha \sum_{i=1}^{m} i^{s-1} e^{-i} x_{t-i} + \alpha\eta_t + \epsilon_t \tag{3-30}$$

where $\eta_t = \sum_{i=m+1}^{\infty} i^{s-1} e^{-i} x_{t-i}$. From equation (3-30), Tsurumi [44] showed that if m is a very large number, then η_t behaves as a constant and the error is very small. Additionally, he suggests to choose a value of m that minimizes the estimated residual variance. Therefore, with m large, the term $\alpha\eta_t$ can be omitted.

Schmidt [39], suggests to replace i^{s-1} of the equation (3-30) by $(i+1)^{s-1}$. Additionally, to increase the flexibility, he proposed that $s-1 = \delta/(1-\delta)$; $\quad 0 \leq \delta < 1$ and a γ parameter. Thus, the equation (3-30) becomes:

$$y_t = \alpha \sum_{i=1}^{m} (i+1)^{\frac{\delta}{1-\delta}} e^{-\gamma i} x_{t-i} + \epsilon_t \tag{3-31}$$

If $e^{-\gamma i} = \lambda^i, \quad 0 \leq \lambda < 1$, then (3-31) becomes in (3-32).

$$y_t = \alpha \sum_{i=1}^{m} (i+1)^{\frac{\delta}{1-\delta}} \lambda^i x_{t-i} + \epsilon_t, \quad 0 \leq \lambda < 1, \quad 0 \leq \delta < 1 \tag{3-32}$$

According to Theil et al. [43], the gamma distribution has the maximum entropy and the Autoregressive Distributed Lag (ADL)-Gamma should be used for situations that have a tendency to show that behavior.

3.3.3. Exponential Lag

Lütkepohl [29] proposed a lag model which avoids the sign changes in the weights of the lags.

According to the equation (3-33), the Exponential Lags family depends of the constant α and the parameter p_k.

$$y_t = \alpha \sum_{i=0}^{\infty} e^{\sum_{k=1}^{m} p_k} x_{t-i} + \epsilon_t, \quad t = 1, 2, 3, ..., T, \quad \epsilon_t \sim \text{White Noise}(0, \sigma^2) \tag{3-33}$$

In order to find the maximum likelihood estimation for (3-33), the sum of the squared error is minimized.

$$S = (\boldsymbol{y(T_0)} - \boldsymbol{f(T_0)})'(\boldsymbol{y(T_0)} - \boldsymbol{f(T_0)}) \tag{3-34}$$

Where $\boldsymbol{y(T_0)} = (y_{T_0}, y_{T_1}, ..., y_T)'$ and

$$f(T_0) = \left[\alpha \sum_{i=0}^{T_0-1} e^{\sum_{k=1}^{m} p_k} x_{T_0-i}, ..., \alpha \sum_{i=0}^{T-1} e^{\sum_{k=1}^{m} P_m} x_{T-i} \right]'$$

Where T_0 is some positive number. It is clear that T_0 has no impact on the asymptotic properties of the estimators. The problem with the ADL-exponential is selecting the polynomial order of m.

According to Judge [20], for a stationary process the x_t values could simply be replaced by the mean of the observed value of x_t.

4. Statistical Model And Results

In this chapter, we present the statistical model to solve the problem proposed in Chapter 1. Each result obtained is adequately referenced to the respective R Software code in Appendix B.

4.1. Statistical Model

Here, we describe the statistical model used to solve the problem proposed in chapter 1. We assume two time series: Water Inflows (A_t) and Energy Spot Price (P_t). Both series are covariance stationary (as we shown below) but possibly fractionally integrated. The proposed model is based on the log-square-volatility of these series, denoted by $log(\sigma^2_{A,t})$ and $log(\sigma^2_{P,t})$ respectively.

The volatilities are defined as conditional standard deviations of a covariance stationary processes, and assumed a non-observable covariance stationary processes themselves. For the case of a generic time series Y_t, if $\mathscr{F}_{t-1}$ denotes the information up to time $t-1$, the volatility is defined as

$$\sigma_{Y,t} = \sqrt{Var(Y_t|\mathscr{F}_{t-1})}, \quad t = 1, 2, ...$$

There exist several procedures for extracting the non-observable volatility processes associated with a time series. The one we apply here is based on GARCH-type models. We first fit a general SARFIMA model to both series. Then the residuals of both models are fitted to a general FIGARCH model, and then the respective volatilities are computed. Finally, a transfer function model or Box-Jenkins model or a Distributed Lag Model as is known in the econometric literature, is fitted to the volatilities. As can be observed, the proposed model is a three-step one. We now proceed to give a more formal description of each of the steps:

- **First step:** fit a SARFIMA processes to the time series A_t and P_t. We say that A_t and P_t follow a $SARFIMA(p, d, q)(P, D, Q)[s]$ model with seasonal component of period s, if it is the solution of the equation (3-6) with $X_n = A_t$ or P_t.

  In case of $d = 0$, $D = 0$, the model (3-6) is a $SARMA(p, q)(P, Q)[s]$ model. We also admit the case $d = D = 1$ which correspond to a $SARIMA(p, 1, q)(P, 1, Q)[s]$ model

(See definitions 3.2 and 3.4).

The process ϵ_n in (3-6) is assumed to be a White Noise $(0, \sigma^2)$ uncorrelated sequence of constant variance, which includes the case of GARCH and FIGARCH types of White Noise. A process ϵ_t is a $GARCH(p, q)$ process if it satisfies the definition 3.7 and a FIGARCH(p,d,q) process if it satisfies the definition 3.8.

- **Second step:** fits models GARCH or FIGARCH (defined in the first step) to the residuals ϵ_t from model for series A_t and P_t. We apply an adequate test for discriminating between GARCH and FIGARCH process. From these fits we obtain the corresponding estimators for $\sigma^2_{A,t}$ and $\sigma^2_{P,t}$. It is worth mention that these two processes are covariance stationary by construction.

- **Third step:** fits a transfer function model to these volatilities processes (defined in the second step). A transfer function model is a statistical model describing the relationship between an output time series, y_t and one or more input time series. In case of one input x_t the model is defined as a linear causal filter plus an error, of the form described by equation (3-12) in which the process ν_t is assumed to be independent from x_t and it follows an ARMA model $\nu_t = \frac{\alpha(L)}{\theta(L)} e_t$ with $e_t \sim i.i.d(0, \sigma^2_e)$ and $\alpha(z), \theta(z)$ are polynomials such that their roots all lie outside the unit circle in the complex plane and have not common roots. Transfer function models are seldom applied in form (3-12) because of the infinite number of parameters. Instead, a more parsimonious form is adopted, of the form (3-11).

  The more concrete model we apply so as to establish a relationship between log-square volatility: $log(\sigma^2_{A,t})$ and $log(\sigma^2_{P,t})$ is

$$log(\sigma^2_{P,t}) = \mu + \frac{\theta_{q'}(L)}{\phi_{p'}(L)} log(\sigma^2_{A,t-b}) + \nu_t \tag{4-1}$$

  Where, $\nu_t = \frac{\theta(L)}{\phi(L)} \epsilon_t$, $\epsilon_t \sim$ White Noise$(0, \sigma^2_{a_t})$

4.2. Model-Building Procedure

We now proceed estimate the parameters of model (4-1). We divided the whole procedure into a three step one, as we described in Section 4.1. Let us remind that the data consist of the daily time series A_t of Water Inflows in GWh and P_t the daily Energy Spot Price in $\$COP/kWh$, as quoted from the Colombian electricity market, between July 1, 1998 and May 31, 2013. In Figure **4-1**, we show the graphs of A_t and P_t series. For the development of this section we use the R software code shown in Appendix B.1.

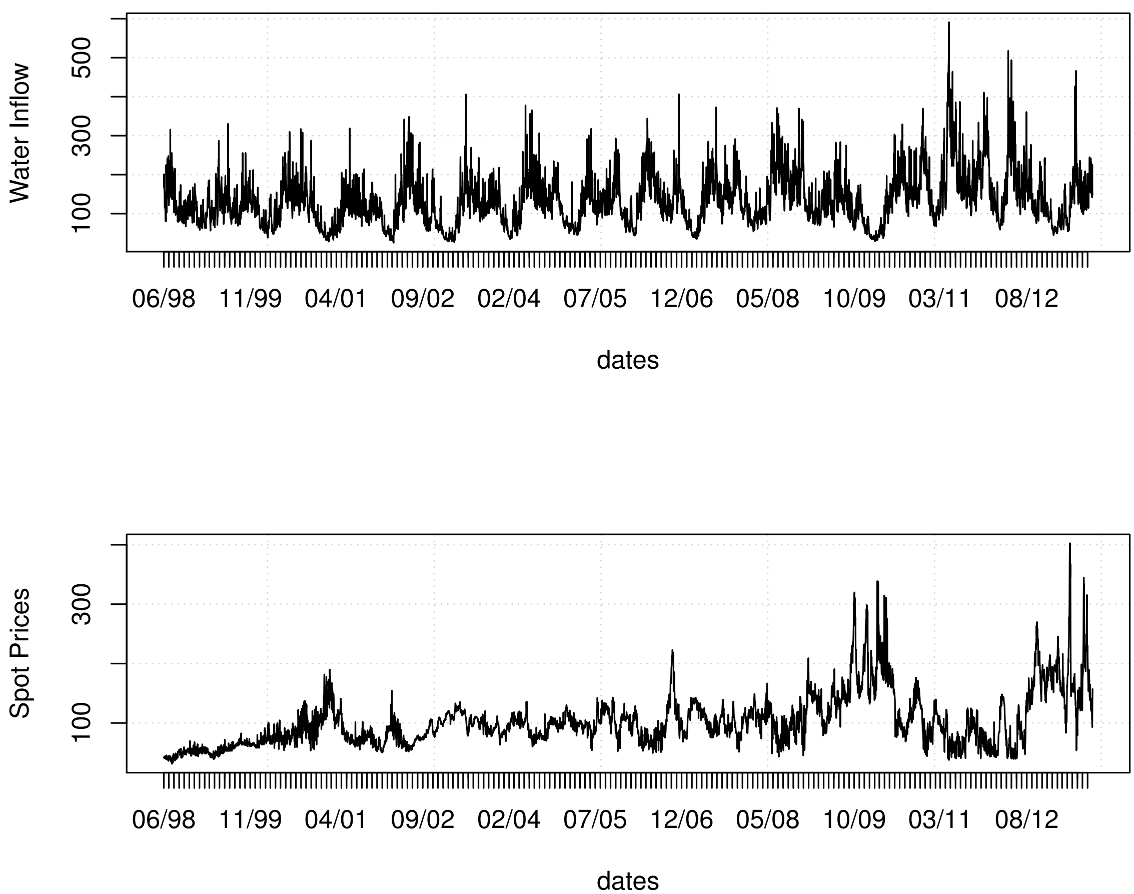


Figure 4-1.: Water Inflows (above) and Energy Price (below)

4.2.1. Step One: Fitting SARFIMA Models for the Series A_t and P_t

Here, we do a preliminary analysis of the Water Inflows (A_t) and the Energy Spot Prices (P_t) time series.

Analysis of the Water Inflows Series

From Figure **4-2**(a), time series A_t displays a strong annual seasonality component. The estimated periodogram appears in Figure **4-2**(c),(d). The dominant peaks are located at periods 365 days and 180 days, observed with more detail in Figure **4-2**(d).

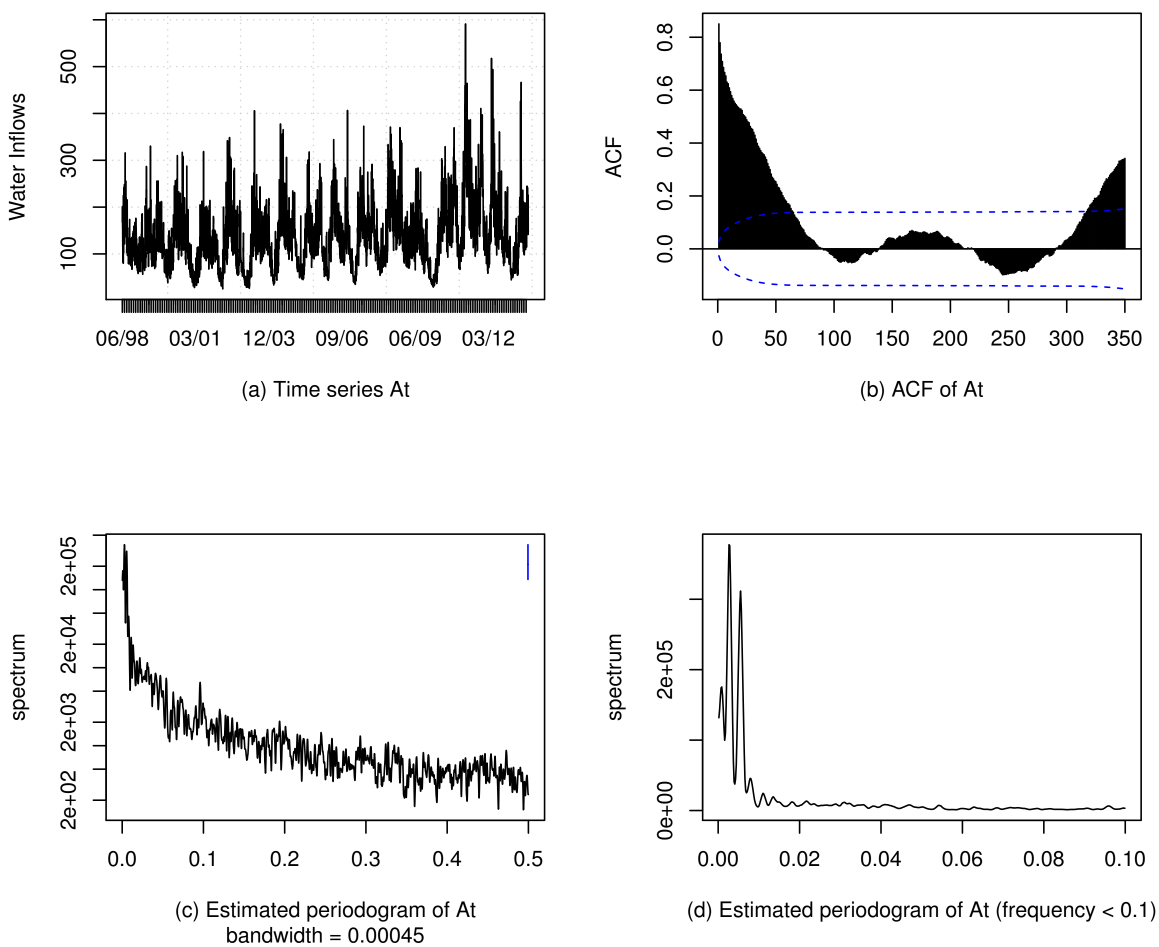


Figure 4-2.: Plot, ACF and Periodogram of A_t

Then, it can be assumed a seasonality with period s = 365 days. The estimated autocorrelation function (ACF) appears in Figure **4-2**(b) and confirms this fact because it displays a periodic-like behavior. For now, we exclude the possibility of long memory because the absences of a peak at frequency zero in the periodogram; however later we will make some formal test regarding the long memory of the series. We did not apply a log transformation because the series displays a stable variance.

In table **4-1** different tests are done to characterize the series[1] where from six unit-root test applied only the KPSS test detected a unit root.

Test	Null Hipothesis Reject if Test value < Critical value	Test value	Critical values	
			1 %	5 %
Augmented Dickey-Fuller	unit root	-12.50	-3.43	-2.86
DF-GLS	unit root	-9.40	-2.57	-1.94
KPSS	stationary	3.08	0.73	0.46
Phillips-Perron	unit root	-23.94	-3.96	-3.41
Schmidt-Phillips	unit root and a linear trend	-32.28	-3.56	-3.02
Zivot-Andrews	unit root with a structural break	-15.03	-5.34	-4.8

Table 4-1.: Unit root test for the Water Inflow time series

The test of long memory Rescaled Range R/S of Lo [27, 16] provides strong evidence against the null hypothesis ($d = 0$), as we show in Table **4-2**. Also, in Table **4-3** the Rescaled Range R/S modified Test (see Giraitis et.al [14, 16]) shows strong evidence of long memory.

Ho: d = 0		
data: A_t	R/S Statistic = 4.8814	Bandwidth q = 4
significance level:	0.05	0.1
critical value:	1.747	1.62

Table 4-2.: Long Memory Test: Rescaled Range R/S of Lo forA_t

The usual test for seasonal unit roots are Canova-Hansen and Hyllebert-Engle-Granger-Yoo, but these are available in the R software only for periods s = 4, 12, not for s = 365 days. Then, we will not apply a formal test for the presence of a seasonal unit root, but we will include a SARIMA model as an admissible model. However, it is known from experience that the hydrological seasonality is 365 days because it is related to terrestrial cycles around

[1] All test used p=3 lags for the approximating AR(p) process and with the option of non-cero mean, from the R package fUnitRoots of Wuertz et al.[47]

Ho: d = 0			
data: A_t	V/S Statistic = 2.4484	Bandwidth q = 4	
significance level:	0.01	0.05	0.1
critical value:	0.2685	0.1869	0.1518

Table 4-3.: Modified Long Memory Test: Rescaled Range R/S Modified of Giraitis et.al [14] forA_t

the sun.

Through Akaike Information Criterion (AIC), the SARIMA model chosen was an $ARIMA(7,1,2)$ (season=0) - See equation (4-2). We argue that the annual seasonal component can in fact be assimilated to a stochastic trend giving rise to an ordinary (not seasonal) unit root. When we applied ordinary (i.e. daily) differentiation a stationary ARMA(7,2) results, although the results in Table **4-1**, where from six unit-root test applied only the KPSS test detected a unit root.

$$\Phi_7(L)(1-L)A_t = \Theta_2(L)\epsilon_t \tag{4-2}$$

The orders p=7, q=2 were established using a minimum AIC criteria, implemented in the R function auto.arima, from the forecast package [19]. This model was fitted using the R `arima` function. In Table **4-4** appears the parameters for the model. Additionally, in Table **4-5** appears the Ljung-Box test, not rejecting the null for uncorrelated residuals.

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	
ar1	-0.2005	0.1138	-1.7616	0.0781	.
ar2	0.4523	0.0610	7.4145	1.221e-13	***
ar3	0.0634	0.0180	3.5068	0.0004	***
ar4	0.0402	0.0167	2.4030	0.0162	*
ar5	0.0295	0.0164	1.8004	0.0717	.
ar6	0.0038	0.0161	0.2375	0.8122	
ar7	0.0425	0.0156	2.7193	0.0065	**
ma1	-0.1796	0.1133	-1.5853	0.1129	
ma2	0.6798	-0.1007	-6.7503	1.475e-11	***
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 4-4.: Parameters of the $A_t \sim ARIMA(7,1,2)$ model

In Table **4-6** we show the respective residuals which have high kurtosis and asymmetry.

Null hypothesis: uncorrelated residuals		
X-squared	df	p-value
73.1414	60	0.1187

Table 4-5.: Ljung-Box Test for the residuals of $A_t \sim ARIMA(7,1,2)$

Residuals of $A \sim ARIMA(7,1,2)$	
Size	5571
Mean	0.0034
St.Dev	32.6375
Skewness	1.4025
Kurtosis	5.7982
Minimum	-130.4896
Maximum	275.0039

Table 4-6.: Residues of $A \sim ARIMA(7,1,2)$

Figure **4-3** displays values $\hat{A}_t$ calculated as simulated values of an $ARIMA(7,1,2)$ with input values the estimated residuals $\hat{\epsilon}_t$ from the $ARIMA(7,1,2)$, versus the observed ones A_t. It is observed that the model fits very well.

Analysis of the Spot Price Series

From Figure **4-1**, it is suspected that the variance P_t does not have stable behavior over time. It is known that, the logarithmic transformation stabilizes the variance of the series without changing its autocorrelation structure and its correlation with other variables. Then we proceeded with the log of the spot prices $log(P_t)$.

The periodogram and the autocorrelation function (ACF) of the log-prices are shown in Figures **4-4**(a) and (b), respectively. It can be observed that the ACF is rather persistent and shows a slow decay. The periodogram shows a sharp peak at $\omega = 0$. But it shows also dominant peaks at frequencies $\omega = k/7, k = 1, 2, 3$. Then, a possible long memory effect and a seasonal component of period s=7 should be included in the model.

Because it is customary in the financial econometrics literature to differentiate log-prices, we applied the same six unit-root test from Table **4-1**, for detecting an ordinary unit root. All the results[2] in Table **4-7** reject the null hypothesis of a unit root, and of stationarity in case of KPSS test. In principle one should not differentiate the log prices.

[2] All test used p=3 lags for the approximating AR(p) process and with the option of non-cero mean, from the R package fUnitRoots of Wuertz et al.[47]

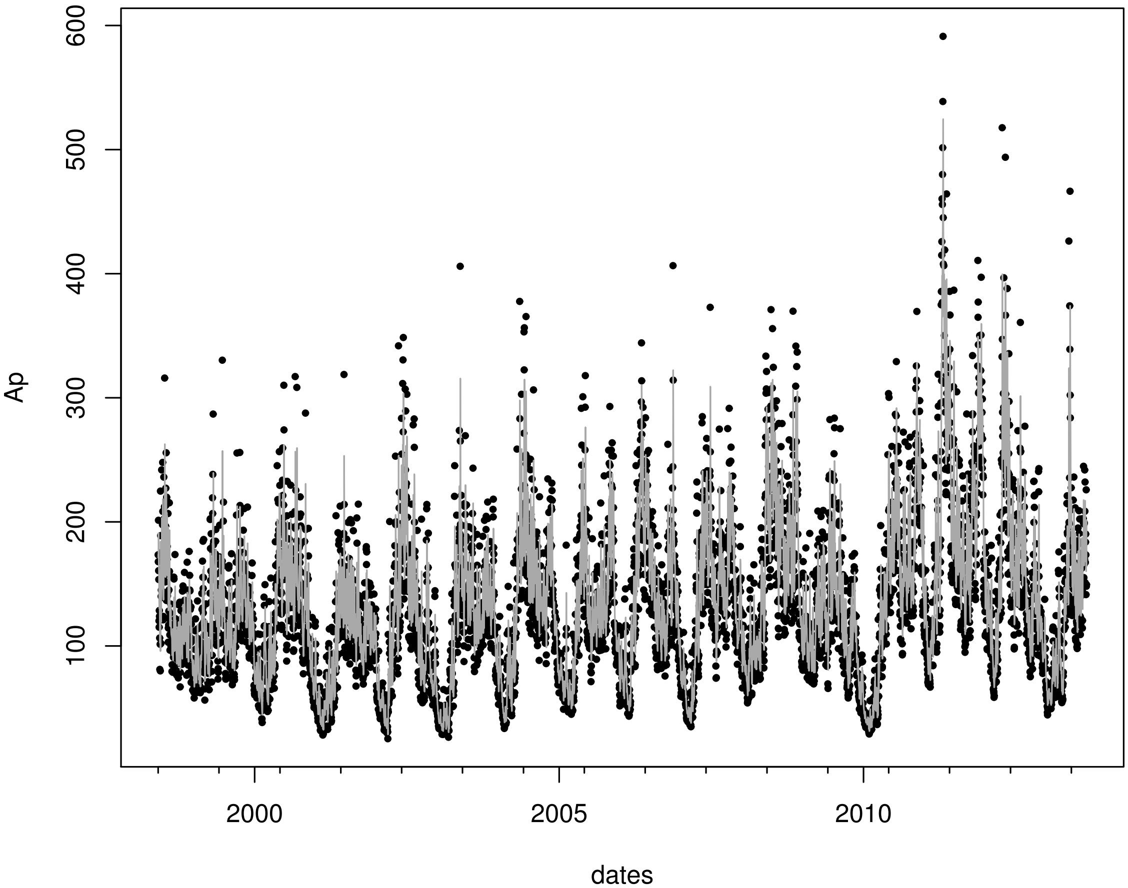


Figure 4-3.: Observed values of A_t (dots) versus simulated values of A_t (continuous)

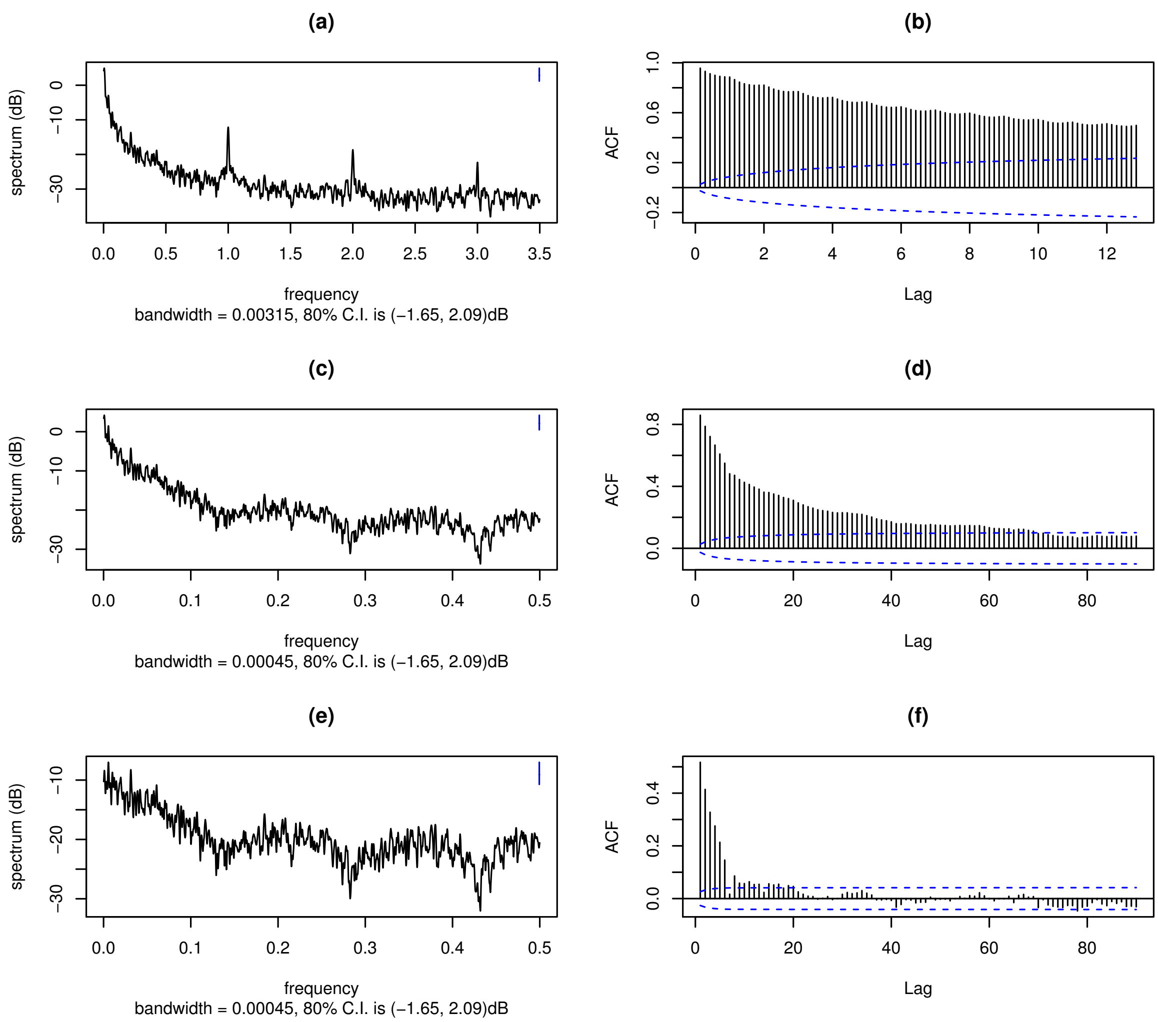


Figure 4-4.: Periodograms and ACF of: $log(P_t) \rightarrow$ (a) and (b). $(1 - L^7)^{0,35} log(P_t) \rightarrow$ (c) and (d). $(1 - L^7)^{0,35}(1 - L)^{0,29} log(P_t) \rightarrow$ (e) and (f).

Test	Null Hipothesis Reject if Test value < Critical value	Test value	Critical values	
			1 %	5 %
Augmented Dickey-Fuller	unit root	-7.2	-3.43	-2.86
DF-GLS	unit root	-2.54	-2.57	-1.94
KPSS	stationary	8.77	0.73	0.46
Phillips-Perron	unit root	-8.92	-3.43	-2.86
Schmidt-Phillips	unit root and a linear trend	-4.62	-3.56	-3.02
Zivot-Andrews	unit root with a structural break	-10.28	-5.34	-4.80

Table 4-7.: Unit root test for the log prices of electricity time series

For examining a possible non-seasonal long memory effect, we applied the Rescaled Range R/S Test of Lo [27, 16] and the Modified Rescaled Range V/S of Giraitis et.al [14, 16]. Both tests reject the null hypothesis Ho: d = 0, see Table **4-8**.

Ho: d = 0			
Data:	R/S Statistic = 2.3485	Bandwidth q = 60	
Significance level:	0.05	0.1	
Critical value:	1.747	1.62	
Data:	V/S Statistic = 0.4478	Bandwidth q = 60	
Significance level:	0.01	0.05	0.1
Critical value:	0.2685	0.1869	0.1518

Table 4-8.: Long Memory Test: R/S of Lo and V/S of Giraitis et.al

Then, we include a fractional non-seasonal differentiation, but is accomplished through the filter (4-3) with $s = 1$ and a fractional order $|d| < \frac{1}{2}$

$$(1 - L^s)^D = \sum_{i=1}^{\infty} \frac{\Gamma(i - D)}{\Gamma(-D)\Gamma(1 + i)} L^{is} \tag{4-3}$$

But for examining the possibility of seasonal fractional integration we applied the same filter (4-3) with $s = 7$ and another fractional order $|d_s| < \frac{1}{2}$. The filter (4-3) was implemented in the R language from an original S-Plus program from Katayama [23, 22] for cases $s = 1$ and $s = 7$. We explored several values for d, $d_s \in (0; 0{,}5)$, such that fitting the following SARFIMA model (4-4) resulted in significant coefficients and uncorrelated residuals ϵ_t.

$$\Phi_p(L)\Phi_P(L^s)(1 - L^s)^{d_s}(1 - L)^d log(P_t) = \Theta_q(L)\Theta_Q(L^s)\epsilon_t \tag{4-4}$$

With $s = 7$ and $d_s = 0{,}35$ the filtered log-price $(1 - L^s)^{d_s} log(P_t)$ presents an autocorrelation function and a periodogram shown in Figures **4-4**(c) and **4-4**(d) respectively; one can observe that the seasonal effect was filtered out but a long memory effects remains. After applying the filter (4-3) with $d = 0{,}29$ and $s = 1$ the long memory effect disappears, as shown in Figures **4-4**(e) and **4-4**(f), leaving a covariance stationary process. Based on a minimum AIC search a $SARMA(1, 3)(0, 2)[7]$ was finally chosen for the series

$$(1 - L^s)^{d_s}(1 - L)^d log(P_t) \tag{4-5}$$

The coefficients of model (4-4) were estimated by maximum likelihood and are displayed in Table **4-9**. The final model for the log-prices is then

$$\Phi_1(L)(1 - L^7)^{0,35}(1 - L)^{0,29} log(P_t) = \Theta_3(L)\Theta_2(L^7)\epsilon_t \tag{4-6}$$

In Table **4-10** we show the Ljung-Box text not rejecting the null of uncorrelated residuals. We adopted the suggestion of Hyndman and Khandakar [19] to set the number of degrees in the Ljung-Box test. In case of periodicity they recommend setting $df = 2s$. Figure **4-5** shows the fitted values of model (4-4) versus the observed log-prices.

	Estimate	Std. Error	z value	Pr(>|z|)	
ar1	0.917009	0.011299	81.1561	<2.2e-16	***
ma1	-0.504215	0.017576	-28.6881	<2.2e-16	***
ma2	-0.048592	0.016033	-3.0307	0.00244	**
ma3	-0.037423	0.015712	-2.3818	0.01723	*
sma1	-0.241719	0.014480	-16.693	<2.2e-16	***
sma2	-0.061141	0.014106	-4.3345	1.46e-05	***
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 4-9.: Estimated Parameters of model 4-4

Null hypothesis: uncorrelated		
X-squared	df	p-value
16.9762	14	0.2574

Table 4-10.: Ljung-Box Test for the residuals of model 4-4

In Table **4-11** we show the respective residuals which have high kurtosis and asymmetry.

We considered necessary to give some comments about the fitted values presented in Figure **4-5**. Given invertibility conditions, the model (4-4) can be re-written in the form

$$log(P_t) = \frac{1}{(1 - L)^{0,29}} \frac{1}{(1 - L^7)^{0,35}} \frac{\Theta_3(L)\Theta_2(L^7)}{\Phi_1(L)} \epsilon_t \tag{4-7}$$

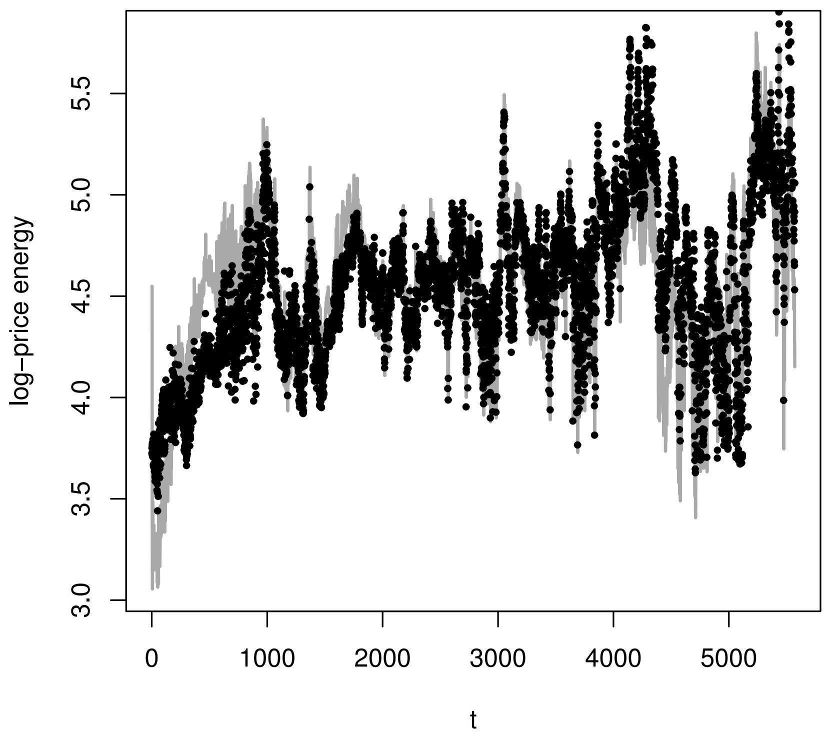


Figure 4-5.: Fitted values from model 4-4 (continuous line) versus observed values (dotted)

Residuals of the Energy Spot Price model (4-6)	
Size	5571
Mean	0.0002
St.Dev	0.1039
Skewness	-0.4136
Kurtosis	7.9078
Minimum	-1.18
Maximum	0.78

Table 4-11.: Residuals of the Energy Spot Price model (4-6)

It was evaluated in three steps. First the filter $\frac{\Theta_3\Theta_2(L^7)}{\Phi_1(L)}$ was applied to the residuals ϵ_t. Then the resulting series was filtered through the inverse of filter (4-3) with $d = 0{,}35$, $\frac{1}{(1-L^7)^{0,35}}$, calculated with the R function `ImpulseVMA()` of library portes of Mahdi and McLeod [30], using a convolution (without the circular option). The resulting series was again filtered through the inverse of filter (4-3) with $d = 0{,}29$, $\frac{1}{(1-L)^{0,29}}$ using the same operations as in the preceding step. The final series is what we call the "fitted values" in Figure **4-5**.

4.2.2. Step Two: Fitting GARCH Models and Volatility Estimation

For the residues ϵ_t of models for A_t and P_t we calculated the Lagrange Multiplier (LM) Test for conditional heteroscedasticity of ARCH type. The results are shown in Table **4-12**.

Null hypothesis: no ARCH effects			
Residuals of	Chi-squared	df	p-value
Model 4-2	245.872	12	< 2.2e-16
Model 4-6	362.441	14	< 2.2e-16

Table 4-12.: ARCH LM-test

According to the discussion in Chapter 3.2.1 and the definition 3.7 and 3.8, we adjust a $GARCH(p,q)$ or $FIGARCH(p,d,q)$ model to $\epsilon_{A,t}$ and $\epsilon_{P,t}$ series, where the subscripts A and P refer to the Water Inflows and Spot Price, respectively.

With the `ugarchfit()` function from the `rugarch` library written by Ghalanos [12], for Water Inflow residues series we found the GARCH(1,1) model presented the best fit. This model is defined by the following system.

$$\begin{aligned}\epsilon_{A,t} &= \sigma_{A,t} e_{A,t}\\ \sigma^2_{A,t} &= \omega + \beta_1 \sigma^2_{A,t-1} + \alpha_1 \epsilon^2_{A,t-1}\end{aligned} \tag{4-8}$$

On the other hand, for log-Price residues series we found the GARCH(2,2) model presented the best fit. This model is defined by the following system.

$$\begin{aligned}\epsilon_{P,t} &= \sigma_{P,t} e_{P,t}\\ \sigma^2_{P,t} &= \omega + \beta_1 \sigma^2_{P,t-1} + \beta_2 \sigma^2_{P,t-2} + \alpha_1 \epsilon^2_{P,t-1} + \alpha_2 \epsilon^2_{P,t-2}\end{aligned} \tag{4-9}$$

If $\sum_{j=1}^{p} \beta_i + \sum_{i=1}^{q} \alpha_i < 1$, then it guarantees that the process (σ_t) is covariance stationary, such that $\epsilon_t = \sigma_t e_t$, and $e_t \sim$ White Noise i.i.d$(0, \sigma^2)$,.

The `ugarchfit()` function of the `rugarch` library allows to model the distribution of errors ϵ_t. To incorporate the skewness and the excess of kurtosis observed in the ϵ_t, which is observed in Tables **4-6** and **4-11**, we use a Normal Inverse Gaussian (NIG) distribution, introduced by Barndorff [1].

A random variable X is said to be distributed according to a NIG distributions if its density function is given by

$$f(x) = \frac{\alpha\delta K_1(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{\pi\sqrt{\delta^2 + (x-\mu)^2}} e^{\delta\gamma + \beta(x-\mu)}, \ -\infty < x < \infty, \tag{4-10}$$

Where the parameters are $\alpha, \delta > 0$, $\mu \in \mathbb{R}$, $0 \leq |\beta| < \alpha$, $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $K_\kappa(.), \kappa \geq 0$ is the modified Bessel function of third class with index κ.

Therefore, if $X \sim NIG(\mu, \alpha, \beta, \delta)$ then $\mathbb{E}(X) = \mu + \delta\beta/\gamma$, $Var(X) = \delta\alpha^2/\gamma^3$, skewness $\gamma_1 = 3\beta/(\alpha\sqrt{\delta\gamma})$, kurtosis $\gamma_2 = 3(1 + 4\beta^2/\alpha^2)/(\delta\gamma)$.

The estimation results are shown in Tables **4-13** and **4-14**. In Table **4-15**, we show a summary of the main statistics for $\sigma_{A,t}$ and $\sigma_{P,t}$.

	Estimate	Std. Error	t value	value Pr(>|t|)	
ω	5.34317	1.491268	3.583	0.00034	***
α_1	0.12482	0.009470	13.181	0.00000	***
β_1	0.87418	0.009816	89.057	0.00000	***
Skew	0.66337	0.015176	43.711	0.00000	***
Shape	1.51269	0.124687	12.132	0.00000	***
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 4-13.: Coefficients of $GARCH(1,1) + NIG(0, \alpha, \beta, 1)$ for A_t

In Figure **4-6** we show the original series A_t and P_t with their respective estimated volatilities. The Spot Price volatility is scaled by a factor 2,5. For example, it can be seen that during 2010 occurred a strong decrease in water inflows which generate an increase in the Spot Price and its volatility. From mid 2012 to mid 2013 we observed a prolonged decline in contributions which resulted in higher prices and volatilities. On the other hand, the years with stable water inflows as 2002-2005 generated a low volatility and stable prices.

4.2.3. Step Three: Fitting a Transfer Function Model to Volatilities

Before checking for a transfer function between the log-square volatility of the Water Inflows and the Energy Spot Prices (see equation (4-1)), we will evaluate if there is a cointegration

	Estimate	Std. Error	t value	value Pr(>|t|)	
ω	0.000090	0.000028	3.2178	0.001292	**
α_1	0.151447	0.018775	8.0666	0.000000	***
α_2	0.006334	0.004068	3.5570	0.001469	**
β_1	0.227398	0.075276	3.0209	0.002520	**
β_2	0.595694	0.075831	7.8556	0.000000	***
Skew	-0.05520	0.023239	-2.3757	0.017517	*
Shape	1.066390	0.096499	11.0508	0.000000	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Table 4-14.: Coefficients of $GARCH(2,2) + NIG(0, \alpha, \beta, 1)$ for P_t

	$\sigma_{A,t}$	$\sigma_{P,t}$
Size	5571	5571
Minimum	7.6651	0.032411
Maximum	109.4405	0.385192
Mean	30.0579	0.097770
Stdev	13.9657	0.043602
Skewness	1.0107	1.307059
Kurtosis	1.7879	2.647669

Table 4-15.: Summary statistics for $\sigma_{A,t}$ and $\sigma_{P,t}$

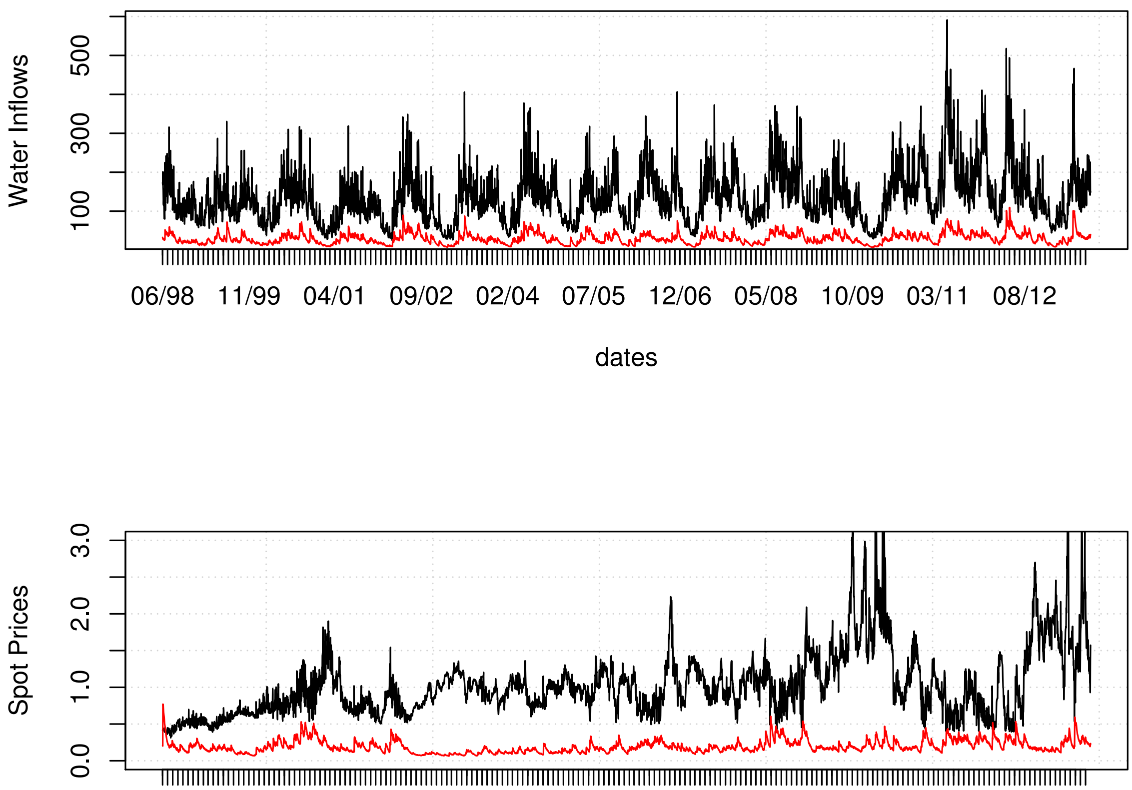


Figure 4-6.: Graph above: Water Inflow A_t. Graph below: Spot Price P_t. Grey color: Original time series. Red color: Estimated volatility $\widehat{\sigma_{A,t}}$ and $\widehat{\sigma_{P,t}}$

model, such as:

$$\begin{aligned} \sigma_{P,t} &= \alpha + \beta\sigma_{A,t} + R_t \\ R_t &= \rho R_{t-1} + \epsilon_t \\ \epsilon_t &\sim N(0, \sigma^2) \end{aligned} \tag{4-11}$$

In Figure **4-7**, we show the result of the evaluation of the cointegration according to equation (4-11). We can conclude that $log(\sigma^2_{A,t})$ does not seem to be integrated, also $log(\sigma^2_{P,t})$ does not seem to be integrated and $log(\sigma^2_{A,t})$ and $log(\sigma^2_{P,t})$ do not appear to be cointegrated.

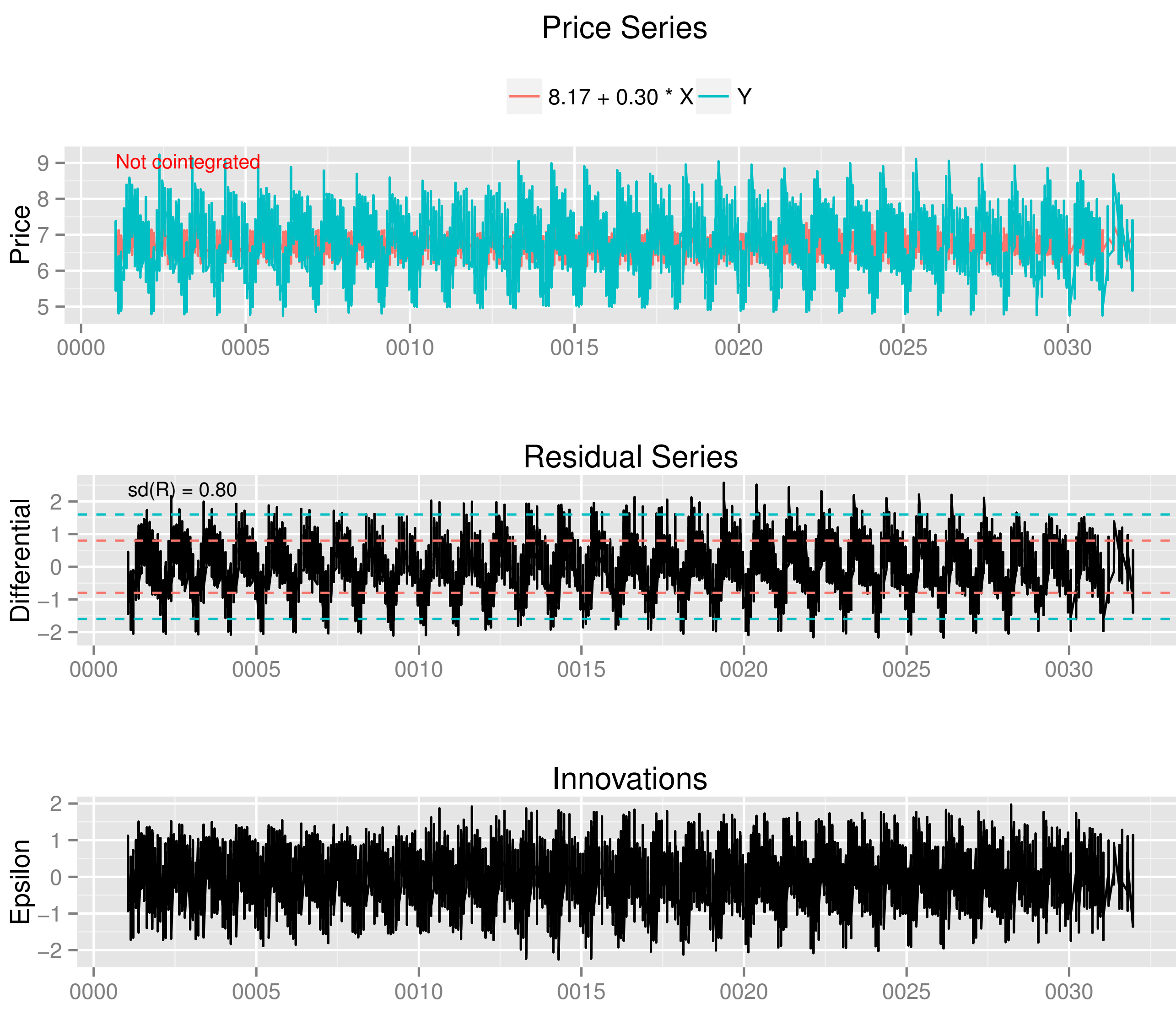


Figure 4-7.: Cointegration between $log(\sigma^2_{A,t})$ and $log(\sigma^2_{P,t})$

The estimation of the transfer function model (4-1)

$$log(\sigma^2_{P,t}) = \mu + \frac{\theta_{q'}(L)}{\phi_{p'}(L)} log(\sigma^2_{A,t-b}) + \nu_t$$

Can be done through several strategies. One which is well known is the Box-Jenkings pre-whitening procedure, based on the cross-correlation of the observed data and in an (sometimes subtle) identification procedure which compares the cross correlation of several simple transfer function models (4-1) with the observed one, choosing as the optimal model the one with the more similar cross-correlation. We briefly review the Box-Jenkins pre-whitening procedure. This procedure is best explained in Section 3.3.1.

If (A_t, P_t) are two processes jointly stationary its cross-covariance function is defined as

$$\gamma_{A,P}(k) = Cov(A_t, P_{t+k}) = \mathbb{E}((A_t - \mu_A)(P_{t+k} - \mu_P)) \text{ for } k = 0, \pm 1, \pm 2, ... \tag{4-12}$$

And it satisfies $\gamma_{A,P}(k) = \gamma_{P_t,A_t}(-k)$. An estimator of $\gamma_{A_t,P_t}(k)$ is

$$\widehat{\gamma}_{A,P}(k) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-k}(A_t - \bar{A})(P_{t+k} - \bar{P}), & k = 0, 1, 2, \dots \\ \frac{1}{T}\sum_{t=1}^{T+k}(A_t - \bar{A})(P_{t-k} - \bar{P}), & k = 0, -1, -2, \dots \end{cases} \tag{4-13}$$

The cross correlation is defined as

$$\rho_{A,P}(k) = \frac{\gamma_{A,P}(k)}{\sigma_{A,t}\sigma_{P,t}} \tag{4-14}$$

Where $\sigma_{A,t} = \sqrt{Var(A_t)}$ and $\sigma_{P,t} = \sqrt{Var(P_t)}$. Using (4-12) well known estimators of variances, an estimator $\hat{\rho}_{A,P}$ of the cross correlation in (4-13) can be defined immediately.

The Box-Jenkins pre-whitening methodology consists of several steps (see Wei [46], Chapter 14). In the first one, a general stationary SARFIMA model (3-6) with $X_n = A_t$ or P_t, is fitted to $log(\sigma^2_{A,t})$ input series. For instance, an ARFIMA of the form.

$$\Phi_p(L)(1 - L)^d log(\sigma^2_{A,t}) = \alpha_0 + \Theta_q(L)U_t \tag{4-15}$$

Where U_t is white noise with variance σ^2_U. One can solve (4-15) for U_t, assuming the invertibility conditions for $\Theta_q(L)$, and we obtain

$$U_t = \frac{\Phi_p(L)(1 - L)^d}{\Theta_q(L)} log(\sigma^2_{A,t}) - \frac{\alpha_0}{\Theta_q(1)} \tag{4-16}$$

And additionally, it can be found a sequence $(\pi_j, j = 0, 1, 2, ...)$ such that the following identity is true, and defines the linear filter $\pi(L)$.

$$\pi(L) = \frac{\Phi_p(L)(1-L)^d}{\Theta_q(L)} = \sum_{j=0}^{\infty} \pi_j L^j \tag{4-17}$$

With $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\pi_0 = 1$.The linear filter $\pi(L)$ is known as prewhitening filter. Applying this filter to the input series, we obtain

$$\alpha_t = \frac{\Phi_p(L)(1-L)^d}{\Theta_q(L)} log(\sigma^2_{A,t}) \tag{4-18}$$

Applying the same prewhitening filter to the output series $log(\sigma^2_{P,t})$ we obtain a filtered output series

$$\beta_t = \frac{\Phi_p(L)(1-L)^d}{\Theta_q(L)} log(\sigma^2_{P,t}) \tag{4-19}$$

Letting $\epsilon_t = \phi_{\sigma_A} \theta^{-1}_{\sigma_A} \nu_t$ the transfer function model model (4-1) becomes

$$\beta_t = \nu(L)\alpha_t + \epsilon_t \tag{4-20}$$

The impulse response weights ν_j for the transfer function can be found as

$$\nu_k = \frac{\sigma_\beta}{\sigma_\alpha} \rho_{\alpha\beta}(k) \tag{4-21}$$

In Figure **4-8** we obtain the Cross Correlation Function (CCF) whit the respective weights ν_k, which according (4-20), it measures the strength of association between $log(\sigma^2_{P,t})$ and $log(\sigma^2_{A,t})$.
From Figure **4-8**, we can conclude that the effect of the Water Inflows volatility begins to manifest in the Spot Price volatility after four days. According tables of typical impulse weights (see table **3-1**) the transfer model identified was

$$log(\sigma^2_{P,t}) = \mu + \Theta'_q(L) L^b log(\sigma^2_{A,t}) + \nu_t \tag{4-22}$$

With $q' = 4$, $b = 4$, $\mu = -4{,}857421$ such that ν_t follows an $ARMA(2, 1)$. The Table **4-16** shows the estimated coefficients of the $ARMA(2, 1)$ and of $\Theta_{q'}(L)$.

Table **4-17** shows the Ljung Box for the residuals of the $ARMA(2, 1)$ model.

The transfer function model can be expressed as

$$log(\sigma^2_{P,t}) = \mu + (1 - \theta L^4) log(\sigma^2_{A,t-4}) + \frac{1 - \theta_1 L}{1 - \phi_1 L + \phi_2 L^2} e_t \tag{4-23}$$

Such that $e_t \sim$ White Noise $(0, \sigma^2)$

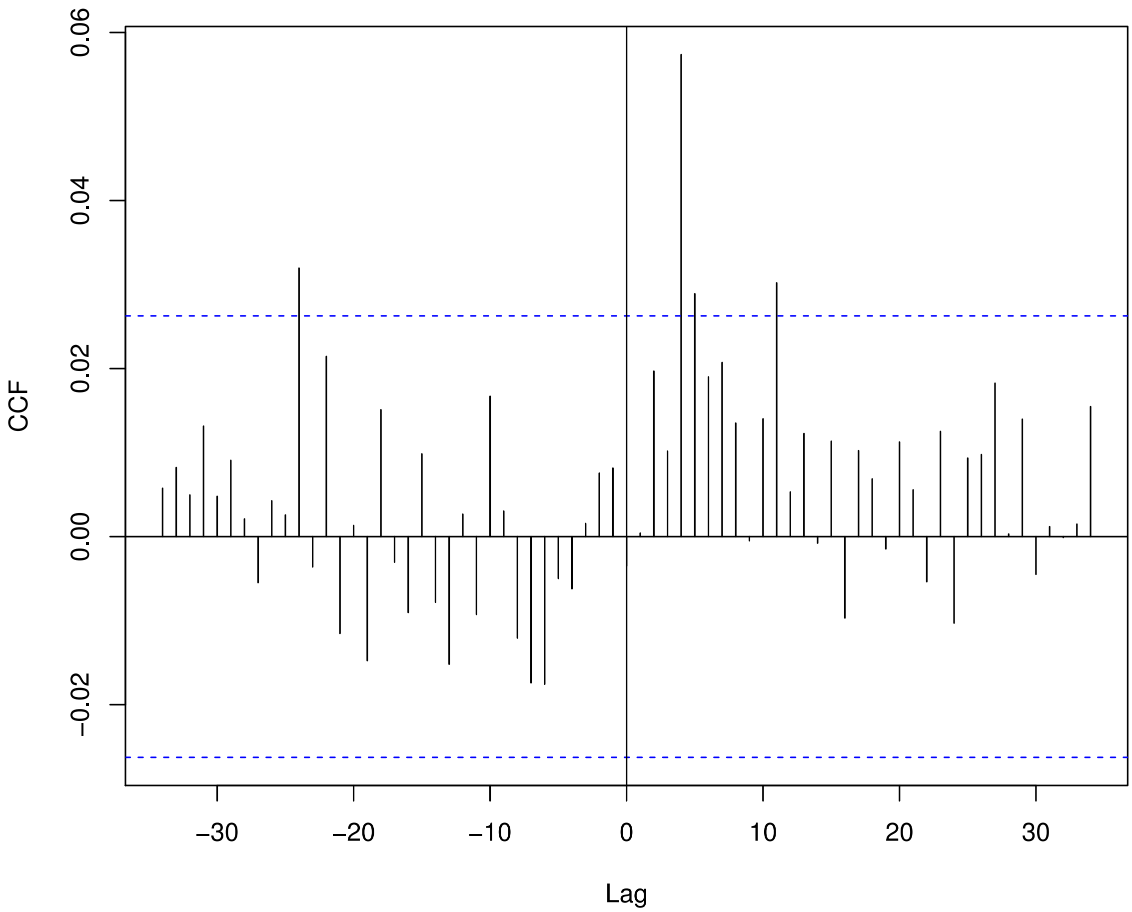


Figure 4-8.: Cross Correlation Function between $log(\sigma^2_{P,t})$ and $log(\sigma^2_{A,t})$

	Estimate	Std. Error	z value	Pr(>|z|)	
ar1	0.296789	0.017251	17.2039	< 2.2e-16	***
ar2	0.678637	0.016600	40.8826	< 2.2e-16	***
ma1	0.210381	0.022587	9.3144	< 2.2e-16	***
intercept	-4.857421	0.223091	-21.7733	< 2.2e-16	***
x-MA0	-0.006052	0.013774	-0.4394	0.6604	
x-MA1	-0.011936	0.015352	-0.7775	0.4368	
x-MA2	0.018629	0.015827	1.1770	0.2392	
x-MA3	-0.019364	0.015354	-1.2612	0.2072	
x-MA4	0.057877	0.013779	4.2004	2.664e-05	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Table 4-16.: Coefficients of the transfer function model

Null hypothesis: uncorrelated residuals		
X-squared	df	p-value
19.4143	14	0.1497

Table 4-17.: Ljung Box Test for the residuals of the $ARMA(2,1)$

4.3. Some Parametric Models Evaluated.

In Section 4.2, we saw that it is possible to establish a parametric ADL-Rational model (See equation (4-23)). However, we will try to adjust other models.

Initially, we will try to make an adjustment by an ADL-Gamma model (it was explained in Section 3.3.2) and then, we will try an adjustment by the ADL-Koyck model (it was explained in Section 3.3.1, page 18).

4.3.1. Model: ADL-Gamma

The ADL Gamma model was proposed by Bollerslev [24] in 1986 and it is shown in the equation (4-24).

$$y_t = \mu + \alpha \sum_{j=0}^{\infty} (j+1)^{\delta/(1-\delta)} \lambda^j x_{t-j} + e_t \tag{4-24}$$

where $\delta, \lambda \in (0,1), \mu, \alpha \in \mathbb{R}, (e_t, t \in \mathbb{Z})$ and $x_t, t \in \mathbb{Z}$ is a covariance stationary process that supports a causal representation, $x_t = \sum_{k=0}^{\infty} \psi_k Z_k$, $Z_k \sim$ White Noise $(0, \sigma_x^2)$ with $k \in \mathbb{Z}$.

We propose a methodology for estimating the parameters of the model (4-24). This methodology is explained in Appendix C. When we use this methodology, we obtain the parameters for the equation (4-24), where $y_t = log(\sigma_{P,t}^2)$ is the log-square volatilities of the Spot Price and $x_t = log(\sigma_{A,t}^2)$ is the log-square volatilities of the Water Inflows.

- $\hat{\mu}$: 2.83051036
- $\hat{\alpha}$: 0.06636876
- $\hat{\lambda}$: 0.09896834
- $\hat{\delta}$: 0.08949369

Graphically, the figure **4-9** contrasts the $log(\sigma_{P,t}^2)$ calculated with the values of $log(\sigma_{P,t}^2)$ evaluated with the estimated parameters $\hat{\mu}, \hat{\alpha}, \hat{\lambda}$ and $\hat{\delta}$ over $log(\sigma_{A,t}^2)$. From this figure, it is evident that the Gamma model can not explain the volatility of the Energy Spot Price from the Water Inflows.

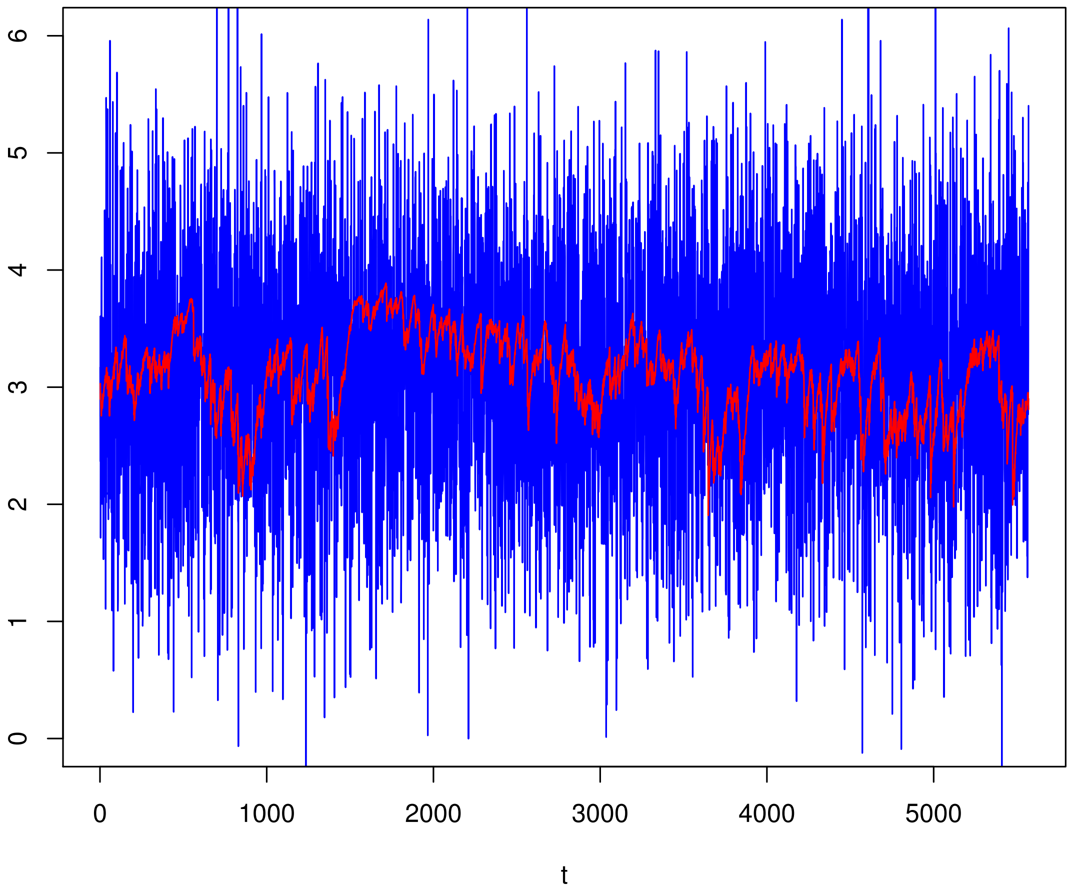


Figure 4-9.: Comparison between the original $log(\sigma^2_{P,t})$ series (red color) and the output series with the estimated parameters $log(\sigma^2_{P,t})|_{\hat{\mu},\hat{\alpha},\hat{\lambda},\hat{\delta}}$ (blue color) after applying the model ADL Gamma on $log(\sigma^2_{A,t})$

4.3.2. Model: ADL-Koyck

According to the ADL-Koyck model, which was explained in Section 3.3.1 (definition 3.9), we execute the R Software code shown in Appendix B.2.

Using the identity (3-28), we obtain the following parameters:

- $\hat{\sigma} = 0{,}7357568$
- $\hat{\mu} = 4{,}5833053$
- $\hat{\lambda} = 0{,}6619485$
- $\hat{\alpha} = -0{,}1312024$

In Figure **4-10** we present a graphical analysis of residuals from the ADL-Koyck model with the parameters estimated. Additionally, the results of the Box-Ljung test are presented in Table **4-18**.

From Figure **4-10** and Table **4-18**, it is obvious that the residuals are not white noise. Through the `auto.arima()` function from the `forecast` library and using the lowest AIC criterion, we see that the residuals follow an $ARIMA(3,1,2)$ model, whose parameters are shown in Table **4-19**.

X-squared	df	p-value
36006.94	12	< 2.2e-16

Table 4-18.: Ljung-Box Test for ADL-Koyck residuals

	Estimate	Std. Error	z value	Pr(>|z|)	
ar1	-0.4935463	0.0343276	-14.3776	< 2.2e-16	***
ar2	0.8973342	0.0064936	138.1870	< 2.2e-16	***
ar3	0.5458028	0.0304531	17.9227	< 2.2e-16	***
ma1	-0.6669545	0.0389377	-17.1287	< 2.2e-16	***
ma2	-0.3301764	0.0388139	-8.5067	< 2.2e-16	***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 4-19.: Parameters of the residuals ADL-Koyck model

However, when we adjust the residues to the Koyck model, the residuals still are not White Noise, as we can see in Figure **4-11** and Table **4-20**. If we do the Ljung Box test with less

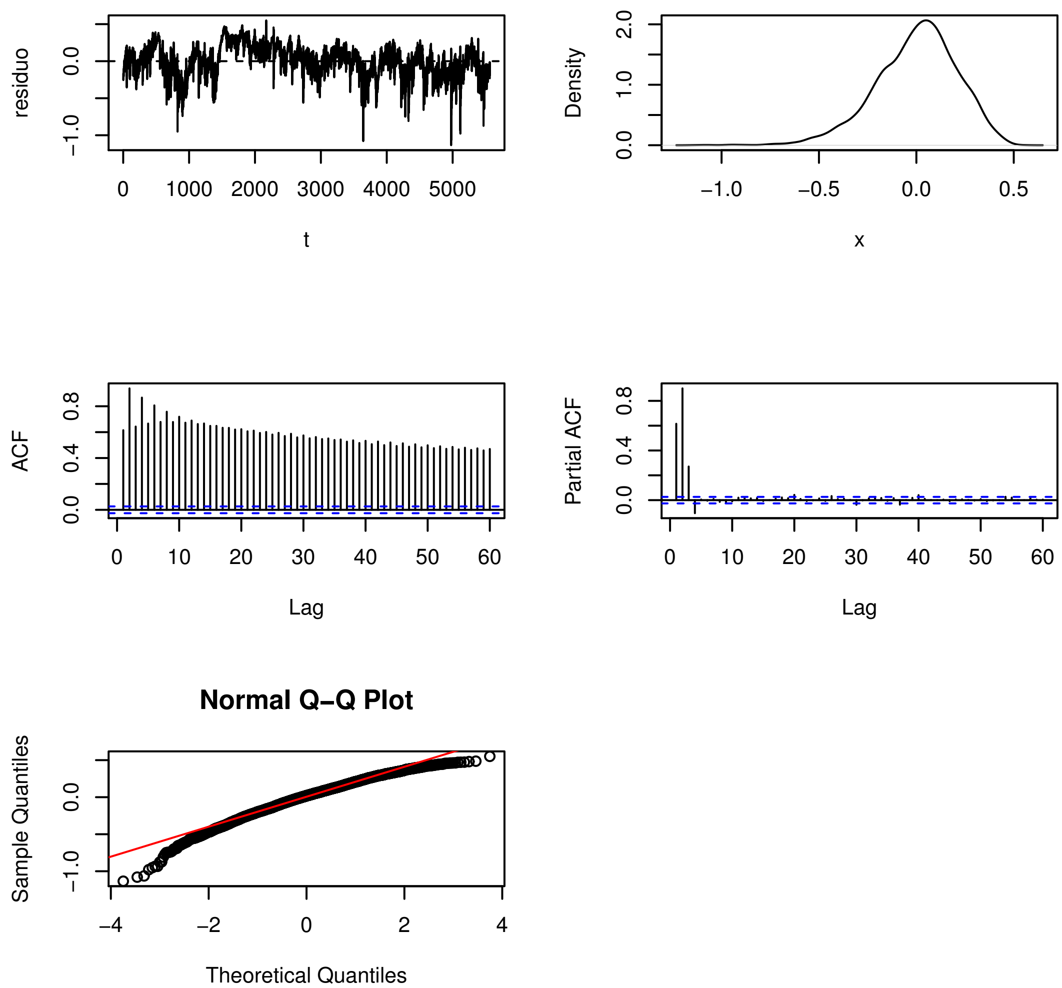


Figure 4-10.: Graphical summary of the ADL Koyck Model residuals with the estimated parameters.

lags, as we show in Table **4-21**, we can see the behavior is approaching to be white noise. This indicates that the model is spurious and is not recommended to use it.

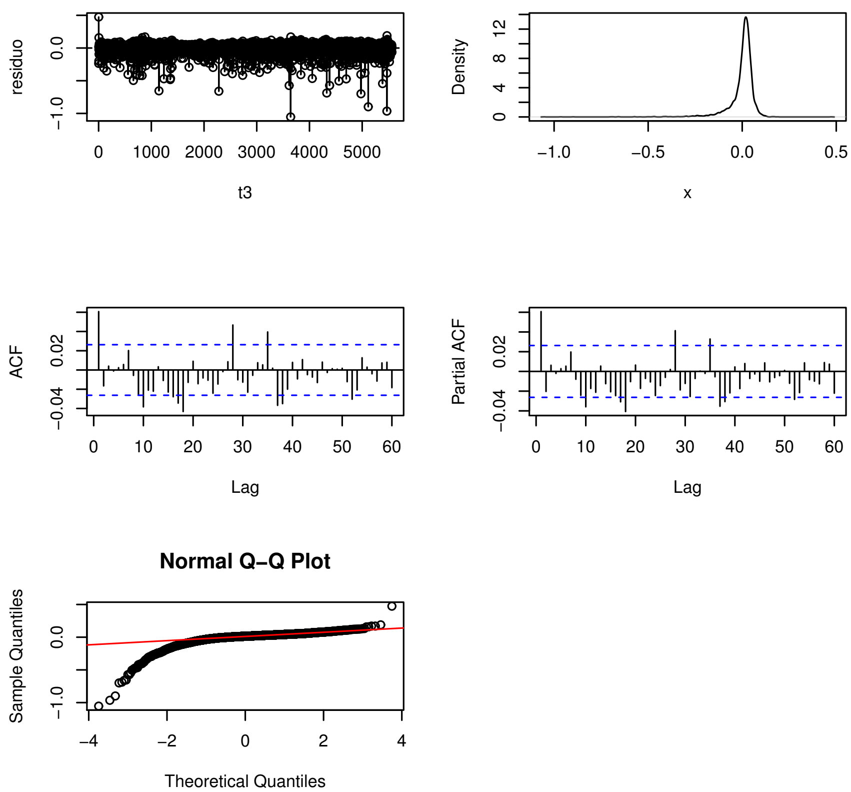


Figure 4-11.: Graphical summary of the Koyck Model when we adjust the residuals to an ARIMA(3,1,2).

X-squared	df	p-value
88.885	30	9.682e-08

Table 4-20.: Ljung-Box Test for ADL-Koyck residuals after setting an ARIMA(3,1,2) model.

In Figure **4-12** we show the Koyck model fitted between the log-square volatility of the Spot Price $log(\sigma^2_{P,t})$ (red color) and the log-square volatility of the Water Inflows $log(\sigma^2_{A,t})$ (blue color) after setting an $ARIMA(3,1,2)$ to the residuals. However, the residuals of the general Koyck model, still not been white noise, as we explained earlier.

Lag	X-squared	df	p-value
25	74.7698	25	7.357e-07
20	68.1616	20	3.625e-07
10	36.6228	10	6.576e-05
5	22.1406	5	0.0004923

Table 4-21.: Ljung-Box Test with different lags for ADL-Koyck residuals after setting an ARIMA(3,1,2) model.

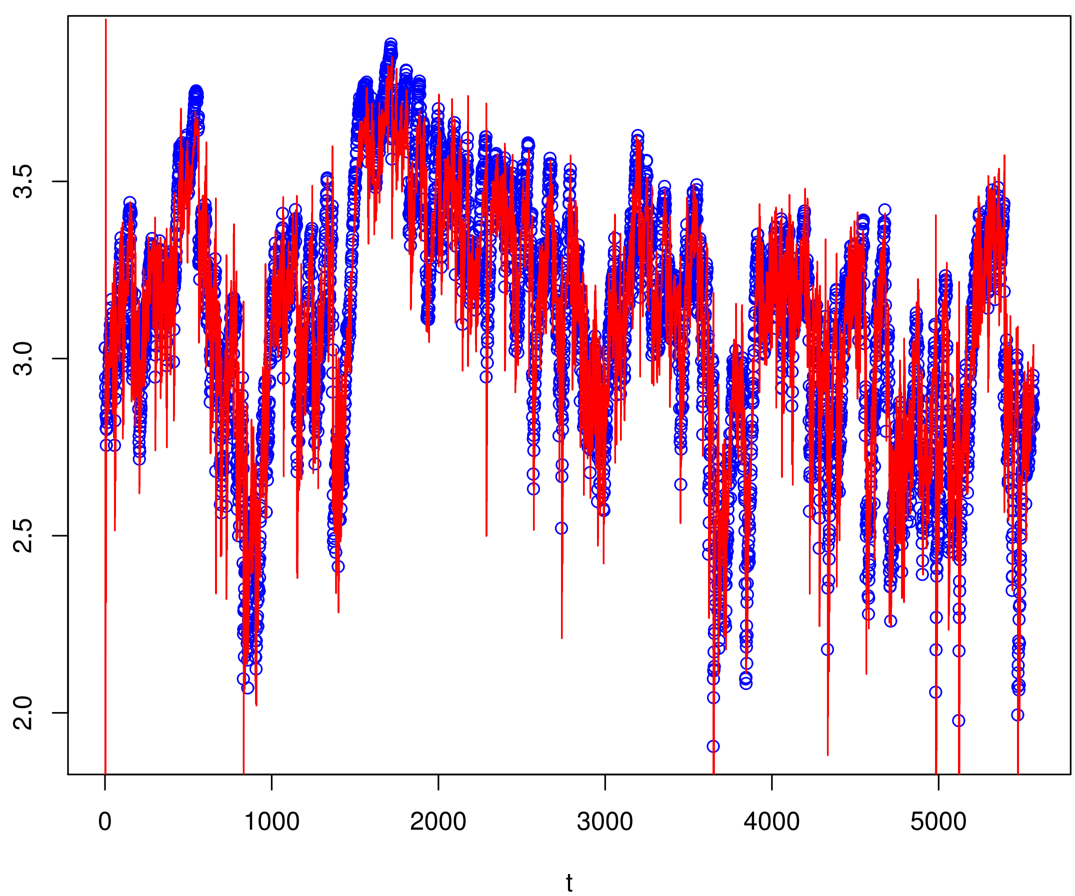


Figure 4-12.: Koyck model fit between the volatility of the spot price $log(\sigma^2_{P,t}$ (red color) and the volatility of the water inflows $log(\sigma^2_{A,t}$ (blue color) after setting an $ARIMA(3,1,2)$ to the residuals.

5. Conclusions and Recommendations

Conclusions refer to the models for water inflows and spot price, and the transfer function model for volatilities.

5.1. Conclusions

- Although the water inflows series displays a strong annual seasonal component we did not apply a SARIMA type model. We do not filtered the seasonal component with a deterministic periodic function but fitted an ARIMA model which rendered significant coefficients and white noise residuals of GARCH type.
- For the electricity spot price we found two stylized facts which should be included into any model
  - A weekly periodicity: it could be explained, in principle, as an effect from the power consumption (residential demand), but confirming this fact is left as a further issue to investigate
  - A seasonal and non seasonal fractional integration: the fitted model was of a SARFIMA type, with a seasonal and non seasonal fractional differentiation filter which rendered a SARMA type model, with residuals of the GARCH type.
- The residuals of A_t were fitted to a GARCH(1,1) process while $log(P_t)$ residuals were fitted to a GARCH(2,2). From these models the corresponding log-square volatilities were estimated.
- There is a significant effect from the hydrology into the prices, and the fitted transfer model 4-23 shows that the water inflows volatility effects are transferred to the energy spot volatility after approximately 4 days. This conclusion confirms the well accepted assumption that hydrology affects the spot prices.
- About the behavior of agents:
  - The agents respond rationally at the behavior of hydrology, suggesting that there is no market power behavior.
  - The delayed response of the Spot Price volatility is due to management regulation reservoirs, which offers coverage to the Water Inflows volatility to agents that have this type of reservoirs.

- The commercial strategy of the electricity companies considered as a fundamental variable current and past hydrological behavior.

5.2. Recommendations

- To model a water inflows series (which has a seasonal component) it can be done through an ARIMA model. The annual seasonal component can in fact be assimilated to a stochastic trend giving rise to an ordinary (not seasonal) unit root.
- Investigate if the weekly periodicity in the Energy Spot Price series is due to the Power Consumption (residential demand).
- Use a GARCH(1,1) model to represent the volatility of Water Inflows and a GARCH(2,2) model to represent the volatility of Energy Spot Price in Colombia.
- Investigate if water inflows can affect log-prices, not through volatilities but from the level series.
- A question remains though, and it is if water inflows can affect log-prices, not through volatilities but from the level series. From (4-6) the model for log-prices follows the system

  $$\Phi_1(L)(1-L^7)^{0,35}(1-L)^{0,29}log(P_t) = \Theta_3(L)\Theta_2(L^7)\epsilon_t`$$

  ' Where, $\epsilon_t = \sigma_t e_t$ and $\sigma^2_{P,t} = \omega + \beta_1\sigma^2_{P,t-1} + \beta_2\sigma^2_{P,t-2} + \alpha_1\epsilon^2_{t-1} + \alpha_2\epsilon^2_{P,t-2}$. The last equation could be modified so as to include the exogenous variable, $log(\sigma^2_{A,t-2})$, such that, for certain γ

  $$\sigma^2_t = \omega + \beta_1\sigma^2_{P,t-1} + \beta_2\sigma^2_{P,t-2} + \alpha_1\epsilon^2_{t-1} + \alpha_2\epsilon^2_{P,t-2} + \gamma log(\sigma^2_{A,t-4})$$

  This model would include the effect of water inflows volatilities into the electricity price. But a model directly relating electricity price and water inflows was not considered because of the strong annual periodicity in water inflows, and the weekly periodicity of log-prices, giving a difficult transfer model with two periods, not divisible ($365 \nmid 7$).
- To present the results of this study to regulatory authorities to show that electricity companies respond rationally to hydrological behavior.

A. Appendix: Time series modeling in R

A.1. Appendix: Box-Jenkins approach for ARIMA(p,d,q) processes using R

The `pi.j()` function of the `afmtools` library, calculates a sequence of coefficients $(\pi_j, j = 0, 1, \ldots, n)$, of $\pi(L)$ filter. This sequence is applied by the `filter()` function of the `stats` library to the series $log(\sigma^2_{P,t})$, $log(\sigma^2_{A,t}) - \mathbb{E}(log(\sigma^2_{A,t}))$ (centered), to calculate W_t and U_t.

Then, the cross-correlation function is applied by the `ccf()` function to find the patterns that could describe the ν_k. That pattern should give us a guide to define the transfer function model $\nu(L)$.

Box and Jenkins said: [4], pag. 380 "The preliminary estimates $\widehat{\nu}_k$ so obtained are again, in general, statistically inefficient but can provide a rough basis for selecting suitable operators...in the transfer function model".

A.2. Time Series Models in R

The estimation of time series models has been widely studied in the literature. Particularly, in the development of this work we have used the maximum likelihood method to estimate the parameters.

The `TSA` package written by Kung and Ripley, [7] is heavily based on the `arima` function of the stats core of R. In this package, the statistical methods to establish the ARIMA model that best fit the data are: maximum likelihood and minimizing the conditional sum of squares. The default (unless there are missing values) is to use conditional sum of squares to find starting values, then R use maximum likelihood.

Moreover, the `auto.arima()` function of the `Forecast` library returns the best ARIMA model according to AIC, AICc or BIC value.

The methods used by the `auto.arima()` function for estimating the best ARIMA model are describing by Hyndman and Khandakar [19]. Moreover, according to Osborn et al. [31]

the method used to analyze seasonal differences is the OCSB test (Osborn , Chui, Smith , Birchenhall). To estimate the distributed lag model for the volatility, we use ordinary least squares by using the `arimax()`. It is important to say that this functions have the ability to involve and analyze seasonal series. Therefore, everything said for ARIMA series also can be extrapolated for series ARMA and SARMA.

In the case of the ARFIMA and SARFIMA series, it is necessary the use of `afmtools` library, written by Contreras-Reyes et al. [8]. This library is specialized in Estimation, Diagnostic and Forecasting functions for ARFIMA models.

A.3. GARCH Models in R

Particularly, the `rugarch` package is specialized in modeling and testing GARCH models. The testing environment is based on a rolling backtest function. Additionally, this package has a lot of distributions to be used for modelling innovations. From this package, we use the `ugarchfit()` function which fits the data to a wide variety of univariate GARCH models.

The `ugarchfit()` function uses a hybrid solution strategy. First of all, it attempts a nonlinear solution, as was proposed by Ye [48]. If this method does not converge then it attempted with an optimization methods using port routines, next with a random initialization and multiple restarts of the nonlinear solutions, and if none of these methods converge, then it try to use the `nloptr` package which offers the possibility of multiple nonlinear optimization methods.

B. Appendix: R codes

B.1. Application of Statistical Model

```
ls()
rm(list=ls())

library(itsmr)
library(forecast)
library(TSA)
library(lmtest)
library(tseries)
library(FitARMA)
library(caschrono)

D = read.table('Precio.Aportes.prn', header = T, stringsAsFactors=FALSE)
attach(D)

Apt = ts(Ap,frequency=360)
Pbt = ts(Pb,frequency=7)

np = length(Pbt)
dates = as.Date(Fecha,format='%d/%m/%Y')

ejex.mes = seq(dates[1],dates[np], 'months')
ejex.ano = seq(dates[1],dates[np],'years')

par(mfrow=c(2,1))
plot(dates,Apt, xaxt='n', panel.first = grid()
,type='l',ylab='Water Inflows')
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)

plot(dates,Pbt, xaxt='n', panel.first = grid()
,type='l',ylab='Energy Prices')
```

```
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)


k4 = kernel('modified.daniell',c(3,3))

par(mfrow=c(2,1))

plot(dates,Ap, xaxt='n', panel.first = grid()
,type='l',ylab='Water Inflows',xlab='(a) Time series At')
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)

acf(Ap,ci.type= 'ma',350,main='(a)')

spec.pgram(Ap, k4,ci = 0.85 ,main='',xlab='(b) Estimated periodogram of At')

B = spec.pgram(Ap, k4, ci = 0.8,plot=FALSE)
plot(B$freq[B$freq < 0.05],B$spec[B$freq < 0.05],type='l',
xlab = '(frequency < 0.05)',
ylab = 'spectrum',main='(b)')
abline(v = 1/180,col = 'red')
abline(v = 1/365,col = 'red')

auto.arima(Ap, max.p=10, max.q=10,
ic=c('aicc','aic', 'bic'), test=c('kpss','adf','pp'), seasonal.test=c('ocsb','ch'),
     allowdrift=TRUE, lambda=NULL, parallel=FALSE, num.cores=2)

Ap_armasubsets=armasubsets(y=Ap,
nar=10,nma=10,
y.name='r',
ar.method='ols')

par(mfrow=c(3,1))
plot(Ap_armasubsets)
acf(Ap)
pacf(Ap)

EACF=eacf(Ap)
armaselect(Ap,nbmod=50)
```

```
Aju_Ap = Arima(Ap,order=c(7,1,2),
include.mean = TRUE)

coeftest(Aju_Ap)

library(fUnitRoots)

adf.test(Apt,k = trunc((length(Apt)-1)^(1/3)))

urdfTest(Apt, lags = (length(Apt)-1)^(1/3), type = c('nc', 'c', 'ct'), doplot = TRUE)

urersTest(Apt, type = c('DF-GLS', 'P-test'), model = c('constant', 'trend'),
lag.max = trunc((length(Apt)-1)^(1/3)), doplot = TRUE)

urkpssTest(Apt, type = c('mu', 'tau'), lags = c('short', 'long', 'nil'),
use.lag = NULL, doplot = TRUE)

urppTest(Apt, type = c('Z-alpha', 'Z-tau'), model = c('constant', 'trend'),
lags = c('short', 'long'), use.lag = NULL, doplot = TRUE)

ur.pp(Apt, type = c('Z-tau'), model = c('trend'),
      lags = c('short', 'long'), use.lag = NULL)

urspTest(Apt, type = c('tau', 'rho'), pol.deg = c(1, 2, 3, 4),
signif = c(0.1), doplot = TRUE)

urzaTest(Apt, model = c('intercept', 'trend', 'both'), doplot = TRUE)

Aju_Ap.hat = fitted(Aju_Ap)

r_Ap.hat = residuals(Aju_Ap)

theta = coef(Aju_Ap)

r_Ap.hat = arima.sim(n = length(Ap),
list(order = c(7,1,2),ar = theta[1:7], ma = theta[8:9]),
innov = r_Ap.hat)+mean(Ap)

par(mfrow=c(1,1))
```

```
plot(dates,Ap,type='p',cex=.5,pch=19)
lines(dates,Aju_Ap.hat,lty=1,col='darkgray')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)

par(mfrow=c(2,2))
t = seq(1,length(r_Ap.hat))
plot(t,r_Ap.hat,type='l',ylab='residuals', main = '(a)')
abline(h=0,lty=2)
plot(density(r_Ap.hat),xlab='x',main= '(b)')
acf(r_Ap.hat,ci.type= 'ma',60,main='(c)')
qqnorm(r_Ap.hat, main='(d)')
qqline(r_Ap.hat,col=2)

r_Ap = residuals(Aju_Ap)

Box.test(x = r_Ap, lag = 60, type='Ljung-Box')

tsdiag(Aju_Ap, gof.lag=60)

dAp = diff(Ap,1,1)

par(mfrow=c(1,2))
acf(dAp,ci.type= 'ma',90,main='')
spec.pgram(dAp, k4, ci = 0.8,main='')

armasubsets_dAp=armasubsets(y=dAp,
nar=14,nma=14,
y.name='r',
ar.method='ols')

par(mfrow=c(3,1))
plot(armasubsets_dAp)
acf(dAp)
pacf(dAp)

EACF=eacf(dAp)
armaselect(dAp,nbmod=50)

(Aju_dAp = Arima(dAp,order=c(7,0,2),
include.mean = TRUE))
```

```
coeftest(Aju_dAp)

require(FinTS)
require(rugarch)

ArchTest(r_Ap,lag=14)

efectis ARCH

spec3 = ugarchspec(
variance.model=list(model='fGARCH', garchOrder=c(1,1),submodel = 'GARCH'),
mean.model=list(armaOrder=c(0,0), include.mean=FALSE),
distribution.model='nig')

z2 = as.numeric(r_Ap)

(fit.rAp.et = ugarchfit(data = z2, spec = spec3))

sigma.rAp.et = fit.rAp.et@fit$sigma

t = seq(1,length(sigma.rAp.et))

par(mfrow=c(1,1))
np = length(sigma.rAp.et)
dates = as.Date(dates,format='%d/%m/%Y')
ejex.mes = seq(dates[1],dates[np], 'months')
ejex.ano = seq(dates[1],dates[np],'years')
plot(dates,sigma.rAp.et, xaxt='n', panel.first = grid(),type='l'
,ylab='daily volatility of Water Inflows', xlab='date')
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)
abline(v=as.Date('2002-10-01'))
abline(v=as.Date('2006-09-01'))
abline(v=as.Date('2009-09-01'))
abline(v=as.Date('2010-06-01'))

G = data.frame(dates=dates,sigma.Ap = sigma.rAp.et)

require(MASS)
```

```
write.matrix(G,'volat.aportes.dat',sep=' ')

library(itsmr)
library(forecast)
library(TSA)

source('filtros.katayama.r')

lPb = log(Pb)

x = lPb
n = length(x)
lag = 7
ds = 0.35
g = pij.SFId.gen(ds, lag, n)
rpbs = convolve(g, rev(x))[1:n]

require(fracdiff)

fdGPH(rpbs, bandw.exp = 0.5)

fdSperio(rpbs)

d = 0.29
h = pij.FId.gen(d, n)
rpb = convolve(h,rev(rpbs))[1:n]

par(mfrow=c(3,1))
ts.plot(lPb)
ts.plot(rpbs)
ts.plot(rpb)

k4 = kernel('modified.daniell', c(3,3))

par(mfrow=c(3,2))

spec.pgram(lPb, k4, taper=0,
log = 'dB', ci = 0.8, main='(a)',xlab='frequency')

acf(lPb,90, ci.type = 'ma', main='(b)')
```

```
spec.pgram(rpbs, k4, taper=0,
log = 'dB', ci = 0.8, main='(c)',xlab='frequency')

acf(rpbs,90, ci.type = 'ma', main='(d)')

spec.pgram(rpb, k4, taper=0,
log = 'dB', ci = 0.8, main='(e)',xlab='frequency')
acf(rpb,90, ci.type = 'ma', main='(f)')

rpbs = ts(rpbs,frequency=7)

lPb = ts(lPb,frequency=7)

rpb = ts(rpb,frequency=7)

auto.arima(rpb)

m.rpb = arima(rpb,order=c(1,0,3),
seasonal = list(order=c(0,0,2),period=7),
include.mean = FALSE)
m.rpb

require(lmtest)

coeftest(m.rpb)

r1 = residuals(m.rpb)
r2 = residuals(m.rpb)[-c(1:5)]

par(mfrow=c(2,2))
t = seq(1,length(r1))
plot(t,r1,type='l',ylab='residuo')
abline(h=0,lty=2)
plot(density(r1),xlab='x',main= '')
acf(r1,ci.type= 'ma',60,main='')
qqnorm(r1)
qqline(r1,col=2)

Box.test(x = r1, lag = 14, type='Ljung-Box')
```

```
Box.test(x = r2, lag = 14, type='Ljung-Box')

tsdiag(m.rpb, gof.lag=50)

y = residuals(m.rpb)

b=coef(m.rpb)

phi <- b[1]
theta <- b[2:4]
sphi <- c(0)
stheta <- b[5:6]
period <- 7

d   <-  0
ds  <-  0

require(polynom)

bs <- polynomial(c(rep(0,period),1))
b1 <- polynomial(c(0,1))

  arpoly <- polynomial (c(1,-phi)   )
  mapoly <- polynomial (c(1,theta)  )
  sarpoly <- polynomial(c(1,-sphi)  )
  smapoly <- polynomial(c(1,stheta) )

  fullarpoly <- arpoly*predict(sarpoly,bs)
  fullmapoly <- mapoly*predict(smapoly,bs)

  mo <- list()
  mo$ar <- -coef(fullarpoly)[-1]
  mo$ma <- coef(fullmapoly)[-1]

  p <- length(mo$ar)
  q <- length(mo$ma)
  n <- length(y)

  if(p>0 || q>0 )
     {x <- arima.sim(model=mo,n=n, innov = y)
```

```
  }else{x <- y} # parameter checking may be necessary here...

  if(d+ds > 0){             # undifference: diffinv is better but initial values need
    pdiff <- (1-b1)^d * (1-bs)^ds
    if(is.null(init))
      init <- numeric( length(coef(pdiff)) - 1 )   # order of pdiff, no function poly.
    x <- filter(x, -coef(pdiff)[-1], method = 'recursive', init =init )
    res <- ts(c(init,x),frequency=period)
  }else{
    res <- ts(x,frequency=period)
  }

ds = 0.35
d = 0.29
nt = 30

library(portes)


g = g[1:141]
ginv = ImpulseVMA(phi=-g[-1],
theta = numeric(0),Trunc.Series = 140)

ginv = as.numeric(ginv)

yest.s = convolve(ginv,rev(res),type='o')[1:length(res)]

h = h[1:170]
hinv = ImpulseVMA(phi=-h[-1],
theta = numeric(0),Trunc.Series = 170)

hinv = as.numeric(hinv)

yest = convolve(hinv,rev(yest.s),type='o')[1:length(yest.s)]

yest[c(1:3)] = mean(yest)

par(mfrow=c(1,1))
t = seq(1,length(lPb))
plot(t,yest-mean(yest)+mean(lPb),type='l',
```

```
col='darkgray',lwd=2,ylab='log-price energy')
points(t,lPb,
col = 'black',pch=19,cex=0.5)
lines(t,yest-mean(yest)+mean(lPb),
col = 'darkgray')

require(FinTS)

require(rugarch)

ArchTest(r1,lag=14)

spec3 = ugarchspec(
variance.model=list(model='fGARCH', garchOrder=c(2,2),submodel = 'GARCH'),
mean.model=list(armaOrder=c(0,0), include.mean=TRUE),
distribution.model='nig')

z1 = as.numeric(r1)

(fit.rpb.et = ugarchfit(data = z1, spec = spec3))


sigma.rpb.et = fit.rpb.et@fit$sigma

t = seq(1,length(sigma.rpb.et))

par(mfrow=c(1,1))
plot(t,sigma.rpb.et,type='l')
np = length(sigma.rpb.et)
fechas = as.Date(Fecha[-c(1:5)],format='%d/%m/%Y')
ejex.mes = seq(fechas[1],fechas[np], 'months')
ejex.ano = seq(fechas[1],fechas[np],'years')
plot(fechas,sigma.rpb.et, xaxt='n', panel.first = grid(),type='l'
,ylab='daily volatility of price', xlab='date')
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)
abline(v=as.Date('2002-10-01'))
abline(v=as.Date('2006-09-01'))
abline(v=as.Date('2009-09-01'))
abline(v=as.Date('2010-06-01'))
```

```
H = data.frame(dates=dates,sigma.Pb = sigma.rpb.et)

require(MASS)
write.matrix(H,'volat.precio.dat',sep='  ')


## rs.test calculates the statistic of the modified R/S test
##
## x: time series
## q: number of lags included for calculation of covariances
##
## significance level: 0.05,     0.1
## critical value:     1.747,    1.62
##
## References: Lo (1991), Long-term Memory in Stock Market Prices,
## Econometrica 59, 1279--1313
## Christoph Helwig Christoph.Helwig at gmx.net

rs.test <- function(x, q, alpha)
{
xbar <- mean(x)
N <- length(x)
r <- max(cumsum(x-xbar)) - min(cumsum(x-xbar))
kovarianzen <- NULL
for (i in 1:q)
{
kovarianzen <- c(kovarianzen,
sum((x[1:(N-i)]-xbar)*(x[(1+i):N]-xbar)))
}
if (q > 0)
s <- sum((x-xbar)^2)/N + sum((1-(1:q)/(q+1))*kovarianzen)*2/N
else
s <- sum((x-xbar)^2)/N
rs <- r/(sqrt(s)*sqrt(N))
method <- 'R/S Test for Long Memory'
names(rs) <- 'R/S Statistic'
names(q) <- 'Bandwidth q'
structure(list(statistic = rs, parameter = q, method = method,
data.name=deparse(substitute(x))), class='htest')
```

```
}

## vs.test calculates the statistic of the modified V/S test
##
## x: time series
## q: number of lags included for calculation of covariances
##
## significance level: 0.01,   0.05,     0.1
## critical value:     0.2685, 0.1869,   0.1518
##
## References: Giraitis, Kokoszka und Leipus (2000), Rescaled variance
## and related tests for long memory in volatility and levels
##
vs.test <- function(x, q, alpha)
{
xbar <- mean(x)
N <- length(x)
v <- sum((cumsum(x-xbar))^2) - (sum(cumsum(x-xbar)))^2/N
kovarianzen <- NULL
for (i in 1:q)
{
kovarianzen <- c(kovarianzen,
sum((x[1:(N-i)]-xbar)*(x[(1+i):N]-xbar)))
}
if (q > 0)
s <- sum((x-xbar)^2)/N + sum((1-(1:q)/(q+1))*kovarianzen)*2/N
else
s <- sum((x-xbar)^2)/N
vs <- v/(s*N^2)
method <- 'V/S Test for Long Memory'
names(vs) <- 'V/S Statistic'
names(q) <- 'Bandwidth q'
structure(list(statistic = vs, parameter = q, method = method,
data.name=deparse(substitute(x))), class='htest')
}

rs.test(x=Apt, q=4, alpha=0.05)

rs.test(x=lPb, q=4, alpha=0.05)
rs.test(x=rpbs, q=4, alpha=0.05)
```

```
rs.test(x=rpb, q=4, alpha=0.05)

## hip nula Ho: d = 0
## significance level: 0.05,     0.1
## critical value:     1.747,  1.62

vs.test(x=Apt, q=4, alpha=0.05)

vs.test(x=lPb, q=4, alpha=0.05)
vs.test(x=rpbs, q=4, alpha=0.05)
vs.test(x=rpb, q=4, alpha=0.05)

## hip nula Ho: d = 0
## significance level: 0.01,   0.05,     0.1
## critical value:     0.2685, 0.1869,   0.1518

basicStats(r_Ap, ci = 0.95)
basicStats(r1, ci = 0.95)

basicStats(sigma.rAp.et, ci = 0.95)
basicStats(sigma.rpb.et, ci = 0.95)

np = length(Pbt)
dates = as.Date(Fecha,format='%d/%m/%Y')

ejex.mes = seq(dates[1],dates[np], 'months')
ejex.ano = seq(dates[1],dates[np],'years')

par(mfrow=c(2,1))
plot(dates,Apt, xaxt='n', panel.first = grid()
,type='l',ylab='Water Inflows')
lines(dates,sigma.rAp.et,lty=1,col='red')
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)

plot(dates,Pbt/100, xaxt='n', panel.first = grid()
,type='l',ylab='Spot Prices', ylim=c(0,3))
lines(dates,sigma.rpb.et*2,lty=1,col='red')
axis.Date(1, at=ejex.mes, format='%m/%y')
axis.Date(1, at=ejex.ano, labels = FALSE, tcl = -0.2)
```

```
r1ap = r_Ap/sigma.rAp.et
r1ap=r1ap[-c(1:5)]

r1pb = r2/sigma.rpb.et

par(mfrow=c(3,2))
t = seq(1,length(r1ap))
plot(t,r1ap,type='l',ylab='residuo')
abline(h=0,lty=2)
plot(density(r1ap),xlab='x',main= '')
acf(r1ap,ci.type= 'ma',60,main='')
pacf(r1ap,60,main='')
qqnorm(r1ap)
qqline(r1ap,col=2)

windows()

par(mfrow=c(3,2))
t = seq(1,length(r1pb))
plot(t,r1pb,type='l',ylab='residuo')
abline(h=0,lty=2)
plot(density(r1pb),xlab='x',main= '')
acf(r1pb,ci.type= 'ma',60,main='')
pacf(r1pb,60,main='')
qqnorm(r1pb)
qqline(r1pb,col=2)

cpgram(r1ap,main='   ')
cpgram(r1pb,main='   ')

Box.test(x = r1ap, lag = 14, type='Ljung-Box')
Box.test(x = r1pb, lag = 14, type='Ljung-Box')

Da = read.table('volat.aportes.dat', header=TRUE, stringsAsFactors=FALSE)

Db = read.table('volat.precio.dat', header=TRUE, stringsAsFactors=FALSE)

fechas = Db$fechas
```

```
sigma.Pb = log(Db$sigma.Pb^2)[-c(1:5)]
sigma.Ap = log(Da$sigma.Ap^2)[-c(1:5)]

length(sigma.Ap)
length(sigma.Pb)

par(mfrow=c(2,1))
t = seq(1,length(sigma.Pb))
plot(t,sigma.Ap,type='l',ylab='daily Water Inflows volatility', xlab='date',main='(a)'
plot(t,sigma.Pb,type='l',ylab='daily log-Price volatility', xlab='date',main='(b)')

y = sigma.Pb
x = sigma.Ap

x = ts(x,frequency=180)
y = ts(y,frequency=7)

require(e1071)

FinTS.stats(x)
FinTS.stats(y)


ndiffs(x=x,test='pp',max.d=3)

ndiffs(x=x,test='adf',max.d=3)

ndiffs(x=x,test='kpss',max.d=3)

ndiffs(x=y,test='pp',max.d=3)

ndiffs(x=y,test='adf',max.d=3)

ndiffs(x=y,test='kpss',max.d=3)

nsdiffs(x)
nsdiffs(y)

auto.arima(x, stationary=TRUE, seasonal=FALSE)
auto.arima(y)
```

```
res.rpb=armasubsets(y=x,
nar=14,nma=14,
y.name='r',
ar.method='ols')

par(mfrow=c(1,1))
plot(res.rpb)

res.rpb=armasubsets(y=y,
nar=14,nma=14,
y.name='r',
ar.method='ols')

par(mfrow=c(1,1))
plot(res.rpb)

x = ts(x,frequency=7)

px = 2; qx = 3;
mx = arima(x,order=c(px,0,qx))
coeftest(mx)

library(portes)

B = ImpulseVMA(phi=numeric(0),
theta = mx$coef[1:px],Trunc.Series = 25)

B = as.numeric(B)

Wn = filter(x=y, filter=B,'conv', sides = 1, circular = TRUE)

Un = filter(x=x, filter=B,'conv', sides = 1, circular = TRUE)

length(Wn)
length(Un)

par(mfrow=c(1,1))

XY.n=ts.intersect(Wn, Un)
```

```
ccf(as.numeric(XY.n[,1]),as.numeric(XY.n[,2]),
main='',ylab='CCF')
abline(v=0)

(mtr2 = arimax(y, order = c(2,0,1),
xtransf = data.frame(x=lag(x,-4)),
transfer = list(c(0,4)),
method='ML'))

coeftest(mtr2)
```

B.2. Setting a ADL-Koyck model

```
x = sigma.rap.et
y = sigma.rpb.et

Ds = data.frame(x=x)
m1 = arimax(x = y, order = c(0,0,1),
include.mean = TRUE, xtransf = Ds,
transfer = list(c(1,0,0)))

m1$coef

r = m1$residuals[-c(1:5)]
t = seq(1,length(r))

par(mfrow=c(3,2))
plot(t,r,type="l",ylab="residuo")
abline(h=0,lty=2)
plot(density(r),xlab="x",main= "")
acf(r,60,main="")
pacf(r,60,main="")
qqnorm(r)
qqline(r,col=2)
Box.test(x = r, lag = 12, type="Ljung-Box")

lambda = -0.0002892297
mu = 9.4510943216
alpha = 0.0062267536
sigma = 0.8370285898
```

```
n = length(y)
y_asterisco = double(n)
et = m1$residuals
y[1] = x[1]
for(j in 2:n){
y_asterisco[j] = (1-lambda)*mu + lambda*y[j-1] + alpha*x[j] + et[j]-lambda*et[j-1]}

y = sigma.rpb.et

windows()
t = seq(1,length(x))
par(mfrow=c(1,1))
plot(t,y,type="l",col="blue")
lines(t,y_asterisco,col="red")

Ds = data.frame(x=x)
(m3 = arimax(x = y_asterisco, order = c(5,1,0),
include.mean = TRUE, xtransf = Ds,
transfer = list(c(1,0,0))))

r3 = m3$residuals[-c(1:5)]
t3 = seq(1,length(r3))
par(mfrow=c(3,2))
plot(t3,r3,type="o",ylab="residuo")
abline(h=0,lty=2)
plot(density(r3),xlab="x",main= "")
acf(r3,30,main="")
pacf(r3,30,main="")
qqnorm(r3)
qqline(r3,col=2)
Box.test(x = r3, lag = 12, type="Ljung-Box")
```

B.3. Accurate Simulation of ADL-Gamma

```
fk = function(k){ (1+k)^(delta/(1-delta))*lambda^k}

n = 500
```

```
x=arima.sim(n = n, list(ar=c(0.8897, -0.4858), ma=c(-0.2279, 0.2488)),
rand.gen = function(n, ...) sqrt(0.1796) * rt(n, df = 5))

delta = 0.3
lambda = 0.5
mu = 0
alpha = 2.5

B = fk(c(0:(n-1)))

et = rnorm(n,0,sd=1)

y = mu+alpha*filter(x=rev(x), filter=B, sides = 1, circular = TRUE)+et

t = seq(1,length(x))
par(mfrow=c(1,1))
plot(t,y,type="l",col="gray")
lines(t,x,col="red")
```

B.4. Parameter estimation for Gamma transfer function

```
mat_alpha=function(equis, ye, delta, lambda)
{
m.alpha = matrix(0,nrow=length(delta),ncol=length(lambda))
suma = rep(0,length(equis))
for(d in 1:length(delta))
{
for(b in 1:length(lambda))
{
for (i in 0:(length(equis)-1))
suma[i+1] = (i+1)^(delta[d]/(1-delta[d]))*lambda[b]^i
Y.asterisco=filter(x=rev(equis), filter=suma, sides = 1, circular = TRUE)
alpha=sum(ye)/sum(Y.asterisco)
m.alpha[d,b]=alpha
}
}
colnames(m.alpha, do.NULL = FALSE)
```

```
colnames(m.alpha) = colnames(m.alpha, do.NULL = FALSE, prefix = "lambda.")
rownames(m.alpha) = rownames(m.alpha, do.NULL = FALSE, prefix = "delta.")
m.alpha
}
```

C. Appendix: Parameter estimation methodology for Gamma transfer function

To simulate an exactly ADL-Gamma we run the R code showed in Appendix B.3. Here were simulated 500 data such that $X_t \sim ARMA(2,2)$. Additionally, we define the parameters $\delta = 0{,}3, \lambda = 0{,}5, \mu = 0, \alpha = 2{,}5$

Then, we generate the figure **C-1** from the R code of Appendix B.3. The red line is the data from the input variable x_t and the blue line is the data of the output variable y_t, according to the model 4-24.

From the values x_t and y_t simulated, we estimate the parameters $\hat{\alpha}$, $\hat{\delta}$ y $\hat{\lambda}$ to test the methodology that we propose.

Considering that $\sum_{t=1}^{t=T}(y_t - \alpha y_t^*)^2 = 0$, then:

$$\sum_{t=1}^{t=T} y_t^2 - 2\alpha y_t y_t^* + \alpha^2 y_t^* = 0 \tag{C-1}$$

It can be shown easily that the solution of equation C-1 is $\alpha = \frac{y_t}{y_t^*}$, where y_t are the actual output values of the transfer function and y_t^* is defined as

$$y_t^* = \sum_{j=0}^{\infty}(j+1)^{\delta/(1-\delta)}\lambda^j x_{t-j} \tag{C-2}$$

where $\delta, \lambda \in (0,1), \mu, \alpha \in \mathbb{R}, (e_t, t \in \mathbb{Z})$

Then, a matrix of (d x b) size is constructed, where d is the number of δ's to be evaluated in equation C-2, such that $\delta \in (0,1)$, according to a defined step size. Equivalently, b is the number of λ's, such that $\lambda \in (0,1)$ to be evaluated in equation C-2 according to a defined step size.

For example: if we choose a step size of 0.1 for δ and λ, then according to solution found for (C-1) and substituting the respective values of δ and λ in the equation C-2, the matrix is evaluated as follows:

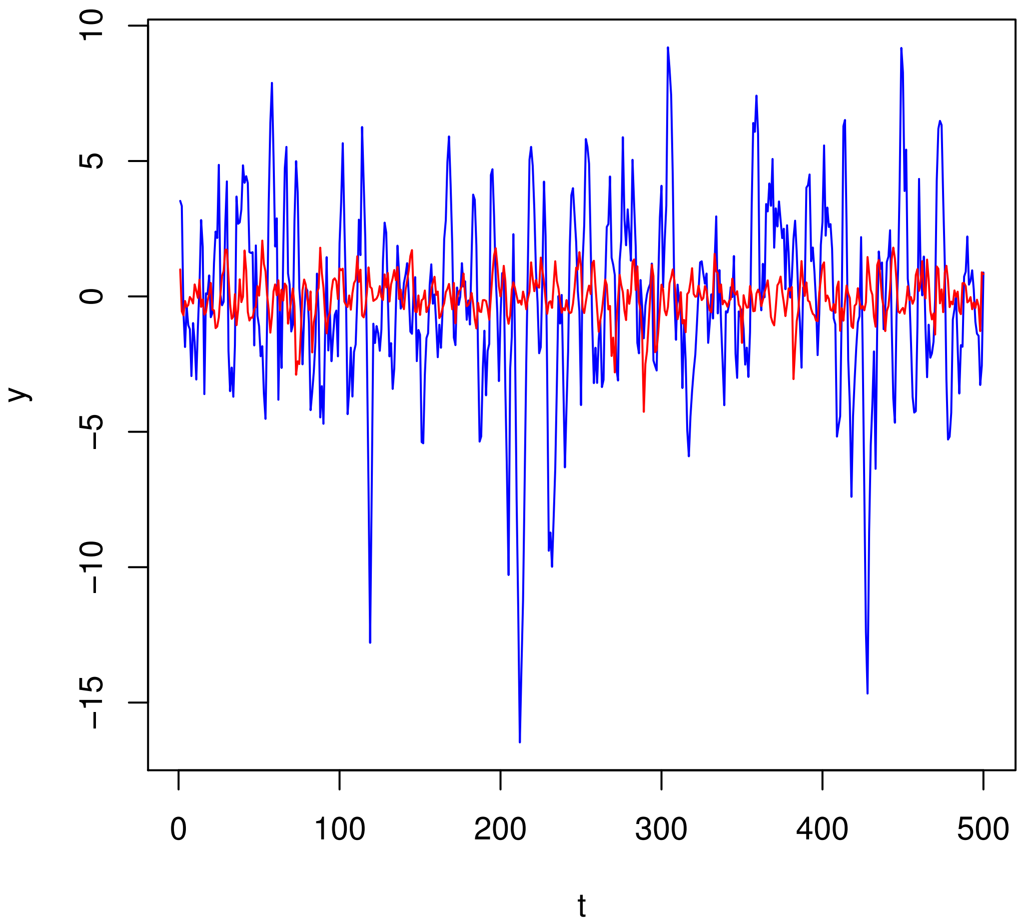


Figure C-1.: Simulation of the model 4-24 with parameters $\delta = 0{,}3, \lambda = 0{,}5, \mu = 0, \alpha = 2{,}5$

$$\begin{vmatrix} \frac{y_t}{y_t^*|_{\delta=0,\lambda=0}} & \frac{y_t}{y_t^*|_{\delta=0,\lambda=0,1}} & \cdots & \frac{y_t}{y_t^*|_{\delta=0,\lambda=0,9}} \\ \frac{y_t}{y_t^*|_{\delta=0,1,\lambda=0}} & \frac{y_t}{y_t^*|_{\delta=0,1,\lambda=0,1}} & \cdots & \frac{y_t}{y_t^*|_{\delta=0,1,\lambda=0,9}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{y_t}{y_t^*|_{\delta=0,9,\lambda=0}} & \frac{y_t}{y_t^*|_{\delta=0,9,\lambda=0,1}} & \cdots & \frac{y_t}{y_t^*|_{\delta=0,9,\lambda=0,9}} \end{vmatrix}$$

According to the above explanation, we run the R code shown in Appendix B.4 then we take the values of x_t and y_t of the previous simulation.

Then, we obtained the following matrix of values of α for different values of δ and λ with step size of 0.1.

```
     lambda.1    lambda.2     lambda.3     lambda.4     lambda.5
```

delta.1	6.048668	5.443800806	4.8389340500	4.234067e+00	3.629201e+00
delta.2	6.048668	5.397685248	4.7517270034	4.111520e+00	3.478011e+00
delta.3	6.048668	5.335509186	4.6356241635	3.950729e+00	3.283013e+00
delta.4	6.048668	5.247672279	4.4745599543	3.732276e+00	3.024468e+00
delta.5	6.048668	5.115626785	4.2390957712	3.422872e+00	2.671433e+00
delta.6	6.048668	4.899420726	3.8711472400	2.963847e+00	2.177520e+00
delta.7	6.048668	4.500194245	3.2490229550	2.257280e+00	1.490785e+00
delta.8	6.048668	3.634564720	2.1389393815	1.213316e+00	6.505585e-01
delta.9	6.048668	1.615412804	0.5433189109	1.911980e-01	6.510858e-02
delta.10	6.048668	0.006982882	0.0003887481	3.200256e-05	2.781568e-06

	lambda.6	lambda.7	lambda.8	lambda.9	lambda.10
delta.1	3.024334e+00	2.419467e+00	1.814600e+00	1.209734e+00	6.048668e-01
delta.2	2.852476e+00	2.236728e+00	1.633525e+00	1.047570e+00	4.887120e-01
delta.3	2.635365e+00	2.011777e+00	1.418127e+00	8.639971e-01	3.682070e-01
delta.4	2.355788e+00	1.732404e+00	1.162964e+00	6.606992e-01	2.493711e-01
delta.5	1.990181e+00	1.385833e+00	8.671190e-01	4.461977e-01	1.423008e-01
delta.6	1.512167e+00	9.677868e-01	5.443801e-01	2.419467e-01	6.048668e-02
delta.7	9.183728e-01	5.112567e-01	2.423841e-01	8.566116e-02	1.475679e-02
delta.8	3.207566e-01	1.391771e-01	4.904021e-02	1.173819e-02	1.082441e-03
delta.9	2.016223e-02	5.259711e-03	1.018379e-03	1.115476e-04	2.944967e-06
delta.10	2.133641e-07	1.209970e-08	3.887951e-10	4.081597e-12	2.528821e-15

Taking as reference the values of the above matrix, the values of δ and λ are searched that minimize the difference with the real y_t value.

$$E_c = \sum_{t=0}^{T} (Y - \alpha \sum_{j=0}^{\infty} (j+1)^{\delta/(1-\delta)} \lambda_{t-j}^{j})^2 \tag{C-3}$$

Where,
E_c = squared error difference between the actual response and the estimated.

Y = real value of the response series.

α = possible value of the parameter α found in the previous step of this procedure.

δ = possible value of the parameter δ.

λ = possible value of the parameter λ.

According to the above matrix of combinations of values of δ and λ, the matrix of the possible combinations of α is calculated, and it is shown in the following matrix

	lambda.1	lambda.2	lambda.3	lambda.4	lambda.5	lambda.6	lambda.7
delta.1	2793.394	2220.9061	1695.2098	1227.3468	842.3478	583.1465	515.2630
delta.2	2793.394	2180.2647	1627.8605	1148.8101	771.6735	544.1616	536.0733
delta.3	2793.394	2126.2217	1541.0241	1052.0622	692.0263	513.7095	589.7559
delta.4	2793.394	2051.3686	1426.0290	932.5004	607.5859	508.0826	704.7814
delta.5	2793.394	1942.1931	1269.6948	787.8158	534.5975	565.1369	935.6143
delta.6	2793.394	1772.3258	1054.6184	630.7786	525.8447	773.3436	1384.0449
delta.7	2793.394	1488.5999	777.8200	543.7005	742.5749	1330.5856	2215.4745
delta.8	2793.394	1013.5422	597.8708	885.1438	1615.8763	2583.1171	3566.4338
delta.9	2793.394	759.0987	1533.2749	2685.3753	3793.9130	4605.7420	5056.2018
delta.10	2793.394	4127.8433	5679.6288	5978.5824	5830.5919	5689.0276	5304.3959

	lambda.8	lambda.9	lambda.10
delta.1	731.7250	1359.499	2552.115
delta.2	840.2731	1570.044	2820.200
delta.3	1007.4279	1853.359	3136.415
delta.4	1267.9130	2236.087	3499.296
delta.5	1676.6358	2747.345	3892.260
delta.6	2310.5646	3400.663	4269.976
delta.7	3235.9757	4142.427	4550.212
delta.8	4362.6893	4760.466	4654.311
delta.9	5167.5442	4868.450	4673.529
delta.10	4720.2810	4752.224	4734.264

From the above matrix, we note that the lower value is located at position $4x6$ equivalent to values of $\hat{\delta} = 0{,}3$ and $\hat{\lambda} = 0{,}5$. These values are consistent with the values simulated.

Looking for the same equivalent position to α in the first matrix, it is noted that $\hat{\alpha} = 2{,}355788 \approx 2{,}36$. An error of 1.047022 is calculated, which is considered low.

In the figure **C-2** we compare the original output series with the output series with the estimated parameters. It shows that the proposed methodology does a good estimation of the parameters for the Gamma transfer function.

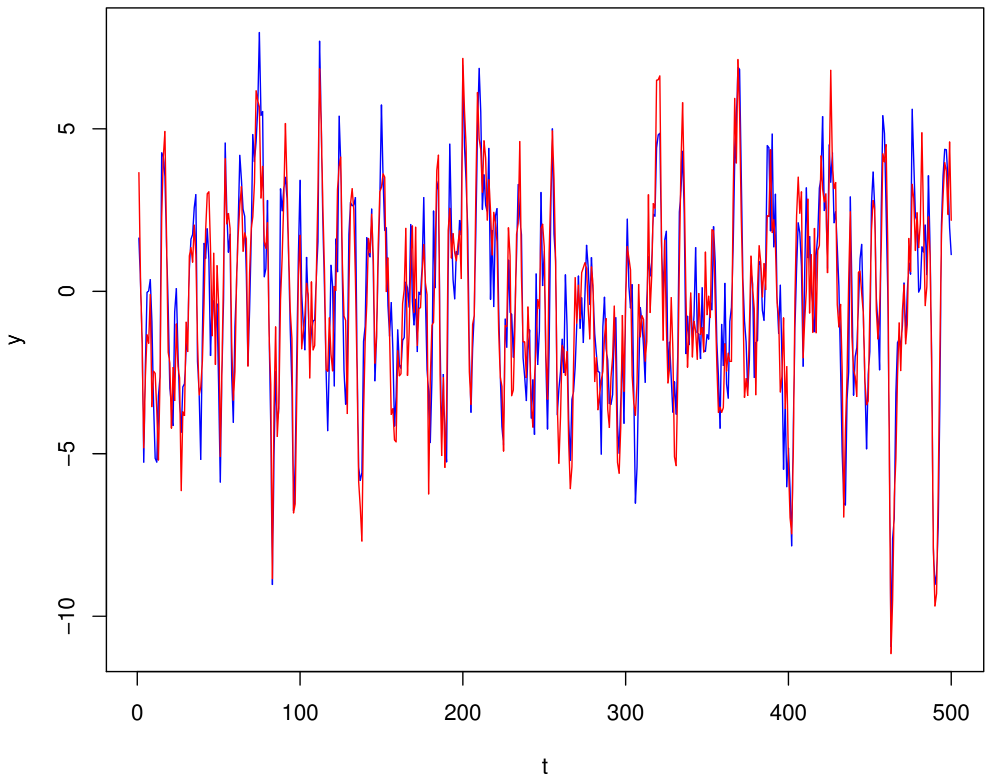


Figure C-2.: Graphical comparison of the series of original output (blue) and the output series with the estimated parameters.

Bibliography

[1] Barndorff Nielsen, O. ; Blaesild, P.: *Encyclopedia of Statistical Sciences*. Vol. 3. Hyperbolic distributions : Johnson N.L., Kotz S. and Read C.B., Eds. John Wiley and Sons, 1983

[2] Bera, Anil K. ; Higgins, Matthew L.: ARCH models: Properties, Estimation and Testing. *Journal of Economic Surveys* 7 (1993), Nr. 4, P. 305–366

[3] Botero, Sergio B. ; Cano, Jovan Alfonso C.: Análisis de series de tiempo para la predicción de los precios de la energía en la Bolsa de Colombia. *Cuadernos de Economía* 27 (2008), P. 173–208

[4] Box, George Edward P. ; Jenkins, Gwilym M.: *Time series analysis: forecasting and control*. Holden-Day, 1976

[5] Box, George Edward P. ; Jenkins, Gwilym M.: *Time series analysis: forecasting and control*. 3rd. Upper Saddle River, NJ, USA : Prentice Hall PTR, 1994

[6] Castaño, Elkin ; Sierra, Jorge: Sobre la existencia de una raíz unitaria en la serie de tiempo mensual del precio de la electricidad en Colombia. *Lecturas de Economía, Departamento de Economía, Universidad de Antioquia* 76 (2012), P. 259–291

[7] Chan, Kung-Sik ; Ripley, Brian: *TSA: Time Series Analysis, R package version 1.01*, 2014. `http://CRAN.R-project.org/package=TSA`

[8] Contreras-Reyes, J. E. ; Goerg, G. M. ; Palma, W.: *afmtools: Estimation, Diagnostic and Forecasting Functions for ARFIMA models. R package version 0.1.7*. `http://cran.r-project.org/src/contrib/Archive/afmtools/`. Version: 2012

[9] Forbes, Kristin J. ; Rigobon, Roberto: No Contagion, Only Interdependence: Measuring Stock Market Comovements. *The Journal of Finance* 57 (2002), Nr. 5, P. 2223–2261

[10] Franco, Carlos J. ; Velásquez, Juan D. ; Cardona, David: Micromundo para simular un mercado eléctrico de corto plazo. *Cuadernos de Economía* XXXI (2012), 12, P. 229–256

[11] Francq, C. ; Zakoian, J.M.: *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, 2011

[12] Ghalanos, A.: *rugarch: Univariate GARCH models. R package version 1.3-3.* `http://CRAN.R-project.org/package=forecast`. Version: 2014

[13] Gil Zapata, Martha M. ; Maya Ochoa, Cecilia: Modelación de la volatilidad de los precios de la energía eléctrica en Colombia. *Revista Ingenierías Universidad de Medellín* 7 (2008), P. 87–114

[14] Giraitis, Liudas ; Kokoszka, Piotr ; Leipus, Remigijus ; Teyssiere, Gilles: Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* 126 (2005), Nr. 2, P. 571–572

[15] Gray, D. ; Malone, S.: *Macrofinancial Risk Analysis*. Wiley, 2008 (The Wiley Finance Series). – ISBN 9780470756324

[16] Helwig, Ch.: *Timeseries: R/S (rescaled range) analysis.* `https://stat.ethz.ch/pipermail/r-help/2001-June/013425.html`. Version: jun 2001

[17] Hipel, K.W. ; McLeod, A.I.: *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier Science, 1994 (Developments in Water Science)

[18] Holan, Scott H. ; Lund, Robert ; Davis, Ginger: The ARMA alphabet soup: A tour of ARMA model variants. *Statistics Surveys* 4 (2010), P. 232–274

[19] Hyndman, Rob J. ; Khandakar, Yeasmin: Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27 (2008), Nr. 3, P. 1–22

[20] Judge, George G. ; Griffiths, William E. ; Hill, Carter ; Lutkepohl, Helmut ; Lee, Tsoung-Chao: *The Theory and Practice of Econometrics, 2nd Edition*. United States : Wiley Series, 1985

[21] Jung, R.C. ; Maderitsch, R.: Structural Breaks in Volatility Spillovers between International Financial Markets: Contagion or Mere Interdependence? *Journal of Banking and Finance* 47 (2014), P. 331–342

[22] Katayama, N.: *S-Plus functions for SARFIMA estimation using CSS.* `http://www2.ipcku.kansai-u.ac.jp/simkatayama/en splus.html`. Version: Dic 2007

[23] Katayama, N.: *Seasonally and Fractionally Differenced Time Series.* `http://www2.ipcku.kansai-u.ac.jp/katayama/en splus.html`. Version: 2007

[24] Koch, Paul D. ; Yang, Shie-Shien: A Method for Testing the Independence of Two Time Series That Accounts for a Potential Pattern in the Cross-Correlation Function. *Journal of the American Statistical Association* (1986), Nr. 394, P. 533–544

[25] Koyck, L.M: *Distributed Lags and Investment Analysis*. Amsterdam : North-Holland Pub., 1954

[26] Lindner, Alexander M.: Stationarity, mixing, distributional properties and moments of GARCH(p, q). : *Handbook of Financial Time Series*, Springer, 2008

[27] Lo, A.: Long-term Memory in Stock Market Prices. *Econometrica* 59 (2001), P. 1279–1313

[28] Lopes, S ; Olbermann, B ; Reisen, V: Non stationary Gaussian ARFIMA processes: Estimation and application. *Brazilian Review of Econometrics* 22 (2002), Nr. 1, P. 103–126

[29] Lütkepohl, Helmut: A Model for Non-Negative and Non-Positive Distributed Lag Functions. *Journal of Econometrics* 16 (1981), Nr. 2, P. 211–219

[30] Mahdi, Esam ; McLeod, A. I.: Improved multivariate portmanteau diagnostic test. *Journal of Time Series Analysis* 33 (2012), Nr. 2, P. 211–222

[31] Osborn, Denise R. ; Chui, A. P. L. ; Smith, Jeremy P. ; Birchenhall, C. R.: Seasonality and the order of integration for consumption. *Oxford Bulletin of Economics and Statistics* 50 (2008), Nr. 4, P. 361–377

[32] Peckham, R: Economies of Contagion: Financial Crisis and Pandemic. *Economy and Society* 42 (2013), Nr. 2, P. 226–248

[33] Pesaran, M.H. ; Pick, A.: Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control* 31 (2007), P. 1245–1277

[34] Poveda, Germán: Caudales mínimos en Colombia : relaciones macroclimáticas, escalamiento y balances hidrológicos. *Seminario Internacional sobre Eventos Extremos Mínimos en Regímenes de Caudales: Diagnóstico, Modelamiento y Análisis* 23-25 de junio (2004)

[35] Poveda, Germán ; Gil, Martha M. ; Quiceno, Natalia: Ciclo Anual de la Hidrología de Colombia en Relación con el ENSO y la NAO. *Bulletin de l'Institut Francais d'Études Andines* 27 (1998), Nr. 3, P. 721–731

[36] Puertas, O ; Carvajal, Y: Incidencia de El Niño-Oscilación del Sur en la precipitación y la temperatura del aire en Colombia, utilizando el Climate Explorer. *Ingeniería y Desarrollo. Universidad del Valle* (2008), Nr. 23, P. 104–118

[37] Quintero Quintero, María del C. ; Isaza Cuervo, Felipe: Dependencia Hidrológica y Regulatoria en la Formación de Precio de la Energía en un Sistema Hidrodominado: Caso Sistema Eléctrico Colombiano. *Revista Ingenierías Universidad de Medellín* 12 (2013), 06, P. 85 – 96

[38] Ramírez, V ; Jaramillo, A: Relación entre el índice oceánico de El Niño y la lluvia, en la región andina central de Colombia. *enicafé* 60 (2009), Nr. 2, P. 161–172

[39] Schmidt, P.: An Argument for the Usefulness of the Gamma Distributed Lag Model. *International Economic Review* 15 (1974), P. 246–250

[40] Siem Jan, Koopman ; Marius, Ooms ; M. Angeles, Carnero: Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices. *Journal of the American Statistical Association* 102 (2007), Nr. 477, P. 16–27

[41] Soofi, A.S. ; Cao, L.: *Modelling and Forecasting Financial Data: Techniques of Non-linear Dynamics*. Springer US, 2002

[42] Tayefi, Maryam ; T.V., Ramanathan: An Overview of FIGARCH and Related Time Series Models. *Astralian Journal Of Statistics* 41 (2012), Nr. 3, P. 175–196

[43] Theil, H. ; Fiebig, D.: A Maximum Entropy Approach to the Specification of Distributed Lags. *Economics Letters* 7 (1981), Nr. 4, P. 339–342

[44] Tsurumi, H.: A Note on Gamma Distributed Lags. *International Economic Review* 12, Nr. 2

[45] Uribe, E. ; Trespalacios, A: Constraste de modelos estocásticos para el precio de la energía en Colombia. *Revista Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín* 3 (2014), Nr. 1, enero junio, P. 41–55

[46] Wei, William W.: *Time Series Analysis. Univeriate and Multivariate Methods. 2nd Ed.* United States : Pearson, Addison Wesley, 2006

[47] Wuertz, Diethelm ; others many: *fUnitRoots: Trends and Unit Roots, R package version 3010.78*, 2013. `http://CRAN.R-project.org/package=fUnitRoots`

[48] Y, Ye: *Interior algorithms for linear, quadratic, and linearly constrained non linear programming*. Kassel, Department of EES Stanford University, PhD Thesis, 1989

[49] Zachmann, G ; Hirschhausen, H C: First evidence of asymmetric cost pass-through of EU emissions allowances: Examining wholesale electricity prices in Germany. *Economics Letters* 99 (2008), Nr. 3, P. 465–469