Universidade de Aveiro Departamento de Matemática

2014

Maria da Conceição Cristo Santos Lopes Costa

Sistemas de Alarme Ótimos e sua Aplicação a Séries Financeiras

Optimal Alarms Systems and its Application to Financial Time Series

Universidade de Aveiro Departamento de Matemática
2014

Maria da Conceição Cristo Santos Lopes Costa

Sistemas de Alarme Ótimos e sua Aplicação a Séries Financeiras

Optimal Alarms Systems and its Application to Financial Time Series

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Matemática, realizada sob a orientação científica de Manuel González Scotto, Professor Auxiliar c/ Agregação do Departamento de Matemática da Universidade de Aveiro e de Isabel Maria Simões Pereira, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro.

Apoio financeiro do FEDER através do COMPETE - Programa Operacional Fatores de Competitividade e de fundos nacionais através do Centro de Investigação e Desenvolvimento em Matemática e Aplicações (Universidade de Aveiro) e da FCT - Fundação para a Ciência e a Tecnologia, no âmbito do projeto PEst-C/MAT/UI4106/2011 com o número COMPETE FCOMP-01-0124-FEDER-022690, até 2013. Em 2014 o apoio financeiro foi garantido por fundos nacionais, no âmbito do projeto PEst-OE/MAT/UI4106/2014.

Ao meu pai, por tudo o que foi.

Ao Rui e à Beatriz, por tudo o que ainda será.

O júri / The jury

Presidente / President

Doutor Vitor Brás de Sequeira Amaral,

Professor Catedrático da Universidade de Aveiro.

Doutor Paulo Eduardo Aragão Aleixo Neves de Oliveira,

Professor Catedrático da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Doutor António Manuel Pacheco Pires,

Professor Catedrático do Instituto Superior Técnico da Universidade Técnica de Lisboa.

Doutor João Carlos Henriques da Costa Nicolau,

Professor Associado com Agregação do Instituto Superior de Economia e Gestão da Universidade de Lisboa.

Doutora Maria Eduarda Rocha Pinto Augusto da Silva,

Professora Associada da Faculdade de Economia da Universidade do Porto.

Doutor Manuel González Scotto,

Professor Auxiliar com Agregação da Universidade de Aveiro. (Orientador)

Doutora Maria Manuela Souto de Miranda,

Professora Auxiliar Aposentada da Universidade de Aveiro.

Agradecimentos

Aos meus orientadores, Professor Manuel Scotto e Professora Isabel Pereira pela capacidade de transformarem as minhas grandes dúvidas em coisas pequenas e, o meu esforço, em algo de valor. Para além da sua orientação científica, o seu apoio e motivação constantes e a total disponibilidade a cada momento do percurso foram determinantes para a conclusão deste trabalho.
À Universidade de Aveiro, em particular ao Departamento de Matemática e ao Centro de Investigação e Desenvolvimento em Matemática e Aplicações por todos os apoios concedidos para o desenvolvimento deste trabalho. Um especial agradecimento ao Professor João Santos pelo empenho pessoal em garantir todas as condições necessárias à concretização deste objectivo.
Aos meus colegas do Departamento de Matemática, por me fazerem sentir em casa. Para além da confiança que me foi sendo transmitida por todas as pessoas com quem trabalhei directamente, tenho a agradecer pessoalmente a Paolo Vettori, Pedro Macedo, Ricardo Almeida e Rita Simões, pela amizade e ajuda incondicionais.
Ao meu marido e à minha filha, que tornaram tudo isto possível e estiveram sempre presentes, mesmo quando eu não estive.

Palavras-chave Autocorrelação, Ergodicidade, Estacionaridade, Máxima Verosimilhança, Heterocedasticidade, Inferência Bayesiana, Memória Longa, Modelos *Observation-driven*, Séries Temporais Não Lineares, Sistemas de Alarme Ótimos, Sobredispersão, Teoria Assintótica, Volatilidade Assimétrica.

Resumo Esta tese centra-se na aplicação de sistemas de alarme ótimos a modelos de séries temporais não lineares. As classes de modelos mais comuns na análise de séries temporais de valores reais e de valores inteiros são descritas com alguma profundidade. É abordada a construção de sistemas de alarme ótimos e as suas aplicações são exploradas.

De entre os modelos com heterocedasticidade condicional é dada especial atenção ao modelo ARCH Fraccionalmente Integrável de Potência Assimétrica, FIAPARCH(p, d, q), e é feita a implementação de um sistema de alarme ótimo, considerando ambas as metodologias clássica e Bayesiana.

Tomando em consideração as características particulares do modelo APARCH(p, q) na aplicação a séries de dados financeiros, é proposta a introdução do seu homólogo para a modelação de séries temporais de contagens: o modelo ARCH de valores INteiros e Potência Assimétrica, INAPARCH(p, q). As propriedades probabilísticas do modelo INAPARCH$(1, 1)$ são extensivamente estudadas, é aplicado o método da máxima verosimilhança (MV) condicional para a estimação dos parâmetros do modelo e estudadas as propriedades assintóticas do estimador de MV condicional. Na parte final do trabalho é feita a implementação de um sistema de alarme ótimo ao modelo INAPARCH$(1, 1)$ e apresenta-se uma aplicação a séries de dados reais.

Keywords Asymmetric Volatility, Asymptotic Theory, Autocorrelation, Bayesian Inference, Ergodicity, Heteroscedasticity, Long Memory, Maximum Likelihood, Observation-driven Models, Overdispersion, Optimal Alarm Systems, Non Linear Time Series, Stationarity.

Abstract This thesis focuses on the application of optimal alarm systems to non linear time series models. The most common classes of models in the analysis of real-valued and integer-valued time series are described. The construction of optimal alarm systems is covered and its applications explored.

Considering models with conditional heteroscedasticity, particular attention is given to the Fractionally Integrated Asymmetric Power ARCH, FIAPARCH(p, d, q) model and an optimal alarm system is implemented, following both classical and Bayesian methodologies.

Taking into consideration the particular characteristics of the APARCH(p, q) representation for financial time series, the introduction of a possible counterpart for modelling time series of counts is proposed: the INteger-valued Asymmetric Power ARCH, INAPARCH(p, q). The probabilistic properties of the INAPARCH$(1, 1)$ model are comprehensively studied, the conditional maximum likelihood (ML) estimation method is applied and the asymptotic properties of the conditional ML estimator are obtained. The final part of the work consists on the implementation of an optimal alarm system to the INAPARCH$(1, 1)$ model. An application is presented to real data series.

. . . if a man riding in an open country should see afar off men and women dancing together,
and should not hear the music according to which they dance and tread out their measures,
he would think them to be fools and madmen,
because they appear in such various motions, and antic gestures and postures.
But if he come nearer,
so as to hear the musical notes, according to which they dance,
and observe the regularity of the exercise,
he will change his opinion of them, . . .

Thomas Manton, 1873

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This introductory chapter has three main sections: *Objectives*, *Overview* and *Organization of the Dissertation*. In the first section we present the objectives of the thesis and clearly state what was expected to achieve with the work developed so far.

Overview section covers the fields of study and presents main subjects related to the contributions in this thesis. In the first part of this section some non-linear models used in the analysis of financial time series are described. Attention is given to the class of models with conditional heteroscedasticity and their properties discussed. The Fractionally Integrated Asymmetric Power ARCH, FIAPARCH(p, d, q) model is discussed in detail. The following sub-section focuses on the implementation of alarm systems. The principles for the construction of optimal alarm systems are stated and discussed and their implementation described. Also, recent alternative approaches regarding level-crossing prediction are presented in order to complement and widen the discussion related to this field. The last part of the *Overview* section is directed to the modelling and analysis of count data. This is an area of growing interest as time series of counts have become available over the last three decades, in a wide variety of contexts. Several model classes developed for the analysis of time series of counts are discussed.

The last section in this chapter, *Organization of the Dissertation*, provides a brief description of the remaining chapters.

1.1 Objectives

A major theme in the analysis of a large variety of random phenomena consists in detecting and warning the occurrence of a catastrophe or some other event connected with an alarm mechanism. An interesting problem, which has not been addressed until recently, is the development of optimal alarm systems for financial time series. This thesis aims to give a contribution towards this direction.

The construction of alarm systems is discussed and optimal alarm systems are implemented for two particular models related to the analysis of financial time series. The first model, considered in Chapter 2, is the Fractionally Integrated Asymmetric Power ARCH, FIAPARCH(p, d, q) model. This model considers a representation for the volatility that has many interesting properties, from which we would like to point out the ability to deal with asymmetric responses of volatility to positive or negative shocks, and the ability to fit from the data the power of the returns for which the dependence on past returns is the strongest.

Asymmetric responses of the volatility for positive or negative shocks have also been observed in time series of counts, and to our knowledge, there was no model for count time series able to deal with this feature. Another goal was set for this work as we proposed to provide an integer-valued counterpart being able to accommodate asymmetric responses relative to the mean of the process. In Chapter 3, a new model is therefore proposed, the INteger-valued

Asymmetric Power ARCH or, in short, INAPARCH(p, q). The probabilistic properties and asymptotic theory related to maximum likelihood estimation are comprehensively studied for the INAPARCH$(1, 1)$ particular case.

To complement the perspective on the application of alarm systems to non-linear time series, the implementation of an optimal alarm system is put through for the INAPARCH$(1, 1)$ model in Chapter 3. The initial goal of this thesis is also addressed in the applications to real data time series in Chapter 2 and Chapter 3, which intend to illustrate the methodologies involved and demonstrate the prediction capability of optimal alarm systems in practice.

1.2 Overview

The introductory chapter intends to allow the thesis to be consistent and self contained. In order to pursue this goal, different areas involved in this work will be addressed in this chapter, accordingly to the following steps, whenever possible. Each subsection will start with a description of the subject and illustrative examples and applications. Then, most important definitions will be presented. Different models, classes of models or methodologies involved, will be presented roughly in the order in which they were first introduced in the literature. With the exception of the *Optimal Alarm Systems* subsection, both *Financial Time Series Models* and *Time Series of Counts* subsections will finalize with a thorough description of the particular class of models that relates more closely to the work developed in this thesis. *Optimal Alarm Systems* subsection will finalize with some recent developments in the field, that, in our perspective, deserve particular attention as they provide alternative perspectives and methodologies regarding the construction of alarm systems. Taking the risk of sounding a little naive, we will try to reach the state of the art regarding theoretical findings and methodologies.

1.2.1 Financial Time Series Models

In this section we will discuss some models usually related to financial time series. The analysis of financial time series has revealed some common features, generally known as *stylized facts*, and we will look into this subject with some detail in this section. In order to accommodate these common features often exhibited by the data, several models have been proposed over the last thirty years. In this section, an overview of these models will be presented. We will start by introducing the general class of multiplicative models and very briefly describe the stochastic volatility models. We will then move on to the class of conditionally heteroscedastic models, starting with the ARCH model of Engle (1982). We will follow on to some of the subsequent generalizations such as the GARCH, (Bollerslev, 1986), the APARCH (Ding et al., 1993), the FIGARCH (Baillie et al., 1996) and the FIAPARCH (Tse, 1998), to name the ones we consider the most important for the work developed in this thesis.

Financial time series are continuous or discrete time processes and time series analysis has always been directed towards the understanding of the mechanism behind the data. One main motivation for the analysis is the search for physical models that can explain, at least to some extent, the empirically observed features of real data. It might seem strange that when we mention *financial time series* and we include in this category so many diverse time series such as series of foreign exchange rates, stock indices or share prices, it should be possible to find any common properties. But, in fact, financial time series share many common properties, particularly after the following transformations. Let P_t denote the price of a financial asset - a stock, an exchange rate or a market index - for $t = 0, 1, 2, \dots$ (t in minutes, hours, days, etc.)[1]. Define

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1,$$

as the simple net return of the asset. Then, the simple gross return is defined as $1 + R_t$ and the **log-return** (or simply **return**), as shall be mentioned in what follows, is

$$X_t = \ln P_t - \ln P_{t-1} = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln\left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right) = \ln(1 + R_t).$$

[1]The common properties mentioned above actually depend on the time scale chosen. Depending on whether the time unit is a second, an hour, a day, a month or a year, qualitative differences are expected between time series and different models may be needed. In what concerns the work of this thesis, the temporal unit may be an hour, a day or a week without any prejudice to what is going to be exposed.

By a Taylor series argument, the (X_t) time series is close to the simple net return time series or relative returns $\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right)$. Normally, relative returns are very small and indistinguishable from the log-returns. Nevertheless, it is preferable to work with the series of log-returns, (X_t), which are free of scale, thus comparable among each other. Also, it is believed that the log-return time series (X_t) can be modelled by a stationary stochastic process in the strict or wide senses (Mikosch, 2003). Stationarity is a basic assumption in time series analysis and it is generally assumed that the transformation given above provides one realization of a stationary process.

The qualitative properties or stylized empirical facts usually observed in series of asset returns are shortly described below following the works of Mikosch (2003) and Cont (2001).

heavy tails : When analysing samples of returns one usually finds sample means not significantly different from zero and sample variances of the order of 10^{-4} or smaller. Sample distributions are roughly symmetric in the center, sharply peaked around zero and with heavy tails on both sides. The shape of a density plot reveals that the normal distribution is not the most appropriate model. The tails of the returns seem to point out to a distribution with power law tails, i.e., for large x and some positive number α (the tail index, that is estimated to be more than two and less than five, for most data sets studied), $P(X_t > x) \sim x^{-\alpha}$. The Pareto distribution or the t-distribution with α degrees of freedom, could model these heavy-tailed unconditional distributions of the returns. Although the precise form of the tails seems difficult to determine, the described behaviour seems to rule out distributions such as the normal or the α-stable distributions with infinite variance, as proposed by Mandelbrot (1963).

conditional heavy tails : Even after modelling the returns with a GARCH-type model the residual time series still exhibit heavy tails, however, less heavy than in the unconditional distribution of returns.

absence of autocorrelations : The linear autocorrelations of the returns are often insignificant, and rapidly decay to zero in a few minutes (for a time lag, h, with $h \geq 15$ minutes, all autocorrelations can be considered zero for all practical purposes).

In standard time series analysis the second-order structure of a stationary time series (Y_t) is fundamental for parameter estimation, model testing and prediction. The autocovariance function (ACVF), γ_Y, and the autocorrelation function (ACF), ρ_Y, are particularly important instruments in the analysis of the second-order dependence structure. For lag $h \in \mathbb{Z}$, these functions are defined as

$$\gamma_Y(h) = \mathrm{Cov}[Y_t, Y_{t+h}] \qquad \text{and} \qquad \rho_Y(h) = \mathrm{Corr}[Y_t, Y_{t+h}].$$

In practice, however, ACVF and ACF have to be estimated. Standard estimators are their sample counterparts, the sample ACVF, $\gamma_{n,Y}$, and sample ACF, $\rho_{n,Y}$, which, for lag $h \in \mathbb{Z}$, are defined by

$$\gamma_{n,Y}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (Y_t - \overline{Y}_n)(Y_{t+h} - \overline{Y}_n) \quad \text{and} \quad \rho_{n,Y}(h) = \frac{\gamma_{n,Y}(h)}{\gamma_{n,Y}(0)}.$$

where $\gamma_{n,Y}(h) = \rho_{n,Y}(h) = 0$ for $|h| \geq n$ and $\overline{Y}_n$ stands for the sample mean. Provided that (Y_t) is stationary, ergodic and $\mathrm{Var}[Y_t] < \infty$, sample ACVF and sample ACF converge asymptotically to ACVF and ACF, respectively.

Recall that the first stylized fact mentioned about log-returns was that they evidence some heavy-tailed distribution. When the marginal distribution of a time series is very heavy-tailed, the rate of convergence of sample ACFV and ACF to their theoretical counterparts can be extremely slow and sample ACVF and ACF may lead to some misinterpretations. Particularly, asymptotic confidence bands can be much wider than the classical $\frac{1}{\sqrt{n}}$-bands, and even wider than the estimated autocorrelations (for details, see Section 9 in Mikosch, 2003 and Davis and Mikosch, 2001). Also, if very long time periods are sampled, on the order of several months or years, the general assumption of stationarity cannot be assumed, and, in that case, sample ACVF and sample ACF should not be taken as estimators of their theoretical counterparts.

Assuming that sample ACF is a reasonable estimator of the ACF, one common feature about series of asset returns is that the sample ACF, $\rho_{n,X}$, is not significant for any lag, except perhaps for the first (which is usually also small in absolute value) showing that the returns are not serially correlated.

slow decay of autocorrelation in the absolute returns : On the other hand, sample ACFs, $\rho_{n,|X|}$, of the absolute returns, $|X_t|$, and ρ_{n,X^2}, of the squares of the returns, X_t^2, are different from zero for a large number of lags and stay almost constant and positive for large lags, meaning that these non-linear simple functions of the returns exhibit significant positive autocorrelation or persistence. The sample ACF of absolute returns, in particular, decays slowly as a function of time lag, roughly as a power law with exponent in the interval $[0.2, 0.4]$. This features are known, in this context, as long memory or long range dependence of absolute returns or their squares.

volatility clustering or dependence in the tails : If we look at a plot of pairwise exceedances of a high threshold (like, for instance, pairs of $|X_t|$ and $|X_{t+1}|$ exceeding the same high sample percentile) it is easily observable that these pairwise exceedances occur in clusters. Large and small values of asset returns occur in clusters and it is said that there is dependence of extremal return values, or dependence in the tails. As this feature is also known as **volatility clustering** it is convenient to introduce now the notion of **volatility**. In econometrics, it is the synonym to standard deviation, σ, and represents a measure for the variation of price of a financial instrument over time. The existence of dependence on the series of non-linear functions of asset returns, like the absolute values, $(|X_t|)$, or the squares, (X_t^2), of the returns, points towards the modelling and prediction of the variability or volatility of the process, instead of the process (X_t) itself. It can then be said that there is correlation in volatility of returns but not in the returns themselves. And this feature motivates the decomposition of the returns as

$$X_t = |X_t|\text{sign}(X_t),$$

where the sequence $(\text{sign}(X_t))$ consists on independent identically distributed (i.i.d.) symmetric random variables. Taking into consideration this decomposition, many models are of the form

$$X_t = \sigma_t Z_t,$$

where (Z_t) is an i.i.d. symmetric sequence, and (σ_t) is the volatility sequence, a stationary non-negative sequence whose dynamics should match the empirically observed dependences. As the volatility, σ_t is not observable, the observable quantities $|X_t|$ and X_t^2 are sometimes interpreted as estimators of σ_t and σ_t^2, respectively. This somehow explains the interest in computing correlations of absolute or square returns and modelling their dependence.

When is it said that a common feature of asset returns is volatility clustering, it means that high-volatility events tend to cluster in time, or, in other words, large price variations are more likely to be followed by large price variations.

gain/loss asymmetry : Although the distribution of the returns is roughly symmetric in the center, we can actually refer to some skewness, particularly in the series related to stock prices and stock index values, for which large downwards movements are frequent and are not paired by large upward movements. This means that downward movements are faster than the upward ones, and this gain-loss asymmetry property refers to the observation that, for stocks or indices, it takes typically longer to gain 5% than to lose 5% (Siven and Lins, 2009). Exchange rates do not usually show this asymmetry in distribution.

leverage effect : It has usually been found that the conditional volatility of stocks responds asymmetrically to positive versus negative shocks: volatility tends to rise higher in response to negative shocks as opposed to positive shocks, or, in a a few words, volatility increases when the stock price falls. This behaviour of asset returns is known as the leverage effect and it is observable that most measures of volatility are negatively correlated with the returns of a particular asset.

aggregational Gaussianity : As the time scale or period over which returns are calculated is increased, their distribution looks more and more like a normal distribution: the *peakedness around zero* and the *heavy-tailedness* turn into bell shapedness. Aggregated returns over periods of time such as a month or a year have an empirical distribution whose estimated probability density resembles more a normal curve than the distribution of hourly or daily returns. Although sometimes in the literature it is stated that returns calculated over longer periods of time have less heavy tails than those calculated over shorter periods of time (which is attributed to the aggregational Gaussianity) this statement is false if the data comes from a distribution with a power law tail. For large classes of strictly stationary time series (X_t) with power law tails, the sums $X_t + \cdots + X_{t+h}$ have the same asymptotic power law tails as X_t itself and, in particular, the same tail index. A possible explanation for the resemblance with the normal distribution is that, the larger the aggregational level, h, the closer the distribution of the returns gets to the normal distribution, by virtue of the central limit theorem (CLT). So, as the aggregational level increases, and due to the CLT, the distribution of the returns tends to look like the normal distribution.

intermittency : Asset returns display, at any time scale, a very high degree of variability. This variability is usually quantified through the presence of irregular bursts in the time series of a wide variety of volatility estimators.

volume/volatility correlation : Since 1970 there were several studies indicating strong positive correlation between volume and volatility (e.g. Karpoff, 1987; Gallant et al., 1992; Yin, 2010) and this was known as the volume/volatility correlation stylized fact. However, recent contributions by Giot et al. (2010), Amatyakul (2010) and Wang and Huang (2012), challenged this stylized fact using the volatility decomposition technique by Barndorff-Nielsen (2004) and Barndorff-Nielsen and Shephard (2006). Using this technique (daily volatility can be decomposed into a continuous component due to small price changes and a jump component due to large price movements) these authors have shown that only the continuous component of the volatility shows a positive contemporal volume-volatility relation, while the jump component shows negative correlation with the trading volume.

asymmetry in time scales : As already mentioned, common properties usually observed in time series of log-returns depend on the time scale chosen. If the temporal unit is small, of the order of seconds or minutes, one speaks of *fine* scales, whereas, when the temporal unit is of the order of weeks or months, it is said that the returns are *coarse-grained*. Not only the properties are different, but also it is believed that coarse-grained measures of volatility predict fine-scale volatility better than the other way round. As explained by Gavrishchaka and Ganguli (2003), if the heterogeneous market hypothesis[2] is taken into account, as seems to be proposed by some empirical studies with high-frequency data (Dacorogna et al., 2001), traders with different time horizons are interested in the volatility on different time grids and short-term traders can react to clusters of coarse volatility, while the level of fine volatility does not affect strategies of long-term traders.

After this long list of qualitative properties of asset returns one easily agrees that the gain in generality doing these observations has necessarily to imply a loss in precision. Anyway, this information is very important as it results from several decades of analysis of different markets and instruments. As is usually said, these properties represent the common denominator among the properties observed in many different studies, in many different sets of assets and markets. Also, due to its qualitative nature, these properties are model-free, meaning that they do not result from any parametric hypothesis about the return process, but, instead, should be viewed as the constraints that any stochastic process should verify if one wants to reproduce with accuracy the statistical properties of asset returns.

The existence of non-linear dependence structure, in the absolute or the

[2]The hypothesis of a homogeneous market assumes that all participants interpret news and react to them in the same way. The hypothesis of a heterogeneous market assumes that different market agents (ranging from intraday dealers or market makers to central banks and large commercial organization) have different time horizons and dealing frequencies.

square values of the returns, demands that the modelling of return data should be done with some non-linear process. To clarify, **non-linear** process means that the process cannot be represented through the equation

$$X_t = \sum_{i=-\infty}^{+\infty} \psi_i Z_{t-i},\ t \in \mathbb{Z},$$

where (Z_t) is **white noise**, i.e., a sequence of random variables (r.v's) with the same distribution, satisfying

- $\mathrm{E}[Z_t] = \mu_Z$, and usually $\mu_Z = 0$;
- $\mathrm{Var}[Z_t] = \sigma_Z^2$;
- $\gamma_Z(h) = 0$ for $h > 0$ and $\gamma_Z(0) = \sigma_Z^2$;
- $\rho_Z(h) = 0$ for $h > 0$ and $\rho(0) = 1$.

As the mean and the variance, μ_Z, and σ_Z^2, respectively, do not depend on t, and the autocovariance for two different time instants depends only on the lag between them - and not on the position of the time instants on the temporal axis - the white noise process is said to be covariance stationary or second order stationary. Moreover, (ψ_i) is a sequence of real numbers satisfying some mild summability condition (as the variance of the linear process is given by $\mathrm{Var}[X_t] = (\sum_{i=0}^{+\infty} \psi_i^2)\sigma_Z^2$, in order for it to be finite, the series $\sum_{i=0}^{+\infty} \psi_i^2 < \infty$).

An important class of processes that satisfy the definition of linear process is the **AutoRegressive Moving Average (ARMA)** processes defined by the recursion

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}. \tag{1.1}$$

Making use of B, the **backshift or lag operator**, $BX_t = X_{t-1}$, it follows that

$$\begin{aligned} X_t - \alpha_1 B X_t - \cdots - \alpha_p B^p X_t &= Z_t + \beta_1 B Z_t + \cdots + \beta_q B^q Z_t \\ (1-\alpha(B))X_t &= (1+\beta(B))Z_t \\ \phi_p(B) X_t &= \theta_q(B) Z_t, \end{aligned}$$

where one can define

$$\begin{aligned} \phi_p(B) &\equiv 1-\alpha(B) = 1 - \sum_{i=1}^{p} \alpha_i B^i, \text{ the autoregressive lag polynomial;} \\ \theta_q(B) &\equiv 1+\beta(B) = 1 + \sum_{j=1}^{q} \beta_j B^j, \text{ the moving average lag polynomial.} \end{aligned}$$

Regarding reader's convenience in further developments in this thesis, Table 1.1 summarises some characteristics of the ARMA processes and their subclasses, the **Autoregressive processes of order** p, **AR(**p**)** and the **Moving Average processes of order** q, **MA(**q**)**. For the theory behind classical time series analysis and linear processes we refer to Murteira et al. (1993) and Brockwell and Davis (1991).

Due to the dependence structure in the absolute or the square values of the returns, linear models in the ARMA family do not seem appropriate to model financial time series. Constrained by the need of modelling non-linear dependency, most models for financial time series used in practice are given, as mentioned before, in the multiplicative form

$$X_t = \sigma_t Z_t, \ t \in \mathbb{Z}, \tag{1.2}$$

where (Z_t) forms an i.i.d. symmetric sequence with zero-mean and unit variance and (σ_t) is a stochastic process such that σ_t and Z_t are independent for fixed t. Normally, it is assumed that $Z_t \sim N(0,1)$, but a heavy-tailed distribution such as the t-distribution could also be considered. (σ_t) represents the volatility process. Moreover, it is also assumed that Z_t is independent of

Table 1.1: General characteristics of ARMA processes.

	AR(p)	MA(q)	ARMA(p,q)
Model in terms of past values of X_t	$\phi_p(B)X_t = Z_t$	$[\theta_q(B)]^{-1}X_t = Z_t$	$[\theta_q(B)]^{-1}\phi_p(B)X_t = Z_t$
Model in terms of past values of Z_t	$X_t = [\phi_p(B)]^{-1}Z_t$	$X_t = \theta_q(B)Z_t$	$X_t = [\phi_p(B)]^{-1}\theta_q(B)Z_t$
Stationarity conditions	Zeros of $\phi_p(B)$ must lie outside the unit circle	Always stationary	Zeros of $\phi_p(B)$ must lie outside the unit circle
Invertibility conditions	Always invertible	Zeros of $\theta_q(B)$ must lie outside the unit circle	Zeros of $\theta_q(B)$ must lie outside the unit circle
ACF	geometric decay	cuts off at q	geometric decay after q

the past values of the process $(X_{t-1}, X_{t-2}, \dots)$. There are two main factors that motivate the choice of this simple multiplicative model, namely that

- in practice, the direction of price changes is well modelled by the sign of Z_t, whereas σ_t provides a good description of the order of magnitude of this change. This representation expresses the belief that the direction of price changes can not be modelled, only their magnitude;

- since σ_t and Z_t are independent for fixed t, the squared volatility σ_t^2 represents the conditional variance of X_t given σ_t. With this representation it is then possible to construct the correlation in the volatility of the returns but not on the returns themselves, as is characteristic in financial time series. The conditional variance, is allowed to change over time and does not coincide with the unconditional variance of the process. This way it is possible to have a stochastic process with a volatility factor whose dynamics can resemble the bursts in volatility of asset returns and match the empirically observed dependences.

Assuming, for instance, that $Z_t \sim N(0,1)$ and that σ_t is a function of the past values $X_{t-1}, X_{t-2}, \ldots$ and $\sigma_{t-1}, \sigma_{t-2}, \ldots$, then, conditionally on the past, X_t is normally distributed, i.e., $X_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$ with $\mathcal{F}_{t-1} = \sigma(X_s, \sigma_s\, s \leq t-1)$. The conditional mean and conditional variance are given by

$$\begin{aligned} \mathrm{E}[X_t|\mathcal{F}_{t-1}] &= \mathrm{E}[\sigma_t Z_t|\mathcal{F}_{t-1}] = \sigma_t \,\mathrm{E}[Z_t|\mathcal{F}_{t-1}] = \sigma_t \cdot 0 = 0, \\ \mathrm{Var}[X_t|\mathcal{F}_{t-1}] &= \mathrm{E}[X_t^2|\mathcal{F}_{t-1}] - (\mathrm{E}[X_t|\mathcal{F}_{t-1}])^2 = \sigma_t^2 \,\mathrm{E}[Z_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 \cdot 1 = \sigma_t^2. \end{aligned}$$

Also, it is easy to show that, as long as $\mathrm{E}[\sigma_t^2] < +\infty$ and σ_t is a function of past values of the process and the volatility itself, then the multiplicative model (1.2) has ACF that is zero for all lags except zero. The lack of serial correlation in many data series of asset returns is also captured by this general class of multiplicative models.

This general class includes the **Autoregressive Conditional Heteroscedasticity (ARCH)** family and also the **Stochastic Volatility (SV)** models. As we will focus our attention on the former set, we will just briefly present the SV models and some of their properties.

All the models in the ARCH family state that the conditional variance depends on the past values of the returns. In the SV model, firstly proposed by Taylor (1980, 1986), the volatility depends on its own past values but is independent of the past values of the returns. As a multiplicative model, the SV process is written as in (1.2) although, in contrast to the models in the ARCH family, there is no feedback between the noise (Z_t) and the volatility process (σ_t), which is a strictly stationary process, independent of the i.i.d. symmetric noise process (Z_t). In this class of models there are two independent sources of randomness and the mutual independence of (σ_t) and

(Z_t) enables the modelling of the dependency structure of the asset returns only through the volatilities σ_t and the modelling of the heavy tails of the distribution of X_t through the interplay between the tails of σ_t and Z_t. Due to this construction, the SV model can explain dependence and heavy tails in a very flexible way and the stochastic volatility can actually be chosen in such a way that the ACF of its absolute or squared values converges to zero arbitrarily slow.

A very common specification for the SV model is as follows

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t &= e^{\frac{h_t}{2}}, \end{aligned}$$

where (Z_t) is a noise process with zero-mean and unit variance. Usually, (σ_t) is given by a parametric model as a Gaussian ARMA process for $\ln(\sigma_t)$, whose simplest form may be written as an AR(1)

$$h_t = \omega + \alpha_1 h_{t-1} + \epsilon_t, \tag{1.3}$$

where (ϵ_t) is a stationary Gaussian sequence, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, independent of Z_t. For this simple formulation it follows that

- $\mathrm{E}[X_t] = \mathrm{E}[\sigma_t Z_t] = \mathrm{E}[\sigma_t]\,\mathrm{E}[Z_t] = 0$;

- $\mathrm{Var}[X_t] = \mathrm{E}[X_t^2] = \mathrm{E}[\sigma_t^2 Z_t^2] = \mathrm{E}[\sigma_t^2]\,\mathrm{E}[Z_t^2] = \mathrm{E}[\sigma_t^2]$ and, supposing $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ and h_t a stationary process, one gets $\mu_h = \mathrm{E}[h_t] = \frac{\omega}{1-\alpha_1}$ and $\sigma_h^2 = \mathrm{Var}[h_t] = \frac{\sigma_\epsilon^2}{1-\alpha_1^2}$. Since h_t is Normal, i.e., $h_t \sim N(\frac{\omega}{1-\alpha_1}, \frac{\sigma_\epsilon^2}{1-\alpha_1^2})$, then σ_t^2 is log-Normal and, finally, one has

  $$\mathrm{Var}[X_t] = \mathrm{E}[X_t^2] = \mathrm{E}[\sigma_t^2] = e^{\mu_h + \frac{\sigma_h^2}{2}}.$$

  It is also easily shown that

  $$\mathrm{E}[X_t^4] = 3e^{\mu_h + 2\sigma_h^2},$$

from which the kurtosis equals $K = 3e^{\sigma_h^2} > 3$, synonym of heavy tails for the distribution of X_t.

- $\gamma_X(h) = 0$ for $h \neq 0$ and X_t is not correlated, as expected (but not independent as there is correlation in $\ln(X_t^2)$). Some SV models have been modified in order that the autocorrelation function of $\ln(X_t^2)$ decays as slowly as pretended though keeping no serial correlation on the X_t.

A general SV model is obtained by allowing an AR(p) or an ARMA(p, q) for $h_t = 2\ln(\sigma_t)$, in (1.3).

Squaring X_t and then taking logarithms, ($\ln(X_t^2) = h_t + 2\ln|Z_t|$), makes it easy to see that the ARMA process $h_t = 2\ln(\sigma_t) = \ln(X_t^2) - 2\ln|Z_t|$ gets perturbed by the extra noise $2\ln|Z_t|$. Estimation becomes more complicated since there is no explicit expression for the likelihood function. In spite of the recent developments in the last years with quasi - maximum likelihood methods and simulation based methods, estimation procedures have been applied in a much more straightforward manner for the models in the ARCH family, making this family a very attractive choice when modelling volatility. This is the reason we are now turning our attention to these models, their properties and estimation procedures.

The first model in the ARCH family was proposed by Engle (1982), who suggested the following representation for the volatility, σ_t, considering the multiplicative model (1.2)

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \ t \in \mathbb{Z}, \tag{1.4}$$

where $\omega > 0$ and $\alpha_i \geq 0$, for $i = 1, \ldots, p$. This model is called the **AutoRegressive Conditional Heteroscedasticity model of order** p, or

ARCH(p). Usually, (Z_t) is an i.i.d. sequence with zero-mean and unit variance such as $N(0,1)$. Is is also common that $Z_t \sim t_n$ or other distribution that better describes the heavy tails of the series of asset returns. The special case $p = 1$ leads to the ARCH(1) model

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \omega + \alpha_1 X_{t-1}^2. \end{aligned} \tag{1.5}$$

For this particular model it follows that

- Unconditional mean, $\mathrm{E}[X_t] = \mathrm{E}[\mathrm{E}[X_t|\mathcal{F}_{t-1}]] = \mathrm{E}[\mathrm{E}[\sigma_t Z_t|\mathcal{F}_{t-1}]] =$

$$\begin{aligned} &= \mathrm{E}[\mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2} Z_t|\mathcal{F}_{t-1}]] \\ &= \mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2}\,\mathrm{E}[Z_t|\mathcal{F}_{t-1}]] = \mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2} \cdot 0] \\ &= 0. \end{aligned}$$

- Unconditional variance, $\mathrm{Var}[X_t] = \mathrm{E}[(X_t - (\mathrm{E}[X_t])^2] = \mathrm{E}[X_t^2] =$

$$\begin{aligned} &= \mathrm{E}[\mathrm{E}[(\sigma_t Z_t)^2|\mathcal{F}_{t-1}]] \\ &= \mathrm{E}[(\omega + \alpha_1 X_{t-1}^2)\,\mathrm{E}[Z_t^2|\mathcal{F}_{t-1}]] = \mathrm{E}[(\omega + \alpha_1 X_{t-1}^2) \cdot 1] \\ &= \omega + \alpha_1\,\mathrm{E}[X_{t-1}^2] \end{aligned}$$

  and if (X_t) is second-order stationary then $\mathrm{E}[X_{t-1}^2] = \mathrm{E}[X_t^2] = \mathrm{Var}[X_t]$, and it is possible to write

$$\mathrm{Var}[X_t] = \frac{\omega}{1 - \alpha_1}.$$

  If the ARCH(p) is to be considered, one will have

$$\mathrm{Var}[X_t] = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i},$$

  instead. Returning to the ARCH(1) case, as $\mathrm{Var}[X_t] \geqslant 0$, it can be concluded that $0 \leqslant \alpha_1 < 1$. Calculating $\mathrm{E}[X_t^4]$ and the kurtosis, K, one obtains $K = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$, meaning that, if X_t is modelled by an ARCH(1) model, it will have heavier tails than with the normal distribution.

- Autocovariance function, $\gamma_X(h) =$

$$
\begin{aligned}
&= \mathrm{E}[(X_t - \mathrm{E}[X_t])(X_{t+h} - \mathrm{E}[X_{t+h}])] = \mathrm{E}[X_t X_{t+h}] \\
&= \mathrm{E}[\mathrm{E}[X_t X_{t+h} | \mathcal{F}_{t+h-1}]] \\
&= \mathrm{E}[X_t\, \mathrm{E}[X_{t+h} | \mathcal{F}_{t+h-1}]] = \mathrm{E}[X_t\, \mathrm{E}[\sqrt{\omega + \alpha_1 X_{t+h-1}^2} Z_{t+h} | \mathcal{F}_{t+h-1}]] \\
&= \mathrm{E}[X_t \sqrt{\omega + \alpha_1 X_{t+h-1}^2}\, \mathrm{E}[Z_{t+h} | \mathcal{F}_{t+h-1}]] \\
&= \mathrm{E}[X_t \sqrt{\omega + \alpha_1 X_{t+h-1}^2} \cdot 0] \\
&= 0.
\end{aligned}
$$

The ACVF, γ_X, is zero for all lags except zero. (X_t) is a sequence of non-correlated variables with mean zero and variance $\frac{\omega}{1-\alpha_1}$. Hence, there is no serial correlation in (X_t), as expected.

The designation of *autoregressive* in the ARCH model comes from the following manipulations. Writing $\nu_t = X_t^2 - \sigma_t^2 = \sigma_t^2 Z_t^2 - \sigma_t^2 = \sigma_t^2(Z_t^2 - 1)$ and substituting in

$$X_t^2 - \sigma_t^2 = X_t^2 - (\omega + \alpha_1 X_{t-1}^2) = \nu_t,$$

one finally gets

$$X_t^2 = \omega + \alpha_1 X_{t-1}^2 + \nu_t, \tag{1.6}$$

which is an AR(1) model in X_t^2. If (Z_t) is an i.i.d. sequence with zero-mean and unit variance and (X_t) is second-order stationary, then (ν_t) constitutes a white noise sequence with $\mathrm{E}[\nu_t] = 0$, $\mathrm{Var}[\nu_t] = \frac{2\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$ and $\gamma_\nu(h) = 0$, for $h > 0$. From equation (1.6) it follows

$$\gamma_{X^2}(h) = \alpha_1 \gamma_{X^2}(h-1),$$

and, after successive substitutions,

$$\rho_{X^2}(h) = \alpha_1^h,$$

which is the ACF of an AR(1).

If we turn to the general ARCH(p) the equation equivalent to (1.6) is

$$\begin{aligned} X_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \nu_t \quad &\Leftrightarrow \quad X_t^2 - \sum_{i=1}^{p} \alpha_i X_{t-i}^2 = \omega + \nu_t \\ &\Leftrightarrow \quad (1 - \sum_{i=1}^{p} \alpha_i B^i) X_t^2 = \omega + \nu_t \\ &\Leftrightarrow \quad \phi(B) X_t^2 = \omega + \nu_t, \qquad (1.7) \end{aligned}$$

where B is the backshift operator. If (Z_t) is an i.i.d. sequence with zero-mean and unit variance and (X_t) is second-order stationary, then (ν_t) constitutes a white noise sequence, and as for the ARCH(1) case there is a formal analogy with an AR(p), ($\phi_p(B)X_t = Z_t$ in Table 1.1). Nevertheless, the resemblance is only formal because there is a fundamental difference between the the right-hand side of both autoregressive equations: in the *true* AR(p) situation, the right-hand side depends only on the well-known sequence (Z_t), which is a white noise sequence with well defined properties, and on the AR(p) in X_t^2, in (1.7), the right-hand side depends on ν_t which is also a white noise sequence, if (X_t) is second-order stationary. Nevertheless, this white noise sequence depends on Z_t and also on X_t, establishing a complicated dependency structure.

Estimation in ARCH(p) models is usually based on the conditional maximum likelihood (ML) method. The likelihood function is given by

$$\begin{aligned} &L(x_1, x_2, \ldots, x_T | \boldsymbol{\theta}) = \\ &\quad = \quad f(x_T|\mathcal{F}_{t-1}, \boldsymbol{\theta}) f(x_{T-1}|\mathcal{F}_{t-2}, \boldsymbol{\theta}) \cdots \\ &\quad \ldots \quad f(x_{m+1}|\mathcal{F}_m, \boldsymbol{\theta}) f(x_m|\mathcal{F}_{m-1}, \boldsymbol{\theta}) \cdots f(x_2|\mathcal{F}_1, \boldsymbol{\theta}) f(x_1|\boldsymbol{\theta}) \\ &\quad = \quad f(x_T|\mathcal{F}_{t-1}, \boldsymbol{\theta}) f(x_{T-1}|\mathcal{F}_{t-2}, \boldsymbol{\theta}) \cdots f(x_{m+1}|\mathcal{F}_m, \boldsymbol{\theta}) f(x_1, x_2, \ldots, x_m|\boldsymbol{\theta}), \end{aligned}$$

with $\boldsymbol{\theta} := (\omega, \alpha_1, \ldots, \alpha_p)$, the vector of parameters to be estimated. The conditional likelihood function, conditionally on the first m observations is

$$
\begin{aligned}
&L(x_{m+1}, x_{m+2}, \ldots, x_T | \boldsymbol{\theta}, x_1, x_2, \ldots, x_m) = \\
&\qquad = f(x_T|\mathcal{F}_{t-1}, \boldsymbol{\theta}) f(x_{T-1}|\mathcal{F}_{t-2}, \boldsymbol{\theta}) \cdots f(x_{m+1}|\mathcal{F}_m, \boldsymbol{\theta}) \\
&\qquad = \prod_{t=m+1}^{T} f(x_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}).
\end{aligned}
$$

If the Z_t are normally distributed then the distributions of $X_t|X_{t-1}$ are also normal, i.e., $X_t|X_{t-1} \sim N(0, \sigma_t^2)$ with σ_t^2 as in (1.4) for the ARCH(p) or as in (1.5) for the particular ARCH(1) case. The conditional log-likelihood function $l(x_{m+1}, \ldots, x_T|\boldsymbol{\theta}, x_1, \ldots, x_m)$, takes the form

$$
\begin{aligned}
&l(x_{m+1}, x_{m+2}, \ldots, x_T | \boldsymbol{\theta}, x_1, x_2, \ldots, x_m) = \\
&\qquad = \ln \left(L(x_{m+1}, x_{m+2}, \ldots, x_T | \boldsymbol{\theta}, x_1, x_2, \ldots, x_m) \right) \\
&\qquad = \ln \left(\prod_{t=m+1}^{T} f(x_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}) \right) \\
&\qquad = \sum_{t=m+1}^{T} \ln \left(f(x_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}) \right) \\
&\qquad = \sum_{t=m+1}^{T} \ln \left(f(x_t|x_{t-1}, \boldsymbol{\theta}) \right) \\
&\qquad = \sum_{t=m+1}^{T} \ln \left(\frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{x_t^2}{2\sigma_t^2} \right) \right).
\end{aligned}
$$

After some simple manipulations it follows

$$
\begin{aligned}
&l(x_{m+1}, x_{m+2}, \ldots, x_T | \boldsymbol{\theta}, x_1, x_2, \ldots, x_m) = \\
&\qquad = \sum_{t=m+1}^{T} \left(\ln(2\pi)^{-\frac{1}{2}} + \ln(\sigma_t)^{-1} - \frac{x_t^2}{2\sigma_t^2} \right) \\
&\qquad = \sum_{t=m+1}^{T} -\frac{1}{2} \ln(2\pi) - \sum_{t=m+1}^{T} \ln(\sigma_t) - \sum_{t=m+1}^{T} \frac{x_t^2}{2\sigma_t^2} \\
&\qquad = -\frac{T-m}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=m+1}^{T} \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=m+1}^{T} \frac{x_t^2}{\sigma_t^2}.
\end{aligned}
$$

As the first term is constant, one can finally write that the explicit form of the conditional log-likelihood function for an ARCH(1), conditionally on the first observation is

$$l(x_2, \ldots, x_T | \omega, \alpha_1, x_1) \propto -\frac{1}{2} \sum_{t=2}^{T} \ln(\omega + \alpha_1 x_{t-1}^2) - \frac{1}{2} \sum_{t=2}^{T} \frac{x_t^2}{\omega + \alpha_1 x_{t-1}^2}. \quad (1.8)$$

For an ARCH(p) the conditional log-likelihood function, conditionally on the first p observations and with σ_t^2 given by (1.4) is

$$l(x_{p+1}, \ldots, x_T | \omega, \alpha_1, \ldots, \alpha_p, x_1, \ldots, x_p) \propto -\frac{1}{2} \sum_{t=p+1}^{T} \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=p+1}^{T} \frac{x_t^2}{\sigma_t^2}. \quad (1.9)$$

If the innovations Z_t are considered t-distributed then the log-likelihood can still be written in an explicit form. In order to maximize the conditional log-likelihood function some non-linear numerical optimization procedure is needed. A common non-linear optimization method is the Newton-Raphson method and several software packages already include estimation procedures for many models in the ARCH family based on the maximum conditional log-likelihood method.

ARCH(p) processes do not fit asset return series as desired unless the order of the model is very high which is a particularly bad choice when sample size is small. The formal resemblance between the ARCH(p) and an AR(p) model, suggested a generalization, considering that in the linear case, an ARMA(p, q) model can be a more parsimonious choice (in terms of number of parameters to estimate) than a *pure* AR(p) or MA(q) model. Then, a generalized ARCH process similar in form to the ARMA representation, could eventually describe volatility with less parameters then an ARCH(p) model.

Bollerslev (1986) suggested a generalization of the ARCH(p) model leading to the **Generalized ARCH model of order** (p, q), or **GARCH**(p, q):

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \ t \in \mathbb{Z} \tag{1.10}$$

with $\omega > 0$ and $\alpha_i \geq 0$, $\beta_j \geq 0$, for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. The GARCH$(1, 1)$ model is obtained by considering the special case $p = 1$ and $q = 1$ in (1.10), i.e.,

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

For this particular model one has

- Unconditional mean, $\mathrm{E}[X_t] = \mathrm{E}[\mathrm{E}[X_t|\mathcal{F}_{t-1}]] = \mathrm{E}[\mathrm{E}[\sigma_t Z_t|\mathcal{F}_{t-1}]] =$

$$\begin{aligned} &= \mathrm{E}[\mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2} Z_t|\mathcal{F}_{t-1}]] \\ &= \mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2}\, \mathrm{E}[Z_t|\mathcal{F}_{t-1}]] \\ &= \mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2} \cdot 0] \\ &= 0. \end{aligned}$$

- Unconditional variance, $\mathrm{Var}[X_t] = \mathrm{E}[(X_t - (\mathrm{E}[X_t])^2] = \mathrm{E}[X_t^2]$. If (X_t) is second-order stationary,

$$\mathrm{Var}[X_t] = \frac{\omega}{1 - (\alpha_1 + \beta_1)}.$$

  If the GARCH(p, q) is to be considered one will have

$$\mathrm{Var}[X_t] = \frac{\omega}{1 - \sum_{i=1}^{m} (\alpha_i + \beta_i)},$$

  instead. For the GARCH$(1, 1)$ case, the kurtosis, K, is given by $K = \frac{3(1-(\alpha_1+\beta_1)^2)}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} > 3$, meaning that, if X_t follows a GARCH$(1, 1)$ model, it will have heavier tails than with the normal distribution.

- ACVF, $\mathrm{E}[(X_t - \mathrm{E}[X_t])(X_{t+h} - \mathrm{E}[X_{t+h}])] = \mathrm{E}[X_t X_{t+h}] =$

$$\begin{aligned}
&= \mathrm{E}[\mathrm{E}[X_t X_{t+h}|\mathcal{F}_{t+h-1}]] = \mathrm{E}[X_t\,\mathrm{E}[X_{t+h}|\mathcal{F}_{t+h-1}]] \\
&= \mathrm{E}[X_t\,\mathrm{E}[\sqrt{\omega + \alpha_1 X_{t-1}^2 + \beta_1\sigma_{t-1}^2} Z_{t+h}|\mathcal{F}_{t+h-1}]] \\
&= \mathrm{E}[X_t\sqrt{\omega + \alpha_1 X_{t-1}^2 + \beta_1\sigma_{t-1}^2}\,\mathrm{E}[Z_{t+h}|\mathcal{F}_{t+h-1}]] \\
&= \mathrm{E}[X_t\sigma_{t+h}\cdot 0] \\
&= 0.
\end{aligned}$$

  The ACVF, γ_X, is zero for all lags except zero. The GARCH(1,1) is covariance stationary.

The ARMA structure is obtained considering again $\nu_t = X_t^2 - \sigma_t^2$ and substituting σ_t^2 from (1.10)

$$\nu_t = X_t^2 - \omega - \sum_{i=1}^{p}\alpha_i X_{t-i}^2 - \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2.$$

Reorganizing,

$$X_t^2 = \omega + \sum_{i=1}^{p}\alpha_i X_{t-i}^2 + \nu_t + \sum_{j=1}^{q}\beta_j(X_{t-j}^2 - \nu_{t-j}) \Leftrightarrow$$

$$X_t^2 - \sum_{i=1}^{p}\alpha_i X_{t-i}^2 - \sum_{i=1}^{q}\beta_j X_{j-i}^2 = \omega + \nu_t - \sum_{j=1}^{q}\beta_j\nu_{t-j} \Leftrightarrow$$

$$\left(1 - \sum_{i=1}^{p}\alpha_i B^i - \sum_{i=1}^{q}\beta_j B^j\right) X_t^2 = \omega + \left(1 - \sum_{j=1}^{q}\beta_j B^j\right)\nu_t \Leftrightarrow$$

$$(1 - \alpha(B) - \beta(B))X_t^2 = \omega + (1 - \beta(B))\nu_t.$$

And this is the ARMA(m, q) representation in X_t^2 of the GARCH(p, q). Introducing $m := \max\{p, q\}$ and $\phi(B)$ representing the autoregressive lag polynomial and $\theta(B)$ the moving average lag polynomial,

$$\begin{aligned}
\phi(B) &\equiv 1 - \alpha(B) - \beta(B) = 1 - \sum_{i=1}^{m}(\alpha_i + \beta_i)B^i, \\
\theta(B) &\equiv 1 - \beta(B) = 1 - \sum_{j=0}^{q}\beta_j B^j,
\end{aligned}$$

the formal similarity with an ARMA model can be evidenced by writing

$$\phi(B)X_t^2 = \omega + \theta(B)\nu_t. \tag{1.11}$$

Estimation in GARCH(p, q) models is usually based on the conditional ML method already described for the ARCH(p) case. If the innovations, Z_t, are considered normally distributed then the conditional log-likelihood function for the GARCH(p, q) is

$$l(x_{m+1}, \ldots, x_T | \boldsymbol{\theta}, x_1, \ldots, x_p) \propto -\frac{1}{2} \sum_{t=m+1}^{T} \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=m+1}^{T} \frac{x_t^2}{\sigma_t^2}, \tag{1.12}$$

with σ_t^2 given by equation (1.10) and with $\boldsymbol{\theta} := (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ the vector of parameters to be estimated.

The GARCH representation for the volatility is a very popular one and indeed there are many reasons for it. Firstly, even a GARCH(1,1) model (with only three parameters to estimate) represents a very reasonable fit for many different financial time series, as long as the financial data is not sampled from a long period in time. For illustration see, for example, Mikosch and Stărică (2000) where the authors showed that the residuals of a GARCH(1,1) fitting behave like an i.i.d. sequence. Secondly, as mentioned above, estimation in a GARCH process is a very straightforward procedure, following the conditional ML method. Yet another reason for the popularity of the GARCH representation of the volatility is the formal similarity with an ARMA process, though, as was already mentioned for the ARCH(p) processes, the noise sequence (ν_t), on the right-hand side of equation (1.11), is not independent of X_t. In fact, as is frequently done in classical time series analysis with linear models, we could try to iterate equation (1.11) in order to obtain X_t^2 as an explicit expression of the noise sequence (ν_t). Stationarity conditions could then be built if (ν_t) was an i.i.d. sequence. But, as defined, $\nu_t = X_t^2 - \sigma_t^2$ is dependent on the left-hand side of equation (1.11), meaning that this simple procedure cannot be applied when trying to find stationarity conditions. Stationarity is not easy to derive in GARCH processes and to understand some important results we need to introduce the stochastic recurrence (or difference) equation (SRE), given by

$$\boldsymbol{Y}_t = \boldsymbol{A}_t \boldsymbol{Y}_{t-1} + \boldsymbol{B}_t, \; t \in \mathbb{Z}, \tag{1.13}$$

where $(\boldsymbol{A}_t, \boldsymbol{B}_t)$ are i.i.d. $\mathbb{R}^2$-valued random pairs, $\boldsymbol{Y}_{t-1}$ is independent of $(\boldsymbol{A}_t, \boldsymbol{B}_t)$ for every time instant t, the $\boldsymbol{A}_t$'s are i.i.d. random $d \times d$ matrices, and the $\boldsymbol{B}_t$'s are i.i.d. d-dimensional vectors. The GARCH(1,1) process can be rewritten as a two-dimensional SRE in the form (1.13) as follows

$$\boldsymbol{Y}_t = \begin{pmatrix} X_t^2 \\ \sigma_t^2 \end{pmatrix}, \quad \boldsymbol{A}_t = \begin{bmatrix} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha_1 & \beta_1 \end{bmatrix}, \quad \boldsymbol{B}_t = \begin{pmatrix} \omega Z_t^2 \\ \omega \end{pmatrix}.$$

Note that σ_t^2 satisfies equation (1.13) with $d = 1$, $\boldsymbol{Y}_t = \sigma_t^2$, $\boldsymbol{A}_t = \alpha_1 Z_{t-1}^2 + \beta_1$ and $\boldsymbol{B}_t = \omega$,

$$\sigma_t^2 \;=\; \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2 + \omega.$$

SRE have been extensively studied and in what regards the existence of a stationary solution for equation (1.13) one can mention the theory of Furstenberg and Kesten (1960) and Kesten (1973) to state that (1.13) has a stationary solution if $\mathrm{E}[\ln^+ \|\boldsymbol{A}\|)] < \infty$, $\mathrm{E}[\ln^+(|\boldsymbol{B}|)] < \infty$ and if the top Lyapunov exponent, γ, is negative

$$\gamma = inf\{n^{-1}\,\mathrm{E}[\ln \|\boldsymbol{A}_1 \cdots \boldsymbol{A}_n\|], \; n \in \mathbb{N}\} < 0,$$

where $\ln^+ x = \ln(max\{1, x\})$ and $|\cdot|$ is any norm in $\mathbb{R}^n$ with $\|\boldsymbol{A}\|$ the corresponding operator norm. The GARCH(1,1) case was treated in Nelson (1990) and Bougerol and Picard (1992) and for this particular case the top Lyapunov exponent can be explicitly obtained, giving $\gamma = \mathrm{E}[\ln(\alpha_1 Z_1^2 + \beta_1)]$. Hence, it can be stated that there exists a unique strictly stationary solution for the equation (1.13) if and only if $\omega > 0$ and $\mathrm{E}[\ln(\alpha_1 Z_1^2 + \beta_1)] < 0$.

Bougerol and Picard (1992) also treated the general GARCH(p, q) case and concluded that equation (1.13) has a unique strictly stationary and ergodic solution if and only if $\omega > 0$ and $\gamma < 0$. In this general case γ has not an explicit form, but it is known that a sufficient condition for $\gamma < 0$ is

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1,$$

as long as the noise sequence (Z_t) is an i.i.d. sequence with mean zero and unit variance. It is also known that a necessary condition for $\gamma < 0$ is that $\sum_{j=1}^{q} \beta_j < 1$.

Ding et al. (1993) questioned the reason why should the volatility depend on the squares of the past values of X_t and σ_t and proposed the **Asymmetric Power ARCH of order** (p, q), **APARCH**(p, q) model defined as

$$\sigma_t^{\delta} = \omega + \sum_{i=1}^{p} \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^{\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{\delta}, \; t \in \mathbb{Z}, \qquad (1.14)$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\delta \geq 0$ and $-1 < \gamma_i < 1$, for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. The APARCH representation has some noteworthy advantages, namely the power of the returns for which the predictable structure in the volatility is the strongest can be determined by the data, and, also, the model allows the detection of asymmetric responses of the volatility for positive or negative shocks. If $\gamma_i > 0$ negative shocks have stronger impact on volatility than positive shocks, as would be expected in the analysis of financial time series, as it is believed that *bad news have stronger impact on volatility than good news.* This is referred to as the leverage effect and, as mentioned in the beginning of the section, reflects the fact that estimated volatility tends to be negatively correlated with the returns. If $\gamma_i < 0$, the reverse happens: positive shocks have stronger impact on volatility than negative shocks. Another advantage of this APARCH representation is that it nests seven other models, as the authors stress out:

1. Engle's ARCH(p) model, (Engle, 1982), when $\delta = 2$, $\gamma_i = 0$ for $i = 1, \ldots, p$ and $\beta_j = 0$ for $j = 1, \ldots, q$;

2. Bollerslev's GARCH(p, q) model, (Bollerslev, 1986), when $\delta = 2$ and $\gamma_i = 0$ for $i = 1, \ldots, p$;

3. Taylor/Schwert's GARCH in standard deviation model, (Taylor, 1986, and Schwert, 1990) when $\delta = 1$ and $\gamma_i = 0$ for $i = 1, \ldots, p$;

4. GJR model of Glosten et al. (1993), when $\delta = 2$ and some subsequent manipulations;

5. Zakoian's TARCH model (Zakoian, 1994) when $\delta = 1$ and $\beta_j = 0$ for $j = 1, \ldots, q$;

6. Higgins and Bera's NGARCH model (Higgins and Bera, 1992) when $\gamma_i = 0$ for $i = 1, \ldots, p$ and $\beta_j = 0$ for $j = 1, \ldots, q$;

7. Geweke (1986) and Pantula (1986)'s log-ARCH model, which is the limiting case of the APARCH model when $\delta \to 0$.

Baillie et al. (1996) proposed the FIGARCH(p, d, q) model in order to accommodate long memory in volatility (accordingly to the most common definition of long memory: autocovariance function, $\gamma_X(h)$, decaying at the hypergeometric rate h^{2d-1}, with $0 < d < 1/2$). When the autoregressive lag polynomial $\phi(B) = 1 - \alpha(B) - \beta(B)$ in the ARMA(m, q) representation of the GARCH(p, q) process, contains a unit root, the GARCH(p, q) process is said to be integrated in variance (Engle and Bollerslev, 1986). The **Integrated GARCH(p, q)** or **IGARCH(p, q)** class of models is given by

$$\phi(B)(1 - B)X_t^2 = \omega + (1 - \beta(B))\nu_t.$$

For this class of models one has $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$ and, consequently, the uncondicional variance in undefined.

The **Fractionally Integrated GARCH(p, d, q)**, or **FIGARCH(p, d, q)** class of models is simply obtained by allowing the differencing operator in the above equation to take non-integer values

$$\phi(B)(1-B)^d X_t^2 = \omega + (1-\beta(B))\nu_t, \tag{1.15}$$

with $1-\beta(B)$ and $\phi(B)$ representing lag polynomials of orders q and p, respectively, and the roots of both $\phi(z)=0$ and $\beta(z)=1$ lying outside the unit circle. d is the fractional differencing parameter and the fractional differencing operator is most conveniently expressed by its Maclaurin series expansion,

$$(1-B)^d = 1 - dB - \frac{d(1-d)}{2!}B^2 - \frac{d(1-d)(2-d)}{3!}B^3 - \cdots = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k, \tag{1.16}$$

where $0 < d < 1/2$. If $1-\beta(B)$ is invertible, then the FIGARCH(p, d, q) model can be expressed as an ARCH(∞)-process writing

$$\begin{aligned} \phi(B)(1-B)^d X_t^2 &= \omega + (1-\beta(B))(X_t^2 - \sigma_t^2) \Leftrightarrow \\ (1-\beta(B))\sigma_t^2 &= \omega + (1-\beta(B))X_t^2 - \phi(B)(1-B)^d X_t^2 \\ \sigma_t^2 &= \frac{\omega}{1-\beta(B)} + \lambda(B)X_t^2, \end{aligned}$$

where $\lambda(B) = 1-(1-\beta(B))^{-1}\phi(B)(1-B)^d$. For the FIGARCH($p, d, q$) model to be well-defined and the conditional variance positive almost surely for all t, all the coefficients in the ARCH(∞) representation must be non-negative. General conditions, however, are difficult to establish. For the FIGARCH(1,d,1) model the infinite series coefficients can be obtained recursively (please refer to the FIAPARCH(1,d,1) model in Chapter 2, Section 2.1, for the explicit form of these recursions) and from this recursions it was shown by Bollerslev and Mikkelsen (1996) that conditions

$$\beta - d \le \phi \le \frac{2-d}{3}, \qquad d\left(\phi - \frac{1-d}{2}\right) \le \beta(\phi - \beta + d), \tag{1.17}$$

are sufficient to ensure non-negativity. In the GARCH(p, q) model that is covariance stationary, shocks to the conditional variance dissipate exponentially, meaning that the effect of a shock on the forecast of the future conditional variance tends to zero at a fast exponential rate. In the IGARCH(p, q) model, shocks to the conditional variance persist indefinitely, meaning that the shocks remain important for all horizon forecasts. In the FIGARCH(p, d, q) model, the differencing parameter introduces a different behaviour: the effect of a shock to the forecast of the future conditional variance is expected to die out at a slow hyperbolic rate. This is the reason why the FIGARCH(p, d, q) process is said to have long memory in volatility.

Tse (1998) modifies the FIGARCH(p, d, q) process to allow for asymmetries, thus originating the **Fractionally Integrated Asymmetric Power ARCH of order (p, d, q), FIAPARCH(p, d, q)** process. Defining $g(X_t) = (|X_t| - \gamma X_t)^\delta$, with $|\gamma| < 1$ and $\delta \geq 0$, the FIAPARCH(p, d, q) model can be written as

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^\delta &= \frac{\omega}{1-\beta(B)} + \lambda(B) g(X_t), \end{aligned} \tag{1.18}$$

where

$$\lambda(B) = 1 - (1-\beta(B))^{-1}\phi(B)(1-B)^d = \sum_{i=1}^{\infty} \lambda_i B^i, \ (\lambda(1) = 1), \tag{1.19}$$

for every $0 < d < 1$, with $\lambda_i \geq 0$, for $i \in \mathbb{N}$, and $\omega > 0$, for the conditional variance to be well defined, so that it is positive almost surely for all t. In order to allow for long memory, the fractional differencing parameter, d, is constrained to lie in the interval $0 < d < 1/2$. Moreover, the polynomials $1 - \beta(B)$ and $\phi(B)$ are assumed to have all their zeros lying outside the unit circle. The fractional differencing operator $(1 - B)^d$ is once again expressed as (1.16).

The FIAPARCH model nests two major classes of ARCH-type models: the APARCH and the FIGARCH models. When $d = 0$ the process reduces to the APARCH(p, q) model, whereas for $\gamma = 0$ and $\delta = 2$ the process reduces to the FIGARCH(p, d, q) model. The FIGARCH representation includes the GARCH (when $d = 0$) and the IGARCH (when $d = 1$), with the implications in terms of impact of a shock on the forecasts of future conditional variances, as discussed above. Considering all the features involved in this specification, Conrad et al. (2011) point out some advantages of the FIAPARCH(p, d, q) class of models, namely

(a) it allows for an asymmetric response of volatility to positive and negative shocks, so being able to traduce the leverage effect;

(b) in this particular class of models it is the data that determines the power of returns for which the predictable structure in the volatility pattern is the strongest;

(c) the models are able to accommodate long memory in volatility, depending on the differencing parameter d.

It is important to mention here that necessary and sufficient conditions for the existence of a stationary solution of the APARCH(p, q) model can be easily obtained from the results derived by Liu (2009). This author introduced a family of GARCH processes, which can be regarded as a class of non-parametric GARCH processes, which include as a special case the APARCH(p, q) model. Liu (2009) obtained necessary and sufficient condition for the existence of a stationary solution of this new family of GARCH processes. Furthermore, Liu (2009) also derived an explicit expression for the stationary solution. In contrast, however, the statistical properties of the general FIGARCH(p, d, q) process remain unestablished. Namely, stationarity is not a certainty as well as the source of long memory on volatility or even its existence are nowadays controversial. For the FIAPARCH process, Tse (1998) also leaves these issues as open questions.

1.2.2 Optimal Alarm Systems

Recently, it has been recognized the potential of optimal alarm systems in detecting and warning the occurrence of catastrophes, and the spectrum of applications of optimal alarm systems is wide and yet to be explored. As examples of applications we can mention the prediction and warning for high water levels at the Danish coast in the Baltic Sea, in Svenson and Holst (1998), or the evaluation of the performance of a water-level predictor as part of a flood warning system, in Beckman et al. (1990). In a different context from that of optimal alarm systems, Guillou et al. (2010), proposed an approach based on Extreme Value Theory (EVT) for the early detection of time clusters[3] in weekly counts of *Salmonella* isolates, reported to the national surveillance system in France. The method checks if each new observation corresponds to an unusual or extremal event and the authors propose its integration in public health surveillance agencies.

Timely public health intervention is fundamental and there are many situations in which surveillance is critical. For instance, atmospheric concentrations of air pollutants constitute real-valued time series that can be analysed under the perspective of up-crossings of some critical levels. Take as examples the studies by Smith et al. (2000) (a combined analysis of daily mortality data from Birmingham, Alabama with PM_{10} − particulate matter

[3] A time cluster is defined in Guillou et al. (2010), as a time interval in which the number of observed events is significantly higher than the expected number of events in a given geographic area. Also, by *event* it is meant any event of interest, generally related to public health, such as cases of illness, admissions to hospitals, number of deaths, etc.

of aerodynamic diameter 10μm or less $-$ and meteorological variables such as minimum and maximum daily temperature and humidity), or Tobías and Scotto (2005) (ground ozone levels, collected in two measurement stations in Barcelona, were analysed under another approach based on EVT, namely, the Peak Over Threshold method; ozone levels higher than a certain threshold can be health-hazards for human health and the analysis revealed that the threshold value was exceeded many times in both stations). For further references on the existent literature on the relation between air pollution and mortality, see Koop and Tole (2004). In this paper the authors discuss the importance of model uncertainty for accurate estimation of the health effects of air pollution, and propose Bayesian model averaging procedures to reduce the uncertainty and inaccuracy of the empirical estimates. They also illustrate their method with a comprehensive data set for Toronto, Canada, taking a certain measure of mortality as the dependent variable and twelve explanatory variables (seven different pollutants and five different weather variables) with all possible associations between them.

The impact of air pollution and other environmental factors on public health is indeed a significant area of environmental economics that can benefit from an accurate prediction of level-crossings. As an example of a potential application of an optimal alarm system for count processes, we can refer the extensive study by Katsouyanni et al. (2001) on the relationship between mortality and air pollution: the daily number of deaths in 29 European cities were analysed in relation to exposure and levels of PM_{10}, black smoke, sulphur dioxide, ozone and nitrogen dioxide, for periods ranging from 1990 to 1997. Other potential applications related to environmental economics are suggested by the works of Touloumi et al. (2004) (short-term health effects of air pollution, considering PM_{10} and nitrogen dioxide, were analysed in a hierarchical modelling approach, using data from 30 cities across Europe), Braga et al. (2001) (counts of daily deaths, modelled through a generalized additive Poisson regression model were related to temperature and relative humidity; data was collected from 12 cities in the United States and several covariates were considered, namely, season, day of the week and barometric pressure), Campbell et al. (2001) (relation between atmospheric pressure and sudden infant death syndrome, in Cook, Chicago) or Schwartz et al. (1997) (association between daily measures of drinking water turbidity and both emergency visits and hospital admissions for gastrointestinal illness at the Children's Hospital of Philadelphia).

Another area of potential application of optimal alarm systems is econometrics and, in particular, risk management. The implementation of probabilistic models for the assessment of market risks or credit risks is mandatory. The paper by Thomas (2000) gives an overview of the history, objectives, techniques and difficulties of credit scoring and behaviour scoring[4]. It also points out how successful the area of forecasting financial risk has become in the last thirty years and how credit scoring would profit even more if it would be possible to change the procedure of estimating the probability of a consumer defaulting to estimating the profit a consumer would bring to the lending organisation. Another example of the implementation of probabilistic models for the assessment, in this case, of market risks, is the forecasting of daily stock volatility in Fuertes et al. (2009), and in all the references therein. Unlike prices or volume, volatility is not directly observable. Nevertheless, forecasting the conditional variance of stock market returns has

[4]Credit scoring relates to the techniques that help organizations decide whether or not grant credit to new applications. Behaviour scoring is the set of tools that aid in the decisions related to existing costumers. What kind of marketing to aim at any particular client? Should the organization agree if the client wants to increase its credit limit? What actions should the firm take if the client starts to fall behind his payments? These are questions that the techniques involved in behaviour scoring aim to answer.

implications in option pricing, portfolio management, Value-at-Risk, and financial market regulation. Yet another socio-economical area that could highly benefit from the perspective of an optimal alarm system is revenue management. It is estimated that efficient revenue management can have profit improvement around 3%-7% in airline, hotel and car rental industries. It enables a company to maximize profit given the same number of units sold, and, one of the core concepts behind it, is the reservation of a portion of capacity for higher value customers at a later date (later booking clients are not as price sensitive as the lower value segment of early booking costumers). Obviously, then, forecasting capability is a competitive advantage in revenue management. Weatherford and Kimes (2003), used data from Choice Hotels and Marriot Hotels to test several forecasting methods of arrival of guests. Brännäs et al. (2002) modelled monthly guest nights in hotels through an integer-valued moving average model. They presented empirical results for a series of Norwegian guests in Swedish hotels and the results indicated strong seasonal patterns in mean check-in and check-out probabilities. Shortly after, Brännäs and Nordström (2004), presented an integer-valued autoregressive model in which the capacity constraint was an integral part. The duration of stay for the visitor and the occupancy probability are measures that can be inferred from the model. The effects of price changes and of the existence of a large festival, on these measures, were empirically assessed with the model. Later on, Brännäs and Nordström (2006), proposed extensions to the basic autoregressive binomial model in which the capacity constraint is an integral part, in order to make it possible to evaluate the effects of festivals and to account for time-variation in the parameters and the capacity constraint. The empirical impact of festivals on tourist accommodation was studied on daily accommodation time series for hotels and cottages in Stockholm and Gothenburg from January 1993 to August 1999. The authors found that festivals had a clear positive impact on tourism demand in both cities. Nevertheless, any of the models mentioned here is directly applicable to calculate in advance the probability of, for instance, the tourism demand being higher than the accommodation capacity.

Actually, except for the references of Svenson and Holst (1998) and Beckman et al. (1990), mentioned early in this section, no other reference addresses the problem of calculating in advance the probability of future up/down-crossings, in the sense of event prediction. It is in this context that the implementation of an (optimal) alarm system reveals to be useful. In what follows, we will start by introducing simple definitions regarding alarm systems and give the historical perspective of the developments in this area. The formalism used throughout this work will also be presented in this section, following naturally after the references more related to it. Finally, aiming at offering a broadening perspective on the wide field of applications of optimal alarm systems, recent contributions on this area will be briefly discussed. Some alternative approaches to the construction of alarm systems will also be addressed.

An **alarm system** is an algorithm which, based on current information, predicts whether a level-crossing event is going to occur at a specified time in the future. As to remind that level-crossings do sometimes have very drastic consequences, the designation **catastrophe** is commonly used. Considering level-crossing events, we can distinguish between one and two-sided cases and also between an exceedance and an up-crossing event. An exceedance is a one-dimensional level-crossing event, where some critical level or threshold u is exceeded by a process at one single time point. An up-crossing is a two-dimensional level-crossing event, involving two adjacent time points: the process is below the critical threshold at the first time point and above the threshold at the second time point. Both cases described refer to one-sided exceedances or up-crossings. The two-sided case involves a level-crossing event that spans many time points, exceeding upper levels and fading below lower levels, symmetric about the mean of the process, as defined in Martin (2012). Only recently the two-sided case has been investigated, as will be mentioned below. Throughout this work a catastrophe will be considered as the up-crossing event

$$C_{t,j} = \{X_{t+j-1} \leq u < X_{t+j}\},$$

for some $j \in \mathbb{N}$ and some real u.

Conceptually, the simplest way, or naive way, of constructing an alarm system is to predict X_{t+h} by a predictor say, $\hat{X}_{t+h}$, which is usually chosen so that the mean square error is minimized, providing

$$\hat{X}_{t+h} = \mathrm{E}[X_{t+h}|X_s, -\infty < s \leq t].$$

An alarm is given every time the least squares predictor exceeds some critical level, i.e., the event that $\hat{X}_{t+h}$ up-crosses $\hat{u}$ foretells that X_{t+h} will up-cross u, for some $\hat{u} < u$ (given that $\mathrm{Var}[\hat{X}_{t+h}] < \mathrm{Var}[X_{t+h}]$ and the detection probability is desired to be reasonably high). This **value prediction** alarm system was defined in Lindgren (1985) as the ability to predict future process values of a stationary linear Gaussian random process, in the sense of least squares. It is very often designated as a **naive alarm system**. This alarm system, however, does not have a good performance on the ability to detect level-crossing events, locate them accurately in time and give as few false alarms as possible. If the intent is to predict whether or not X_{t+h} will exceed some critical threshold u for some $h > 0$, one can not only use the single value $\hat{X}_{t+h}$ for each particular h, but must also consider the predicted change rate. In this **event prediction** based alarm system, the prediction capability must be judged by the system's ability to accurately predict a level-crossing event. An event prediction is constructed on a defined prediction horizon and involves level-crossings of a given critical level by the random process.

At each moment, the algorithm of the alarm system signals whether or not a catastrophe is bound to happen h time steps ahead. An alarm is a **false alarm** if, after an alarm signal, no catastrophe occurs at the specified time; a catastrophe is said to be **undetected** if the catastrophe occurs without the previous alarm signalling. The success of the alarm system is measured by its false alarm rate and by its detection probability and the definition of *optimal* alarm involves a compromise between these two characteristics. In what follows, whenever we mention **optimal alarm system**, it is meant as defined in Grage et al. (2010).

Definition 1.1. ***Optimal alarm system:*** *An optimal alarm system for a specified set of available data is defined as a system which, for a given probability of detecting a catastrophe, has the highest probability of correct alarm.*

Following these considerations, Lindgren (1985) and de Maré (1980) set the principles for the construction of optimal alarm systems. Establishing the analogy between alarm systems and hypothesis testing, de Maré (1980), developed a general context optimal alarm system based on a likelihood-ratio argument. The alarm problem can be thought of as a hypothesis test where one has to choose whether to give an alarm or not. de Maré showed that the Neyman-Pearson lemma gives a condition for this test to be optimal. Lindgren (1985) restated this condition, giving an explicit formulation of the optimal alarm system in terms of the pair predicted value/predicted growth rate, for a Gaussian stationary process. The optimal alarm system is bound to give an alarm when the prediction exceeds a variable alarm level that adjusts according to the expected growth rate of the process. The optimal alarm condition is then, fundamentally, an alarm region (or decision boundary) that is defined by the likelihood ratio between predicted value and growth rate.

Crucial to the measurement of the success of the alarm system are the **operating characteristics**, *probability of correct alarm* and *probability of detected catastrophe*, introduced by Lindgren (1975b) and developed by Beckman (1987). In Svensson et al. (1996) a comparison between a naive alarm system (based only on the predictions of the process values) and an optimal alarm system (that depends on the expected growth rate of the process) was carried out. The operating characteristics mentioned before were calculated and the optimal catastrophe predictor was found to perform much better than the naive predictor. For stationary stochastic processes, Svensson et al. (1996), showed that the likelihood ratio criterion, as introduced by de Maré, is equivalent to a conditional inequality that compares the conditional probability of catastrophe with the level P_b, the border probability. Formulated in these terms, the alarm region is parametrized in terms of predicted future process values (conditional probability of level-crossing) and P_b. P_b turns out to be an extremely important parameter as it effectively defines the interval spanned by the alarm region. Consequently, the border probability is a key parameter in the control of the trade-off between the number of false alarms and undetected events. A detailed analysis of the border probability and it's effect on the resulting operating characteristics of the system must be carried out when designing, in practice, an optimal alarm system. In Svenson and Holst (1998), the principles of optimal prediction of level-crossings were applied to the sea levels of the Baltic sea. Other basic results regarding optimal prediction of level-crossings were obtained by Lindgren (1975a,b, 1980), Svensson (1998) and Svensson and Holst (1997).

It is worth to mention that the alarm system introduced by Lindgren and de Maré, ignores the sampling variation of the model parameters. General practice in all of the above mentioned references involves carrying out the calculations needed for the construction of the optimal alarm system considering that the parameters in the stationary process are known, and, then, in appropriate time, replacing them by their proper estimates. The stationarity assumption is required and the variation of the parameter estimates is not considered when computing, for instance, the operating characteristics of the alarm system. Giving heed to this issue, Amaral-Turkman and Turkman (1990) suggested the formulation of a Bayesian predictive approach. In this work the authors have basically kept all the notions and principles of Lindgren (1985) but replaced the probabilities by their predictive counterparts. Particular calculations were carried out for stationary autoregressive processes of order 1, AR(1). The computational burden, however, was not solved until the work by Antunes et al. (2003) where the operating characteristics of the alarm system were numerically obtained. Further extensions and generalizations were also proposed as the authors extended the application of the optimal alarm systems to random walks and autoregressive models of order p. Also, the authors introduced what they defined as **on-line alarm systems**, where posterior probabilities are updated at each time point, as opposed to **off-line alarm systems**, where the alarm systems are constructed for unconditional events, which, by the assumption of stationarity, are assumed to have the same probability over time. In practice, a process is continuously observed in time and this information should be used to update the probability of the events under consideration, in particular, the probability of the up-crossing event. Following this on-line event prediction perspective, the assumption of stationarity can be relaxed.

All the references cited before relate to real-valued stochastic processes. Extending the areas of application of optimal alarm systems, Monteiro et al. (2008) addressed the development of alarm systems for time series of counts represented through integer-valued autoregressive models where the parameters are functions of covariates of interest and vary on time. The Doubly Stochastic INteger-valued AutoRegressive process of order 1, DSINAR(1), was considered and both classical and Bayesian methodologies were used in the construction of the optimal alarm system. An empirical application was done, considering the number of sunspots (areas of reduced surface temperature that appear visibly as dark-spots on the photosphere of the sun) on the surface of the sun. Though the optimality conditions were met, the authors reported a rather high number of false alarms, both in the simulation study and in the working example, a similar result to the one found by Svenson and Holst (1998) in the analysis of high water levels at the Danish coast in the Baltic sea.

As the optimal alarm systems constructed is this work follow closely the approaches of Antunes et al. (2003) and Monteiro et al. (2008), we will take here the opportunity to present the theoretical fundamentals of the method, concerning basic definitions and operating characteristics. The development of an optimal alarm system for the continuous case, with application to the FIAPARCH(1,d,1) model, will be done in Chapter 2. The development of an optimal alarm system for the discrete case, with application to the INAPARCH(1,1) model will be done in Chapter 3. In order to maintain the generality of the presentation, the details and particularities of the applications will be postponed until implementation is done and several questions regarding methodology arise.

Let (X_t) be a discrete parameter stochastic process with parameter space $\Theta \subset \mathbb{R}^k$, for some fixed $k \in \mathbb{N}$. The time sequel $\{1, 2, \ldots, t-1, t, t+1, \ldots\}$ is divided into three sections, $\{1, 2, \ldots, t-q\}$, $\{t-q+1, \ldots, t\}$, and $\{t+1, \ldots\}$, namely, the past, the present and the future. For some $q > 0$, the sets $D_t = \{X_1, \ldots, X_{t-q}\}$, $\mathbf{X_2} = \{X_{t-q+1}, \ldots, X_t\}$ and $\mathbf{X_3} = \{X_{t+1}, \ldots\}$ represent, respectively, the data or informative experience, the present experiment and the future experiment, at time point t.

Definition 1.2. *The event of interest, $C_{t,j}$, is defined as a catastrophe and is any event in the σ-field generated by $\mathbf{X_3}$.*

As already mentioned, throughout this work the catastrophe will be considered as the up-crossing event of the fixed level u, at time point $t + j$,

$$C_{t,j} = \{X_{t+j-1} \leq u < X_{t+j}\},$$

for some $j \in \mathbb{N}$ and for some real u. In a particular application, the down-crossing event of the fixed level u, at time point $t + j$,

$$C_{t,j} = \{X_{t+j-1} \geq u > X_{t+j}\},$$

for some $j \in \mathbb{N}$ and for some real u, will be considered, instead.

Definition 1.3. *Any event $A_{t,j}$ in the σ-field generated by $\mathbf{X_2}$, predictor of $C_{t,j}$, will be an event predictor or alarm.*

It is said that an alarm is given at time t, for the catastrophe $C_{t,j}$, if the observed value of $\mathbf{X_2}$ belongs to the predictor event or alarm region. In addition, the alarm is said to be correct if the event $A_{t,j}$ is followed by the event $C_{t,j}$. Thus, the probability of correct alarm will be defined as the probability of catastrophe conditional on the alarm being given. Conversely, a false alarm is defined as the occurrence of $A_{t,j}$ without $C_{t,j}$. If an alarm is given when the catastrophe occurs, it is said that the catastrophe is detected and the probability of detection will be defined as the probability of an alarm being given conditional on the occurrence of the catastrophe.

Definition 1.4. *The alarm region $A_{t,j}$ is said to have size $\alpha_{t,j}$ if $\alpha_{t,j} = P(A_{t,j}|D_t)$.*

Note that $\alpha_{t,j}$ can be understood as the proportion of time spent in the alarm state.

Definition 1.5. *The alarm region $A_{t,j}$ is optimal of size $\alpha_{t,j}$ if*

$$P(A_{t,j}|C_{t,j}, D_t) = \sup_{B \in \sigma_{\mathbf{X_2}}} P(B|C_{t,j}, D_t), \tag{1.20}$$

where the supreme is taken over all sets $B \in \sigma_{\mathbf{X_2}}$ such that $P(B|D_t) = \alpha_{t,j}$.

The alarm region $A_{t,j}$ is optimal, if it has the highest detection probability, among all regions with the same alarm size. The optimality condition could be defined in another way. We could define an alarm region to be optimal if it satisfies Definition 1.1, i.e., if it has the highest probability of correct alarm (or, it gives the least number of false alarms) for a given probability of detection. However, as stated in Lemma 1, in Antunes et al. (2003), these definitions lead to the same alarm region.

Definition 1.6. *An optimal alarm system of size $(\alpha_{t,j})$ is a family of alarm regions $(A_{t,j})$ in time, satisfying (1.20).*

The following lemma is equivalent to Lemma 4.1. of Lindgren (1985) and follows closely the notation of Antunes et al. (2003) and Monteiro et al. (2008).

Lemma 1.7. *Let $p(\mathbf{x_2}|D_t)$ and $p(\mathbf{x_2}|C_{t,j}, D_t)$ be the predictive density of $\mathbf{X_2}$ and the predictive density of $\mathbf{X_2}$ conditional on the event $C_{t,j}$, respectively. Then, the alarm system $(A_{t,j})$ with alarm region given by*

$$A_{t,j} = \left\{ \mathbf{x_2} \in \mathbb{R}^q : \frac{p(\mathbf{x_2}|C_{t,j}, D_t)}{p(\mathbf{x_2}|D_t)} \geq k_{t,j} \right\},$$

or, equivalently,

$$A_{t,j} = \left\{ \mathbf{x_2} \in \mathbb{R}^q : \frac{P(C_{t,j}|\mathbf{x_2}, D_t)}{P(C_{t,j}|D_t)} \geq k_{t,j} \right\},$$

for a fixed $k_{t,j}$ such that $P(\mathbf{X_2} \in A_{t,j}|D_t) = \alpha_{t,j}$, is optimal of size $\alpha_{t,j}$.

If (X_t) is an integer-valued process, simple adaptations of the previous lemma are required. In the discrete case, $p(\mathbf{x_2}|D_t)$ represents the predictive probability of $\mathbf{X_2}$ and $p(\mathbf{x_2}|C_{t,j}, D_t)$, the predictive probability of $\mathbf{X_2}$ conditional on the event $C_{t,j}$. In this case, $\mathbf{x_2} \in \mathbb{N}_0^q$, also. This lemma ensures that the alarm region defined above renders the highest detection probability. Moreover, to enhance the fact that the optimal alarm system depends on the choice of $k_{t,j}$, it is important to stress that in view of the fact that $P(C_{t,j}|D_t)$ does not depend on $\mathbf{x_2}$, the alarm region can be rewritten in the form

$$A_{t,j} = \{\mathbf{x_2} \in \mathbb{R}^q : P(C_{t,j}|\mathbf{x_2}, D_t) \geq k\}, \tag{1.21}$$

where $k = k_{t,j}P(C_{t,j}|D_t)$ is chosen in some optimal way to accommodate conditions over the operating characteristics of the alarm system.

Definition 1.8. *The following probabilities are called the operating characteristics of an alarm system:*

1. $P(A_{t,j}|D_t)$ *- Alarm size,*
2. $P(C_{t,j}|A_{t,j}, D_t)$ *- Probability of correct alarm,*
3. $P(A_{t,j}|C_{t,j}, D_t)$ *- Probability of detecting the event,*
4. $P(\overline{C}_{t,j}|A_{t,j}, D_t)$ *- Probability of false alarm,*
5. $P(\overline{A}_{t,j}|C_{t,j}, D_t)$ *- Probability of undetected event.*

The choice of k will depend on a compromise between maximizing the probabilities of correct alarm and of detecting the event. As it is not possible, in general, to maximize both alarm characteristics simultaneously, some criteria must be met in order that the alarm system achieves a satisfactory behaviour. Several criteria have already been proposed in the literature. We will address this issue and discuss some criteria already proposed, further on, when dealing with the application of the alarm system to particular situations.

To complement this section, we will now present some recent contributions regarding the construction of alarm systems. Different perspectives and methodologies related to this field are given, as a reinforcement to the original claim that the spectrum of applications of optimal alarm systems is wide and yet to be explored.

It is known that neural network algorithms are often successfully used to produce predictions of non-Gaussian time series, generally based on the minimization of a quadratic loss function. When one wants to predict whether or not the time series will exceed a certain fixed level, as is the case in event prediction, the mean square error is not very useful as a loss function. Several modifications of the standard network algorithms have been proposed in the literature to improve performance in warning for exceedance. Take, as examples, the prediction of episodes of poor air quality in Nunnari (2006), or the forecasting of ozone peaks and exceedance levels in Dutot et al. (2007); see also Cawley et al. (2007) for a review of the existing methodologies related to artificial neural networks for estimating predictive uncertainty and dealing with decision making processes. In Grage et al. (2010) the authors investigated to what degree an artificial neural network can approximate an optimal alarm system. The authors applied two neural network models to Gaussian as well as non-Gaussian stochastic processes and compared their behaviour with the behaviour of a naive and an optimal catastrophe predictor. In all cases, the network models were much better than the naive predictor but not quite as good as the optimal predictor (which, as already mentioned, for a Gaussian stationary process, can be explicitly specified in terms of the predicted value of the process itself and of its derivative). Anyway, the authors were able to show that a neural network can be trained to approximate an optimal alarm system arbitrarily well.

Another recent approach to level-crossing prediction was taken by Martin (2010) who combined the use of Kalman filtering[5] with the design of an optimal alarm system for a zero-mean stationary linear dynamical system driven by Gaussian noise, in state-space form. The author found a negligible loss in accuracy by using approximations to the theoretical optimal predictor and the major advantage of much less computational complexity. The negligibility of the loss in accuracy was demonstrated by comparing approximations of the optimal level-crossing predictor to baseline methods: the approximations clearly outperformed the baseline methods. To our knowledge, and to the exception of Kerr (1982) who evaluated tightened upper bounds on the false alarm and correct detection probabilities in an optimized Kalman filter-based failure detection algorithm, the level-crossing event considered in Martin (2010), was the first reference of a two-sided level-crossing event. Recall that two-sided level-crossing means spanning many time steps and exceeding upper and lower levels, symmetric about the mean of the process. A more detailed extreme value analysis was done in a subsequent paper by Martin (2012). The author considered as the level-crossing event at least one exceedance outside the threshold envelope $[-L, L]$, within the specified step-ahead prediction window. In the construction of the alarm system,

[5] The Kalman filter is a recursive filter that estimates the internal state of a linear dynamic system from a series of noisy measurements. A system's state, represented by some linear dynamical model can be inferred through measured data. Sensor noise, approximations in the equations and external factors not accounted for, introduce uncertainty about the inferred values. The Kalman filter averages a prediction of a system's state with a new measurement using a weighted average based on the model's covariance, a procedure that results in a new state estimate lying between the predicted and the measured state, and having a better estimated uncertainty than either the predicted or the measured state. The designation of recursive filter comes from the fact that this process is repeated every time step, with the new estimate and its covariance serving as information for the prediction used in the following iteration.

four different situations were considered: two baseline alternatives and two different approximations to the optimal alarm system (namely the *closed-form* and the *root-finding* approximations). For each of these situations, the alarm region, the alarm probability and the true positive and false positive rates (conditional probabilities that relate to the operating characteristics of the alarm system) were obtained, in explicit form, whenever possible. An extreme value analysis was carried out considering the limiting cases when $L \to 0$ (small value level-crossing prediction) and $L \to \infty$ (large-value level-crossing prediction). The author concluded that, given the assumed technical conditions, level-crossings of a linear Gaussian process can be predicted with the greatest accuracy for extremely high levels or very low measurement noise (although intuitive, this last result was actually finally supported by rigorous theoretical proof).

Recently, Das and Kratz (2010, 2012) developed an alarm system as a strategy for capital allocation in insurance institutions. Although based on the Cramér Lundberg model[6], and different in principle from the construction of an optimal alarm system, there are some similarities with it, in particular, in what concerns the alarm time being dependent on a critical value of a conditional probability. An alarm is given at some time point, when the conditional probability of ruin, given survival up to the alarm time, say α, is high, in the absence of any intervention. Also, it is required that the

[6] The Cramér Lundberg model, or the classical compound-Poisson risk model, represents the theoretical foundation of ruin theory. It describes the risk of an insurance company experiencing two opposing cash flows: incoming premium collection and outgoing claim settlement. A simple model for the risk, V_t, can be written as $V_t = u_t + ct - \sum_{i=1}^{N_t} X_i$, where it is assumed that premiums arrive at constant rate c, and claims, X_i's, arrive according to a Poisson process with intensity λ and are i.i.d. non-negative random variables with distribution F and mean μ, forming a compound Poisson process. u_t denotes the capital function at time t. The ruin time of the risk process is formally defined as $T = \inf\{t > 0 : V_t < 0\}$, with $T = \infty$ if there is no ruin.

probability of non-ruin before the alarm, say β, is sufficiently high. In this paper, then, the conditions for giving an alarm are constructed on two intuitive requirements or empirical properties of the alarm system: at the alarm time, the probability of ruin in *not so distant future*, say d, is *substantial* if no action is taken, and, the probability of the system getting ruined before the alarm time is *minimal*. The time window d, the probabilities α and β and the initial capital u constitute the parameters of the alarm time model. An alarm system consisting of a sequence of alarms is defined as a natural extension of the single alarm, and capital addition occurs whenever an alarm is given. To test the effectiveness of this method, the authors compared the survival probabilities of the proposed alarm system and of an alternative system, with no alarms but with a higher initial capital (equivalent in total to the capital added in the alarm case). The simulations revealed the better performance of the alarm system in the long run, meaning, higher survival probability over finite horizon.

Following the works of Cirillo et al. (2010) and Cirillo and Hüsler (2011) a Bayesian non-parametric approach to catastrophe prediction was proposed in Cirillo et al. (2013). This innovative approach uses urn processes, which are a very large family of probabilistic models in which objects of real interest are represented as coloured balls in one or more urns or boxes. The probabilities of certain events are expressed in terms of sampling, replacing and adding balls to the urns. The particular construction in Cirillo et al. (2013), is part of the class of **reinforced urn processes**, RUP, introduced in Muliere et al. (2000), as reinforced random walks on a state space of urns. Using this models, several recent applications in level-crossing or catastrophe prediction can be found in the literature. In the first above mentioned reference, Cirillo et al. (2010), presented a general recursive model constructed by the means of interacting Polya urns[7], to model the dependence among failures both within and between k groups of failing systems (systems whose probability of failure is not negligible in a fixed time horizon). The examples of failing systems range from financial portfolios and credit risk to electrical and mechanical systems or even the world wide web itself. An application is presented to credit risk modelling. Another example of the application of RUP in risk modelling can be found in Amerio et al. (2004), where a stochastic model for credit default for debt issuers belonging to the same Moody's rated class is proposed. In the second reference mentioned in this paragraph, attention is given to systems that are subject to shock of random magnitude, at random times. If the systems break down when some shock overcomes a given resistance level, extreme shock models, as introduced in Gut and Hüsler (1999), become appropriate to describe the situation. Cirillo and Hüsler (2011) proposed an alternative approach to extreme shock models using reinforced urn processes, providing the predictive distribution of system's defaults under a Bayesian non-parametric perspective. Yet other applications of RUP can be found in Bulla (2005), related to survival analysis, or in Mezzetti et al. (2007), in determining the maximum tolerated dose in clinical trials for new drug development. For a survey of processes with reinforcement and other applications, see Pemantle (2007).

The urn-based alarm system proposed in Cirillo et al. (2013) can be con-

[7]Polya urn was introduced by Eggenberger and Pólya (1923) to model the diffusion of infectious diseases and study self-reinforcing phenomena. In its simplest version, an urn containing balls of two different colors is considered. Every time the urn is sampled, the color of the chosen ball is registered and the ball is put back into the urn, together with another ball of the same color. In this way, the more a given color has been sampled in the past, the more likely it will be sampled in the future. The reinforcement scheme can be generalized, for instance, introducing s balls of the same color, or considering a random or time-varying reinforcement.

stantly updated because of reinforcement, according to Bayesian paradigm. Towards a better understanding of the capabilities of the construction proposed, we are going to, very briefly, introduce the fundamentals of the urn-based alarm system.

- $V = \mathbb{N}_0^+ \times \{0, 1, 2, \ldots, L\}$ is a state space, whose elements (n, l) represent levels of risk l, at time instant t. The level of risk L corresponds to a catastrophe. $v \in V$ represents a general state in V and (n, l) represents a specific couple $(n, l) \in V$.

- Every state $v \in V$ is endowed with a Polya urn $U(v)$, i.e., a urn that is sampled, with replacement and reinforcement: every time a ball of a given colour is sampled, the ball is replaced and $s(v) > 0$ extra balls of the same color are added to the urn. Thanks to the reinforcement mechanism the process is able to learn from the past, as will be explained.

- Every urn $U(v)$ is characterized by a set of 4 colours, $C = (c_1, c_2, c_3, c_4)$ and a reinforcement $s(v)$. $m_v(c)$ represents the number of balls of colour c in urn $U(v)$. Notice that $m_v(c)$ may be equal to 0 for some c.

- The composition of each urn $U(v), v \in V$ is given by the sum of all the balls in the urn.

- There is a function $d : V \times C \longrightarrow V$ that represents a rule of motion, one of the most important elements in this construction.

Fix an initial state $(0, 0)$ and recursively define a reinforced random walk (X_n) on V, starting in $(0, 0)$, $X_0 = (0, 0)$. For all $n \geq 1$, if $X_{n-1} = v \in V$, a ball is sampled from $U(v)$, its colour registered and the ball is returned to the urn together with $s(v)$ balls of the same colour. According to the rule of motion d, an element of the state space will be attributed to X_n, $X_n = d(v, c)$. At this moment it is worth mentioning that the process (X_n), taking values on a finite-valued state space, can be viewed as the simplified version of another underlying process (Y_n), that may be characterized by a much larger space of states. For instance, consider that the underlying process (Y_n) can take values on a scale that goes from 0 to $+\infty$. As long as it is possible to define only $L+1$ different regions of risk for the process (Y_n), it is possible to establish a direct correspondence between the region of (Y_n) and the value (X_n) assumes in the scale from 0, no risk, to L, catastrophe. Of course, this data categorization implies loss of information, that should be minimized the larger the number of risk categories and the smoother the underlying process (Y_n). Moving on, the rule of motion establishes which level of risk will be occupied at the next time step, given the colour of the ball sampled from urn $U(v)$. Suppose we are in $X_{n-1} = (n, l) \in V$ and sample a ball of colour c from $U((n, l))$. Then, $X_n = d((n, l), c)$ will be something like $X_n = (n + 1, l')$, where l' is a new level of risk that is determined by the sampled colour c (for instance, we can have, $l' = l$ if $c = c_1$; $l' = l + 1$ if $c = c_2$; $l' = l - 1$ if $c = c_3$ and $l' = L$ if $c = c_4$). Of course, as the levels of risk are finite, from 0 to L, limiting conditions must be imposed on urn composition to avoid nonsense situations: take the example given and consider that $X_{n-1} = (n, 0)$, then, the urn $U(n, 0)$ must have zero c_3 balls, as the risk level cannot decrease and no c_3 ball can be sampled.

Consider a finite sequence $\psi = ((0, 0), \ldots, (n, l))$ of elements of the state space V. Such sequence is said admissible for the process (X_n) if its elements can effectively be visited by (X_n) in the order given by ψ. From the probability $P((X_n) = \psi)$, calculated in Cirillo et al. (2013), the authors concluded that the process (X_n) is partially interchangeable, in the sense of Diaconis and Freedman (1980). If it is also recurrent, then, as shown by Diaconis and Freedman (1980), it can be expressed as a mixture of Markov Chains, and many properties from Markov Chain theory can be applied here.

In this construction it is assumed that every time the process X_n reaches a level of risk equal to L, i.e., a catastrophe happens, the process is reset to $(0,0)$. Restarting the process after each catastrophe enables the past information to be assessed trough the updated urns: at every cycle, if a state v has already been visited in the past, the composition of the urn $U(v)$ has been changed through reinforcement and future forecasts are dependent on this information. If a state w is, on the contrary, visited for the first time, then the urn $U(w)$ is unchanged and it only contains the prior information related to its initial composition. Given the urn construction, the probabilities associated with every feasible path, during a particular cycle, clearly depend on the updates of the different urns thus far, according to the Bayesian principle. Also, even the prior knowledge of the process under investigation can be translated into the initial composition of the urns, when the process is initialized for the first time: if no prior knowledge exists, all urns can be initialized with exactly the same number of balls of the same colour; if, on the contrary, it is known that the risk of catastrophe increases over time, one can put more c_4 balls in urn $U((n,l))$, as n grows. Increasing or decreasing the reinforcement, will give a larger or smaller weight to past data in computing the posterior distributions. The process is rather flexible, and, according to the Bayesian paradigm, the model is constantly updated, on the basis of the available information.

The authors presented an empirical application to the monthly number of sunspots as collected from the Royal Observatory of Belgium, from January 1900 to December 1990. Three threshold values were chosen in order to determine the levels of risk of the alarm system and a catastrophe was taken as having a number of sunspots greater or equal to 180. Data until December 1974 was used as running-in for Bayesian learning and catastrophe prediction was performed on the period 1975 - 1900. At the beginning of the first cycle, all urns started with the same composition, containing $45c_1$, $32c_2$, $22c_3$ and $1c_4$ balls (all urns had a total of 100 balls and the probability of instant catastrophe is 1%). At every time point n (i.e., a particular month), it was predicted the probability of having a catastrophe in $n+k$, with $k = 1$, $k = 2$ and $k = 5$, given an alarm threshold of 0.05 (meaning that an alarm is given when the probability of catastrophe in $n+k$ is equal or greater to 0.05). For this alarm threshold, the probability of detection is about 80%, considering shorter forecasting intervals, and reduces to only 40% when $k = 5$. To reduce the number of false alarms, the alarm threshold was increased to 0.2. The number of false alarms indeed decreased, thus increasing the probability of the alarm being correct, but the probability of detection also decreased (to 60% when $k = 1$ or $k = 2$ and 20% when $k = 5$). This compromise between these two fundamental operating characteristics of the alarm system was also found in several references mentioned earlier, namely, for instance, in Monteiro et al. (2008). We will get back to this subject when analysing our particular results that are also in agreement with this behaviour.

An interesting analysis yet in the paper by Cirillo et al. (2013), was the acting on the reinforcement $s(v)$ to try to enhance the predictive power of the urn-based alarm system. Assuming $s(v) = s$ and letting s vary in the interval $[0; 10]$, it was looked for the value of s that minimizes the number of false alarms for $k = 0.05$, considering an alarm threshold of 0.05. The value $s = 2.2$ was found as the only minimum in the interval chosen and the corresponding operating characteristics were definitely improved: the number of false alarms actually decreased and the detection probability increased for the longer forecasting interval to around 60%.

1.2.3 Time Series of Counts

Time series of counts arise in many contexts, often as counts of events or individuals in consecutive intervals or at consecutive points in time. In particular, time series of small non-negative observed counts have become available in a wide variety of contexts over the last three decades, approximately. Statistical quality control, computer science, economics and finance, medicine and epidemiology, and environmental sciences are just some of the fields from which discrete time series emerged.

One of the most famous data sets in the field of time series of counts is the United States polio incidence counts, consisting on monthly data from January 1970 to December 1983. The counts range from 0 to 14 and have a sample mean of 1.33 and variance of 3.5. The data set is constituted by small non-negative counts exhibiting **overdispersion** (variance-mean ratio being considerably greater than one). It was introduced in the literature in the seminal work of Zeger (1988) and has been analysed in several other works ever since, e.g., Davis et al. (1999), Fahrmeir and Tutz (2001), or more recently Jung and Tremayne (2011), just to mention a few. In the same paper by Davis et al. (1999), another time series of counts was considered: the daily number of asthma presentations from January 1, 1990 to December 31, 1993, at the Cambelltown hospital in Sydney, Australia. This particular data series was meant to be part of a larger study relating atmospheric pollution and the number of asthma cases presented at various emergency departments in the South West of Sydney. Also in this context of environmental factors and public health effects many other count time series have emerged: just consider the examples already given in the optimal alarm systems subsection. Yet another interesting example in this field is the daily number of deaths in Évora district, Portugal, registered from 1980 to 1997, appearing in Gomes (2005), as an illustration of the model proposed by the author, the generalized DSINAR(1). As covariates, the author used maximum and minimum daily temperatures registered in Évora district, in the same time period.

Referring again to the previous subsection, other examples of count time series can be given, such as the monthly guest nights in hotels modelled through an INteger-valued Moving Average (INMA) model in Brännäs et al. (2002), or, through an INteger-valued AutoRegressive (INAR) model in Brännäs and Nordström (2004) and in Brännäs and Nordström (2006). Other integer time series arising in international tourism demand are given in the Tourism Satellite Account (TSA) analysis for Sweeden, regarding the years 1992-1993, in the work of Nordström (1996), or in the forecasting of international tourism demand in Spain, in Garcia-Ferrer and Queralt (1997).

Time series of counts originating from economics, in particular, from the financial area, include the discrete transaction price movements on financial markets and the number of transactions in stocks. As examples of the first case, we can mention, e.g. Liesenfeld et al. (2006) and Rydberg and Shephard (2003). The dynamics of price movements were analysed considering the transaction data of two shares traded at the New York Stock Exchange (NYSE) over a period of one trading month, in the former paper, and the transaction data for the IBM stock at NYSE in 1995, in the later one. As explained in Rydberg and Shephard (2003), the price movements are restricted to take on integer multiples of a smallest non-zero price change, called a **tick**. The tick size depends on the institutional setting and, when normed, price movements can be thought of as being integers, explaining why transaction price movements can be included in the literature of time series of counts. In both studies, a model considering the decomposition of price movements was considered; in the first case, only the direction of the price change and the size of the price change were addressed, while, in the second case a binary process on $\{0, 1\}$ modelling activity (the price moves or not), was also included. As examples of time series of counts of number of transactions in stocks, the works by Quoreshi (2006) and Brännäs and Quoreshi (2010) can be mentioned. Each transaction refers to a trade between a buyer and a seller in a volume of stocks, for a given price. This kind of data is referred to as **tick-by-tick** data. The counts are usually small and there are frequent zero counts, even for frequently traded stocks if the counts are recorded in short time intervals of, for instance, one minute length. Sometimes, aggregated data over five minutes or one hour intervals are considered. For a discussion on the relation between intradaily price dynamics and size of the observation interval, see Chiang and Wang (2004). Quoreshi (2006) proposed a Bivariate INteger-valued Moving Average (BINMA) model to the tick-by-tick data for Ericsson B and Astrazeneca, collected for the period of November, 5 to December, 12, 2002, and aggregated into five minutes intervals. Brännäs and Quoreshi (2010), modelled the number of transactions per minute in Ericsson B, in the period 2-22 July, 2002. An INMA model was proposed.

In social sciences, time series of counts may arise, for instance, in the analysis of series of claims for wage loss benefit. A well-known series is given by the monthly counts of claimants collecting short-term disability benefits from the Workers' Compensation Board (WCB) of British Columbia, Canada. This low count time series was introduced in Freeland (1998) and modelled through the Poisson INAR(1) model. Further analysis was carried out in Freeland and McCabe (2004), and, McCabe and Martin (2005) extended the analysis by presenting a Bayesian methodology for producing coherent forecasts of low count time series. This methodology was tested, once again, in the Canadian wage loss claims data. Time series of monthly counts of fatal accidents, severe injury accidents, minor injury accidents and vehicle damage accidents were modelled by Johansson (1996), considering extended Poisson and negative binomial count data models, in a paper investigating the effect of lowered speed limit on the number of accidents on Swedish motorways. A somewhat related example can be mentioned in the work of Pokropp et al. (2006), where the number of children injured in traffic accidents in Germany was analysed together with explanatory variables representing seasonal effects in what concerns periods of high or low traffic activities and weather conditions.

Time series of event counts are also common in political science and other social science applications. Political communication studies generally involve an outcome that is a count variable, such as the number of stories printed on a subject, or television stories devoted to a subject, in a given day or any other time span. Brandt and Williams (2001) proposed a Poisson autoregressive model that makes assumptions to address all of the attributes of time-dependent media count data. Fogarty and Monogan (2013) provided a comprehensive review on this subject and applied the Poisson autoregressive model to several count time series, previously reported and analysed, namely, the Peake and Eshbaugh-Soha (2008)'s data on television news attention to energy policy from 1969 to 1983; the Flemming et al. (1997) study of the number of stories related to free speech and censorship (number of stories on the subject listed in the *Reader's Guide*); and, the Ura (2009)'s data on *USA Today*'s coverage of homosexuality (adjusted to monthly total number of news stories about the topic).

Several examples of count time series arising in fields related to computer sciences can be given. In Weiß (2008b), the time series of counts of number of downloads of the program *CWß TeXpert*, a free TeX editor for Windows, was analysed for the period between June 1, 2006 and February 28, 2007.

Several models were estimated, considering either a negative binomial or a generalized Poisson marginal distribution. Criteria that help select an appropriate model were discussed. Another concrete example is provided in Weiß (2007) with a computer pool of n machines which are either occupied (state 1) or not (state 0). The number of machines occupied at time t (which consists on the number of machines which have been occupied before and remain occupied at time instant t plus the number of machines newly occupied) was modelled through the binomial AutoRegressive model of order 1, binomial AR(1). Besides this particular application, the author suggests that the binomial AR(1) process, suitable for processes of counts with finite range in $\mathbb{N}_0$, could also be applied to, e.g., hotel rooms in a certain hotel being occupied at day t, clerks in a counter room serving a costumer, telephones in a call center being occupied, and many other time series for which a finite range of counts should be a constraint of the model.

Communications networks usually have thousands of network elements, such as routers and switches. The monitoring and control of the network is based on reported statistics by the network elements, which happens on a regular basis. Number of packets handled, number of packets dropped, and numbers of processing errors of various types may be reported every minute or second for every network element. Either if the error counts become to high or the number of packets received becomes too low, there is suspicion of network malfunctioning. Addressing this particular application in the area of statistical process control, Lambert and Liu (2006), applied a control chart methodology as an adaptive count thresholding procedure to monitor streams of network counts, in real and simulated data. The methodology herein was developed for counts from communication networks, but it could be relevant for other kinds of counts, with unspecified cyclical patterns, or trends, and missing data. A related example in the frontier between statistical process control and computer science is the approach to cyber attack detection by Ye et al. (2001). As stressed by the authors, a computer and network system must be protected to assure security goals such as availability, confidentiality, and integrity, by using a variety of techniques for prevention, detection, isolation, assessment, reaction, and vulnerability testing. In this paper, a multilevel, multiscale process model of a computer and network system was developed to capture the security-related system behaviour. The Exponentially Weighted Moving Average (EWMA) model for univariate dynamic processes, was applied to measures such as the total number of all method requests per unit time, and the total number of a particular outcome per unit time.

Although originating from different fields, common features of time series of counts are usually observed, namely, a rather pronounced dependence structure (time series of counts are generally autocorrelated) and extra binomial variation (or overdispersion, relative to the mean of the series). A wide range of modelling approaches have been developed and many different models have been used in the analysis of count data, e.g., static regression models, including some generalizations such as Stochastic Autoregressive Mean (SAM) models and Generalized Linear ARMA (GLARMA) models, autoregressive conditional mean models, and INteger-valued ARMA (INARMA) models, just to name some well-known classes. The fact that (as mentioned in Cameron and Trivedi (1998) and reinforced by Jung and Tremayne (2011), more than a decade latter) still no dominant model has emerged, led us to the decision of presenting a summary survey of the available models for discrete-valued time series. Particular emphasis will be given to the class of models related to the work presented in this dissertation. Though wide-ranging, it will certainly not be exhaustive and many alternative classifications are possible.

The most popular classification scheme is due to Cox et al. (1981) and Davis et al. (1999) who introduced two classes of models, **observation-driven** and **parameter-driven** models, in order to deal with autocorrelation and overdispersion in data. In the later class, serial correlation (and overdispersion, if existent) is introduced via a latent dynamic process, or *intensity process*, while, in the former case, it is assumed that the process depends on its own past history (the conditional mean function of the observed counts given past observations depends on lagged observations). The difference between these two classes of models is better understood through the state-space model representation. A **state-space model** for a time series $(Y_t) \equiv (Y_t : t \in \mathbb{Z})$ consists on two equations:

i. **Observation equation**

$$Y_t = g_t X_t + W_t,\ t \in \mathbb{Z},$$

where g_t is a constant, possibly dependent on t, and W_t is a white noise sequence with variance σ_W^2.

ii. **State equation**

$$X_{t+1} = f_t X_t + V_t,\ t \in \mathbb{Z},$$

where f_t is a constant, possibly dependent on t, and V_t is a white noise sequence with variance σ_V^2, not correlated with W_t.

The observation equation remains the same for both the observation-driven and parameter-driven models. The difference between them is in the state equation which describes the latent, non-observable, intensity process. For the parameter-driven models the observation process does not depend on its past history; only depends on the accompanying intensity process that defines the properties and structure of the observation process. For observation-driven models data autocorrelation is implicit within the model formulation:

the observation process is explained by the data itself (at time t, the observation process depends on the past history); overdispersion is introduced by the state equation. Hence, if in the state equation there is a dependency on the past values of the observable process, the data autocorrelation is implicit in the model formulation and the model considered is an observation-driven one.

Stationarity and ergodicity of the observable process (Y_t) are established by the intensity process. If the intensity process is stationary and ergodic so is the observable process. For parameter-driven models these properties are, in general, easy to derive since in many cases the intensity process is Markovian not depending on (Y_t). For observation-driven models, however, the intensity process depends on the observable process (Y_t) and the stability behaviour (stationarity and ergodicity) is difficult to obtain.

A second classification scheme that is popular in the literature of Generalized Linear Models (GLM) distinguishes between **conditional models** and **marginal models**. The former class considers conditional distributions of observed counts given lagged values and is conceptually equivalent to the observation-driven class of models. In marginal models, the regression coefficients are meant to describe the marginal response to changing covariates, i.e., marginal distributions and associations between responses are modelled separately from conditioning on covariates[8].

As it is not possible to classify all models according to these schemes, the

[8] An illustrative example from the analysis of data from longitudinal studies, in Lee and Nelder (2004), considers a marginal gender contrast that compares the mean among men to that among women. If a conditional gender contrast was to be put forth, the comparison should be done between the mean among men and the mean among women, holding the same value of a random effect.

classes of models will be presented roughly in the order they were first introduced. Also, it is worth mentioning that some particular model specifications can have different designations by different authors in the field. For instance, the above mentioned stochastic autoregressive model, introduced by Zeger (1988), is referred to as the *SAM Model* in Jung et al. (2006) and in Jung and Tremayne (2011), but is denoted as a *Serially Correlated Error Model* in the textbook of Cameron and Trivedi (1998), and included in the section under the designation of *Regression Models* in the survey by McKenzie (2003). Or, even in a particular class, the distinction between models can be done in different ways: considering the class of INARMA models, that could also be classified as *Models based on Thinning*, the distinction between the models can be done according to the thinning operations involved, following Weiß (2008a), or, as in McKenzie (2003), according to the model's intended marginal distribution.

For a comprehensive account on the developments in this field, or, in some particular model class, refer to the following reviews. The monograph of MacDonald and Zucchini (1997), which constitutes the first survey of the different approaches thus far, the textbook of Cameron and Trivedi (1998) and the review by McKenzie (2003), with strong emphasis on models based on thinning operations, provide an excellent overview of the historical developments in the field. For a discussion of models within the framework of GLM, consider the works of Fahrmeir and Tutz (2001) and Kedem and Fokianos (2002). For recent developments involving the classes of INAR and INMA processes, see the survey of thinning operations by Weiß (2008a). For recent overviews of the last developments in the field, see Jung and Tremayne (2006, 2011), Fokianos (2011) and Tjøstheim (2012). While Jung and Tremayne (2006, 2011) consider different classes of models, ranging from static regression models to integer autoregressive models (going through autoregressive conditional mean models) and cover aspects of model specification, parameter estimation and inference, the last two contributions focus on Poisson regression models for count time series and on the issues of stationarity, ergodicity and asymptotic inference.

Based on the aforementioned comprehensive reviews, we are now able to provide a brief summary of model classes developed for the analysis of time series of counts.

For many years, a common approach to modelling discrete time series was to consider a continuous modelling approach, what could be justified in virtue of the CLT, in the case the count series were constituted by large numbers. However, when one is facing a series of low counts, which happens quite often as remarked in many of the aforementioned examples, the approximation by continuous r.v's is not valid and alternatives are mandatory.

1. **Markov Chains and Higher-Order Markov Chains**
   Until the late 1970's there were remarkably few models able to deal with discrete time series and Markov chains represented the only general class suitable. Markov models do present two major drawbacks in applications: tendency to be overparametrized and a limited correlation structure. The works of Pegram (1980) and Raftery (1985a) deserve particular attention as they represent attempts to simplify the structure of higher-order Markov chains, reparametrizing them in terms of fewer parameters. Consider a k^{th} order Markov chain over a finite set of states denoted by $\{1, 2, \ldots, m\}$ and denote the transition probabilities by $\{p(s_0|s_1, s_2, \ldots, s_k)\}$. This model may require as many as $(m-1)m^k$ parameters. Raftery (1985a) introduced a class of models that later designated by Mixture Transition Distribution (MTD)

models, proposing that

$$p(s_0|s_1, s_2, \ldots, s_k) = \sum_{i=1}^{k} \lambda_i q(s_0|s_i),$$

where $\sum_{i=1}^{k} \lambda_i = 1$ and $\{q(j|i) : j = 1, 2, \ldots, m\}$ is a probability distribution for each value of $i = 1, 2, \ldots, m$. This model has only $m(m-1) + (k-1)$ parameters and is more suitable for systems which tend to revert to previously occupied states. Each unit increase in the order of dependence requires only one more parameter. MTD models can be used as models for discrete time series with particular marginal distributions and Raftery (1985b) already introduced the particular binomial and Poisson versions.

2. **Markov Regression Models**

   Zeger and Qaqish (1988) proposed a quasi-likelihood approach to regression with time series of data. As serial observations are unlikely to be independent they proposed Markov models in which the expected response at a given time depends not only on the associated covariates, but also on past outcomes. The authors refer to this observation-driven models as Markov regression models.

3. **Hidden Markov Models**

   A hidden Markov model (Y_t) is a particular kind of dependent mixture, accordingly to MacDonald and Zucchini (1997), who described in detail this kind of models in their book. Suppose $\boldsymbol{Y^{(t)}}$ and $\boldsymbol{S^{(t)}}$ represent the histories from time 1 to time t. Then, the simplest model of this kind can be summarized by

$$\begin{aligned} P(S_t|\boldsymbol{S^{(t-1)}}) &= P(S_t|S_{t-1}),\ t = 2, 3, \ldots \\ P(Y_t|\boldsymbol{Y^{(t-1)}}, \boldsymbol{S^{(t)}}) &= P(Y_t|S_t),\ t \in \mathbb{N}. \end{aligned}$$

The model consists of two parts: firstly, an unobserved *parameter process* (S_t) satisfying the Markov property, and secondly the *state-dependent process* (Y_t) such that when S_t is known, the distribution of Y_t depends only on the current state S_t and not on previous states or observations. Note that hidden-Markov models are parameter-driven models.

A particular example of this class is the Poisson-hidden Markov model where it is assumed that (S_t) is an irreducible, homogeneous Markov chain on a set of states $\{1, 2, \ldots, m\}$ and that, conditional on $S_1 = s_1, S_2 = s_2, \ldots, S_n = s_n$, $Y_1, Y_2, \ldots, Y_n$ are independent r.v's with Poisson distribution and mean λ_{s_t}. Thus, the process (Y_t) chooses its current marginal distribution according to the state of the Markov chain at time t, S_t. In this model there are m means and $m(m-1)$ transition probabilities, adding up to a total of m^2 parameters. This model is able to accommodate overdispersion.

Other distributions can be used instead of the Poisson. Also, other designations are used for the general model, such as *Markov-dependent mixture*, *Markov-switching model*, or *Markov mixture model*.

4. **Discrete ARMA (DARMA) Models**

Structurally based on the well-known ARMA processes, the *Discrete ARMA models*, hereafter, DARMA models, were the first real attempt at a class of general yet simple models for discrete time series. They were introduced in a series of papers by Jacobs and Lewis (1978a,b,c) and here we will only present the simplest case, representative of the class.

The Discrete AutoRegressive model of order (1), DAR(1), is defined by the following recursion

$$Y_t = V_t Y_{t-1} + (1 - V_t) Z_t,$$

where (V_t) are i.i.d. binary r.v's with $P(V_t = 1) = \alpha$ and (Z_t) forms an i.i.d. sequence with distribution given by π. If Y_0 is sampled from π then (Y_t) is a stationary process with marginal distribution π. In this model, the current observation Y_t is defined as a mixture of two independent r.v's; namely, it is either the last observation, Y_{t-1}, with probability α, or another independent r.v., Z_t, sampled from the same distribution. This is actually a very simple formulation and also a very general one, since π can be any distribution. Naturally, when dealing with the discrete case, the sample space for π must be a subset of the integers.

All DARMA models are constructed as mixtures of i.i.d. r.v's sharing the distribution π. Consequently, all correlations are positive. In fact, for the DAR(1) case defined above, the ACF of (Y_t) is $\rho_Y(h) = \alpha^h$, for $h \in \mathbb{N}_0$, which matches the ACF of an AR(1). Although simple and general, due to the nature of their construction these models represent somewhat unusual processes with compromised practical applications: notice that the recursion above implies that dependence in the model is realized by runs of constant values in the sample path, and, the larger the value of α, the longer the runs. For continuous r.v's this behaviour is extremely unlikely and, at least questionable in the discrete case. For this reason, DARMA models fell out of favour as new approaches to time series of counts were developed.

5. **Generalized Linear Model (GLM) Framework**

   As some of the subsequent model specifications were developed in the context of generalized linear models, we will very briefly present here the terminology and some fundamental concepts involved, following Kedem and Fokianos (2002). Let (Y_t) be a time series of interest, called the *response*, and let $\boldsymbol{Z}_{t-1} = (Z_{(t-1)1} \; \cdots \; Z_{(t-1)p})'$ be the corresponding p-dimensional vector of past explanatory variables or covariates, $t = 1, 2, \ldots, N$. $\boldsymbol{Z}_t$ will be referred to as the *covariate process*. $\mathcal{F}_{t-1}$ is the σ-field generated by $(Y_s, \boldsymbol{Z}_s;\ s \leq t-1)$. Sometimes, it is also convenient to think of $\boldsymbol{Z}_{t-1}$ as already including past values of the response, $Y_{t-1}, Y_{t-2}, \ldots$. Let $\mu_t = \mathrm{E}[Y_t|\mathcal{F}_{t-1}]$, denote the conditional expectation of the response given the past. A fundamental question is how to relate μ_t to the covariates. In classical theory of linear models it is assumed that the conditional expectation of the response given the past of the process is a linear function of the covariates. When data is not normal the linear relationship is compromised. When the observations follow a distribution from an exponential family, generalized linear models do provide an answer to this issue, including the classical linear model under normality as a special case.

Time series following GLM are defined by the following assumptions, regarding the *random and systematic components*:

i. Random Component

The conditional distribution of the response given the past belongs to the exponential family of distributions in *natural* or *canonical* form, i.e.,

$$f(y_t, \theta_t, \phi|\mathcal{F}_{t-1}) = \exp\left(\frac{y_t\theta_t - b(\theta_t)}{\alpha_t(\phi)} + c(y_t, \phi)\right),\ t = 1, \ldots, N.$$

The parametric function $\alpha_t(\phi)$ is of the form $\frac{\phi}{\omega_t}$ where ϕ is a *dispersion* parameter and ω_t is a known parameter designated by *weight* or *prior weight*. The parameter θ_t is called the *natural* parameter of the distribution.

ii. Systematic Component

For $t = 1, \ldots, N$, there is a monotone function $g(\cdot)$ such that

$$g(\mu_t) = \eta_t = \sum_{i=1}^{p} \beta_i Z_{(t-1)i} = \boldsymbol{Z'}_{t-1}\boldsymbol{\beta}.$$

The function $g(\cdot)$ is the **link function** of the model while η_t is called the linear predictor of the model. The link function

$$g(\mu_t) \equiv \theta_t(\mu_t) = \eta_t = \boldsymbol{Z'}_{t-1}\boldsymbol{\beta},$$

is called the **canonical link function**.

When dealing with observations of counts the distribution of choice is the Poisson distribution. For a Poisson distribution with mean μ_t, the conditional density may be written as

$$f(y_t, \theta_t, \phi|\mathcal{F}_{t-1}) = \exp\left((y_t \ln \mu_t - \mu_t) - \ln y_t!\right), \ t = 1, \ldots, N,$$

and $E[Y_t|\mathcal{F}_{t-1}] = \mu_t$, $b(\theta_t) = \mu_t = exp(\theta_t)$, $\phi = 1$, and, $\omega_t = 1$. The canonical link function is

$$g(\mu_t) \equiv \theta_t(\mu_t) = \ln \mu_t = \eta_t = \boldsymbol{Z'}_{t-1}\boldsymbol{\beta}.$$

For the purpose of illustration, consider the simple example with $\boldsymbol{Z}_{t-1} = (1 \ X_t \ Y_{t-1})'$, where $\boldsymbol{Z}_{t-1}$ already includes past values of the response, Y_{t-1}. Then

$$\ln \mu_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1}$$

with (X_t) representing some covariate process, or a possible trend or seasonal component.

5.1 **Static Regression Models**

Static regression models represent a natural starting point for the analysis of time series of counts, as much the same way as the Poisson distribution is the natural candidate for the distribution of the response process. Moreover, to adequately capture the dependence structure in data, sometimes is sufficient to use an appropriate set of time-varying exogenous regressors, as stated in Cameron and Trivedi (1998). This assumptions define the *static Poisson regression model*. Conditional on the available information on the covariates up to time t, i.e., conditional on $\boldsymbol{Z}_{t-1}$, the observed counts, Y_t, are assumed to follow a Poisson distribution with time-varying parameter μ_t

$$Y_t|\boldsymbol{Z}_{t-1} \sim Po(\mu_t),$$

which is to say

$$f(y_t, \mu_t|\mathcal{F}_{t-1}) = \frac{\exp(\mu_t)\mu_t^{y_t}}{y_t!},\ t = 1, \ldots, N,$$

where $\mathcal{F}_{t-1}$ denotes all the available information to the observer up to time t (past values of the response series and past values of the covariates). Following the general theory of GLM,

$$\mu_t(\boldsymbol{\beta}) = h(\boldsymbol{Z'}_{t-1}\boldsymbol{\beta}) = \exp(\boldsymbol{Z'}_{t-1}\boldsymbol{\beta}),\ t = 1, \ldots, N,$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters and $h(\cdot)$ is the **inverse link function**. Consistent parameter estimators for $\boldsymbol{\beta}$ are straightforwardly obtained by Maximum Likelihood Estimation (MLE).

For this Poisson model the conditional expectation of the response is equal to its conditional variance

$$\mathrm{E}[Y_t|\mathcal{F}_{t-1}] = \mathrm{Var}[Y_t|\mathcal{F}_{t-1}] = \mu_t,\ t = 1, \ldots, N.$$

Only the Poisson specification was mentioned here, but, in the presence of overdispersion, an alternative specification is the Negbin2 model of Cameron and Trivedi (1998). Static regression models can cope with both positive and negative serial correlation in data.

5.2 **Stochastic Autoregressive Mean (SAM) Models**

The benchmark model in the class of parameter-driven specifications is the dynamic regression model by Zeger (1988) who extended the static Poisson regression model by specifying a serially correlated multiplicative error term. Classified as a marginal model in the GLM framework, this model extends the general linear models by incorporating a latent autoregressive process in the conditional mean function. This latent autoregressive process evolves independently of the past observed counts and is able to introduce autocorrelation and overdispersion into the model. Such as with the static regression model, both positive and negative serial correlation can be modelled.

In the standard version of the model, conditional on the available information on the covariates up to time t, $\boldsymbol{Z}_{t-1}$, and on ξ_t, a latent, non-negative stochastic process, the observed counts, Y_t, are assumed to follow a Poisson distribution with time-varying parameter μ_t

$$Y_t|\boldsymbol{Z}_{t-1}, \xi_t \sim Po(\mu_t \xi_t),$$

where $\mu_t(\boldsymbol{\beta}) = \exp(\boldsymbol{Z'}_{t-1}\boldsymbol{\beta})$ and $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters. Additional assumptions concerning $\lambda_t = \ln(\xi_t)$ are necessary, and a convenient specification (e.g., Chan and Ledolter, 1995; Kuk and Cheng, 1997; Jung and Liesenfeld, 2001) is the Gaussian first-order autoregressive form $\lambda_t = \delta\lambda_{t-1} + \nu\epsilon_t$, where $\epsilon_t \sim N(0, 1)$. To ensure stationarity of λ_t it is assumed that $|\delta| < 1$. Note that for $\delta = 0$ and $\nu \to 0$ the latent process ξ_t vanishes and a standard Poisson static regression model is obtained. For a complete description of the statistical properties of the SAM model see Davis et al. (1999).

The main drawback with the class of SAM models is that their efficient estimation is not straightforward because the dynamic latent process leads to a likelihood function which depends on high-dimensional integrals and is not available in closed form. Non-standard likelihood-based estimation procedures are needed and usually Monte Carlo (MC) simulation techniques are adopted.

5.3 Generalized Linear ARMA (GLARMA) Models

Another extension of the GLM framework is the GLARMA class of models proposed by Davis et al. (1999, 2003, 2005) and Shephard (1995). The GLM framework was extended to allow for serial correlation as well as overdispersion in the data by specifying the logarithm of the conditional mean process as a linear function of previous counts. The GLARMA class therefore belongs to the class of observation-driven models.

The general GLARMA(p, q) is defined by the following specifications

$$Y_t|\mathcal{F}_{t-1} \sim Po(\mu_t),\ \ W_t := \log(\mu_t) = \boldsymbol{Z'}_{t-1}\boldsymbol{\beta} + \omega_t,$$

where

$$\omega_t = \sum_{i=1}^{p} \alpha_i(\omega_{t-i} + e_{t-i}) + \sum_{i=1}^{q} \beta_i e_{t-i}$$

and $e_t = \frac{Y_t - \mu_t}{\mu_t^{\lambda}}$, $\lambda \geqslant 0$ and $e_t = \omega_t = 0$ for $t \leqslant 0$. Since the conditional mean $\mathrm{E}[Y_t|\mathcal{F}_{t-1}]$ depends on the whole past, the process (Y_t) is not Markovian. However, the mean process $\log(\mu_t)$ is q^{th} order Markov. Unless $\boldsymbol{Z'}_{t-1}\boldsymbol{\beta}$ is constant, $log(\mu_t)$ is not a time-homogeneous process.

Being an observation-driven model the stochastic properties of the model are difficult to obtain: there are no explicit expressions for the unconditional mean and unconditional variance and neither the ACF nor the unconditional distribution are known. Nevertheless, as already mentioned, serial correlation and overdispersion are able to be captured by the GLARMA representation. The models can cope with both positive and negative serial correlation and it is straightforward to include covariates. An additional advantage of the GLARMA models is that their efficient estimation by ML is easy to implement.

In Davis et al. (2003, 2005) a simpler version of the general GLARMA model was extensively analysed. It was assumed that the mean process $\log(\mu_t)$ follows a linear model in the explanatory variables, $\boldsymbol{Z'}_{t-1}$, with residuals having a moving average structure

$$Y_t|\mathcal{F}_{t-1} \sim Po(\mu_t), \;\; W_t := \log(\mu_t) = \boldsymbol{Z'}_{t-1}\boldsymbol{\beta} + \sum_{i=1}^{q} \beta_i e_{t-i},$$

where $e_t = \frac{Y_t - \mu_t}{\mu_t^\lambda}$, $\lambda \geqslant 0$. Considering $q = 1$ and $\boldsymbol{Z'}_{t-1}\boldsymbol{\beta} = \beta$, the mean process (or state process, in the state-space representation for observation-driven models) reduces to

$$W_t = \beta + \beta_1(Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}}.$$

In this particular case, many desirable properties of the state process (W_t) were obtained. Firstly, the model structure for W_t is now Markovian with mean $\mathrm{E}[W_t] = \mathrm{E}[\mathrm{E}[W_t|W_{t-1}]] = \beta$. Making use of Markov chain theory and under the condition $0.5 \leqslant \lambda \leqslant 1$, Davis et al. (2003) demonstrated the existence of a stationary solution for the (W_t) process. In the particular case $\lambda = 1$ the uniqueness of the stationary distribution was also demonstrated. Also, making use of the fact that the process (W_t) satisfies Doeblin's condition and is strongly aperiodic it was concluded that it is uniformly ergodic (Meyn and Tweedie, 1994).

6. **INteger-valued ARMA (INARMA) Models**

The class of models introduced in this section was originally developed having in mind the autocorrelation structure and other attractive properties of the ARMA models defined by the recursion in (1.1). Although typical mathematical operations are well defined for counts in $\mathbb{N}_0$, recursion (1.1) cannot be applied to the integer-valued case because multiplying an integer by a real number usually results in a non-integer value. However, if the scalar multiplication is to be replaced by a different operation with similar properties, the discreteness of the counts may be preserved. A general form of this class of models can be given by

$$Y_t = R_t(Y_{t-1}, \alpha) + Z_t,$$

where $\alpha \in [0, 1]$ provides a measure of the relationship between previous counts Y_{t-1} and the current observation Y_t, (Z_t) is a i.i.d. integer-valued innovation process independent of R_t, and R_t represents a convenient **thinning operation**[9]. Due to this construction this class of models is also referred to as *Models based on Thinning* (e.g. McKenzie, 2003). This class of observation-driven models has received wide atten-

[9]The basic idea behind the probabilistic operation of thinning is that a count represents the random size of an imaginary population and the thinning operation randomly deletes some of the members of this population, (Weiß, 2008a). As the size of the *thinned* population is always integer-valued the thinning operation always leads to integer values. Not to give the wrong impression, it is worth mentioning that with the development of thinning operations, thinning does not necessarily mean shrinking: in the context of *generalized thinning*, by Latour (1998), the random operation of thinning can be interpreted as a reproduction process.

tion in recent years and theoretical models covering a wide range of possible correlation structures combined with equidispersed and overdispersed discrete marginal distributions are available in the literature. As the focus of our work does not involve this particular class of processes, we will briefly refer to a few key concepts and specifications.

The most popular thinning operation is the *binomial thinning* first introduced by Steutel and Van Harn (1979) when adapting the notions of *self-decomposability* and *stability* for integer-valued time series (the authors succeeded in treating this concepts as special cases of *infinite divisibility* and demonstrating that many important distributions e.g. Poisson, negative binomial or generalized Poisson belong to the class of Discrete Self-Decomposable (DSD) distributions).

Let Y be a discrete r.v. with support in $\mathbb{N}_0$ or any subset $\{0, 1, \ldots, n\}$ and $\alpha \in [0, 1]$. Define

$$\alpha \circ Y := \sum_{i=1}^{Y} B_i \tag{1.22}$$

where the B_i are i.i.d. Bernoulli-distributed r.v's, $B_i \sim B(1, \alpha)$, independent of Y. Then it is said that $\alpha \circ Y$ arises from Y by **binomial thinning** and $\circ$ is the **binomial thinning operator**. The interpretation of the binomial thinning operation can be done considering a population of size Y at time t. If the same population is observed again at, say $t + 1$, the population may have shrinked, because some individuals may have died between times t and $t+1$. If the individuals survive independently of each other and if the probability of surviving in between t and $t+1$ is equal to α for all individuals, then the number of survivors is given by $\alpha \circ Y$.

Conveniently replacing the scalar multiplication in the ARMA recursion (1.1) by binomial thinning leads to the family of INARMA models. The cornerstone of the INARMA models, the INAR(1), was proposed by McKenzie (1985) and extensively studied by e.g. Al-Osh and Alzaid (1987), Alzaid and Al-Osh (1988), McKenzie (1988) and da Silva and Oliveira (2004). Replacing the scalar multiplication in the classical AR(1) model with the binomial thinning operator results in the recursion

$$Y_t = \sum_{i=1}^{Y_{t-1}} B_i + Z_t \equiv \alpha \circ Y_{t-1} + Z_t, \tag{1.23}$$

which defines the INAR(1) process for (Y_t) when (Z_t) is a i.i.d. innovation process with range $\mathbb{N}_0$, $\alpha \in [0, 1]$ and all thinning operations are independent of each other and of (Z_t). It should be clear that the thinning operation is performed at each time t, i.e., it would be more precise to write $\circ_t$ in the right-hand side of (1.23), instead. Nevertheless, an additional assumption of the INAR(1) process is that the thinning operations at each time t and Z_t are independent of $(Y_s)_{s<t}$. In this model, and relating to the interpretation given above, Y_t may describe the number of individuals in the population at time t, in which case $\alpha \circ Y_{t-1}$ represents the number of survivors from time $t-1$ and Z_t represents the number of immigrants. A practical example is the work by Brännäs et al. (2002), where the number of guest nights in hotels was modelled following the INAR(1) formulation and considering Y_t as the number of costumers at time t, $\alpha \circ Y_{t-1}$ the number of costumers retained in the service from the last period, and Z_t the number of new costumers.

Al-Osh and Alzaid (1987) studied the INAR(1) in the case the innovations are Poisson-distributed, i.e. (Z_t) is i.i.d. according to $Po(\lambda)$ with $\mu_Z = \sigma_Z^2 = \lambda$. This turned out to be a necessary and sufficient condition for the count process (Y_t) to have marginal Poisson distribution according to $Po(\frac{\lambda}{1-\alpha})$. The Poisson INAR(1) model was thus introduced and the process was shown to be a stationary Markov process. It is interesting to note here that the Poisson distribution seems to play a similar role for the discrete AR(1) case to that of the normal distribution for the real-valued AR(1) classical case. A drawback of this particular specification is that the Poisson INAR(1) model is not able to cope with overdispersion. Further properties of this model can be found in Freeland (1998), Freeland and McCabe (2004) and Weiß (2007).

All distributions that are DSD in the sense of Steutel and van Harn can be marginal distributions for the stationary solution to equation (1.23). Besides the Poisson distribution, many of the most usual distributions on the non-negative integers are included, such as the geometric and the negative binomial distributions; see McKenzie (1986, 1987) for further discussion and other INAR(1) specifications with alternative marginals. However, as explained in Weiß (2008a), the INAR(1) is best suited for Poisson marginals, and, modifications of the concept of binomial thinning or even alternative thinning operations should be considered if one is interested, for instance, in dealing with overdispersion in data, or, modeling a time series with a finite range of counts. Regarding adaptations to the binomial thinning concept, though maintaining the same structure, we would like to mention the following contributions

i. Brännäs and Hellström (2001) and Ristić et al. (2013) for introducing a dependency structure in the Bernoulli variables B_i in equation (1.22);

ii. Latour (1998) by the definition of the *generalized thinning* operator as the random operation $\alpha \circ_\beta Y := \sum_{i=1}^{Y} B_i$, where the r.v's B_i are i.i.d and independent on Y but now allowed to have the full range $\mathbb{N}_0$, with mean α and variance β;

iii. Kim and Park (2008) for introducing the *signed binomial thinning* operator in order to include negative integers in the range of applications.

Other random operations besides binomial thinning are worth mentioning. In binomial thinning the α parameter is a real number. Joe (1996), Zheng et al. (2007) and Gomes and Canto e Castro (2009) suggested allowing the α to be random itself thus extending the previous thinning concept. Zheng et al. (2007) used the **random coefficient thinning** thus defined to generalize the INAR(1) and introduced the RCINAR(1) model; Zheng et al. (2006) generalized the higher order INAR(p) model, introducing the RCINAR(p). In these two references Zheng and co-workers illustrated the performance of the models in the analysis of the well-known polio data and in the analysis of a series of epileptic seizure counts. A different approach to generalize the binomial thinning operator is due to Al-Osh and Aly (1992) and is referred to as **iterated thinning** because it can be understood as two nested thinning operations: a first usual binomial thinning operation is applied and can be interpret as selecting a random number out of Y individuals and, afterwards, each of the selected individuals experiences a second random experiment, independently of the other selected individuals. The last alternative to binomial thinning mentioned here arises in relation to the following issue: if the standard INAR(1) model is used to model a process with Generalized Poisson[10]. (GP) marginals, the distribution of the corresponding innovations cannot be obtained explicitly. However, using the quasi-binomial distribution[11] to define

[10] For properties of the Generalized Poisson distribution see Consul (1989) and Ambagaspitiya and Balaskrishnan (1994)

[11] For properties of the quasi-binomial distribution refer to Consul and Mittal (1975)

a generalized thinning operation, Alzaid and Al-Osh (1993) were able to obtain a count process (Y_t) with GP marginals. The thinning operation thus introduced is designated by **quasi-binomial thinning** and the AR(1)-like process related to it as the QINAR(1) (Alzaid and Al-Osh, 1993).

Regarding this family of INARMA models it is left to say that other specifications and higher order members have also been discussed in the literature. The INAR(p) was proposed in, at least, three different formulations: Alzaid and Al-Osh (1990), Du and Li (1991) and Franke and Subba Rao (1995). The INteger-valued Moving Average (INMA) model of order 1 was introduced in McKenzie (1986) and that of order q, INMA(q), was proposed and analysed in Al-Osh and Alzaid (1988) and McKenzie (1988). Subsequent developments regarding estimation were done by Brännäs and Hall (2001). An example of an empirical application to the number of transactions in stocks can be found in Quoreshi (2006), where a BINMA model is proposed. Ohter models related to the INMA(q) were introduced by Aly and Bouzar (1994, 2005), Zhu and Joe (2003) and Neal and Subba Rao (2007). The general class of INARMA(p, q) models was first addressed in McKenzie (1986) and a full INARMA(p, q) model was put forth in Dion et al. (1995).

7. **Autoregressive Conditional Poisson (ACP) Models**

   In this work, focus is put on models in which the count variable is assumed to be Poisson-distributed, conditioned on the past, or, in other words, the conditional distribution of the count variable, given the past, is assumed to be Poisson with time-varying mean λ_t, satisfying some autoregressive mechanism. Models within this class consist on

and Shenton (1986).

two processes: one observable process of counts and one accompanying intensity process, usually not observable. They are observation-driven processes.

The class of ACP models was first introduced by Heinen (2003). The author defined this class of models by adapting the autoregressive conditional duration model of Engle and Russell (1998) to the integer-valued case, assuming a conditional Poisson distribution. Due to its analogy to the conventional GARCH model, the ACP model has also been referred to as the INGARCH model by Ferland et al. (2006).

An INteger-valued GARCH process of orders (p,q), INGARCH(p,q) in short, is defined to be an integer-valued process (Y_t) such that, conditioned on the past experience, Y_t is Poisson-distributed with mean λ_t, and λ_t is obtained recursively from the past values of the observable process (Y_t) and (λ_t) itself,

$$Y_t|\mathcal{F}_{t-1} \sim Po(\lambda_t), \;\; \lambda_t = \gamma_0 + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=1}^{q} \delta_j \lambda_{t-j},\; t \in \mathbb{Z}$$

with $\gamma_0 > 0$, $\gamma_i \geq 0$, and $\delta_j > 0$. Ferland et al. (2006) showed that if $\sum_{i=1}^{p} \gamma_i + \sum_{j=1}^{q} \delta_j < 1$ then the process (Y_t) is strictly stationary with finite first- and second-order moments. Weiß (2009) derived the variance and ACF for the INGARCH(p,q) models. The particular case $p = q = 1$ was addressed by Ferland et al. (2006) who obtained the following results.

- If $\gamma_1 + \delta_1 < 1$ then the process (Y_t) is strictly stationary and possesses moments of any order.
- The unconditional mean is given by

$$\mu := \mathrm{E}[Y_t] = \frac{\gamma_0}{1-(\gamma_1+\delta_1)}.$$

- The unconditional variance is given by

$$\sigma^2 := \mathrm{Var}[Y_t] = \mu \left(1 + \frac{\gamma_1^2}{1 - (\gamma_1 + \delta_1)^2}\right).$$

  Note that the result above implies that $\mathrm{Var}[Y_t] \geqslant \mathrm{E}[Y_t]$ leading to overdispersion. Moreover if the $\gamma_1 = 0$ (which means that the conditional mean does not depend on past values of the observable process) then the variance equals the expected value and the process is equidispersed.

- Finally, the autocovariance function of the INGARCH$(1,1)$ is given by

$$\gamma_Y(k) := \mathrm{Cov}[Y_t, Y_{t-k}] = \mu \frac{\gamma_1[1 - \delta_1(\gamma_1 + \delta_1)](\gamma_1 + \delta_1)^{k-1}}{1 - (\gamma_1 + \delta_1)^2},\ k \in \mathbb{N}.$$

This model was also analyzed by Fokianos et al. (2009) and Fokianos and Tjøstheim (2012) under the designation of Poisson Autoregression. Linear and non-linear models for λ_t were considered. For the linear model case the representation considered is as follows

$$Y_t | \mathcal{F}_{t-1}^{Y,\lambda} \sim Po(\lambda_t),\ \ \lambda_t = d + a\lambda_{t-1} + bY_{t-1},\ t \in \mathbb{N},$$

where it is assumed that the parameters d, a, b are positive, and λ_0 and Y_0 are fixed. It is worth to mention that though this representation corresponds exactly to the INGARCH$(1,1)$ model, the approach followed by Fokianos et al. (2009) is different in the sense that it is shown that the linear model can be rephrased as follows

$$Y_t = N_t(\lambda_t),\ \ \lambda_t = d + a\lambda_{t-1} + bY_{t-1},\ t \in \mathbb{N}$$

with λ_0 and Y_0 fixed. For each time point t, the authors introduced a Poisson process of unit intensity, $N_t(\cdot)$, so that $N_t(\lambda_t)$ represents the number of such events in the time interval $[0, \lambda_t]$. Following this rephrasing a perturbation is introduced in order to demonstrate ϕ-irreducibility and, as a consequence, geometric ergodicity follows.

The non-linear case is considered a generalization of the previous situation in which the conditional mean $\mathrm{E}[Y_t|\mathcal{F}_{t-1}^{Y,\lambda}] = \lambda_t$, is a non-linear function of both the past values of λ_t and the past values of the observations, $\lambda_t = f(\lambda_{t-1}) + b(Y_{t-1})$. Sufficient conditions to prove geometric ergodicity were also derived by the authors.

It is important to stress the fact that the assumptions made in Fokianos et al. (2009) that turned into sufficient conditions for geometric ergodicity could not be fulfilled for the non-linear model proposed in this work. Hence, a different approach had to be adopted. To this extend, we turned our attention to the work of Doukhan et al. (2012), Neumann (2011), and Franke (2010). For completeness and reader's convenience some of the results obtained by the above authors are briefly summarized below.

Neumann (2011) considered a class of observation-driven Poisson count processes satisfying

$$N_t|\mathcal{F}_{t-1}^{N,\lambda} \sim Po(\lambda_t), \;\; \lambda_t = f(\lambda_{t-1}, N_{t-1}), \, t \in \mathbb{N},$$

for some function $f : [0, +\infty[\times \mathbb{N}_0 \to [0, +\infty[$, where

$$\mathcal{F}_{t-1}^{N,\lambda} = \sigma(\lambda_1, \ldots, \lambda_t, N_1, \ldots, N_t)$$

is the σ-algebra generated by past and present values of count and intensity processes (N_t) and (λ_t), respectively, at time t. It is assumed that the function f satisfies the contractive condition

$$|f(\lambda, y) - f(\lambda', y')| \leq k_1|\lambda - \lambda'| + k_2|y - y'|, \; \forall\, \lambda, \lambda' \geq 0, \forall\, y, y' \in \mathbb{N}_0,$$

where k_1 and k_2 are non-negative constants such that $k := k_1 + k_2 < 1$. Under the mentioned contractive condition it follows that the bivariate process (N_t, λ_t) has a unique stationary distribution. Supposing that the bivariate chain (N_t, λ_t) is in its stationary regime and obeys the contractive condition, the following is also true, namely

- The count process (N_t) is β-mixing or absolutely regular. Since β-mixing implies strong-mixing it can be concluded that (N_t) is ergodic.
- The intensities λ_t can be expressed as measurable functionals of past values of the count variables.
- The bivariate process (N_t, λ_t) is ergodic.

Franke (2010) introduced a class of models for time series of counts which include as special cases the INGARCH-type models and also the log linear models for conditionally distributed data. Starting from a sequence of independent Poisson processes $(N_t(\cdot), t \in \mathbb{Z})$, a Functional INGARCH$(p, q)$ or FINGARCH(p, q) process is defined as a process satisfying the recursion

$$Y_t = N_t(\lambda_t), \;\; \lambda_t = g(\lambda_{t-1}, \ldots, \lambda_{t-p}, Y_{t-1}, \ldots, Y_{t-p}),\; t \in \mathbb{Z},$$

for some measurable function $g : [0, +\infty[^p \times \mathbb{N}_0^q \to [0, +\infty[$. Assuming that g is Lipschitz in each argument with Lipschitz constants summing up to a constant less than 1, i.e.,

$$|g(\lambda, y) - g(\lambda', y')| \leq \sum_{i=1}^{p} a_i |\lambda_i - \lambda'_i| + \sum_{i=1}^{q} b_i |y - y'|,$$

for $\lambda, \lambda' \in [0, +\infty[^p$, $\forall\, y, y' \in \mathbb{N}_0^q$ with $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j =: L < 1$, important results are established. Firstly, if $g(\lambda, y)$ satisfies the Lipschitz condition then there exists a strictly stationary FINGARCH(p, q) process, (Y_t), satisfying above definition and having a finite mean. Suppose that (Y_t) is a stationary FINGARCH$(1, 1)$ process satisfying definition above for $p = q = 1$ and that $g(\lambda, y)$ satisfies the Lipschitz condition which in this case simplifies to

$$|g(\lambda, y) - g(\lambda', y')| \leq a|\lambda - \lambda'| + b|y - y'|, \ \forall \lambda, \lambda' \geq 0, \forall y, y' \in \mathbb{N}_0$$

with $a \equiv a_1$, $b \equiv b_1$, and $L < 1$. By these conditions it follows that (Y_t) is θ-weak dependent with geometrically decreasing coefficients for some $c > 0$, with $\theta_t \leq cL^t$.

For the general stationary FINGARCH(p, q), if $g(\lambda, y)$ satisfies the Lipschitz condition, then Franke (2010) showed that (Y_t) is θ-weak dependent with geometrically decreasing coefficients, for some $c > 0$, and $\theta_t \leq c(L^{\frac{1}{max(p,q)}})^t$.

Doukhan et al. (2012) assumed that (Y_t) is a count time series and (λ_t) a sequence of mean processes, namely

$$Y_t|\mathcal{F}_{t-1} \sim Po(\lambda_t), \ \ \lambda_t = f(\lambda_{t-1}, \dots, N_{t-1}, \dots), \, t \in \mathbb{Z},$$

where $\mathcal{F}_t$ represents the σ-algebra generated by $(Y_s, s \leq t)$ and f is some function defined on $[0, +\infty[^{\infty} \times \mathbb{N}_0^{\infty} \to [0, +\infty[$. Note that this formulation allows for models with any order. Following the notion of τ-dependence as introduced by Dedecker and Prieur (2004) it is assumed that for any vectors, say, $\mathbf{x}$ and **x'** in $\mathbb{R}_+^{\mathbb{N}_*} \times \mathbb{N}_*^{\mathbb{N}_*}$ with $\mathbb{N}_* = \{1, 2, \dots\}$, there exists a sequence $(\alpha_j : j \in \mathbb{N})$ of non-negative real numbers such that

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq \sum_{l=1}^{\infty} \alpha_l \|x_l - x_l'\|.$$

If $\sum_{l=1}^{\infty} \alpha_l < 1$ then it follows that there exists a τ-weakly dependent strictly stationary process (Y_t, λ_t) which has finite moments up to any positive order and such that the decay of the coefficients $\tau(\cdot)$ ensures the conditions needed for obtaining the asymptotic properties of the maximum likelihood estimator.

1.3 Organization of the Dissertation

This thesis focuses on the application of optimal alarm systems to non-linear time series models. Non-linear models related to financial time series are considered.

In Chapter 2 particular attention is given to the FIAPARCH(p, d, q) model and an optimal alarm system is implemented, considering both classical and Bayesian methodologies. The expressions for the alarm characteristics of the alarm system are obtained and a simulation study is carried out in order to illustrate the method. Regarding a better performance of the alarm system, different criteria are analysed and a compromise between operating characteristics is achieved. Section 2.2 covers the estimation of the FIAPARCH$(1, d, 1)$ model by classical and Bayesian methodology. Last section of Chapter 2 includes a real data application with the daily returns of the São Paulo Stock Market, the IBOVESPA returns data set.

In Chapter 3 the class of Autoregressive Conditional Poisson models is addressed and a new model is proposed. As explained in Section 3.1, although asymmetric responses of the volatility for positive or negative shocks have also been observed in time series of counts, no model presented in the introductory section 1.2.3 is able to address this issue. The INteger-valued Asymmetric Power ARCH, INAPARCH(p, q), is thus introduced in Section 3.2 as an integer-valued counterpart of the APARCH representation for the volatility. With this thesis we also expect to contribute to the modelling of asymmetric overdispersion in time series of counts. The probabilistic properties of the INAPARCH(1,1) model are extensively studied in Section 3.2. Parameter estimation and asymptotic theory regarding conditional maximum likelihood estimation are developed in Section 3.3. A simulation study is presented in Section 3.4 to test and illustrate the methodology.

Relating to the aforementioned goal of this thesis in what concerns the application of optimal alarm systems to non-linear time series models, the implementation of an alarm system to the INAPARCH model is also addressed. Expressions for the INAPARCH(1,1) particular case are obtained in Section 3.5. Last section of Chapter 3 presents another real data application of optimal alarm systems, now considering time series of counts. The number of intra-day transactions in stocks is analysed in Section 3.6, for the Glaxosmithkline and Astrazeneca data sets.

Chapter 2

Optimal Alarm Systems for FIAPARCH Processes

This chapter is organized as follows: in Section 2.1, basic theoretical concepts related to optimal alarm systems are presented and implemented for the particular case of FIAPARCH processes. Expressions for the alarm characteristics of the alarm system are given. Estimation of the FIAPARCH$(1, d, 1)$ model by classical and Bayesian methodology is covered in Section 2.2. In Section 2.3, the results are illustrated through a simulation study. A real-data example is given in Section 2.4, considering the IBOVESPA data set containing the daily returns of the São Paulo Stock Market during the period 04/07/1994 to 02/10/2008.

2.1 Introduction

Let $(X_t)_{t\in\mathbb{N}}$ be a discrete parameter stochastic process with parameter space $\Theta \subset \mathbb{R}^k$, for some fixed $k \in \mathbb{N}$. The time sequence $\{1, 2, \dots, t-1, t, t+1, \dots\}$ will be divided into three sections, $\{1, 2, \dots, t-q\}$, $\{t-q+1, \dots, t\}$, and $\{t+1, \dots\}$, namely, the past, the present and the future, such that for some

$q > 0$ the following subsets will be defined

Data or informative experience: $D_t = \{X_1, X_2, \ldots, X_{t-q}\}$

Present experiment: $\mathbf{X_2} = \{X_{t-q+1}, \ldots, X_t\}$

Future experiment: $\mathbf{X_3} = \{X_{t+1}, \ldots\}$

Any event of interest, $C_{t,j}$, in the σ-algebra generated by $\mathbf{X_3}$ is defined as a catastrophe. In this work, the catastrophe shall be considered as the up-crossing event of some fixed level u,

$$C_{t,j} = \{X_{t+j-1} \leqslant u < X_{t+j}\} \text{ for some } j \in \mathbb{N}. \tag{2.1}$$

The alarm region of optimal size $\alpha_{t,j}$ is given by

$$\begin{aligned} A_{t,j} &= \left\{ \mathbf{x_2} \in \mathbb{R}^q : \frac{P(C_{t,j}|\mathbf{x_2}, D_t)}{P(C_{t,j}|D_t)} \geq k_{t,j} \right\} \\ &= \{\mathbf{x_2} \in \mathbb{R}^q : P(C_{t,j}|\mathbf{x_2}, D_t) \geq k\}, \end{aligned} \tag{2.2}$$

where $k = k_{t,j} P(C_{t,j}|D_t)$.

In this chapter, the construction of an optimal alarm system will be carried out for the FIAPARCH(p, d, q) model. As described in the introductory chapter, it can be written as

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^\delta &= \frac{\omega}{1 - \beta(B)} + \left[1 - (1 - \beta(B))^{-1} \phi(B)(1 - B)^d\right] g(X_t), \end{aligned} \tag{2.3}$$

where

$$g(X_t) = (|X_t| - \gamma X_t)^\delta$$

with $0 < d < 1$, $\omega > 0$, $|\gamma| < 1$, and $\delta \geq 0$. All zeros of the polynomials $1 - \beta(B)$ and $\phi(B)$ are assumed to lye outside the unit circle. If the fractional differencing parameter, d, lies in the interval $0 < d < 1/2$, long memory in volatility is expected to occur. The fractional differencing operator $(1-B)^d$ is yet again expressed as (1.16).

The simplest version of the FIAPARCH(p, d, q) model, which appears to be particularly useful in practice, occurs when both $1 - \beta(B)$ and $\phi(B)$ are polynomials of degree 1, $\beta(B) = \beta B$ and $\phi(B) = \phi B$ with $|\beta| < 1$. This is the FIAPARCH(1,d,1) model, and the volatility σ_t, takes the form

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^\delta &= \frac{\omega}{1-\beta B} + \left[1 - (1-\beta B)^{-1}\phi B(1-B)^d\right] g(X_t). \end{aligned}$$

Necessary and sufficient conditions for the non-negativity of the conditional variance for the FIAPARCH(1,d,1) resemble the ones obtained by Conrad and Haag (2006) for the FIGARCH(1,d,1), namely

- Case I: $0 < \beta < 1$,
  either $\lambda_1 \geq 0$ and $\phi \leq \frac{1-d}{2}$
  or for $i > 2$ with $\frac{i-2-d}{i-1} < \phi \leq \frac{i-1-d}{i}$ it holds that $\lambda_{i-1} \geq 0$.
- Case II: $-1 < \beta < 0$,
  either $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\phi \leq \frac{1-d}{2}\left(\beta + \frac{2-d}{3}\right) / \left(\beta + \frac{1-d}{2}\right)$
  or $\lambda_{i-1} \geq 0$, $\lambda_{i-2} \geq 0$ and $\frac{i-3-d}{i-2}\left(\beta + \frac{i-2-d}{i-1}\right) / \left(\beta + \frac{i-3-d}{i-2}\right) < \phi \leq \frac{i-2-d}{i-1}\left(\beta + \frac{i-1-d}{i}\right) / \left(\beta + \frac{i-2-d}{i-1}\right)$ with $i > 3$.

As previously stated in Chapter 1, Section 1.2.1, in the FIGARCH(1,d,1) model of Baillie et al. (1996) the conditional volatility has an infinite series representation in terms of X_t^2. In the FIAPARCH(1,d,1) model, X_t^2 is replaced by $g(X_t)$, implying that the impact of a shock on the forecast of future conditional variance should also decay at a slow hyperbolic rate, as in the FIGARCH$(1,d,1)$ case. Statistical properties such as stationarity and ergodicity are still subject of discussion. Nevertheless, the coefficients λ_i in $\lambda(B) = \sum_{i=1}^{\infty} \lambda_i B^i$ remain unaltered. The infinite series coefficients can be obtained recursively as

$$\lambda_i = \begin{cases} \phi - \beta + d & i = 1 \\ \beta\lambda_{i-1} + [\frac{i-1-d}{i} - \phi]\delta_{i-1} & i \geq 2 \end{cases}$$

with $\delta_1 = d$ and $\delta_i = \delta_{i-1}\frac{i-1-d}{i}$ for $i \geq 2$. Usually, when estimating the model parameters, a finite truncation at some particular lag is imposed.

Moving on to the construction of the alarm system, the first step consists on the calculation of the probability of catastrophe conditional on D_t and $\mathbf{x_2}$, i.e., $P(C_{t,j}|\mathbf{x_2}, D_t, \boldsymbol{\theta})$, and the probability of catastrophe conditional on D_t, $P(C_{t,j}|D_t, \boldsymbol{\theta})$.

$$\begin{aligned} P(C_{t,j}|\mathbf{x_2}, D_t, \boldsymbol{\theta}) &= P(X_{t+j-1} \leqslant u < X_{t+j}|\mathbf{x_2}, D_t, \boldsymbol{\theta}) \\ &= P(X_{t+j-1} \leqslant u, X_{t+j} > u|\mathbf{x_2}, D_t, \boldsymbol{\theta}) \\ &= \int_{C_{t,j}} \int \ldots \int f_1 dx_{t+1} \ldots dx_{t+j-2} dx_{t+j-1} dx_{t+j} \end{aligned}$$

where

$$f_1 \equiv f_{X_{t+1},\ldots,X_{t+j-2},X_{t+j-1},X_{t+j}|x_1,\ldots,x_t,\boldsymbol{\theta}}(x_{t+1},\ldots,x_{t+j-2},x_{t+j-1},x_{t+j})$$

and with the integration region, $C_{t,j}$, being the catastrophe region as in (2.1). If $Z_t \sim N(0,1)$ then

$$P(C_{t,j}|\mathbf{x_2}, D_t, \boldsymbol{\theta}) = \int_u^{+\infty} \int_{-\infty}^{u} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \prod_{k=1}^{j} \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\}$$

$$dx_{t+1} \ldots dx_{t+j-2} dx_{t+j-1} dx_{t+j}. \tag{2.4}$$

$$\begin{aligned} P(C_{t,j}|D_t,\boldsymbol{\theta}) &= P(X_{t+j-1} \leqslant u < X_{t+j}|D_t,\boldsymbol{\theta}) \\ &= P(X_{t+j-1} \leqslant u, X_{t+j} > u|D_t,\boldsymbol{\theta}) \\ &= \int_{C_{t,j}} \int \cdots \int f_2 dx_{t-q+1} \ldots dx_{t+j-2} dx_{t+j-1} dx_{t+j}, \end{aligned}$$

where

$$f_2 \equiv f_{X_{t-q+1},\ldots,X_{t+j-2},X_{t+j-1},X_{t+j}|x_1,\ldots,x_{t-q},\boldsymbol{\theta}}(x_{t-q+1},\ldots,x_{t+j-2},x_{t+j-1},x_{t+j}).$$

Once again, considering the integration region $C_{t,j}$, the catastrophe region in (2.1), and assuming $Z_t \sim N(0,1)$, then

$$\begin{aligned} P(C_{t,j}|D_t,\boldsymbol{\theta}) &= \int_u^{+\infty} \int_{-\infty}^{u} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{k=1}^{q+j} \frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}} \exp\left\{-\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}}\right\} \\ &\quad dx_{t-q+1} \ldots dx_{t+j-2} dx_{t+j-1} dx_{t+j}. \end{aligned}$$

After calculating these probabilities it is then possible to move on to the operating characteristics of the alarm system.

1. Alarm size

Since $\mathbf{X_2} = \{X_{t-q+1}, X_{t-q+2}, \ldots, X_{t-1}, X_t\}$, the size of the alarm region is given by:

$$\begin{aligned} \alpha_{t,j} &= P(A_{t,j}|D_t,\boldsymbol{\theta}) \\ &= \int_{A_{t,j}} \prod_{k=1}^{q} \frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}} \exp\left\{-\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}}\right\} dx_{t-q+1} \ldots dx_t, \end{aligned}$$

with $A_{t,j}$ being the alarm region which depends on the value of $k_{t,j}$ chosen.

2. Probability of correct alarm

$$P(C_{t,j}|A_{t,j},D_t,\boldsymbol{\theta}) = \frac{P(C_{t,j} \cap A_{t,j}|D_t,\boldsymbol{\theta})}{P(A_{t,j}|D_t,\boldsymbol{\theta})},$$

where

$$P(C_{t,j} \cap A_{t,j}|D_t, \boldsymbol{\theta}) = P(X_{t+j-1} \leq u < X_{t+j} \cap \mathbf{X_2} \in A_{t,j}|D_t, \boldsymbol{\theta})$$
$$= \int_u^{+\infty} \int_{-\infty}^{u} \int_{A_{t,j}} \prod_{k=1}^{q+j} \frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}} \exp\left\{ -\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}} \right\}$$
$$dx_{t-q+1} \dots dx_{t+j-2} dx_{t+j-1} dx_{t+j}.$$

Thus, $P(C_{t,j}|A_{t,j}, D_t, \boldsymbol{\theta}) =$

$$= \int_u^{+\infty} \int_{-\infty}^{u} \int_{A_{t,j}} \prod_{k=1}^{q+j} \frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}} \exp\left\{ -\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}} \right\}$$
$$dx_{t-q+1} \dots dx_{t+j-2} dx_{t+j-1} dx_{t+j} \times$$
$$\times \left[\int_{A_{t,j}} \prod_{k=1}^{q} \frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}} \exp\left\{ -\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}} \right\} dx_{t-q+1} \dots dx_t \right]^{-1}.$$

3. Probability of detecting the event

$$\begin{aligned} P(A_{t,j}|C_{t,j}, D_t, \boldsymbol{\theta}) &= \frac{P(A_{t,j} \bigcap C_{t,j}|D_t, \boldsymbol{\theta})}{P(C_{t,j}|D_t, \boldsymbol{\theta})} \\ &= \frac{P(\mathbf{X_2} \in A_{t,j}, X_{t+j-1} \leqslant u < X_{t+j}|D_t, \boldsymbol{\theta})}{P(C_{t,j}|D_t, \boldsymbol{\theta})} \end{aligned}$$

Since the numerator in this expression is the same as the numerator in the expression for the probability of correct alarm, and, given the probability of catastrophe, $P(C_{t,j}|D_t,\boldsymbol{\theta})$,

$$= \int_u^{+\infty}\int_{-\infty}^{u}\int_{A_{t,j}}\prod_{k=1}^{q+j}\frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}}\exp\left\{-\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}}\right\}$$

$$dx_{t-q+1}\dots dx_{t+j-2}dx_{t+j-1}dx_{t+j}\times$$

$$\times \quad \left[\int_u^{+\infty}\int_{-\infty}^{u}\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}\prod_{k=1}^{q+j}\frac{1}{\sqrt{2\pi}\sigma^2_{t-q+k}}\exp\left\{-\frac{x^2_{t-q+k}}{2\sigma^2_{t-q+k}}\right\}\right.$$

$$\left. dx_{t-q+1}\dots dx_{t+j-2}dx_{t+j-1}dx_{t+j}\right]^{-1}.$$

4. Probability of false alarm

$$P(\overline{C_{t,j}}|A_{t,j},D_t,\boldsymbol{\theta}) = 1 - P(C_{t,j}|A_{t,j},D_t,\boldsymbol{\theta}).$$

5. Probability of not detecting the event

$$P(\overline{A_{t,j}}|C_{t,j},D_t,\boldsymbol{\theta}) = 1 - P(A_{t,j}|C_{t,j},D_t,\boldsymbol{\theta}).$$

The application of the alarm system to the FIAPARCH$(1,d,1)$ model will be carried out for the particular case $q=1$ and $j=2$ in Lemma 1.7. The event of interest (i.e. the catastrophe) is defined as the up-crossing of some fixed level u two steps ahead, that is

$$C_{t,2} = \{(x_{t+1},x_{t+2})\in\mathbb{R}^2 : x_{t+1}\le u < x_{t+2}\}. \tag{2.5}$$

The alarm region of optimal size $\alpha_{t,2}$ is given by

$$A_{t,2} = \left\{x_t\in\mathbb{R} : \frac{P(C_{t,2}|x_t,D_t,\boldsymbol{\theta})}{P(C_{t,2}|D_t,\boldsymbol{\theta})}\ge k_{t,2}\right\}$$

$$= \{x_t\in\mathbb{R} : P(C_{t,2}|x_t,D_t,\boldsymbol{\theta})\ge k\}, \tag{2.6}$$

where $k = k_{t,2}P(C_{t,2}|D_t, \boldsymbol{\theta})$.

As already stated, the first step in the construction of the alarm system consists on the calculation of the probability of catastrophe conditional on D_t and x_t, i.e. $P(C_{t,2}|x_t, D_t, \boldsymbol{\theta})$ and $P(C_{t,2}|D_t, \boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\omega, \beta, \phi, \gamma, \delta, d)$. For this purpose, note that

$$\begin{aligned} P(C_{t,2}|x_t, D_t, \boldsymbol{\theta}) &= P(X_{t+1} \leq u < X_{t+2}|x_1, \ldots, x_t, \boldsymbol{\theta}) \\ &= \int_{C_{t,2}} f_{X_{t+1}, X_{t+2}|x_1, \ldots, x_t, \boldsymbol{\theta}}(x_{t+1}, x_{t+2}) dx_{t+1} dx_{t+2} \end{aligned}$$

with the integration region, $C_{t,2}$, being the catastrophe region as in (2.5). If $Z_t \sim N(0,1)$ then

$$P(C_{t,2}|x_t, D_t, \boldsymbol{\theta}) = \int_u^{+\infty} \int_{-\infty}^u \prod_{k=1}^2 \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_{t+1} dx_{t+2}. \tag{2.7}$$

Moreover, $P(C_{t,2}|D_t, \boldsymbol{\theta}) =$

$$\begin{aligned} &= P(X_{t+1} \leq u < X_{t+2}|x_1, \ldots, x_{t-1}, \boldsymbol{\theta}) \\ &= \int_{C_{t,2}} \int f_{X_t, X_{t+1}, X_{t+2}|x_1, \ldots, x_{t-1}, \boldsymbol{\theta}}(x_t, x_{t+1}, x_{t+2}) dx_t dx_{t+1} dx_{t+2}. \end{aligned}$$

Again, by assuming $Z_t \sim N(0,1)$ it follows that

$$P(C_{t,2}|D_t, \boldsymbol{\theta}) = \int_u^{+\infty} \int_{-\infty}^u \int_{-\infty}^{+\infty} \prod_{k=0}^2 \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t dx_{t+1} dx_{t+2}. \tag{2.8}$$

Having calculated these probabilities it is then possible to determine the alarm region and calculate the alarm characteristics of the alarm system.

1. Alarm size

$$\begin{aligned} \alpha_{t,2} &= P(A_{t,2}|D_t, \boldsymbol{\theta}) \\ &= \int_{A_{t,2}} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{x_t^2}{2\sigma_t^2}\right\} dx_t, \end{aligned}$$

with $A_{t,2}$ being the alarm region which depends on the value of $k_{t,2}$ chosen.

2. Probability of correct alarm

$$P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}) \quad = \quad \frac{P(C_{t,2} \cap A_{t,2}|D_t, \boldsymbol{\theta})}{P(A_{t,2}|D_t, \boldsymbol{\theta})},$$

where $P(C_{t,2} \cap A_{t,2}|D_t, \boldsymbol{\theta}) =$

$$\begin{aligned} &= P(X_{t+1} \leq u < X_{t+2} \cap X_t \in A_{t,2}|D_t, \boldsymbol{\theta}) \\ &= \int_u^{+\infty} \int_{-\infty}^u \int_{A_{t,2}} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t dx_{t+1} dx_{t+2}. \end{aligned}$$

Thus $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}) =$

$$= \frac{\int_u^{+\infty} \int_{-\infty}^u \int_{A_{t,2}} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t dx_{t+1} dx_{t+2}}{\int_{A_{t,2}} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{x_t^2}{2\sigma_t^2}\right\} dx_t}.$$

3. Probability of detecting the event

$$P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta}) =$$

$$\begin{aligned} &= \frac{P(A_{t,2} \cap C_{t,2}|D_t, \boldsymbol{\theta})}{P(C_{t,2}|D_t, \boldsymbol{\theta})} \\ &= \frac{\int_u^{+\infty} \int_{-\infty}^u \int_{A_{t,2}} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t dx_{t+1} dx_{t+2}}{\int_u^{+\infty} \int_{-\infty}^u \int_{-\infty}^{+\infty} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi\sigma_{t+k}^2}} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t dx_{t+1} dx_{t+2}}. \end{aligned}$$

4. Probability of false alarm

$$P(\overline{C_{t,2}}|A_{t,2}, D_t, \boldsymbol{\theta}) = 1 - P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}).$$

5. Probability of not detecting the event

$$P(\overline{A_{t,2}}|C_{t,2}, D_t, \boldsymbol{\theta}) = 1 - P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta}).$$

2.2 Estimation procedures

In this section we consider the estimation of the operating characteristics of the alarm system. From the classical framework the method considered is the well-known Quasi-Maximum Likelihood Estimation procedure (QMLE) assuming conditional normality. The QMLE estimates are obtained maximizing the conditional log-likelihood function with respect to $\boldsymbol{\theta} = (\omega, \beta, \phi, \gamma, \delta, d)$, recurring to a routine available within the OxMetrics5 program. The robust standard errors by Bollerslev and Wooldridge (1992) were also calculated. According to these authors this estimator is generally consistent, has a normal limiting distribution and provides asymptotic standard errors that are valid under non-normality. Nevertheless, the authors state that the QMLE estimator is not asymptotically efficient under non-normality and care should be taken, since as Engle and González-Rivera (1991) proved, GARCH estimates are consistent but asymptotically inefficient with the degree of inefficiency increasing with the degree of departure from normality. The impact of violations in conditional normality, however, remains unknown for the FIGARCH and FIAPARCH case. Baillie et al. (1996) suggested that the FIGARCH estimates obtained via QMLE are consistent and asymptotically normal[1]. Furthermore, they also demonstrated the suitability of the QMLE procedure in the estimation of samples with sizes of 1500 and 3000.

From the Bayesian perspective we need to start with a prior distribution for the vector of parameters $\boldsymbol{\theta}$. Assuming independence between all the pa-

[1]In fact, the consistency and asymptotic normality of the QMLE estimator had been formally established for the IGARCH(1,1) process. Baillie et al. (1996) followed a dominance-type argument to extend this result to the FIGARCH$(1, d, 0)$ case and refer the need for a formal proof of consistency and asymptotic normality for the general IGARCH(p, q) and FIAGARCH(p, d, q) cases.

rameters involved, the prior distribution of $\boldsymbol{\theta}$, say $h(\boldsymbol{\theta})$, will be proportional to

$$h(\boldsymbol{\theta}) \propto I_{\{\omega>0\}} I_{\{-1<\beta<1\}} I_{\{\phi\geqslant 0\}} I_{\{-1<\gamma<1\}} I_{\{\delta\geqslant 0\}} I_{\{0<d<1/2\}}.$$

The posterior distribution $h(\boldsymbol{\theta}|D_t)$ is proportional to $L(D_t|\boldsymbol{\theta})h(\boldsymbol{\theta})$,

$$\begin{aligned} h(\boldsymbol{\theta}|D_t) &\propto L(D_t|\boldsymbol{\theta})h(\boldsymbol{\theta}) \\ &\propto \prod_{n=2}^{t-1} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{x_n^2}{2\sigma_n^2}\right\} \times \\ &\times I_{\{\omega>0\}} I_{\{-1<\beta<1\}} I_{\{\phi\geqslant 0\}} I_{\{-1<\gamma<1\}} I_{\{\delta\geqslant 0\}} I_{\{0<d<1/2\}}. \end{aligned}$$

Hence, the probability of catastrophe conditional on D_t and $\mathbf{x}_2 = \{x_t\}$, takes the form

$$P(C_{t,2}|x_t, D_t) = \int_{\Theta} P(C_{t,2}|x_t, D_t, \boldsymbol{\theta}) h(\boldsymbol{\theta}|D_t) d\boldsymbol{\theta} \tag{2.9}$$

with Θ being the parameter space. On the other hand, the probability of catastrophe conditional on D_t, will be given by

$$P(C_{t,2}|D_t) = \int_{\Theta} P(C_{t,2}|D_t, \boldsymbol{\theta}) h(\boldsymbol{\theta}|D_t) d\boldsymbol{\theta}, \tag{2.10}$$

where $P(C_{t,2}|x_t, D_t, \boldsymbol{\theta})$ and $P(C_{t,2}|D_t, \boldsymbol{\theta})$ are calculated through (2.7) and (2.8), respectively. However, due to the complexity of expressions (2.7) and (2.8), analytical calculations are not possible. Nonetheless, since by (2.9) and (2.10)

$$P(C_{t,2}|x_t, D_t) = \mathrm{E}_{\boldsymbol{\theta}|D_t}[P(C_{t,2}|x_t, D_t, \boldsymbol{\theta})]$$

and

$$P(C_{t,2}|D_t) = \mathrm{E}_{\boldsymbol{\theta}|D_t}[P(C_{t,2}|D_t, \boldsymbol{\theta})],$$

their respective Monte Carlo approximations can be used, that is

$$\widehat{P}(C_{t,2}|x_t, D_t) = \frac{1}{m} \sum_{i=1}^{m} P(C_{t,2}|x_t, D_t, \boldsymbol{\theta_i})$$

and

$$\widehat{P}(C_{t,2}|D_t) = \frac{1}{m} \sum_{i=1}^{m} P(C_{t,2}|D_t, \boldsymbol{\theta_i}),$$

where the observations $\boldsymbol{\theta_i} = (\omega_i, \beta_i, \phi_i, \gamma_i, \delta_i, d_i)$ with $i = 1, 2, \ldots, m$ constitute a sample of the posterior distribution $h(\boldsymbol{\theta}|D_t)$. A similar procedure is applied to approximate the operating characteristics.

2.3 Simulation results

In this section we present a simulation study to illustrate the performance of the alarm system constructed for the FIAPARCH$(1, d, 1)$ model. We considered two sets of parameters, namely, $\boldsymbol{\theta_1} = (0.40, 0.28, 0.10, 0.68, 1.27, 0.30)$ and $\boldsymbol{\theta_2} = (0.80, 0.52, 0.37, 0.76, 1.40, 0.20)$. The choice of the parameters in $\boldsymbol{\theta_1}$ is very similar to those appearing in the real-data example presented in Section 2.4. Figure 2.1 below shows a simulated sample path for the particular choice of $\boldsymbol{\theta_1}$.

Parameter estimates, $\hat{\boldsymbol{\theta_1}}$ and $\hat{\boldsymbol{\theta_2}}$, and their corresponding standard errors were obtained for both samples, following the QMLE procedure of Bollerslev and Wooldridge (1992). Robust standard errors are estimated from the product $A(\hat{\boldsymbol{\theta_i}})^{-1}B(\hat{\boldsymbol{\theta_i}})A(\hat{\boldsymbol{\theta_i}})^{-1}$, where $A(\hat{\boldsymbol{\theta_i}})$ and $B(\hat{\boldsymbol{\theta_i}})$ denote the Hessian and the outer product of the gradients evaluated at $\hat{\boldsymbol{\theta_i}}$, $i = 1, 2$, respectively. The OxMetrics5 program was used.

Moreover, Bayesian estimates of $\boldsymbol{\theta}$ were also obtained for both samples. Since the standard Gibbs methodology is difficult to implement to FIAPARCH models partially due to the non-standard forms of the full conditional densities, the Metropolis-Hastings algorithm was implemented in the software Matlab. In addition, a multivariate t-distribution was used as the proponent one. The sampler algorithm ran 100000 iterations including a *burn-in* period of 40000 observations which are discarded for the posterior analysis, as suggested by Vrontos et al. (2000). Furthermore, only every twentieth iteration is stored in order to obtain an, approximately, independent and identically

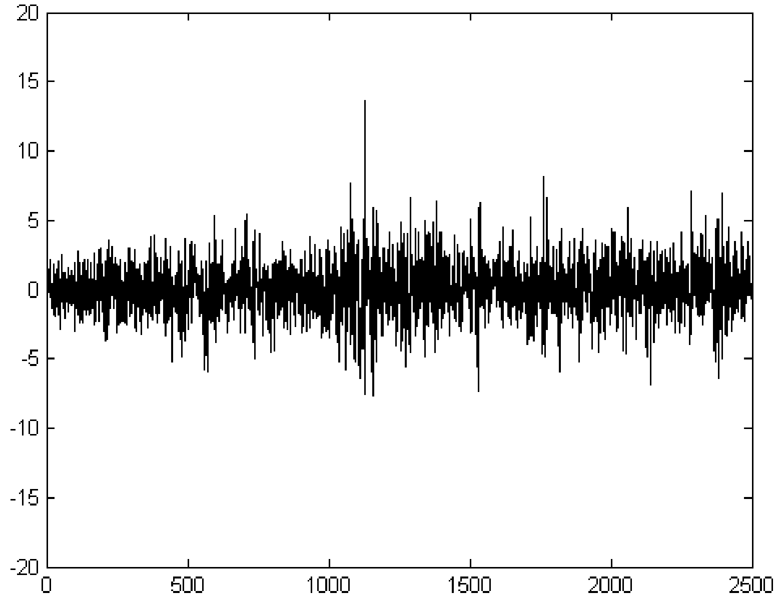


Figure 2.1: FIAPARCH$(1, d, 1)$ process: simulated sample path with $\boldsymbol{\theta_1} = (0.40, 0.28, 0.10, 0.68, 1.27, 0.30)$.

distributed sample. The estimates were taken as the means of the posterior distribution. The convergence of the Markov chain was analyzed through the R criterion of Gelman and Rubin (1992), the Z-score test of Geweke (1992) and by graphical methods.

Parameter estimates obtained with both classical and Bayesian procedures are presented in Table 2.1, for both samples, with standard deviations given in parenthesis.

The analysis of the alarm system is carried out at $t = 2000$, i.e., $\mathbf{x}_2 = \{x_{2000}\}$. The event of interest is the two step ahead catastrophe defined by the upcrossing of the fixed level u, at time $t+2$: $C_{2000,2} = \{(x_{2001}, x_{2002}) \in \mathbb{R}^2 :$

Table 2.1: Parameters and Estimates.

	Sample 1			Sample 2		
	True Parameters	QMLE	Bayesian Estimates	True Parameters	QMLE	Bayesian Estimates
ω	0.40	0.3181	0.2876	0.80	0.5016	0.4244
		(0.0737)	(0.0669)		(0.1605)	(0.1299)
ϕ	0.10	0.2004	0.2253	0.37	0.4357	0.4200
		(0.0919)	(0.0900)		(0.0872)	(0.0596)
γ	0.68	0.6734	0.6743	0.76	0.5536	0.4953
		(0.1050)	(0.1210)		(0.1384)	(0.1345)
β	0.28	0.3936	0.4069	0.52	0.6449	0.6665
		(0.1168)	(0.1076)		(0.0967)	(0.0583)
δ	1.27	1.2164	1.3732	1.40	1.4641	1.5036
		(0.2450)	(0.2117)		(0.2928)	(0.2313)
d	0.30	0.3116	0.2978	0.20	0.3107	0.3542
		(0.0636)	(0.0580)		(0.0715)	(0.0750)

$x_{2001} \leq u < x_{2002}\}$. In a first stage, two values of u were chosen, accordingly to the sample quantiles, namely the 90th percentile ($Q_{0.90}$), and the 95th percentile ($Q_{0.95}$). The choice of these values is justified by the fact that we are interested in relatively rare events. For both fixed levels u, the probabilities $P(C_{t,2}|x_t, D_t, \boldsymbol{\theta})$ and $P(C_{t,2}|D_t, \boldsymbol{\theta})$ were numerically approximated as described in the previous section. In order to compute the optimal alarm region for each case, one has to obtain the region for several values of k, according to expression (2.6) and then, for each value of k, compute the operating characteristics of the alarm system, i.e., the size of the region, $\alpha_{t,2}$, the probability of correct alarm, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ and the probability of detection, $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$. For every fixed value of k the region has to be obtained through a systematic search in a three dimensional region for (x_t, x_{t+1}, x_{t+2}). We considered a thin grid of values of x_t in

$[-100, 100]$ and determined, for each value of x_t, whether $P(C_{t,2}|x_t, D_t, \boldsymbol{\theta})$ exceeded k. This procedure was repeated for k ranging from $P(C_{t,2}|D_t, \boldsymbol{\theta})$ to $P(C_{t,2}|D_t, \boldsymbol{\theta}) + n \times 0.005$, with $n \in \mathbb{R}^+$, for Sample 1. For Sample 2 the k step considered was 0.002 instead of 0.005. This procedure is repeated for both the classical (using the true values of the parameters and their QMLE estimates) and the Bayesian approach. Results are shown in Table 2.2 for Sample 1 and in Table 2.3 for Sample 2. Just note that in Table 2.2 and Table 2.3, $P(C_{t,2}|A_{t,2})$ and $P(A_{t,2}|C_{t,2})$ are also conditioned on the past and should be written as $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ and $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$, respectively. However, due to the limited space available, the conditioning on D_t and $\boldsymbol{\theta}$ was omitted.

Table 2.2: Operating Characteristics for Sample 1, at time point $t = 2000$.

$\mu = Q_{0.95}$=3.136									
	True Parameters $P(C_{t,2}\|D_t) = 0.0328$			QMLE $P(C_{t,2}\|D_t) = 0.0340$			Bayesian Estimates $P(C_{t,2}\|D_t) = 0.0236$		
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0350	0.4789	0.0335	0.4903	0.5353	0.0346	0.5446	0.1904	0.0267	0.2155
0.0400	0.2998	0.0345	0.3155	0.3255	0.0355	0.3400	0.1902	0.0257	0.2074
0.0450	0.2072	0.0349	0.2209	0.2971	0.0359	0.3133	0.1211	0.0264	0.1354
0.0500	0.2067	0.0344	0.2173	0.2102	0.0363	0.2247	0.0718	0.0283	0.0862
0.0600	0.1377	0.0347	0.1458	0.1413	0.0360	0.1496	0.0397	0.0318	0.0535
0.0700	0.0864	0.0363	0.0957	0.0896	0.0373	0.0983	0.0203	0.0391	0.0337
0.0800	0.0509	0.0390	0.0605	0.0535	0.0398	0.0625	0.0097	0.0555	0.0227
0.0900	0.0282	0.0439	0.0377	0.0300	0.0454	0.0401	0.0042	0.0982	0.0177
0.1000	0.0146	0.0558	0.0248	0.0158	0.0563	0.0262	0.0017	0.2061	0.0151

μ=$Q_{0.90}$=2.293									
	True Parameters $P(C_{t,2}\|D_t) = 0.0827$			QMLE $P(C_{t,2}\|D_t) = 0.0844$			Bayesian Estimates $P(C_{t,2}\|D_t) = 0.0693$		
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0850	0.5303	0.0832	0.5339	0.6042	0.0846	0.6055	0.1904	0.0722	0.1984
0.0900	0.3209	0.0837	0.3250	0.3490	0.0853	0.3528	0.1902	0.0719	0.1974
0.0950	0.2960	0.0844	0.3021	0.3033	0.0849	0.3050	0.1211	0.0717	0.1252
0.1000	0.2069	0.0843	0.2109	0.2117	0.0864	0.2167	0.1211	0.0713	0.1245
0.1100	0.1377	0.0852	0.1420	0.2101	0.0859	0.2137	0.0718	0.0730	0.0757
0.1200	0.0864	0.0862	0.0901	0.1413	0.0864	0.1446	0.0397	0.0773	0.0442
0.1300	0.0509	0.0887	0.0546	0.0535	0.0905	0.0573	0.0203	0.0825	0.0242
0.1400	0.0282	0.0888	0.0302	0.0535	0.0904	0.0572	0.0042	0.1123	0.0069
0.1500	0.0146	0.0965	0.0170	0.0158	0.1054	0.0197	0.0017	0.2474	0.0062

Table 2.3: Operating Characteristics for Sample 2, at time point $t = 2000$.

$\mu = Q_{0.95}$=7.155									
	True Parameters			QMLE			Bayesian Estimates		
	$P(C_{t,2}\|D_t) = 0.0746$			$P(C_{t,2}\|D_t) = 0.0893$			$P(C_{t,2}\|D_t) = 0.0817$		
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0900	0.0372	0.1386	0.0691	0.1830	0.0971	0.1989	0.0681	0.1126	0.0938
0.0920	0.0285	0.1592	0.0608	0.1508	0.1005	0.1697	0.0520	0.1229	0.0782
0.0940	0.0216	0.1870	0.0541	0.1226	0.1038	0.1425	0.0391	0.1362	0.0652
0.0960	0.0161	0.2293	0.0496	0.0777	0.1155	0.1005	0.0289	0.1638	0.0580
0.0980	0.0119	0.2825	0.0451	0.0606	0.1254	0.0851	0.0151	0.2347	0.0433
0.1000	0.0087	0.3712	0.0432	0.0466	0.1396	0.0728	0.0106	0.3094	0.0403
0.1020	0.0063	0.4785	0.0401	0.0353	0.1587	0.0627	0.0074	0.4080	0.0368
0.1040	0.0044	0.6359	0.0378	0.0264	0.1826	0.0539	0.0050	0.5759	0.0354
μ=$Q_{0.90}$=5.432									
	True Parameters			QMLE			Bayesian Estimates		
	$P(C_{t,2}\|D_t) = 0.1267$			$P(C_{t,2}\|D_t) = 0.1439$			$P(C_{t,2}\|D_t) = 0.1353$		
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1440	0.0216	0.2321	0.0395	0.1508	0.1532	0.1605	0.0520	0.1736	0.0667
0.1460	0.0119	0.3284	0.0309	0.1226	0.1575	0.1341	0.0289	0.2110	0.0451
0.1480	0.0087	0.4083	0.0280	0.0777	0.1697	0.0916	0.0211	0.2378	0.0370
0.1500	0.0063	0.5163	0.0255	0.0606	0.1780	0.0749	0.0106	0.3403	0.0267
0.1520	0.0044	0.6781	0.0238	0.0353	0.2082	0.0511	0.0074	0.4435	0.0241
0.1540	0.0031	0.9331	0.0229	0.0264	0.2292	0.0420	0.0034	0.8073	0.0201

Results are very different for the two samples. For the first sample analysed and considering the true values of the parameters, the probability of the alarm being correct, does not exceed 5.6% in the $u = Q_{0.95}$ case, or 9.7% in the $u = Q_{0.90}$ case. The probability of detection for this sample, ranges from 2.4% to 49.0% for $u = Q_{0.95}$, or from 1.7% to 53.4% for $u = Q_{0.90}$. The results obtained with the QML estimates do not differ considerably, in particular in what concerns the probability of correct alarm. Regarding the probability of detecting the event, we can say the alarm system behaves better in this case since the detection probability reaches 54.5% for $u = Q_{0.95}$ and 60.6% for $u = Q_{0.90}$. Considering now the Bayesian approach, the probability of detection is the lowest obtained. It does not even reach 22%. On the other hand, the estimation procedure involved in the Bayesian approach seems to be able to produce higher probabilities of correct alarm, depending on an accurate choice of k. The probability of correct alarm ranges from lower values than in the classical approach to more than the double of these values, with increasing k, reaching 24.7% in the $u = Q_{0.90}$ case. Furthermore, note that as the probability of correct alarm increases, the probability of detecting the event decreases, as expected. This can be justified by the fact that as k increases, the size of the alarm region decreases, which implies that the number of alarms should decrease, so as the probability of detection, $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$. However, as the number of alarms decreases, the probability of false alarms also decreases and therefore the probability of the alarm being correct, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$, increases.

In the second sample, results are quite different: although the general tendencies of rising the probability of correct alarm and decreasing the detection probability with the increase in k are followed, the ranges of variation are different. While in the first sample, the widest ranges of variation were observed for the detection probability, in the second sample, the widest ranges of variation were obtained for the probability of the alarm being correct. For instance, considering the true values of the parameters, the probability of correct alarm reaches very high values: it ranges from 13.9% to 63.6% in the $\mu = Q_{0.95}$ case and from 23.2% to 93.3% in the $\mu = Q_{0.90}$ case. Considering QMLE, the probability of correct alarm only reaches 18.3% in the $\mu = Q_{0.95}$ case and 22.9% in the $\mu = Q_{0.90}$ case. Also with the Bayesian approach the range of variation can be considered very large: from 11.3% to 57.6% in the $\mu = Q_{0.95}$ case and from 17.4% to 80.7% in the $\mu = Q_{0.90}$ case. On the other hand, quite small ranges of variation are observed for the detection probability, which decreases from 6.9% to 3.8%, considering $\mu = Q_{0.95}$ and from 4.0% to 2.3%, considering $\mu = Q_{0.90}$, in the situation were the true values of the parameters are considered. The widest ranges of variation in the detection probability are obtained when considering the QML estimates, ranging from 19.9% to 5.4% in the $\mu = Q_{0.95}$ case, and ranging from 16.1% to 4.2% in the $\mu = Q_{0.90}$ case. For the Bayesian estimates, the probability of detecting the event does not even reach 10%.

As already discussed, it is not possible, in general, to maximize both probabilities, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ and $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$, simultaneously. Hence, a compromise should be reached by the proper choice of k. In doing so, several criteria have been already proposed. Svensson et al. (1996), for example, suggested that k should be chosen so that the probability of correct alarm and the probability of detecting the event are approximately equal, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}) \simeq P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$. On the other hand, Antunes et al. (2003) suggested that k should be chosen so that the alarm size is about twice the probability of having a catastrophe given the past values of the process, $P(C_{t,2}|D_t, \boldsymbol{\theta}) \simeq \frac{1}{2}P(A_{t,2}|D_t, \boldsymbol{\theta})$, stating that in this situation the system will be spending twice the time in the alarm state than in the catastrophe region. We analysed both criteria in this work and from hereafter, the former criterion will be designated by **Criterion 2** and the last by **Criterion 1**. Also, from hereafter, we will consider only sample 1, simulated with $\boldsymbol{\theta_1} = (0.40, 0.28, 0.10, 0.68, 1.27, 0.30)$, as this choice of parameters is very similar to the ones estimated from the real-data example presented in Section 2.4.

In order to test the alarm system, three extra values of the series were simulated, $(\mathbf{x}_2, \mathbf{x}_3) = (x_t, x_{t+1}, x_{t+2})$. This procedure was repeated 10000 times with the same informative experience, D_t. With the alarm regions calculated before for $u = Q_{0.90} = 2.293$ and for the two criteria already mentioned, we observed, for each of the 10000 samples, whether an alarm was given or not and whether a catastrophe occurred or not. Results are given in Table 2.4.

Criterion 1 tends to provide better estimates for the probability of correct alarm and detection probability than Criterion 2. The probability of the

Table 2.4: Results at time point $t = 2000$. Percentages in parenthesis.

	Criterion	Alarms		Catastrophes	
		False	Total	Detected	Total
True Parameters	1	1112 (0.8330)	1335	223 (0.2059)	1083
	2	651 (0.8314)	783	132 (0.1273)	1037
QMLE Approach	1	1163 (0.8526)	1364	201 (0.1963)	1024
	2	380 (0.8260)	460	80 (0.0771)	1037
Bayesian Approach	1	1161 (0.8401)	1382	221 (0.2103)	1051
	2	668 (0.8477)	788	120 (0.1204)	997

alarm being correct is even higher than the theoretical expected values in Table 2.2 for the cases in which the true parameters or the QML estimates are considered: it is approximately 16.7% and 14.7% respectively, for Criterion 1. The estimated detection probability, even though lower than maximum theoretical values in Table 2.2, is also higher than expected, considering that some criterion was being pursued. Take, for instance, the case in which the Bayesian estimates are considered and note that the probability of correct alarm of around 16.0% with Criterion 1 corresponds to a detection probability of 21.0%, a much higher value than the one expected from inspection of Table 2.2. It seems that, in practice, estimated operating characteristics tend to reach higher values than what was expected theoretically.

Finally, we illustrate how the on-line prediction performs in practice. The event to predict is $C_{t,2} = \{(x_{t+1}, x_{t+2}) \in \mathbb{R}^2 : x_{t+1} \leq u < x_{t+2}\}$, $t = 2000, \ldots, 2010$, again with $u = Q_{0.90} = 2.293$. Alarm regions and corresponding operating characteristics are presented in Table 2.5 for Criterion 1 and in Table 2.6 for Criterion 2.

Table 2.5: Operating characteristics at different time points for Criterion 1.

Approach	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
True Parameters	2000	0.0827	0.1100	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1377	0.0852	0.1420
	2001	0.1047	0.1047	$[-\infty,-1.5]\cup[5.5,+\infty]$	0.1848	0.1093	0.1929
	2002	0.0936	0.0936	$[-\infty,-2.0]\cup[9.5,+\infty]$	0.1209	0.0980	0.1265
	2003	0.0923	0.1073	$[-\infty,-1.5]\cup[7.5,+\infty]$	0.2167	0.0947	0.2224
	2004	0.0897	0.0977	$[-\infty,-1.5]\cup[8.0,+\infty]$	0.2076	0.0914	0.2116
	2005	0.0879	0.0979	$[-\infty,-1.5]\cup[7.5,+\infty]$	0.2036	0.0893	0.2069
	2006	0.0803	0.0953	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1311	0.0831	0.1356
	2007	0.0687	0.0887	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.1286	0.0716	0.1340
	2008	0.0573	0.0873	$[-\infty,-2.0]\cup[9.5,+\infty]$	0.1194	0.0614	0.1279
	2009	0.0508	0.0758	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.1045	0.0522	0.1075
	2010	0.0545	0.0845	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.0924	0.0566	0.0960
QMLE	2000	0.0844	0.1200	$[-\infty,-2.0]\cup[10.5,+\infty]$	0.1413	0.0864	0.1446
	2001	0.1047	0.1097	$[-\infty,-1.5]\cup[6.0,+\infty]$	0.1867	0.1123	0.2002
	2002	0.0969	0.0969	$[-\infty,-2.0]\cup[9.5,+\infty]$	0.1230	0.1005	0.1276
	2003	0.0946	0.1096	$[-\infty,-1.5]\cup[7.5,+\infty]$	0.2202	0.0972	0.2262
	2004	0.0919	0.1019	$[-\infty,-1.5]\cup[7.5,+\infty]$	0.2110	0.0943	0.2165
	2005	0.0900	0.1000	$[-\infty,-1.5]\cup[7.5,+\infty]$	0.2066	0.0917	0.2104
	2006	0.0821	0.0971	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.1340	0.0843	0.1376
	2007	0.0697	0.0897	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.1314	0.0723	0.1363
	2008	0.0594	0.0894	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1217	0.0619	0.1269
	2009	0.0506	0.0756	$[-\infty,-2.0]\cup[8.0,+\infty]$	0.1059	0.0528	0.1104
	2010	0.0544	0.0844	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.0930	0.0566	0.0966
Bayesian	2000	0.0693	0.0950	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.1211	0.0717	0.1252
	2001	0.0911	0.0911	$[-\infty,-1.5]\cup[6.0,+\infty]$	0.1685	0.0939	0.1736
	2002	0.0820	0.0820	$[-\infty,-2.0]\cup[9.5,+\infty]$	0.1047	0.0845	0.1078
	2003	0.0794	0.0994	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1297	0.0820	0.1340
	2004	0.0764	0.0914	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1218	0.0797	0.1271
	2005	0.0715	0.0915	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1176	0.0779	0.1282
	2006	0.0680	0.0830	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1144	0.0711	0.1196
	2007	0.0576	0.0776	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1121	0.0598	0.1165
	2008	0.0498	0.0748	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.1038	0.0513	0.1068
	2009	0.0419	0.0669	$[-\infty,-2.0]\cup[9.0,+\infty]$	0.0902	0.0441	0.0948
	2010	0.0447	0.0747	$[-\infty,-2.0]\cup[9.5,+\infty]$	0.0790	0.0467	0.0825

Regarding the probability of correct alarm, results presented in Tables 2.5 and 2.6 are very similar: it ranges from around 5 to 11% considering the true values of the parameters or the QML estimates and from approximately 4 to 9% considering the Bayesian estimates. The main difference between Table 2.5 and Table 2.6 resides in the values of the detection probability that are always higher for Criterion 1, reaching values near 23% in some time instants, in the classical approach. With Criterion 2, the probability of detecting the event is always about half the value obtained with Criterion

Table 2.6: Operating characteristics at different time points for Criterion 2.

Approach	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
	2000	0.0827	0.1200	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0864	0.0862	0.0901
	2001	0.1047	0.1247	$[-\infty,-2.0]\cup[10.5,+\infty]$	0.1153	0.1088	0.1198
	2002	0.0936	0.1036	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0717	0.1001	0.0767
	2003	0.0923	0.1223	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0958	0.0949	0.0985
	2004	0.0897	0.1147	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0872	0.0924	0.0899
True Parameters	2005	0.0879	0.1129	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0835	0.0906	0.0862
	2006	0.0803	0.1053	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0805	0.0831	0.0832
	2007	0.0687	0.0987	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0783	0.0726	0.0827
	2008	0.0573	0.1023	$[-\infty,-2.5]\cup[13.0,+\infty]$	0.0705	0.0630	0.0774
	2009	0.0508	0.0908	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0582	0.0531	0.0608
	2010	0.0545	0.0945	$[-\infty,-2.5]\cup[11.0,+\infty]$	0.0487	0.0593	0.0530
	2000	0.0844	0.1300	$[-\infty,-3.0]\cup[13.5,+\infty]$	0.0535	0.0905	0.0573
	2001	0.1047	0.1297	$[-\infty,-2.0]\cup[10.5,+\infty]$	0.1174	0.1104	0.1238
	2002	0.0969	0.1069	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0735	0.1027	0.0780
	2003	0.0946	0.1246	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0992	0.0974	0.1021
	2004	0.0919	0.1169	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0904	0.0947	0.0932
QMLE	2005	0.0900	0.1150	$[-\infty,-2.5]\cup[11.0,+\infty]$	0.0863	0.0929	0.0891
	2006	0.0821	0.1121	$[-\infty,-2.5]\cup[12.5,+\infty]$	0.0831	0.0850	0.0860
	2007	0.0697	0.0997	$[-\infty,-2.5]\cup[11.0,+\infty]$	0.0808	0.0731	0.0847
	2008	0.0594	0.0994	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0723	0.0637	0.0776
	2009	0.0506	0.0956	$[-\infty,-2.5]\cup[13.0,+\infty]$	0.0593	0.0529	0.0619
	2010	0.0544	0.0994	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0491	0.0590	0.0533
	2000	0.0693	0.1100	$[-\infty,-2.5]\cup[12.5,+\infty]$	0.0718	0.0730	0.0757
	2001	0.0911	0.1011	$[-\infty,-2.0]\cup[8.5,+\infty]$	0.1002	0.0943	0.1037
	2002	0.0820	0.0820	$[-\infty,-2.0]\cup[9.5,+\infty]$	0.1047	0.0845	0.1078
	2003	0.0794	0.1094	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0793	0.0835	0.0835
	2004	0.0764	0.1014	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0724	0.0813	0.0771
Bayesian	2005	0.0715	0.1065	$[-\infty,-2.5]\cup[13.5,+\infty]$	0.0689	0.0794	0.0766
	2006	0.0680	0.0930	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0663	0.0726	0.0707
	2007	0.0576	0.0876	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0643	0.0619	0.0692
	2008	0.0498	0.0848	$[-\infty,-2.5]\cup[12.0,+\infty]$	0.0576	0.0536	0.0619
	2009	0.0419	0.0769	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0470	0.0461	0.0517
	2010	0.0447	0.0847	$[-\infty,-2.5]\cup[11.5,+\infty]$	0.0388	0.0476	0.0413

1, leading us to conclude that, overall, Criterion 1 provides better estimates for the operating characteristics.

2.4 Exploring the IBOVESPA returns data set

In this section, we model the data set IBOVESPA which contains daily returns of the São Paulo Stock Market during the period 04/07/1994 to 02/10/2008 (www.ipeadata.gov.br). Data consists on the closing rates of stocks, I_t, being the log-returns calculated as $y_t = \ln(I_t/I_{t-1})$, $t = 1, \ldots, n$. The results obtained from this procedure were then multiplied by 100 just to ensure the stability of posterior calculations. Sáfadi and Pereira (2010) proved that the FIAPARCH$(1, d, 1)$ provides a good fit for this kind of data sets. To fit a FIAPARCH$(1, d, 1)$ model for the log-returns we proceeded as follows: first, the AR(10) model

$$y_t = 0.0689 + 0.0645 y_{t-10} + x_t,$$

is fitted using the least squares method, in order to eliminate serial dependence. The time series plots of both the IBOVESPA daily returns and the residuals (x_t), hereafter designated by x-returns, are exhibited in Figure 2.2 below.

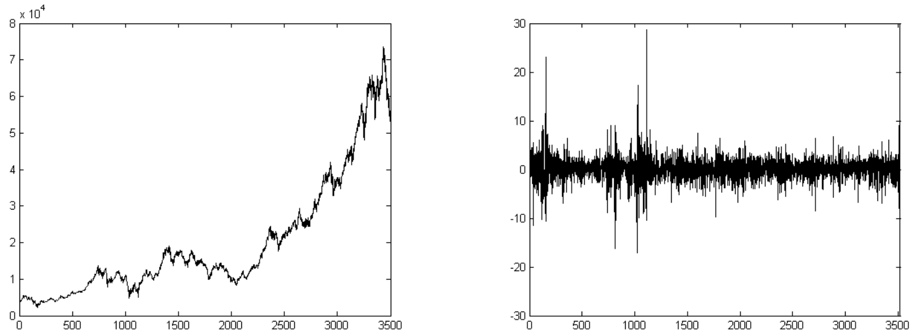


Figure 2.2: Plot of the IBOVESPA data, I_t, (left) and the x-returns, x_t, (right) from 04/07/1994 to 02/10/2008.

This is, indeed, the set of data reported to show the common features of financial time series mentioned in Section 1.2.1, that is weak dependence without any evident pattern on the series level and significant dependence on squared and absolute returns.

The FIAPARCH$(1, d, 1)$ model was fitted to the series of x-returns by means of the QMLE procedure and the Bayesian approach described in Section 2.2. In both cases the adequacy of the fit was checked through the analysis of the standardized residuals. Table 2.7 presents the estimates obtained for both procedures.

Table 2.7: Parameter estimates. Standard deviations in parenthesis.

	QMLE	Bayesian Estimates
ω	0.3903 (0.1092)	0.4227 (0.0576)
ϕ	0.0957 (0.1334)	0.1289 (0.0397)
γ	0.6782 (0.1363)	0.7813 (0.1108)
β	0.2794 (0.1693)	0.3246 (0.0568)
δ	1.2744 (0.1274)	1.2218 (0.1008)
d	0.2952 (0.0642)	0.3020 (0.0258)

Since the IBOVESPA x-returns are related to the daily changes of the stock indexes of S. Paulo Stock Market, we considered that the event of interest is given by

$$C_{t,2} = \{(x_{t+1}, x_{t+2}) \in \mathbb{R}^2 : x_{t+1} \geq u > x_{t+2}\}$$

with $t = 3450, \ldots, 3516$, corresponding to July, August and September of 2008, and $u = Q_{0.25} = -1.219$. Note that, the down-crossing event $C_{t,2}$ can be viewed as related with a stock market crash. Moreover, the choice of k was done according only to Criterion 1: $P(C_{t,2}|D_t, \boldsymbol{\theta}) \simeq \frac{1}{2}P(A_{t,2}|D_t, \boldsymbol{\theta})$. Two reasons justify this choice. First, Criterion 2 is difficult to implement since $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ may never get so close to $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$ or when it does, some operating characteristics may show not so good results (at least as compared with those obtained with Criterion 1). Secondly, Criterion 1 results in better estimates of the operating characteristics. For the time period considered, the total number of alarms, the total number of catastrophes, the number of false alarms and the number of detected events was counted. Results are presented in Table 2.8, considering the QMLE procedure.

Table 2.8: Results of the alarm system with $u = -1.219$. Percentages in parenthesis.

Month	Alarms		Catastrophes	
	False	Total	Detected	Total
July	1 (0.50)	2	1 (0.16)	6
August	1 (0.50)	2	1 (0.20)	5
September	0 (0.00)	3	3 (0.27)	11
Trimester	2 (0.28)	7	5 (0.22)	22

A closer look at Table 2.8 reveals that the estimate of the probability of the alarm being correct is 50% in July and August and raises to 100% in September. In addition, the estimate of the probability of detecting a catastrophe remains around 20% during the time period considered. We noticed that this on-line prediction system exhibits an adaptive behaviour, that is, as long as the available information is integrated within the informative experience, the system adapts itself in order to produce the minimum number of false alarms. This fact explains on one hand the high estimate of the probabilities of the alarm given being correct and on the other hand that the system produces few alarms, so the probability of detection can not be very high.

Chapter 3

Integer-valued Asymmetric Power ARCH Model

This chapter is organized as follows: in Section 3.1, relevant background information and the reasons for the introduction of the INGARCH-type model proposed in this work are presented. Definitions are given in Section 3.2 and probabilistic properties of the proposed model are discussed. Parameter estimation is covered in Section 3.3. In Section 3.4, results are illustrated through a simulation study. The implementation of an optimal alarm system to the INAPARCH(1,1) model is done in Section 3.5. Finally, in Section 3.6, an application is presented to two data series concerning the number of transactions in stocks.

3.1 Introduction

The analysis of continuous-valued financial time series like log-return series of foreign exchange rates, stock indices or share prices, has revealed some common features: sample means not significantly different from zero, sample variances of the order 10^{-4} or smaller and sample distributions roughly symmetric in the center, sharply peaked around zero but with a tendency to negative asymmetry. In particular, it has usually been found that the conditional volatility of stocks responds asymmetrically to positive versus negative shocks: volatility tends to rise higher in response to negative shocks as opposed to positive shocks, which is known as the leverage effect. To account for asymmetric responses in the volatility, Ding et al. (1993) introduced the Asymmetric Power ARCH or, in short, APARCH(p, q) in which

$$Y_t = \sigma_t Z_t, \; \sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta, \; t \in \mathbb{Z}, \quad (3.1)$$

where (Z_t) is an i.i.d. sequence with zero mean, $\omega > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $\delta \geqslant 0$, $-1 < \gamma_i < 1$. As mentioned in Section 1.2.1, the APARCH representation in (3.1) has some noteworthy advantages, from which we would like to point out the fact that the model allows the detection of asymmetric responses of the volatility for positive or negative shocks.

Asymmetric responses on the volatility are also commonly observed in the analysis of time series representing the number of intra-day transactions in stocks, in which the numbers are typically quite small, as reported in Brännäs and Quoreshi (2010). It is worth mentioning that none of the models covered in Section 1.2.3 is able to cope with the presence of asymmetric overdispersion in data. In order to account for this feature, we propose a counterpart of the APARCH model for the analysis of time series of counts.

3.2 Integer-valued APARCH(p, q) Processes

Definition 3.1 (INAPARCH(p, q) model)**.** *An INteger-valued APARCH(p, q) is defined to be an integer-valued process (Y_t), such that, conditioned on the past, the distribution of Y_t is Poisson with mean value λ_t satisfying the recursive equation*

$$\lambda_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|Y_{t-i} - \lambda_{t-i}| - \gamma_i (Y_{t-i} - \lambda_{t-i}))^\delta + \sum_{j=1}^{q} \beta_j \lambda_{t-j}^\delta, \; t \in \mathbb{Z}$$

with $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $|\gamma_i| < 1$ *and* $\delta \geq 0$.

3.2.1 First and second order moments of INAPARCH(1,1)

In deriving the first and second order moments of the INAPARCH$(1, 1)$ the particular case $\delta = 1$ is considered. Note that the unconditional mean must obey the relation

$$\mathrm{E}[Y_t] = \omega + \beta_1 \, \mathrm{E}[Y_{t-1}] + \alpha_1 \, \mathrm{E}[|Y_{t-1} - \lambda_{t-1}|],$$

in which

$$\begin{aligned} \mathrm{E}[|Y_{t-1} - \lambda_{t-1}|] &= \mathrm{E}[Y_{t-1}] \left(1 - 2\phi_1 + \frac{H(\lambda_{t-1} - 2)}{H(\lambda_{t-1} - 1)} \phi_1 - \right. \\ &- \left. \frac{1 - H(\lambda_{t-1} - 2)}{1 - H(\lambda_{t-1} - 1)} (1 - \phi_1) \right), \end{aligned} \tag{3.2}$$

where $H(\lambda_{t-1} - i) := 1 - F_P(\lambda_{t-1} - i)$, for $i = 1, 2$, and $F_P(\cdot)$ stands for the Poisson distribution function. In the expression above, ϕ_1 represents the probability of Y_{t-1} being greater or equal to the conditional mean λ_{t-1}, at the same time point, $\phi_1 = P(Y_{t-1} - \lambda_{t-1} \geqslant 0)$. After suitable substitution, it comes that

$$\begin{aligned} \mathrm{E}[Y_t] &= \omega + \mathrm{E}[Y_{t-1}] \left\{ \beta_1 + \alpha_1 \left(1 - 2\phi_1 + \frac{H(\lambda_{t-1} - 2)}{H(\lambda_{t-1} - 1)} \phi_1 - \right. \right. \\ &- \left. \left. \frac{1 - H(\lambda_{t-1} - 2)}{1 - H(\lambda_{t-1} - 1)} (1 - \phi_1) \right) \right\}. \end{aligned}$$

By the property $\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X|Y]] + \mathrm{Var}[\mathrm{E}[X|Y]]$, it follows that the unconditional variance, σ_Y^2, is larger than the unconditional mean μ_Y, leading to a process with overdispersion. Moreover, the unconditional variance must obey the relation

$$\begin{aligned} \mathrm{Var}[Y_t] &= \mathrm{E}[Y_t] - \mathrm{E}[Y_{t-1}](\beta_1^2 - \alpha_1^2 + \alpha_1^2\gamma_1^2) + \mathrm{Var}[Y_{t-1}](\beta_1^2 + 2\alpha_1^2\gamma_1^2) - \\ &- \alpha_1^2 (\mathrm{E}[|Y_{t-1} - \lambda_{t-1}|])^2, \end{aligned}$$

where the expectation of the absolute value of the difference between the observable counts and the conditional mean, at the same time point, $t-1$, is given by (3.2).

Regarding the autocovariance function the following relation must be satisfied for $k > 1$,

$$\mathrm{E}[Y_{t-k}Y_t] = \omega\,\mathrm{E}[Y_{t-k}] + \beta_1\,\mathrm{E}[Y_{t-k}Y_{t-1}] + \alpha_1\,\mathrm{E}[Y_{t-k}|Y_{t-1} - \lambda_{t-1}|],$$

where

$$\begin{aligned}\mathrm{E}[Y_{t-k}|Y_{t-1} - \lambda_{t-1}|] &= \mathrm{E}[Y_{t-k}Y_{t-1}]\left\{1 - 2\phi_1 + \frac{H(\lambda_{t-1}-2)}{H(\lambda_{t-1}-1)}\phi_1 - \right.\\ &- \left.\frac{1-H(\lambda_{t-1}-2)}{1-H(\lambda_{t-1}-1)}(1-\phi_1)\right\}.\end{aligned}$$

Thus, for $k > 1$,

$$\mathrm{E}[Y_{t-k}Y_t] = \omega\,\mathrm{E}[Y_{t-k}] + \mathrm{E}[Y_{t-k}Y_{t-1}] \times$$

$$\times \quad \left\{\beta_1 + \alpha_1\left(1 - 2\phi_1 + \frac{H(\lambda_{t-1}-2)}{H(\lambda_{t-1}-1)}\phi_1 - \frac{1-H(\lambda_{t-1}-2)}{1-H(\lambda_{t-1}-1)}(1-\phi_1)\right)\right\}.$$

3.2.2 Stationarity and Ergodicity of the INAPARCH(1,1)

The analysis of weak dependence properties of a process are fundamental for the establishment of standard asymptotics and valid inference, and prediction. Following the work of Doukhan et al. (2012) (see also Davis et al., 2003; Neumann, 2011; Franke, 2010) we will establish the existence and uniqueness of a stationary solution, and ergodicity for the $p = q = 1$ case. The INAPARCH$(1,1)$ process is defined as an integer-valued process (Y_t) such that

$$\begin{aligned}&Y_t|\mathcal{F}_{t-1} \sim Po(\lambda_t)\\ &\lambda_t^\delta = \omega + \alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^\delta + \beta\lambda_{t-1}^\delta,\ t \in \mathbb{Z}\end{aligned} \tag{3.3}$$

with $\alpha \equiv \alpha_1$, $\beta \equiv \beta_1$ and $\gamma \equiv \gamma_1$. The γ parameter should reflect the leverage effect relative to the conditional mean of the process (Y_t).

Proposition 3.2. *Under the conditions in Definition 3.1, the bivariate process* (Y_t, λ_t) *has a stationarity solution.*

Proof. For a general Markov chain and according to Theorem 12.0.1(i) of Meyn and Tweedie (1994), if (X_t) is a weak Feller chain and if for any $\epsilon > 0$ there exists a compact set $C \subset X$ such that

$$P(x, C^c) < \epsilon, \forall x \in X,$$

then (X_t) is bounded in probability and thus there exists at least one stationary distribution for the chain. We will show that the chain is bounded in probability and therefore admits at least one stationary distribution. First note that the chain is weak Feller (cf., Davis et al., 2003). Define $C := [-c, c]$ then,

$$\begin{aligned} & P(\lambda, C^c) = \\ = \; & P(\lambda_t^\delta \in C^c | \lambda_{t-1} = \lambda) \\ = \; & P\left(|\omega + \alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^\delta + \beta\lambda_{t-1}^\delta| > c | \lambda_{t-1} = \lambda\right) \end{aligned}$$

which, by Markov's inequality

$$\begin{aligned} \leqslant \; & \frac{\mathrm{E}\left[|\omega + \alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^\delta + \beta\lambda_{t-1}^\delta| | \lambda_{t-1} = \lambda\right]}{c} \\ \leqslant \; & \frac{\mathrm{E}\left[|\omega| + |\alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^\delta| + |\beta\lambda_{t-1}^\delta| | \lambda_{t-1} = \lambda\right]}{c} \\ \leqslant \; & \frac{\mathrm{E}\left[|\omega|\right] + \mathrm{E}\left[|\alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^\delta| | \lambda_{t-1} = \lambda\right]}{c} + \\ + \; & \frac{\mathrm{E}\left[|\beta\lambda_{t-1}^\delta| | \lambda_{t-1} = \lambda\right]}{c} \\ = \; & \frac{\mathrm{E}\left[|\omega|\right] + \mathrm{E}\left[|\alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^\delta| | \lambda_{t-1} = \lambda\right] + \mathrm{E}\left[|\beta\lambda^\delta|\right]}{c} \end{aligned}$$

and since $\alpha, \beta, \delta, \lambda > 0$

$$P(\lambda, C^c) \leqslant \frac{\omega}{c} + \frac{\alpha}{c}\,\mathrm{E}\left[|(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^{\delta}||\lambda_{t-1} = \lambda\right] + \frac{\beta\lambda^{\delta}}{c}.$$

In view of the fact that $|\gamma| < 1$ and $|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}) \geqslant 0$, the expression above simplifies to

$$P(\lambda, C^c) \leqslant \frac{\omega + \beta\lambda^{\delta}}{c} + \frac{\alpha}{c}\,\mathrm{E}\left[(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^{\delta}|\lambda_{t-1} = \lambda\right].$$

Since by definition

$$\begin{aligned}
&\mathrm{E}\left[(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^{\delta}|\lambda_{t-1} = \lambda\right] = \\
&= \sum_{y_{t-1}=0}^{+\infty} (|y_{t-1} - \lambda| - \gamma(y_{t-1} - \lambda))^{\delta} P(Y_{t-1} = y_{t-1}|\lambda_{t-1} = \lambda) \\
&= \sum_{y_{t-1}=0}^{+\infty} (|y_{t-1} - \lambda| - \gamma(y_{t-1} - \lambda))^{\delta} \frac{e^{-\lambda}\lambda^{y_{t-1}}}{(y_{t-1})!},
\end{aligned}$$

then

$$P(\lambda, C^c) \leqslant \frac{\omega + \beta\lambda^{\delta}}{c} + \frac{\alpha}{c} e^{-\lambda} \sum_{y_{t-1}=0}^{+\infty} \frac{\lambda^{y_{t-1}}}{(y_{t-1})!}(|y_{t-1} - \lambda| - \gamma(y_{t-1} - \lambda))^{\delta}.$$

By d'Alembert's criterion, the series

$$\sum_{y_{t-1}=0}^{+\infty} \frac{\lambda^{y_{t-1}}}{(y_{t-1})!}(|y_{t-1} - \lambda| - \gamma(y_{t-1} - \lambda))^{\delta},$$

is absolutely convergent. Being convergent, the series has a finite sum and so it can be written that

$$P(\lambda, C^c) \leqslant \frac{\omega + \beta\lambda^{\delta}}{c} + \frac{\alpha}{c} e^{-\lambda} \sum_{y_{t-1}=0}^{+\infty} \frac{\lambda^{y_{t-1}}}{(y_{t-1})!}(|y_{t-1} - \lambda| - \gamma(y_{t-1} - \lambda))^{\delta} < \epsilon.$$

Thus, for any $\epsilon > 0$ just choose c large enough so that

$$\frac{1}{c}\left(\omega + \beta\lambda^{\delta} + \alpha e^{-\lambda} \sum_{y_{t-1}=0}^{+\infty} \frac{\lambda^{y_{t-1}}}{(y_{t-1})!}(|y_{t-1} - \lambda| - \gamma(y_{t-1} - \lambda))^{\delta}\right) < \epsilon,$$

leading to conclude that the series has at least one stationary solution.

In proving uniqueness we proceed as follows: note that the INAPARCH$(1,1)$ model belongs to the class of observation-driven Poisson count processes considered in Neumann (2011)

$$Y_t|\mathcal{F}_{t-1}^{Y,\lambda} \sim Po(\lambda_t); \;\; \lambda_t = f(\lambda_{t-1}, Y_{t-1}),\, t \in \mathbb{N}$$

with

$$f(\lambda_{t-1}, Y_{t-1}) = (\omega + \alpha(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1}))^{\delta} + \beta\lambda_{t-1}^{\delta})^{\frac{1}{\delta}}.$$

Thus, the result follows if the function f above satisfies the following contractive condition

$$|f(\lambda, y) - f(\lambda', y')| \leq k_1|\lambda - \lambda'| + k_2|y - y'| \quad \forall\, \lambda, \lambda' \geq 0,\, \forall\, y, y' \in \mathbb{N}_0, \tag{3.4}$$

where k_1 and k_2 are nonnegative constants such that $k := k_1 + k_2 < 1$. For the INAPARCH$(1,1)$ model the contractive condition simplifies to

$$\begin{aligned} |f(\lambda_{t-1}, Y_{t-1}) - f(\lambda'_{t-1}, Y'_{t-1})| &\leq \|\frac{\partial f}{\partial \lambda_{t-1}}\|_\infty |\lambda_{t-1} - \lambda'_{t-1}| + \\ &+ \|\frac{\partial f}{\partial Y_{t-1}}\|_\infty |Y_{t-1} - Y'_{t-1}|, \end{aligned}$$

where for the Euclidean space $\mathbb{R}^d$ and $h : \mathbb{R}^d \to \mathbb{R}$, $\|h\|_\infty$ is defined by $\|h\|_\infty = \sup_{x\in\mathbb{R}^d} |h(x)|$. The partial derivatives are given by

$$\begin{aligned} \frac{\partial f}{\partial Y_{t-1}} &= \alpha\lambda_t^{1-\delta} \left(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1})\right)^{\delta-1} (I_{Y_{t-1}} - \gamma) \\ \frac{\partial f}{\partial \lambda_{t-1}} &= \beta\lambda_{t-1}^{\delta-1}\lambda_t^{1-\delta} - \alpha\lambda_t^{1-\delta} \left(|Y_{t-1} - \lambda_{t-1}| - \gamma(Y_{t-1} - \lambda_{t-1})\right)^{\delta-1} (I_{Y_{t-1}} - \gamma) \end{aligned}$$

where

$$I_{Y_{t-1}} = \begin{cases} 1 & Y_{t-1} > \lambda_{t-1} \\ -1 & Y_{t-1} < \lambda_{t-1} \end{cases},$$

that is

- Case I: $Y_{t-1} > \lambda_{t-1}$,

$$\begin{aligned}
\frac{\partial f}{\partial Y_{t-1}} &= \alpha(1-\gamma)^{\delta}\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1} > 0\\
\frac{\partial f}{\partial \lambda_{t-1}} &= \beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1} - \alpha(1-\gamma)^{\delta}\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\\
&\leqslant \beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}.
\end{aligned}$$

- Case II: $Y_{t-1} < \lambda_{t-1}$,

$$\begin{aligned}
\frac{\partial f}{\partial Y_{t-1}} &= \alpha(-1-\gamma)^{\delta}\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\\
&= \alpha(-1)^{\delta}(1+\gamma)^{\delta}\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\\
\frac{\partial f}{\partial \lambda_{t-1}} &= \beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1} - \alpha(-1-\gamma)^{\delta}\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\\
&= \beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1} - \alpha(-1)^{\delta}(1+\gamma)^{\delta}\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}.
\end{aligned}$$

Having in mind that $-1 < \gamma < 1$, then both $1-\gamma$ and $1+\gamma$ take values in $]0, 2[$. Moreover, for $\delta \geqslant 2$

$$\left|\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| < \delta.$$

Thus, for $Y_{t-1} > \lambda_{t-1}$ and taking $\delta \geqslant 2$ it follows that

$$\left|\frac{\partial f}{\partial Y_{t-1}}\right| = \alpha(1-\gamma)^{\delta}\left|\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| < \alpha(1-\gamma)^{\delta}\delta < \alpha 2^{\delta}\delta.$$

On the other hand, for $Y_{t-1} < \lambda_{t-1}$ and $\delta \geqslant 2$

$$\begin{aligned}
\left|\frac{\partial f}{\partial Y_{t-1}}\right| &= \alpha|(-1)^{\delta}|(1+\gamma)^{\delta}\left|\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right|\\
&= \alpha(1+\gamma)^{\delta}\left|\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| < \alpha 2^{\delta}\delta,
\end{aligned}$$

leading to obtain

$$\|\frac{\partial f}{\partial Y_{t-1}}\|_{\infty} = \alpha 2^{\delta}\delta, \text{ for } \delta \geqslant 2.$$

Note that by stationarity, when $Y_{t-1} > \lambda_{t-1}$, it follows that

$$\left|\frac{\partial f}{\partial \lambda_{t-1}}\right| < \beta 2^{\delta-1}.$$

For the case $Y_{t-1} < \lambda_{t-1}$

$$\begin{aligned}
\left|\frac{\partial f}{\partial \lambda_{t-1}}\right| &= \left|\beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1} - \alpha(-1)^\delta(1+\gamma)^\delta\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| \\
&= \left|\beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| + \left|\alpha(1+\gamma)^\delta\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| \\
&= \beta\left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\delta-1} + \alpha(1+\gamma)^\delta\left|\left(\frac{Y_{t-1}-\lambda_{t-1}}{\lambda_t}\right)^{\delta-1}\right| \\
&< \beta 2^{\delta-1} + \alpha 2^\delta \delta,
\end{aligned}$$

for $\delta \geqslant 2$, since δ and $\delta - 1$ have opposite parities.

Finally, a majorant for the partial derivative in order to λ_{t-1}, can be taken by

$$\|\frac{\partial f}{\partial \lambda_{t-1}}\|_\infty = \beta 2^{\delta-1} + \alpha 2^\delta \delta.$$

Hence, if

$$\alpha 2^{\delta+1}\delta + \beta 2^{\delta-1} < 1, \tag{3.5}$$

for $\delta \geqslant 2$, then the contractive condition holds. This concludes the proof.

■

Neumann (2011) proved that the contractive condition in (3.4) is, indeed, sufficient to ensure uniqueness of the stationary distribution and ergodicity of (Y_t, λ_t). The results are quoted below.

Proposition 3.3. *Suppose that the bivariate process (Y_t, λ_t) satisfies (3.3) and (3.5) for $\delta \geq 2$. Then the stationary distribution is unique and* $\mathrm{E}[\lambda_1] < \infty$.

Proposition 3.4. *Suppose that the bivariate process (Y_t, λ_t) is in its stationarity regime and satisfies (3.3) and (3.5) for $\delta \geq 2$. Then the bivariate process (Y_t, λ_t) is ergodic and* $\mathrm{E}[\lambda_1^2] < \infty$.

Furthermore, following Theorem 2.1. in Doukhan et al. (2012), it can be shown that if the process (Y_t, λ_t) satisfies (3.3) and (3.5) for $\delta \geq 2$, then there exists a solution of (3.3) which is a τ-weakly dependent strictly stationary process with finite moments up to any positive order and is ergodic.

3.3 Estimation

In this section, we estimate the parameters of the INAPARCH(p, q) model. The conditional maximum likelihood method can be applied in a very straightforward manner. Note that by the fact that the conditional distribution is Poisson, the conditional likelihood function, given the starting value λ_0 and the observations $y_1, \ldots, y_n$, takes the form

$$L(\boldsymbol{\theta}) := \prod_{t=1}^{n} \frac{e^{-\lambda_t(\boldsymbol{\theta})} \lambda_t^{y_t}(\boldsymbol{\theta})}{y_t!} \tag{3.6}$$

with $\boldsymbol{\theta} := (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma_1, \ldots, \gamma_p, \delta) \equiv (\theta_1, \theta_2, \ldots, \theta_{2p+q+2})$ denoting the unknown parameter vector. The log-likelihood function is given by

$$\ln(L(\boldsymbol{\theta})) = \sum_{t=1}^{n} [y_t \ln(\lambda_t) - \lambda_t - \ln(y_t!)] = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}). \tag{3.7}$$

The score function is the vector defined by

$$S_n(\boldsymbol{\theta}) := \frac{\partial \ln(L(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{3.8}$$

The auxiliary calculations presented below are needed for the calculation of the first order derivatives of the general INAPARCH(p, q) model.

$$\frac{\partial \ell_t}{\partial \theta_i} = \frac{\partial \lambda_t}{\partial \theta_i} \left(\frac{y_t}{\lambda_t} - 1 \right), \ i = 1, \ldots, 2 + 2p + q,$$

where

$$\frac{\partial \lambda_t}{\partial \theta_i} = \frac{\lambda_t}{\delta \lambda_t^{\delta}} \frac{\partial (\lambda_t^{\delta})}{\partial \theta_i}, \ i = 1, \ldots, 2 + 2p + q.$$

Thus, the first derivatives are given by the following expressions

$$\frac{\partial\lambda_t}{\partial\omega} = \frac{\lambda_t}{\delta\lambda_t^\delta}\left(\delta\sum_{i=1}^{p}\alpha_i g_{t-i}^{\delta-1}(I_{t-i}+\gamma_i)\frac{\partial\lambda_{t-i}}{\partial\omega} + \sum_{j=1}^{q}\beta_j\frac{\partial\lambda_{t-j}^\delta}{\partial\omega} + 1\right),$$

$$\frac{\partial\lambda_t}{\partial\alpha_i} = \frac{\lambda_t}{\delta\lambda_t^\delta}\left(\delta\sum_{k=1}^{p}\alpha_k g_{t-k}^{\delta-1}(I_{t-k}+\gamma_k)\frac{\partial\lambda_{t-k}}{\partial\alpha_i} + \sum_{j=1}^{q}\beta_j\frac{\partial\lambda_{t-j}^\delta}{\partial\alpha_i} + g_{t-i}^\delta\right),$$

$$\frac{\partial\lambda_t}{\partial\gamma_i} = \frac{\lambda_t}{\delta\lambda_t^\delta}\left(\delta\sum_{k=1}^{p}\alpha_k g_{t-k}^{\delta-1}(I_{t-k}+\gamma_k)\frac{\partial\lambda_{t-k}}{\partial\gamma_i} + \right.$$
$$\left. + \sum_{j=1}^{q}\beta_j\frac{\partial\lambda_{t-j}^\delta}{\partial\gamma_i} - \delta\alpha_i g_{t-i}^{\delta-1}(y_{t-i}-\lambda_{t-i})\right),$$

for $i = 1, \ldots, p$,

$$\frac{\partial\lambda_t}{\partial\beta_j} = \frac{\lambda_t}{\delta\lambda_t^\delta}\left(\delta\sum_{i=1}^{p}\alpha_i g_{t-i}^{\delta-1}(I_{t-i}+\gamma_i)\frac{\partial\lambda_{t-i}}{\partial\beta_j} + \sum_{k=1}^{q}\beta_k\frac{\partial\lambda_{t-k}^\delta}{\partial\beta_j} + \lambda_{t-j}^\delta\right),$$

for $j = 1, \ldots, q$,

$$\frac{\partial\lambda_t}{\partial\delta} = \frac{\lambda_t}{\delta\lambda_t^\delta}\left\{\sum_{i=1}^{p}\alpha_i g_{t-i}^\delta\left(\frac{\delta}{g_{t-i}}(I_{t-i}+\gamma_i)\frac{\partial\lambda_{t-i}}{\partial\delta} + \ln(g_{t-i})\right) + \right.$$
$$\left. + \sum_{j=1}^{q}\beta_j\frac{\partial\lambda_{t-j}^\delta}{\partial\delta} - \frac{\lambda_t^\delta}{\delta}\ln(\lambda_t^\delta)\right\},$$

where $g_{t-i} = |y_{t-i}-\lambda_{t-i}| - \gamma_i(y_{t-i}-\lambda_{t-i})$ and

$$I_t = \begin{cases} -1 & y_t > \lambda_t \\ 1 & y_t < \lambda_t \end{cases}.$$

Thus, for the INAPARCH$(1,1)$ model the score function can then be explicitly written as

$$S_n(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{t=1}^n \left(\frac{y_t}{\lambda_t} - 1\right) \frac{\partial \lambda_t}{\partial \omega} \\ \sum_{t=1}^n \left(\frac{y_t}{\lambda_t} - 1\right) \frac{\partial \lambda_t}{\partial \alpha} \\ \sum_{t=1}^n \left(\frac{y_t}{\lambda_t} - 1\right) \frac{\partial \lambda_t}{\partial \gamma} \\ \sum_{t=1}^n \left(\frac{y_t}{\lambda_t} - 1\right) \frac{\partial \lambda_t}{\partial \beta} \\ \sum_{t=1}^n \left(\frac{y_t}{\lambda_t} - 1\right) \frac{\partial \lambda_t}{\partial \delta} \end{bmatrix}$$

with

$$\frac{\partial \lambda_t}{\partial \omega} = \frac{\lambda_t}{\delta \lambda_t^\delta} \left(\delta(\alpha(I_{t-1} + \gamma) g_{t-1}^{\delta-1} + \beta \lambda_{t-1}^{\delta-1}) \frac{\partial \lambda_{t-1}}{\partial \omega} + 1 \right),$$

$$\frac{\partial \lambda_t}{\partial \alpha} = \frac{\lambda_t}{\delta \lambda_t^\delta} \left(\delta(\alpha(I_{t-1} + \gamma) g_{t-1}^{\delta-1} + \beta \lambda_{t-1}^{\delta-1}) \frac{\partial \lambda_{t-1}}{\partial \alpha} + g_{t-1}^\delta \right),$$

$$\frac{\partial \lambda_t}{\partial \gamma} = \frac{\lambda_t}{\delta \lambda_t^\delta} \left(\delta(\alpha(I_{t-1} + \gamma) g_{t-1}^{\delta-1} + \beta \lambda_{t-1}^{\delta-1}) \frac{\partial \lambda_{t-1}}{\partial \gamma} - \alpha \delta g_{t-1}^{\delta-1} (y_{t-1} - \lambda_{t-1}) \right),$$

$$\frac{\partial \lambda_t}{\partial \beta} = \frac{\lambda_t}{\delta \lambda_t^\delta} \left(\delta(\alpha(I_{t-1} + \gamma) g_{t-1}^{\delta-1} + \beta \lambda_{t-1}^{\delta-1}) \frac{\partial \lambda_{t-1}}{\partial \beta} + \lambda_{t-1}^\delta \right),$$

$$\begin{aligned} \frac{\partial \lambda_t}{\partial \delta} = \frac{\lambda_t}{\delta \lambda_t^\delta} \Big(& \delta(\alpha(I_{t-1} + \gamma) g_{t-1}^{\delta-1} + \beta \lambda_{t-1}^{\delta-1}) \frac{\partial \lambda_{t-1}}{\partial \delta} + \alpha g_{t-1}^\delta \ln(g_{t-1}) + \\ & + \beta \lambda_{t-1}^\delta \ln(\lambda_{t-1}) \Big) - \frac{\lambda_t}{\delta} \ln(\lambda_t). \end{aligned}$$

The solution of the equation $S_n(\boldsymbol{\theta}) = 0$ is the conditional maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, if it exists. To study the asymptotic properties of the maximum likelihood estimator we proceed as follows: first it can be shown that the score function, evaluated at the true value of the parameter, say $\boldsymbol{\theta_0}$, is asymptotically normal. The score function has martingale difference terms defined by

$$\frac{\partial \ell_t}{\partial \theta_i} = \left(\frac{y_t}{\lambda_t} - 1 \right) \frac{\partial \lambda_t}{\partial \theta_i}.$$

The partial derivatives defined above can be rewritten after repeated substitution by

$$\frac{\partial\lambda_t}{\partial\omega} = \frac{\lambda_t}{\delta\lambda_t^{\delta}}\left(1 + \sum_{i=1}^{t-1}\prod_{j=1}^{i}\lambda_{t-j}^{1-\delta}(\alpha(I_{t-j}+\gamma)g_{t-j}^{\delta-1} + \beta\lambda_{t-j}^{\delta-1})\right),$$

$$\frac{\partial\lambda_t}{\partial\alpha} = \frac{\lambda_t}{\delta\lambda_t^{\delta}}\left(g_{t-1}^{\delta} + \sum_{i=1}^{t-2}g_{t-(i+1)}^{\delta}\prod_{j=1}^{i}\lambda_{t-j}^{1-\delta}(\alpha(I_{t-j}+\gamma)g_{t-j}^{\delta-1} + \beta\lambda_{t-j}^{\delta-1})\right),$$

$$\begin{aligned}\frac{\partial\lambda_t}{\partial\gamma} &= -\frac{\lambda_t}{\lambda_t^{\delta}}\left(\alpha g_{t-1}^{\delta-1}(y_{t-1}-\lambda_{t-1}) + \sum_{i=1}^{t-2}\alpha g_{t-(i+1)}^{\delta-1}(y_{t-(i+1)}-\lambda_{t-(i+1)})\times\right.\\ &\left.\times\prod_{j=1}^{i}\lambda_{t-j}^{1-\delta}(\alpha(I_{t-j}+\gamma)g_{t-j}^{\delta-1} + \beta\lambda_{t-j}^{\delta-1})\right),\end{aligned}$$

$$\frac{\partial\lambda_t}{\partial\beta} = \frac{\lambda_t}{\delta\lambda_t^{\delta}}\left(\lambda_{t-1}^{\delta} + \sum_{i=1}^{t-2}\lambda_{t-(i+1)}^{\delta}\prod_{j=1}^{i}\lambda_{t-j}^{1-\delta}(\alpha(I_{t-j}+\gamma)g_{t-j}^{\delta-1} + \beta\lambda_{t-j}^{\delta-1})\right),$$

$$\begin{aligned}\frac{\partial\lambda_t}{\partial\delta} &= \frac{\lambda_t}{\delta\lambda_t^{\delta}}(\alpha g_{t-1}^{\delta}\ln(g_{t-1}) + \beta\lambda_{t-1}^{\delta}\ln(\lambda_{t-1}) - \frac{\lambda_t}{\delta}\ln(\lambda_t) -\\ &-\sum_{i=1}^{t-1}\frac{\lambda_{t-i}}{\delta}\ln(\lambda_{t-1})\prod_{j=1}^{i}\lambda_{t-j+1}^{1-\delta}(\alpha(I_{t-j}+\gamma)g_{t-j}^{\delta-1} + \beta\lambda_{t-j}^{\delta-1}) +\\ &+\frac{\lambda_t}{\delta\lambda_t^{\delta}}\sum_{i=1}^{t-2}\Big(\alpha g_{t-(i+1)}^{\delta}\ln(g_{t-(i+1)}) + \beta\lambda_{t-(i+1)}^{\delta}\ln(\lambda_{t-(i+1)})\Big)\times\\ &\times\prod_{j=1}^{i}\lambda_{t-j}^{1-\delta}(\alpha(I_{t-j}+\gamma)g_{t-j}^{\delta-1} + \beta\lambda_{t-j}^{\delta-1}).\end{aligned}$$

It follows that at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$

$$\mathrm{E}\left[\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}|\mathcal{F}_{t-1}\right] = 0,$$

since $\mathrm{E}\left[\frac{y_t}{\lambda_t} - 1|\mathcal{F}_{t-1}\right] = 0$ and

$$\mathrm{E}\left[\left(\frac{y_t}{\lambda_t} - 1\right)^2 |\mathcal{F}_{t-1}\right] = \mathrm{Var}\left[\frac{y_t}{\lambda_t} - 1|\mathcal{F}_{t-1}\right] = \frac{1}{\lambda_t},$$

where $\mathcal{F}_{t-1}$ represents the σ-algebra generated by $(Y_s, s \leq t-1)$. It can also easily be shown that, for $\delta \geqslant 2$

$$\mathrm{E}\left[\lambda_t^{2-2\delta}|\mathcal{F}_{t-1}\right] < +\infty;$$

$$\mathrm{E}\left[\lambda_t^{1-\delta}|\mathcal{F}_{t-1}\right] < +\infty;$$

$$\mathrm{E}\left[\lambda_t^{2-\delta}\ln(\lambda_t)|\mathcal{F}_{t-1}\right] < \mathrm{E}[\ln(\lambda_t)|\mathcal{F}_{t-1}] < \mathrm{E}[\lambda_t|\mathcal{F}_{t-1}] < +\infty;$$

$$\mathrm{E}\left[\lambda_t^2\ln^2(\lambda_t)|\mathcal{F}_{t-1}\right] < +\infty;$$

$$\mathrm{E}\left[\lambda_t\ln(\lambda_t)|\mathcal{F}_{t-1}\right] < +\infty.$$

Thus, it can be concluded that $\mathrm{Var}\left[\frac{\partial \ell_t}{\partial \boldsymbol{\theta}}|\mathcal{F}_{t-1}\right] < +\infty$ and that $\partial \ell_t / \partial \boldsymbol{\theta}$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$. The application of a central limit theorem for martingales guarantees the desired asymptotic normality.

It is worth to mention here that in Section 3.2 it was concluded that the process has finite moments up to any positive order and is τ-weak dependent, which implies ergodicity. This is sufficient to state that the Hessian matrix converges in probability to a finite limit. Finally, all third derivatives are bounded by a sequence that converges in probability[1]. Given these three conditions, it is then concluded that the conditional maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, is consistent and asymptotically normal,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{d} \mathcal{N}(0, G^{-1}(\boldsymbol{\theta}))$$

[1] In Appendix A, the third derivative of $\ell_t(\boldsymbol{\theta})$ in order to ω is provided for illustration. Other third order derivatives are obtained straightforwardly.

with variance-covariance matrix, $G(\boldsymbol{\theta})$, given by

$$G(\boldsymbol{\theta}) = \mathrm{E}\left[\frac{1}{\lambda_t}\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)'\right].$$

A consistent estimator of $G(\boldsymbol{\theta})$ is given by $G_n(\hat{\boldsymbol{\theta}})$, where

$$G_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \mathrm{Var}\left[\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n} \frac{1}{\lambda_t(\boldsymbol{\theta})}\left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'.$$

The diagonal entries of the Hessian matrix are related to the expressions presented next and all other entries are calculated in a very straightforward manner

$$\begin{aligned}\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \omega^2} = & \left(\frac{\delta-1}{\delta\lambda_t^{\delta}} - \frac{y_t}{\lambda_t^{\delta+1}}\right)\left(1+\delta(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1}+\beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \omega}\right)\frac{\partial \lambda_t}{\partial \omega}+ \\ & + \left(\frac{y_t}{\lambda_t^{\delta}} - \frac{1}{\lambda_t^{\delta-1}}\right)\left\{(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1}+\beta\lambda_{t-1}^{\delta-1})\frac{\partial^2 \lambda_{t-1}}{\partial \omega^2}+\right. \\ & \left. + (\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2}+\beta(\delta-1)\lambda_{t-1}^{\delta-2})\left(\frac{\partial \lambda_{t-1}}{\partial \omega}\right)^2\right\},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \alpha^2} = & \left(\frac{\delta-1}{\delta\lambda_t^{\delta}} - \frac{y_t}{\lambda_t^{\delta+1}}\right)\left(g_{t-1}^{\delta}+\delta(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1}+\beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \alpha}\right)\frac{\partial \lambda_t}{\partial \alpha}+ \\ & + \left(\frac{y_t}{\lambda_t^{\delta}} - \frac{1}{\lambda_t^{\delta-1}}\right)\left\{(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1}+\beta\lambda_{t-1}^{\delta-1})\frac{\partial^2 \lambda_{t-1}}{\partial \alpha^2}+\right. \\ & + (\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2}+\beta(\delta-1)\lambda_{t-1}^{\delta-2})\left(\frac{\partial \lambda_{t-1}}{\partial \alpha}\right)^2+ \\ & \left. + +2(I_{t-1}+\gamma)g_{t-1}^{\delta-1}\frac{\partial \lambda_{t-1}}{\partial \alpha}\right\},\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \gamma^2} = &\left(\frac{\delta-1}{\delta\lambda_t^{\delta}} - \frac{y_t}{\lambda_t^{\delta+1}}\right)\Big(-\alpha\delta g_{t-1}^{\delta-1}(y_{t-1}-\lambda_{t-1}) + \\
&+ \delta(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \gamma}\Big)\frac{\partial \lambda_t}{\partial \gamma} + \\
&+ \left(\frac{y_t}{\lambda_t^{\delta}} - \frac{1}{\lambda_t^{\delta-1}}\right)\Big\{(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial^2 \lambda_{t-1}}{\partial \gamma^2} + \\
&+ (\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2} + \beta(\delta-1)\lambda_{t-1}^{\delta-2})\left(\frac{\partial \lambda_{t-1}}{\partial \gamma}\right)^2 + \\
&+ 2\gamma\Big(g_{t-1}^{\delta-1} - (\delta-1)(I_{t-1}+\gamma)g_{t-1}^{\delta-2}(y_{t-1}-\lambda_{t-1})\Big)\frac{\partial \lambda_{t-1}}{\partial \gamma} + \\
&+ \alpha(\delta-1)g_{t-1}^{\delta-2}(y_{t-1}-\lambda_{t-1})^2\Big\},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \beta^2} = &\left(\frac{\delta-1}{\delta\lambda_t^{\delta}} - \frac{y_t}{\lambda_t^{\delta+1}}\right)\left(\lambda_{t-1}^{\delta} + \delta(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \beta}\right)\frac{\partial \lambda_t}{\partial \beta} + \\
&+ \left(\frac{y_t}{\lambda_t^{\delta}} - \frac{1}{\lambda_t^{\delta-1}}\right)\Big\{(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial^2 \lambda_{t-1}}{\partial \beta^2} + \\
&+ (\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2} + \beta(\delta-1)\lambda_{t-1}^{\delta-2})\left(\frac{\partial \lambda_{t-1}}{\partial \beta}\right)^2 + \\
&+ 2\lambda_{t-1}^{\delta-1}\frac{\partial \lambda_{t-1}}{\partial \beta}\Big\},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \delta^2} = &\left(\left(\frac{\delta-1}{\delta\lambda_t^{\delta}} - \frac{y_t}{\lambda_t^{\delta+1}}\right)\frac{\partial \lambda_t}{\partial \delta} - \frac{y_t-\lambda_t}{\delta\lambda_t^{\delta}}\left(\frac{1}{\delta} + \ln(\lambda_t)\right)\right) \times \\
&\times \Big(\delta(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \delta} + \\
&+ \alpha g_{t-1}^{\delta}\ln(g_{t-1}) + \beta\lambda_{t-1}^{\delta}\ln(\lambda_{t-1})\Big) + \\
&+ \left(\frac{y_t}{\lambda_t^{\delta}} - \frac{1}{\lambda_t^{\delta-1}}\right)\Big\{(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial^2 \lambda_{t-1}}{\partial \delta^2} + \\
&+ (\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2} + \beta(\delta-1)\lambda_{t-1}^{\delta-2})\left(\frac{\partial \lambda_{t-1}}{\partial \delta}\right)^2 + \\
&+ \frac{2}{\delta}\Big(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1}(1+\delta\ln(g_{t-1})) + \beta\lambda_{t-1}^{\delta-1}(1+\delta\ln(\lambda_{t-1}))\Big)\frac{\partial \lambda_{t-1}}{\partial \delta} + \\
&+ \frac{\alpha}{\delta}g_{t-1}^{\delta}(\ln(g_{t-1}))^2 + \frac{\beta}{\delta}\lambda_{t-1}^{\delta}(\ln(\lambda_{t-1}))^2\Big\}.
\end{aligned}
$$

3.4 Simulation

In this section, a simulation study is carried out to illustrate the theoretical findings given in the section above for the INAPARCH$(1,1)$ model. The simulation study contemplates five different combinations for $\boldsymbol{\theta}$, which are displayed in Table 3.1 below. For each set of parameters time series of length 500 with 300 independent replicates from the INAPARCH$(1,1)$ model were simulated. A sample path and its corresponding sample ACF are presented in Figure 3.1, for the combination of parameters C2. The remaining cases are presented in the Appendix A.

Note that for C1-C4 cases, condition (3.5) holds, whereas for case C5 this condition fails. The simulation study was computed using Matlab and the programs developed are provided in Appendix B. The results are summarized in Table 3.1. The bias of the conditional ML estimates is presented in Figure 3.2 for the combination of parameters C2 and in Appendix A for the remaining cases. Numbers one to five below the boxplots refer to the estimated parameters, in the order appearing in Table 3.1.

Considering the conditional ML estimates in Table 3.1 and the boxplots of the bias in Figure 3.2, a few conclusions can be drawn. Firstly, the α parameter seems to be conveniently estimated, i.e., the point estimates follow the theoretical values in a coherent way, even for very small values such as for the combinations of parameters C1, C2 and C3. The observed bias is also quite small. On the other hand, the β parameter is always overestimated, there is a tendency to underestimate the ω parameter and there is a very high degree of variability, in particular for the δ parameter.

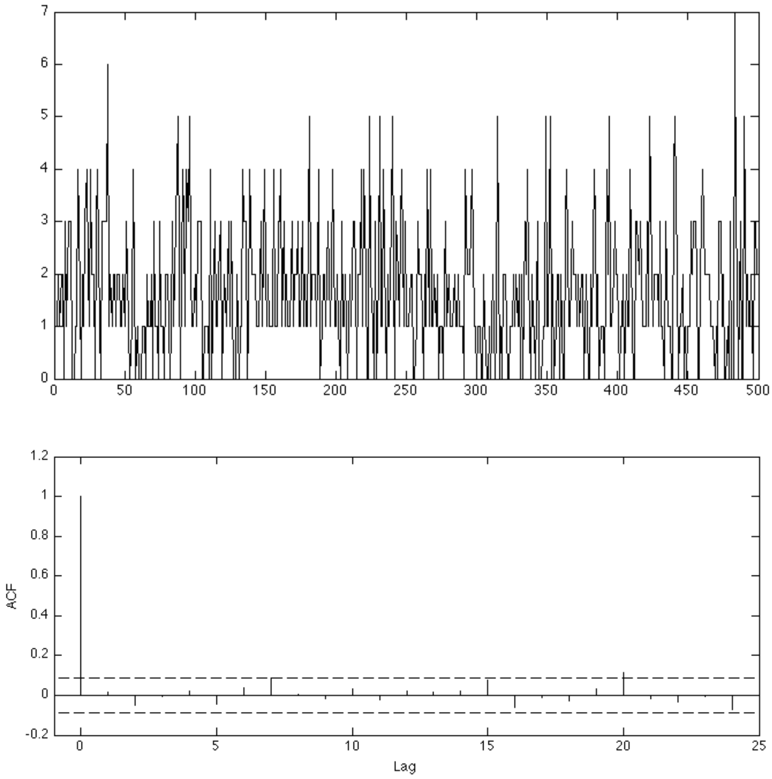


Figure 3.1: Sample path for the INAPARCH$(1,1)$ process. Combination of parameters C2 (top) and its corresponding autocorrelation function (bottom).

3.4.1 Log-likelihood analysis

For C2 and C4 cases, 300 samples were simulated considering values of δ varying from 2.0 to 3.0 (i.e., six different situations for each case). After preliminary data analysis with the construction of boxplots and histograms (presented in Appendix A and that can confirm the presence of overdispersion) the log-likelihood was studied in the following manner: for each set of 300 samples the log-likelihood was calculated, varying the δ parameter in the range 2.0 to 3.0. It was expected that the log-likelihood was maximum for the δ value used to simulate that particular set of 300 samples.

Table 3.1: Parameter estimates and standard errors (se) in parentheses.

Parameter Values							Point estimates and (se)				
Case	ω	α	γ	β	δ	$2^\delta(2\alpha\delta+\frac{\beta}{2})$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\delta}$
C1	2.30	0.01	0.68	0.10	2.00	0.36	1.8510	0.0641	0.6356	0.1850	1.9245
							(0.4825)	(0.0685)	(0.3180)	(0.2246)	(0.7156)
C2	2.30	0.03	0.68	0.06	2.00	0.60	1.9067	0.0755	0.6174	0.1452	1.9170
							(0.5142)	(0.0698)	(0.3351)	(0.1981)	(0.6860)
C3	2.30	0.01	0.68	0.10	3.00	0.88	1.9674	0.0571	0.5922	0.1572	2.9588
							(0.4229)	(0.0684)	(0.2914)	(0.1813)	(0.7183)
C4	2.30	0.05	0.68	0.08	2.00	0.96	1.8931	0.0880	0.7005	0.1753	1.9535
							(0.5294)	(0.0722)	(0.3070)	(0.2102)	(0.7154)
C5	2.30	0.30	0.68	0.10	2.00	5.00	2.2724	0.3082	0.7489	0.1294	2.0401
							(0.7519)	(0.1290)	(0.2229)	(0.1318)	(0.6510)

Results are presented in Figure 3.3 and Table 3.2 for Case 2. Case 2 was chosen for representation herein just because for this case the first three values for the δ parameter lie inside the region that obeys condition (3.5) and the last 3 lie outside this region. Nevertheless, same behaviour was observed for both Case 2 and Case 4 (represented in Appendix A) and the δ value for which the calculated log-likelihood was maximum was exactly what was expected for both cases and all 6 different situations. In Table 3.2, it can be observed that the mean log-likelihood is maximum for the δ value corresponding to the δ value used for the simulation of the respective set of samples. In Figure 3.3, the numbers 1 to 6 in the x-axis correspond to the δ values of $\{2.00, 2.20, 2.40, 2.60, 2.80, 3.00\}$, respectively, and it can be seen that the results are in accordance with Table 3.2. For each situation, the median log-likelihood is maximum for the expected δ value and variability is comparable not only between different δ values for the same set of 300 samples, but also between different sets of samples.

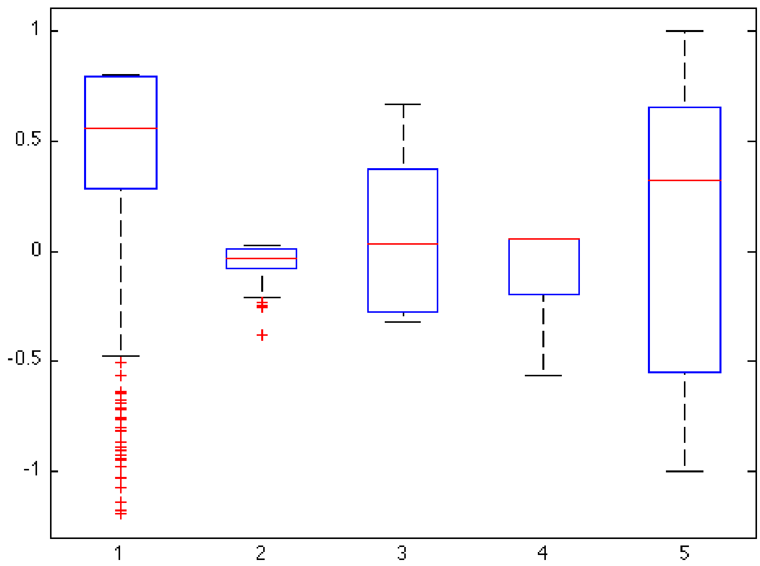


Figure 3.2: Bias of the conditional ML estimates, for the combination of parameters C2.

Table 3.2: Maximum likelihood estimation results for Case 2.

Samples simulated with	Log-likelihood for varying δ					
$\boldsymbol{\theta}$=$(2.30, 0.03, 0.68, 0.06, \delta)$	2.0	2.2	2.4	2.6	2.8	3.0
$\delta = 2.00$	-785.4787	-786.1563	-787.6991	-789.6634	-791.8038	-793.9828
$\delta = 2.20$	-775.2089	-774.5939	-775.0658	-776.1291	-777.5016	-779.0191
$\delta = 2.40$	-766.7914	-765.1027	-764.6847	-764.9993	-765.7337	-766.7013
$\delta = 2.60$	-760.1167	-757.5743	-756.4490	-756.1685	-756.3958	-756.9265
$\delta = 2.80$	-755.0275	-751.7783	-750.0676	-749.2947	-749.1024	-749.2715
$\delta = 3.00$	-751.1783	-747.3026	-745.0736	-743.8653	-743.3025	-743.1530

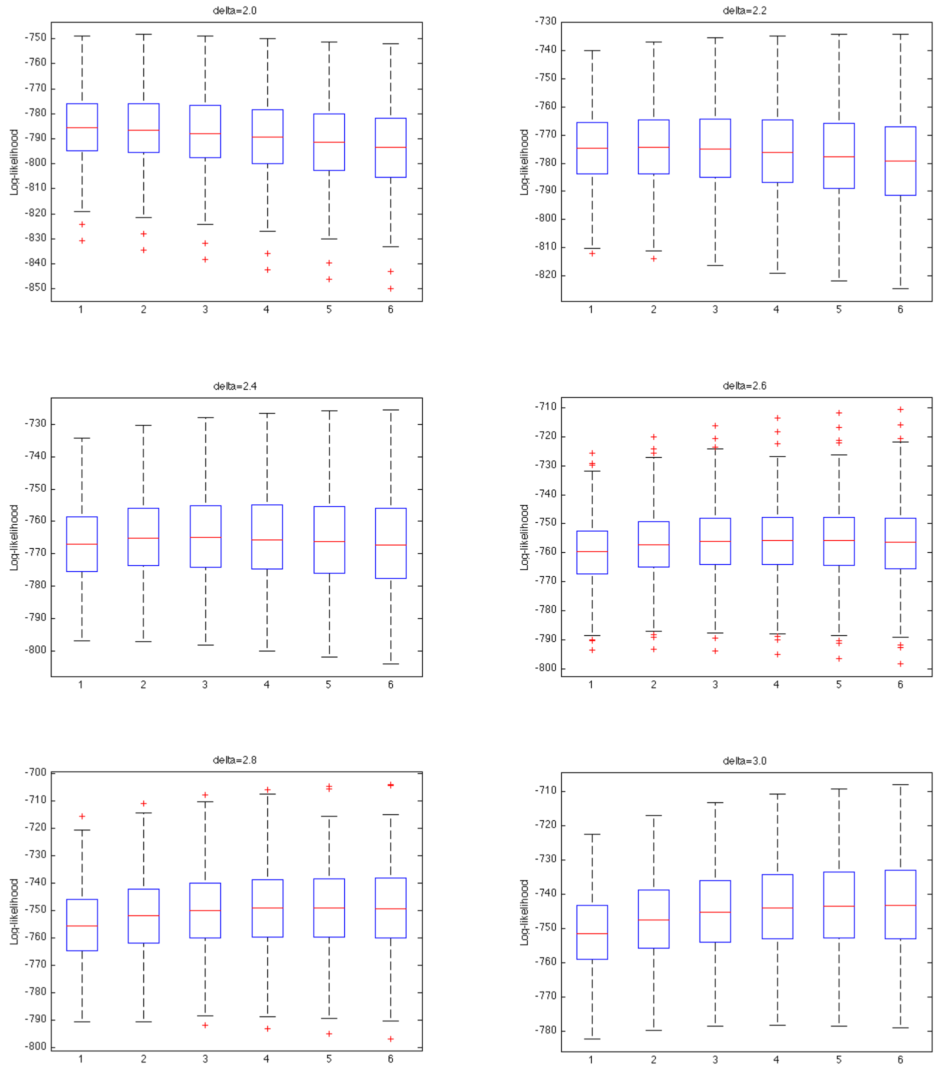


Figure 3.3: Log-likelihood for varying δ, Case 2.

3.5 Optimal alarm systems: application to the INAPARCH (1,1) model

Let $(Y_t)_{t\in\mathbb{N}}$ be a count time process with parameter space $\Theta \subset \mathbb{R}^k$ for some $k \in \mathbb{N}$. The time sequence $\{1, 2, \ldots, t-1, t, t+1, \ldots\}$ will be divided in three sections: $\{1, 2, \ldots, t-q\}$, $\{t-q+1, \ldots, t\}$, $\{t+1, \ldots\}$, namely, past, present and future, such that for some $q > 0$ the following subsets will be defined

Data or informative experience: $D_t = \{Y_1, Y_2, \dots, Y_{t-q}\}$

Present experiment: $\mathbf{Y_2} = \{Y_{t-q+1}, \dots, Y_t\}$

Future experiment: $\mathbf{Y_3} = \{Y_{t+1}, \dots\}$

Any event of interest, $C_{t,j}$, in the σ-algebra generated by $\mathbf{Y_3}$ is defined as a catastrophe. In this work, the catastrophe shall be considered as the upcrossing event of some fixed level u,

$$C_{t,j} = \{Y_{t+j-1} \leqslant u < Y_{t+j}\} \text{ for some } j \in \mathbb{N}.$$

The alarm region of optimal size $\alpha_{t,j}$ is given by

$$\begin{aligned} A_{t,j} &= \left\{\mathbf{y_2} \in \mathbb{N}^q : \frac{P(C_{t,j}|\mathbf{y_2}, D_t)}{P(C_{t,j}|D_t)} \geq k_{t,j}\right\} \\ &= \{\mathbf{y_2} \in \mathbb{N}^q : P(C_{t,j}|\mathbf{y_2}, D_t) \geq k\}, \end{aligned} \tag{3.9}$$

where $k = k_{t,j} P(C_{t,j}|D_t)$.

The first step in the construction of the alarm system consists on the calculation of both probabilities: the probability of catastrophe conditional on D_t and $\mathbf{y_2}$, i.e., $P(C_{t,j}|\mathbf{y_2}, D_t, \boldsymbol{\theta})$, and the probability of catastrophe conditional on D_t, $P(C_{t,j}|D_t, \boldsymbol{\theta})$.

$P(C_{t,j}|\mathbf{y_2}, D_t, \boldsymbol{\theta}) =$

$$
\begin{aligned}
&= P(Y_{t+j-1} \leqslant u < Y_{t+j}|\mathbf{y_2}, D_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}, Y_{t+j} > u|y_1, ..., y_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}|\mathbf{y_2}, D_t, \boldsymbol{\theta}) P(Y_{t+j} > u|y_{t+j-1}, \mathbf{y_2}, D_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}|\mathbf{y_2}, D_t, \boldsymbol{\theta}) \left(1 - P(Y_{t+j} \leqslant u|Y_{t+j-1}, \mathbf{y_2}, D_t, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}|\mathbf{y_2}, D_t, \boldsymbol{\theta}) \times \\
&\times \left(1 - \sum_{y_{t+j}=0}^{u} P(Y_{t+j} = y_{t+j}|y_{t+j-1}, \mathbf{y_2}, D_t, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+j-1}=0}^{u} p(y_{t+j-1}|y_t, \boldsymbol{\theta}) \left(1 - \sum_{y_{t+j}=0}^{u} p(y_{t+j}|y_{t+j-1}, \boldsymbol{\theta})\right),
\end{aligned}
$$

$P(C_{t,j}|D_t, \boldsymbol{\theta}) =$

$$
\begin{aligned}
&= P(Y_{t+j-1} \leqslant u < Y_{t+j}|D_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}, Y_{t+j} > u|D_t \boldsymbol{\theta}) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}|D_t, \boldsymbol{\theta}) P(Y_{t+j} > u|y_{t+j-1}, D_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}|D_t, \boldsymbol{\theta})(1 - P(Y_{t+j} \leqslant u|y_{t+j-1}, D_t, \boldsymbol{\theta})) \\
&= \sum_{y_{t+j-1}=0}^{u} P(Y_{t+j-1} = y_{t+j-1}|D_t, \boldsymbol{\theta}) \\
&\times \left(1 - \sum_{y_{t+j}=0}^{u} P(Y_{t+j} = y_{t+j}|y_{t+j-1}, D_t, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+j-1}=0}^{u} p(y_{t+j-1}|y_{t-q}, \boldsymbol{\theta}) \left(1 - \sum_{y_{t+j}=0}^{u} p(y_{t+j}|y_{t+j-1}, \boldsymbol{\theta})\right).
\end{aligned}
$$

After calculating these probabilities it is then possible to move on to the operating characteristics of the alarm system:

1. Alarm size

Since $\mathbf{Y_2} = \{Y_{t-q+1}, Y_{t-q+2}, \ldots, Y_{t-1}, Y_t\}$, the size of the alarm region is given by:

$$\begin{aligned}
\alpha_{t,j} &= P(A_{t,j}|D_t, \boldsymbol{\theta}) \\
&= \sum_{\mathbf{y_2}\in A_{t,j}} P(\mathbf{Y_2} = \mathbf{y_2}|D_t, \boldsymbol{\theta}) \\
&= \sum_{\mathbf{y_2}\in A_{t,j}} p(y_t|y_{t-1}, \boldsymbol{\theta}) p(y_{t-1}|y_{t-2}, \boldsymbol{\theta}) \cdots p(y_{t-q+1}|y_{t-q}, \boldsymbol{\theta}) \\
&= \sum_{\mathbf{y_2}\in A_{t,j}} \prod_{i=0}^{q-1} p(y_{t-i}|y_{t-i-1}, \boldsymbol{\theta}) \\
&= \sum_{\mathbf{y_2}\in A_{t,j}} \prod_{i=1}^{q} p(y_{t-i+1}|y_{t-i}, \boldsymbol{\theta}) \\
&= \sum_{\mathbf{y_2}\in A_{t,j}} \prod_{i=1}^{q} \frac{e^{-\lambda_{t-i+1}} \lambda_{t-i+1}^{y_{t-i+1}}}{(y_{t-i+1})!}.
\end{aligned}$$

2. Probability of correct alarm

$$\begin{aligned}
P(C_{t,j}|A_{t,j}, D_t, \boldsymbol{\theta}) &= \frac{P(C_{t,j} \bigcap A_{t,j}|D_t, \boldsymbol{\theta})}{P(A_{t,j}|D_t, \boldsymbol{\theta})} \\
&= \frac{P(Y_{t+j-1} \leqslant u < Y_{t+j}, \mathbf{Y_2} \in A_{t,j}|D_t, \boldsymbol{\theta})}{P(\mathbf{Y_2} \in A_{t,j}|D_t, \boldsymbol{\theta})} \\
&= \frac{\displaystyle\sum_{\mathbf{y_2}\in A_{t,j}} P(\mathbf{Y_2} = \mathbf{y_2}|D_t, \boldsymbol{\theta}) P(C_{t,j}|\mathbf{Y_2} = \mathbf{y_2}, D_t, \boldsymbol{\theta})}{\displaystyle\sum_{\mathbf{y_2}\in A_{t,j}} P(\mathbf{Y_2} = \mathbf{y_2}|D_t, \boldsymbol{\theta})} \\
&= \frac{\displaystyle\sum_{\mathbf{y_2}\in A_{t,j}} \prod_{i=1}^{q} p(y_{t-i+1}|y_{t-i}, \boldsymbol{\theta}) P(C_{t,j}|\mathbf{Y_2} = \mathbf{y_2}, D_t, \boldsymbol{\theta})}{\displaystyle\sum_{\mathbf{y_2}\in A_{t,j}} \prod_{i=1}^{q} p(y_{t-i+1}|y_{t-i}, \boldsymbol{\theta})}
\end{aligned}$$

and, given the probability of catastrophe, $P(C_{t,j}|\mathbf{y_2}, D_t, \boldsymbol{\theta})$,

$$= \sum_{\mathbf{y_2}\in A_{t,j}} \left[\prod_{i=1}^{q} p(y_{t-i+1}|y_{t-i}, \boldsymbol{\theta}) \sum_{y_{t+j-1}=0}^{u} p(y_{t+j-1}|y_t, \boldsymbol{\theta}) \times \right.$$
$$\left. \times \left(1 - \sum_{y_{t+j}=0}^{u} p(y_{t+j}|y_{t+j-1}, \boldsymbol{\theta}) \right) \right] \left[\sum_{\mathbf{y_2}\in A_{t,j}} \prod_{i=1}^{q} p(y_{t-i+1}|y_{t-i}, \boldsymbol{\theta}) \right]^{-1}.$$

3. Probability of detecting the event

$$\begin{aligned} P(A_{t,j}|C_{t,j}, D_t, \boldsymbol{\theta}) &= \frac{P(A_{t,j} \bigcap C_{t,j}|D_t, \boldsymbol{\theta})}{P(C_{t,j}|D_t, \boldsymbol{\theta})} \\ &= \frac{P(\mathbf{Y_2} \in A_{t,j}, Y_{t+j-1} \leqslant u < Y_{t+j}|D_t, \boldsymbol{\theta})}{P(C_{t,j}|D_t, \boldsymbol{\theta})} \\ &= \frac{\sum_{\mathbf{y_2}\in A_{t,j}} P(\mathbf{Y_2} = \mathbf{y_2}|D_t, \boldsymbol{\theta}) P(C_{t,j}|\mathbf{Y_2} = \mathbf{y_2}, D_t, \boldsymbol{\theta})}{P(C_{t,j}|D_t, \boldsymbol{\theta})}. \end{aligned}$$

Since the numerator in this expression is the same as the numerator in the expression for the probability of correct alarm, and, given the probability of catastrophe, $P(C_{t,j}|D_t, \boldsymbol{\theta})$,

$$= \sum_{\mathbf{y_2}\in A_{t,j}} \left[\prod_{i=1}^{q} p(y_{t-i+1}|y_{t-i}, \boldsymbol{\theta}) \sum_{y_{t+j-1}=0}^{u} p(y_{t+j-1}|y_t, \boldsymbol{\theta}) \times \right.$$
$$\left. \times \left(1 - \sum_{y_{t+j}=0}^{u} p(y_{t+j}|y_{t+j-1}, \boldsymbol{\theta}) \right) \right]$$
$$\times \left[\sum_{y_{t+j-1}=0}^{u} p(y_{t+j-1}|y_{t-q}, \boldsymbol{\theta}) \left(1 - \sum_{y_{t+j}=0}^{u} p(y_{t+j}|y_{t+j-1}, \boldsymbol{\theta}) \right) \right]^{-1}.$$

4. Probability of false alarm

$$P(\overline{C_{t,j}}|A_{t,j}, D_t, \boldsymbol{\theta}) = 1 - P(C_{t,j}|A_{t,j}, D_t, \boldsymbol{\theta}).$$

5. Probability of not detecting the event

$$P(\overline{A_{t,j}}|C_{t,j}, D_t, \boldsymbol{\theta}) = 1 - P(A_{t,j}|C_{t,j}, D_t, \boldsymbol{\theta}).$$

The application to the INAPARCH$(1,1)$ model will be done for the particular case $q=1$ and $j=2$. Thus, the time sequel is divided in the following manner:

$$D_t = \{y_1, y_2, \ldots, y_{t-1}\} \qquad \mathbf{y_2} = \{y_t\} \qquad \mathbf{y_3} = \{y_{t+1}, y_{t+2}, \ldots\}.$$

The event of interest or the catastrophe is defined as the up-crossing of some fixed level u two steps ahead,

$$C_{t,2} = \{(y_{t+1}, y_{t+2}) \in \mathbb{N}^2 : y_{t+1} \leqslant u < y_{t+2}\}.$$

The optimal alarm region of size α_2 is given by

$$\begin{aligned} A_{t,2} &= \left\{ y_t \in \mathbb{N} : \frac{P(C_{t,2}|y_t, D_t)}{P(C_{t,2}|D_t)} \geqslant k_{t,2} \right\} \\ &= \{y_t \in \mathbb{N} : P(C_{t,2}|y_t, D_t) \geqslant k\}, \end{aligned}$$

where $k = k_{t,2}P(C_{t,2}|D_t)$. As already mentioned, the first step in the construction of the alarm system consists on the calculation of $P(C_{t,2}|y_t, D_t, \boldsymbol{\theta})$ and $P(C_{t,2}|D_t, \boldsymbol{\theta})$.

$$P(C_{t,2}|y_t, D_t, \boldsymbol{\theta}) =$$

$$\begin{aligned} &= P(Y_{t+1} \leqslant u < Y_{t+2}|y_t, D_t, \boldsymbol{\theta}) \\ &= \sum_{y_{t+1}=0}^{u} P(Y_{t+1} = y_{t+1}, Y_{t+2} > u|y_1, \ldots, y_t, \boldsymbol{\theta}) \\ &= \sum_{y_{t+1}=0}^{u} P(Y_{t+1} = y_{t+1}|y_t, D_t, \boldsymbol{\theta}) P(Y_{t+2} > u|y_{t+1}, y_t, D_t, \boldsymbol{\theta}) \\ &= \sum_{y_{t+1}=0}^{u} P(Y_{t+1} = y_{t+1}|y_t, D_t, \boldsymbol{\theta}) \left(1 - P(Y_{t+2} \leqslant u|y_{t+1}, y_t, D_t, \boldsymbol{\theta})\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{y_{t+1}=0}^{u} P(Y_{t+1}=y_{t+1}|y_t, D_t, \boldsymbol{\theta})\times \\
&\times \left(1-\sum_{y_{t+2}=0}^{u} P(Y_{t+2}=y_{t+2}|y_{t+1}, y_t, D_t, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+1}=0}^{u} p(y_{t+1}|y_t, \boldsymbol{\theta})\left(1-\sum_{y_{t+2}=0}^{u} p(y_{t+2}|y_{t+1}, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+1}=0}^{u} \frac{e^{-\lambda_{t+1}}\lambda_{t+1}^{y_{t+1}}}{(y_{t+1})!}\left(1-\sum_{y_{t+2}=0}^{u} \frac{e^{-\lambda_{t+2}}\lambda_{t+2}^{y_{t+2}}}{(y_{t+2})!}\right)
\end{aligned}$$

$P(C_{t,2}|D_t, \boldsymbol{\theta}) =$

$$\begin{aligned}
&= P(Y_{t+1}\leqslant u < Y_{t+2}|D_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+1}=0}^{u} P(Y_{t+1}=y_{t+1}, Y_{t+2}>u|D_t\boldsymbol{\theta}) \\
&= \sum_{y_{t+1}=0}^{u} P(Y_{t+1}=y_{t+1}|D_t, \boldsymbol{\theta})P(Y_{t+2}>u|y_{t+1}, D_t, \boldsymbol{\theta}) \\
&= \sum_{y_{t+1}=0}^{u} P(Y_{t+1}=y_{t+1}|D_t, \boldsymbol{\theta})\left(1-P(Y_{t+2}\leqslant u|y_{t+1}, D_t, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+1}=0}^{u} P(Y_{t+1}=y_{t+1}|D_t, \boldsymbol{\theta})\left(1-\sum_{y_{t+2}=0}^{u} P(Y_{t+2}=y_{t+2}|y_{t+1}, D_t, \boldsymbol{\theta})\right) \\
&= \sum_{y_{t+1}=0}^{u} p(y_{t+1}|y_{t-1}, \boldsymbol{\theta})\left(1-\sum_{y_{t+2}=0}^{u} p(y_{t+2}|y_{t+1}, \boldsymbol{\theta})\right) \\
&= \sum_{y_t} p(y_t|y_{t-1}, \boldsymbol{\theta})\sum_{y_{t+1}=0}^{u} p(y_{t+1}|y_t, \boldsymbol{\theta})\left(1-\sum_{y_{t+2}=0}^{u} p(y_{t+2}|y_{t+1}, \boldsymbol{\theta})\right) \\
&= \sum_{y_t} \frac{e^{-\lambda_t}\lambda_t^{y_t}}{(y_t)!}\sum_{y_{t+1}=0}^{u} \frac{e^{-\lambda_{t+1}}\lambda_{t+1}^{y_{t+1}}}{(y_{t+1})!}\left(1-\sum_{y_{t+2}=0}^{u} \frac{e^{-\lambda_{t+2}}\lambda_{t+2}^{y_{t+2}}}{(y_{t+2})!}\right).
\end{aligned}$$

Having calculated these probabilities it is then possible to explicit all the operating characteristics.

1. Alarm size

   Since $\mathbf{y_2} = \{y_t\}$, the alarm size is simply

$$\begin{aligned}\alpha_{t,2} &= P(A_{t,2}|D_t, \boldsymbol{\theta}) \\ &= \sum_{y_t \in A_{t,2}} P(Y_t = y_t|D_t, \boldsymbol{\theta}) \\ &= \sum_{y_t \in A_{t,2}} p(y_t|y_{t-1}, \boldsymbol{\theta}) \\ &= \sum_{y_t \in A_{t,2}} \frac{e^{-\lambda_t}\lambda_t^{y_t}}{(y_t)!},\end{aligned}$$

   with $A_{t,2}$ being the alarm region which depends on the choice of $k_{t,2}$.

2. Probability of correct alarm

$$\begin{aligned}P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}) &= \frac{P(C_{t,2} \bigcap A_{t,2}|D_t, \boldsymbol{\theta})}{P(A_{t,2}|D_t, \boldsymbol{\theta})} \\ &= \frac{P(Y_{t+1} \leqslant u < Y_{t+2}, Y_t \in A_{t,2}|D_t, \boldsymbol{\theta})}{P(Y_t \in A_{t,2}|D_t, \boldsymbol{\theta})} \\ &= \frac{\sum\limits_{y_t \in A_{t,2}} P(Y_t = y_t|D_t, \boldsymbol{\theta}) P(C_{t,2}|Y_t = y_t, D_t, \boldsymbol{\theta})}{\sum\limits_{y_t \in A_{t,2}} P(Y_t = y_t|D_t, \boldsymbol{\theta})} \\ &= \frac{\sum\limits_{y_t \in A_{t,2}} p(y_t|y_{t-1}, \boldsymbol{\theta}) P(C_{t,2}|y_t, D_t, \boldsymbol{\theta})}{\sum\limits_{y_t \in A_{t,2}} p(y_t|y_{t-1}, \boldsymbol{\theta})}\end{aligned}$$

   and given $P(C_{t,2}|y_t, D_t, \boldsymbol{\theta})$ it follows that

$$\begin{aligned}P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}) = \sum_{y_t \in A_{t,2}} \Bigg[\frac{e^{-\lambda_t}\lambda_t^{y_t}}{(y_t)!} \sum_{y_{t+1}=0}^{u} \frac{e^{-\lambda_{t+1}}\lambda_{t+1}^{y_{t+1}}}{(y_{t+1})!} \times \\ \times \left(1 - \sum_{y_{t+2}=0}^{u} \frac{e^{-\lambda_{t+2}}\lambda_{t+2}^{y_{t+2}}}{(y_{t+2})!}\right)\Bigg] \left[\sum_{y_t \in A_{t,2}} \frac{e^{-\lambda_t}\lambda_t^{y_t}}{(y_t)!}\right]^{-1}.\end{aligned}$$

3. Probability of detecting the event

$$\begin{aligned}
P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta}) &= \frac{P(A_{t,2} \bigcap C_{t,2}|D_t, \boldsymbol{\theta})}{P(C_{t,2}|D_t, \boldsymbol{\theta})} \\
&= \frac{P(Y_t \in A_{t,2}, Y_{t+1} \leqslant u < Y_{t+2}|D_t, \boldsymbol{\theta})}{P(C_{t,2}|D_t, \boldsymbol{\theta})} \\
&= \frac{\displaystyle\sum_{y_t \in A_{t,2}} P(Y_t = y_t|D_t, \boldsymbol{\theta}) P(C_{t,2}|Y_t = y_t, D_t, \boldsymbol{\theta})}{P(C_{t,2}|D_t, \boldsymbol{\theta})}.
\end{aligned}$$

Once again, the numerator in this expression is the same as the numerator in the expression for the probability of correct alarm, and, given the probability of catastrophe, $P(C_{t,2}|D_t, \boldsymbol{\theta})$, the above expression can be rewritten as

$$\begin{aligned}
= & \sum_{y_t \in A_{t,2}} \left[\frac{e^{-\lambda_t} \lambda_t^{y_t}}{(y_t)!} \sum_{y_{t+1}=0}^{u} \frac{e^{-\lambda_{t+1}} \lambda_{t+1}^{y_{t+1}}}{(y_{t+1})!} \left(1 - \sum_{y_{t+2}=0}^{u} \frac{e^{-\lambda_{t+2}} \lambda_{t+2}^{y_{t+2}}}{(y_{t+2})!} \right) \right] \times \\
\times & \left[\sum_{y_t} \frac{e^{-\lambda_t} \lambda_t^{y_t}}{(y_t)!} \sum_{y_{t+1}=0}^{u} \frac{e^{-\lambda_{t+1}} \lambda_{t+1}^{y_{t+1}}}{(y_{t+1})!} \left(1 - \sum_{y_{t+2}=0}^{u} \frac{e^{-\lambda_{t+2}} \lambda_{t+2}^{y_{t+2}}}{(y_{t+2})!} \right) \right]^{-1}.
\end{aligned}$$

4. Probability of false alarm

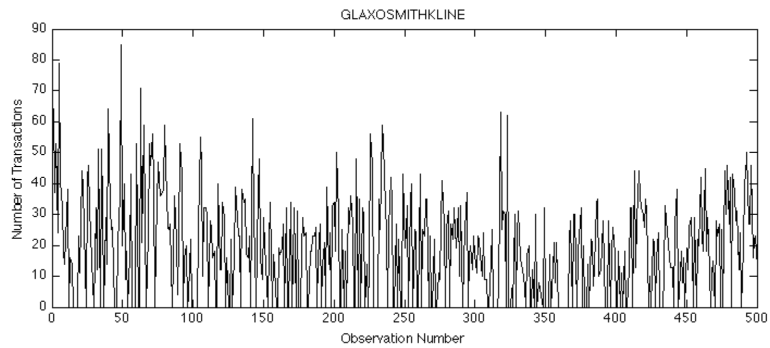
$$P(\overline{C_{t,2}}|A_{t,2}, D_t, \boldsymbol{\theta}) = 1 - P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta}).$$

5. Probability of not detecting the event

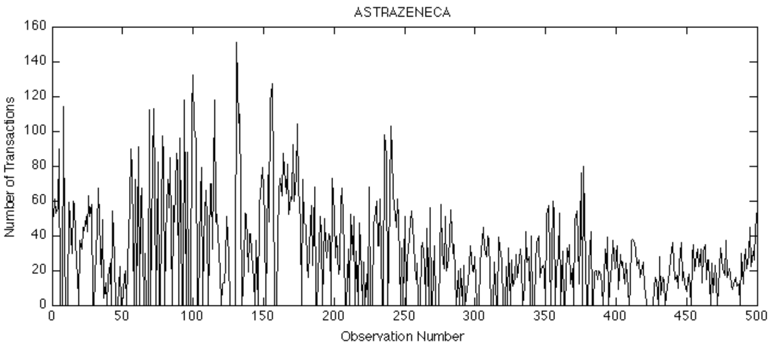
$$P(\overline{A_{t,2}}|C_{t,2}, D_t, \boldsymbol{\theta}) = 1 - P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta}).$$

3.6 Application to the number of transactions in stocks

Finally, the conditional maximum likelihood estimation procedure was applied to estimate two time series of count data, generated from stock transactions. The tick-by-tick data for Glaxosmithkline and Astrazeneca have been downloaded from www.dukascopy.com, and treated in order to fill in the zero counts during the trading periods considered and delete all trading during the first and the last five minutes of each day (trading mechanisms may be different during the opening and closing of the stock exchange market). The data consists on the number of transactions per minute during one trading day (September 19, 2012, for Glaxosmithkline and September 21, 2012, for Astrazeneca), corresponding to 501 observations for each series. The series are presented in Figure 3.4 and the estimation results in Table 3.3 with standard errors in parentheses.



(mean 18.86, variance 277.00, maximum 85)



(mean 30.83, variance 819.14, maximum 151)

Figure 3.4: Time series plots for Glaxosmithkline and Astrazeneca.

The estimated value of the γ parameter ($\hat{\gamma} = -0.3269$ for the Glaxosmithkline series and $\hat{\gamma} = -0.2787$ for the Astrazeneca series) is negative for both

Table 3.3: Maximum likelihood estimation results for Glaxosmithkline and Astrazeneca time series.

Time series	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\delta}$
Glaxosmithkline	0.3781	0.1392	-0.3269	0.8791	0.9826
	(0.0685)	(0.0074)	(0.0843)	(0.0073)	(0.0005)
Astrazeneca	2.4862	0.2824	-0.2787	0.7501	1.0598
	(0.1087)	(0.0062)	(0.0363)	(0.0044)	(0.0008)

series meaning that for these time series, there is evidence that positive shocks have stronger impact on overdispersion than negative shocks. The estimated value of the δ parameter ($\hat{\delta} = 0.9826$ for the Glaxosmithkline series and $\hat{\delta} = 0.9826$ for the Astrazeneca series) both fail the condition $\delta \geq 2$. It is worth mentioning that this is not a surprising result since in the estimation of the Standard & Poor 500 stock market daily closing price index by Ding et al. (1993) the δ estimate obtained did not satisfy the sufficient condition for the process to be covariance stationary, which was also $\delta \geq 2$. We believe emphasis should be put on finding necessary instead of sufficient conditions for stationarity and this will remain as future work.

The application of the alarm system was done to the aforementioned data series. As these are real data series, only the maximum likelihood estimates were considered in this application. The analysis was done for the time instants $t = 450$ to $t = 460$. A preliminary study was done in order to chose the value of the fixed value u. The probabilities $P(C_{t,2}|y_t, D_t, \boldsymbol{\theta})$ and $P(C_{t,2}|D_t, \boldsymbol{\theta})$ and also the alarm region were calculated for different values of u, for all the time instants mentioned. As a result of this preliminary study and in order to have reasonable probabilities of catastrophe, two different values of u were chosen for each data series: the 39^{th} percentile ($Q_{0.39}$) and the 50^{th} percentile ($Q_{0.50}$). It is worth mentioning that these data series have many zero counts and the probability of catastrophe for higher percentiles is very low. Hence, the fixed levels u considered in this application cannot be understood as a catastrophe in the sense that it should be related to a relatively rare event, but, it is simply a fixed level for which the probability of up-crossing is not negligible.

In order to obtain the optimal alarm region for each case, it is necessary to obtain the alarm region for several values of k, according to expression (1.21), Chapter 1. For each value of k the operating characteristics alarm size, $\alpha_{t,2}$, probability of correct alarm $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ and probability of detecting the event $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$ are then calculated. For every fixed value of k, the alarm region has to be obtained through a systematic search in a three dimensional region for $\{y_t, y_{t+1}, y_{t+2}\}$. We considered y_t taking all the integer values from 0 to 150 and determined, for each value of y_t, if $P(C_{t,2}|y_t, D_t, \boldsymbol{\theta})$ exceeds or not k. This procedure is repeated for all the values of k tested. The results concerning time instants $t = 450, \ldots, 460$ for the Astrazeneca series are shown in Tables 3.4, 3.5 and 3.6. The results concerning time instants $t = 450, \ldots, 460$ for the Glaxosmithkline series are shown in Tables 3.7, 3.8 and 3.9.

The step and range of variation in k were chosen for each case in order to have as many different situations as possible.

Considering Tables 3.4, 3.5 and 3.6 for the Astrazeneca series and the crossing of the fixed level $u = Q_{0.39} = 19$, the alarm size ranges from values in the interval $[0.31, 0.51]$ for the lowest value of k to around 1×10^{-5} for the highest k. The variation with k in the probability of detecting the event has the same amplitude, but, because this operating characteristic is always slightly higher than the alarm size, it starts, for the lowest value of k, taking values in the interval $[0.40, 0.53]$. It is not surprising that the probability of detection has the same behaviour as the alarm size, because as the alarm size decreases with the increase in k, the number of alarms decreases, leading directly to a lower probability of detecting the event. On the other hand, as k increases, the probability of the alarm being correct increases, starting in values around 5% or 8% for the first time instants considered, and reaching values around 16%. This behaviour is not also unexpected: as the number of alarm decreases the probability of false alarm also decreases, and, consequently, the probability of the alarm being correct is expected to increase.

Still in Tables 3.4, 3.5 and 3.6 are the results for the Astrazeneca series with $u = Q_{0.50} = 25$. Alarm size always starts at lower values than in the previous situation for corresponding time instants. The only exception is $t = 451$ for which has exactly the same value. This operation characteristic takes values starting in the interval $[0.29, 0.40]$ and decreases, as k increases, until 10^{-12}. The probability of detecting the event follows the behaviour of the alarm size, although being always slightly higher: starts taking values in $[0.45, 0.53]$ and decreases until 10^{-10}, for the highest value of k and for the first time instants analysed. The probability of the alarm being correct also has a different range of variation for the first and the last time instants considered. Considering the first time instants, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ increases from $0.2\% - 0.5\%$ until around 11%; considering the last time instants, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ starts in the interval $[0.01, 0.04]$ and increases with k but also does not exceed 12.6%.

The behaviour of the alarm system for the Glaxosmithkline series is similar to what was described for the Astrazeneca series, not only in what concerns the general tendencies of the operating characteristics but also in what concerns the comparison of the level crossings of the 39 and 50^{th} percentiles.

Considering Tables 3.7, 3.8 and 3.9 for the Glaxosmithkline series and the crossing of the fixed level $u = Q_{0.39} = 13$, there is a particular time instant for which the variation of the operating characteristics is more significant. For $t = 452$, the alarm size and the detection probability range from 40% to around 2×10^{-4}, as k increases. Simultaneously, the probability of the alarm being correct increases from 15% to 20%. For all other time instants the operating characteristics follow exactly the same tendency but in a shorter range. For the first value of k, the alarm size takes values in $[0.38, 0.85]$, depending on the time instant, and, as k increases, the alarm size decreases, reaching values ranging from 0.8% to 38.6%. The probability of detecting the event has the same behaviour as the alarm size and is always very similar to the value of the alarm size although slightly higher. On the other hand, as k increases, the probability of the alarm being correct increases, but this variation does not exceed the range $[0.17, 0.20]$.

Considering now the fixed level crossing of the 50^{th} percentile for the Glaxosmithkline series, also in Tables 3.7, 3.8 and 3.9, the general tendencies of the operation characteristics are the same as in the previous case, but in different ranges. For instance, the alarm size ranges from around 40%, for the lowest value of k, to very small sizes, of the order of 10^{-16}, for the highest value of k. Also, for corresponding time instants, alarm size is always smaller in the case of the crossing of the fixed level $u = Q_{0.50} = 18$, and this difference is bigger for the last time instants analysed. The probability of detecting the event has the same behaviour and similar value as the alarm size. Regarding the probability of the alarm being correct, the range of variation is wider than in the previous case of the 39^{th} percentile: as k increases, $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ increases from 0.02 for $t = 452$ to around 0.18 for some time instants.

One last remark about Tables 3.4 to 3.9 is that the first time instants seem to have lower probability of catastrophe for both level crossings. As a coincidence this seems to happen for both time series, influencing also the resulting operating characteristics of the alarm system.

As is obvious from the remarks above it is not possible to maximize simultaneously $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$ and $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta})$. A compromise must be reached between these operating characteristics by a proper choice of k. Several criteria have already been proposed in the literature. For instance, as already mentioned in Chapter 2, Section 2.3, Antunes et al. (2003) suggested that k should be chosen so that the alarm size is about twice the probability of having a catastrophe given the past values of the process, $P(C_{t,2}|D_t, \boldsymbol{\theta}) \simeq \frac{1}{2}P(A_{t,2}|D_t, \boldsymbol{\theta})$, meaning that in this situation the system spends twice the time in the alarm state than in the catastrophe region. The first criterion used in this application is a variation of the former. Since the alarm size is given by $P(A_{t,2}|D_t, \boldsymbol{\theta})$ and as was seen above, because the probability of detecting the event has the same behaviour of the alarm size with the variation with k, taking also similar values, we decided to substitute the alarm size with the detection probability. Moreover, we also found that the probability of correct alarm is always of the same order of the probability of catastrophe given past values of the process, $P(C_{t,2}|D_t, \boldsymbol{\theta})$: the difference between these two probabilities never exceeds 0.02. As such, we also substituted $P(C_{t,2}|D_t, \boldsymbol{\theta})$ by $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$, the probability of correct alarm. Therefore, our Criterion 1 relates directly to operating characteristics and is $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta}) \simeq 2P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$.

Another criterion found in the literature and also mentioned in Chapter 2, Section 2.3, is the one suggested by Svensson et al. (1996) in which k should be chosen so that the probability of correct alarm and the probability of detecting the event are approximately equal. Our Criterion 1 is already related with these two operating characteristics. Also, because the probability of detection is directly dependent on the alarm size, it can be chosen to be as high as desired. Thus, it seems wise to look for the best set of operating characteristics in a different perspective, looking towards minimizing the number of false alarms, which is the same as maximizing the probability of the alarm being correct. As the probability of the alarm being correct increases, the detection probability decreases and, in order not to have too small detection probability we state the Criterion 2 as: Maximum $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$, as long as $P(A_{t,2}|C_{t,2}, D_t, \boldsymbol{\theta}) \geq 0.001$.

The online prediction is illustrated in Tables 3.10, 3.11, 3.12 and 3.13. The informative experience evolves as the time instant varies from $t = 450$ to $t = 460$. The probability of catastrophe given the past experience, the alarm region and respective operating characteristics are presented, for each criteria. Tables 3.10 and 3.11 refer to the fixed level crossings $u = Q_{0.39} = 19$ and $u = Q_{0.50} = 25$, respectively, for the Astrazeneca series. Tables 3.12 and 3.13 refer to the fixed level crossings $u = Q_{0.39} = 13$ and $u = Q_{0.50} = 18$, respectively, for the Glaxosmithkline series.

One general remark regarding the online prediction system is that Criterion 2, which tends to minimize the number of false alarms, is always satisfied for a higher value of k, when compared with Criterion 1. Only exceptions are two cases for which both criteria are simultaneously satisfied: $t = 452$, level crossing of the 50^{th} percentile for the Astrazeneca series and $t = 458$, level crossing of the 39^{th} percentile for the Glaxosmithkline series. This observation is not surprising since the probability of correct alarm increases with the increase in k.

In order to test the alarm system, three extra values of both time series were simulated: $(\mathbf{y_2}, \mathbf{y_3}) = (y_t, y_{t+1}, y_{t+2})$. This procedure was repeated 100 000

times with the same informative experience, D_t, for each series. Considering the alarm regions obtained before for $u = Q_{0.39}$ and for $u = Q_{0.50}$ and for the two criteria already mentioned, it was observed for each of the 100 000 samples whether an alarm was given or not and whether a catastrophe occurred or not. The operating characteristics can then be estimated with these counts. This procedure was repeated for several time instants and results are presented in Tables 3.14 and 3.15 for the Astrazeneca series and in Tables 3.16 and 3.17 for the Glaxosmithkline series. The time instants were chosen for their better set of operating characteristics and particularly for the higher values of $P(C_{t,2}|A_{t,2}, D_t, \boldsymbol{\theta})$.
Regarding these results several conclusions can be outlined:

- First of all, considering the fixed level crossing of the 39^{th} percentile, the application overall overestimates the theoretical operating characteristics. This overestimation is more noticeable for the probability of correct alarm, whose theoretical values are around a half of the estimated ones.

- Considering the fixed level crossing of the 50^{th} percentile, the estimates obtained with the application are very similar to the theoretical values of the operating characteristics. Notice, for instance, that in the Glaxosmithkline series, the estimated and theoretical probability of detecting the event for the time instants $t = 454$, $t = 459$ and $t = 460$ differ only on the fourth decimal place.

- Overall, Criterion 1 seems to provide better estimates of the operating characteristics, even when one considers the fixed level crossing of the 39^{th} percentile, which, as already mentioned, provides estimates somewhat far from the theoretical values.

- Alarm size and probability of detection are the operating characteristics better estimated with this application. Particularly, the alarm size (not shown directly on Tables 3.14 to 3.17, but easily obtainable) always follows the theoretical value to the third decimal place, considering any of the fixed level crossings treated above.

Table 3.4: Operating characteristics at time points $t = 450, \ldots, 453$ for the Astrazeneca series.

$u = Q_{0.39} = 19$				$u = Q_{0.50} = 25$			
t=450				t=450			
$P(C_{t,2}\|y_t, D_t) = 0.0787$				$P(C_{t,2}, y_t, D_t) = 0.0049$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0655	0.3783	0.0805	0.4645	0.0036	0.3254	0.0059	0.5227
0.0755	0.1982	0.0904	0.2735	0.0136	0.0113	0.0177	0.0549
0.0855	0.0853	0.1037	0.1350	0.0236	9.2994×10^{-4}	0.0300	0.0076
0.0955	0.0525	0.1111	0.0891	0.0336	2.2149×10^{-4}	0.0379	0.0023
0.1055	0.0324	0.1178	0.0583	0.0436	2.0778×10^{-5}	0.0520	2.9600×10^{-4}
0.1155	0.0195	0.1244	0.0370	0.0536	3.7483×10^{-6}	0.0625	6.4171×10^{-5}
0.1255	0.0064	0.1367	0.0133	0.0636	1.5318×10^{-6}	0.0679	2.8507×10^{-5}
0.1355	0.0035	0.1423	0.0075	0.0736	2.3750×10^{-7}	0.0791	5.1493×10^{-6}
0.1455	9.2971×10^{-4}	0.1518	0.0022	0.0836	3.3476×10^{-8}	0.0903	8.2845×10^{-7}
0.1555	2.1996×10^{-4}	0.1586	$0.5.3245 \times 10^{-4}$	0.0936	4.3072×10^{-9}	0.1010	1.1921×10^{-7}
t=451				t=451			
$P(C_{t,2}\|y_t, D_t) = 0.0463$				$P(C_{t,2}\|y_t, D_t) = 0.0017$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0604	0.3124	0.0777	0.4018	0.0031	0.3124	0.0051	0.5148
0.0704	0.1817	0.0860	0.2587	0.0131	0.0089	0.0166	0.0475
0.0804	0.1060	0.0944	0.1656	0.0231	6.7534×10^{-4}	0.0285	0.0062
0.0904	0.0434	0.1080	0.0777	0.0331	7.0772×10^{-5}	0.0406	9.2566×10^{-4}
0.1004	0.0265	0.1148	0.0503	0.0431	1.3647×10^{-5}	0.0501	2.2018×10^{-4}
0.1104	0.0156	0.1214	0.0314	0.0531	2.3638×10^{-6}	0.0605	4.6046×10^{-5}
0.1204	0.0089	0.1280	0.0189	0.0631	3.6962×10^{-7}	0.0715	8.5042×10^{-6}
0.1304	0.0026	0.1400	0.0061	0.0731	1.4087×10^{-7}	0.0772	3.4982×10^{-6}
0.1404	0.0013	0.1455	0.0032	0.0831	1.9058×10^{-8}	0.0884	5.4209×10^{-7}
0.1504	3.2775×10^{-4}	0.1543	8.3707×10^{-4}	0.0931	2.3533×10^{-9}	0.0992	7.5153×10^{-8}
0.1604	2.9144×10^{-5}	0.1612	7.7785×10^{-5}	0.1031	2.6624×10^{-10}	0.1094	9.3727×10^{-9}
t=452				t=452			
$P(C_{t,2}\|y_t, D_t) = 0.0473$				$P(C_{t,2}\|y_t, D_t) = 0.0018$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0381	0.3489	0.0504	0.4615	0.0013	0.2994	0.0023	0.5272
0.0481	0.1260	0.0636	0.2103	0.0113	0.0012	0.0141	0.0124
0.0581	0.0721	0.0715	0.1353	0.0213	5.1216×10^{-5}	0.0254	9.8759×10^{-4}
0.0681	0.0266	0.0852	0.0594	0.0313	3.5946×10^{-6}	0.0371	1.0123×10^{-4}
0.0781	0.0153	0.0924	0.0370	0.0413	5.3168×10^{-7}	0.0464	1.8723×10^{-5}
0.0881	0.0085	0.0996	0.0221	0.0513	7.0761×10^{-8}	0.0567	3.0420×10^{-6}
0.0981	0.0045	0.1071	0.0127	0.0613	8.5163×10^{-9}	0.0677	4.3710×10^{-7}
0.1081	0.0012	0.1214	0.0037	0.0713	9.3114×10^{-10}	0.0791	5.5846×10^{-8}
0.1181	5.5679×10^{-4}	0.1283	0.0019	0.0813	2.9740×10^{-10}	0.0847	1.9109×10^{-8}
0.1281	2.5937×10^{-4}	0.1347	9.1651×10^{-4}	0.0913	2.8375×10^{-11}	0.0959	2.0652×10^{-9}
0.1381	5.1216×10^{-5}	0.1463	1.9651×10^{-4}	0.1013	2.4841×10^{-12}	0.1065	2.0066×10^{-10}
t=453				t=453			
$P(C_{t,2}\|y_t, D_t) = 0.0745$				$P(C_{t,2}\|y_t, D_t) = 0.0044$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0388	0.3510	0.0512	0.4630	0.0014	0.3025	0.0024	0.5291
0.0488	0.1286	0.0644	0.2132	0.0114	0.0012	0.0143	0.0129
0.0588	0.0744	0.0721	0.1381	0.0214	5.5180×10^{-5}	0.0256	0.0010
0.0688	0.0276	0.0858	0.0611	0.0314	3.9202×10^{-6}	0.0374	1.0761×10^{-4}
0.0788	0.0159	0.0929	0.0382	0.0414	5.8460×10^{-7}	0.0467	2.0048×10^{-5}
0.0888	0.0089	0.1002	0.0229	0.0514	7.8444×10^{-8}	0.0570	3.2816×10^{-6}
0.0988	0.0047	0.1076	0.0132	0.0614	9.5189×10^{-9}	0.0680	4.7499×10^{-7}
0.1088	0.0012	0.1219	0.0039	0.0714	1.0494×10^{-9}	0.0794	6.1147×10^{-8}
0.1188	5.9269×10^{-4}	0.1287	0.0020	0.0814	3.3654×10^{-10}	0.0850	2.1002×10^{-8}
0.1288	2.7720×10^{-4}	0.1351	9.6467×10^{-4}	0.0914	3.2377×10^{-11}	0.0963	2.2872×10^{-9}
0.1388	5.5180×10^{-5}	0.1466	2.0833×10^{-4}	0.1014	2.8579×10^{-12}	0.1068	2.2394×10^{-10}

Table 3.5: Operating characteristics at time points $t = 454, \ldots, 457$ for the Astrazeneca series.

$u = Q_{0.39} = 19$				$u = Q_{0.50} = 25$			
t=454 $P(C_{t,2}\|y_t, D_t) = 0.1395$				t=454 $P(C_{t,2}\|y_t, D_t) = 0.0210$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0575	0.3806	0.0715	0.4732	0.0028	0.3062	0.0047	0.5108
0.0675	0.1731	0.0833	0.2509	0.0128	0.0077	0.0160	0.0436
0.0775	0.0973	0.0924	0.1564	0.0228	5.5741×10^{-4}	0.0277	0.0055
0.0875	0.0620	0.0993	0.1070	0.0328	5.6373×10^{-5}	0.0397	7.9245×10^{-4}
0.0975	0.0234	0.1130	0.0460	0.0428	1.0613×10^{-5}	0.0491	1.8459×10^{-4}
0.1075	0.0137	0.1198	0.0284	0.0528	1.7947×10^{-6}	0.0594	3.7780×10^{-5}
0.1175	0.0077	0.1264	0.0169	0.0628	2.7394×10^{-7}	0.0703	6.8271×10^{-6}
0.1275	0.0022	0.1387	0.0053	0.0728	3.7923×10^{-8}	0.0816	1.0972×10^{-6}
0.1375	0.0011	0.1443	0.0028	0.0828	1.3621×10^{-8}	0.0872	4.2115×10^{-7}
0.1475	2.6742×10^{-4}	0.1534	7.1367×10^{-4}	0.0928	1.6415×10^{-9}	0.0982	5.7129×10^{-8}
0.1575	5.5663×10^{-5}	0.1596	1.5452×10^{-4}	0.1028	1.8124×10^{-10}	0.1084	6.9663×10^{-9}
t=455 $P(C_{t,2}\|y_t, D_t) = 0.1563$				t=455 $P(C_{t,2}\|y_t, D_t) = 0.0623$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1085	0.4239	0.1223	0.4778	0.0111	0.3056	0.0168	0.4647
0.1135	0.3056	0.1273	0.3583	0.0211	0.0541	0.0283	0.1380
0.1185	0.2127	0.1323	0.2594	0.0311	0.0131	0.0391	0.0463
0.1235	0.1305	0.1388	0.1669	0.0411	0.0044	0.0477	0.0190
0.1285	0.1236	0.1395	0.1588	0.0511	0.0013	0.0572	0.0068
0.1335	0.0832	0.1442	0.1107	0.0611	3.5349×10^{-4}	0.0676	0.0022
0.1385	0.0541	0.1488	0.0741	0.0711	8.5315×10^{-5}	0.0784	6.0369×10^{-4}
0.1435	0.0347	0.1526	0.0488	0.0811	1.8622×10^{-5}	0.0892	1.4997×10^{-4}
0.1485	0.0216	0.1559	0.0310	0.0911	3.6916×10^{-6}	0.0998	3.3236×10^{-5}
0.1535	0.0130	0.1585	0.0189	0.1011	1.5874×10^{-6}	0.1048	1.5017×10^{-5}
0.1585	0.0074	0.1603	0.0109	0.1111	2.7443×10^{-7}	0.1140	2.8236×10^{-6}
t=456 $P(C_{t,2}\|y_t, D_t) = 0.1464$				t=456 $P(C_{t,2}\|y_t, D_t) = 0.0250$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1520	0.5094	0.1572	0.5270	0.0325	0.3961	0.0418	0.5101
0.1530	0.4373	0.1581	0.4547	0.0425	0.1186	0.0552	0.2018
0.1540	0.3734	0.1588	0.3902	0.0525	0.0534	0.0647	0.1063
0.1550	0.3734	0.1588	0.3902	0.0625	0.0227	0.0744	0.0519
0.1560	0.3300	0.1592	0.3456	0.0725	0.0088	0.0844	0.0228
0.1570	0.2673	0.1599	0.2813	0.0825	0.0031	0.0945	0.0089
0.1580	0.2229	0.1604	0.2352	0.0925	0.0017	0.0994	0.0053
0.1590	0.2229	0.1604	0.2352	0.1025	5.2170×10^{-4}	0.1087	0.0017
0.1600	0.1306	0.1612	0.1385	0.1125	1.4272×10^{-4}	0.1170	5.1454×10^{-4}
0.1610	0.0738	0.1615	0.0784	0.1225	1.7229×10^{-5}	0.1260	6.6871×10^{-5}
t=457 $P(C_{t,2}\|y_t, D_t) = 0.1603$				t=457 $P(C_{t,2}\|y_t, D_t) = 0.0395$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1156	0.4194	0.1293	0.4690	0.0130	0.3654	0.0185	0.5180
0.1206	0.2596	0.1358	0.3050	0.0230	0.0673	0.0304	0.1571
0.1256	0.2318	0.1372	0.2751	0.0330	0.0173	0.0416	0.0553
0.1306	0.1605	0.1418	0.1969	0.0430	0.0061	0.0503	0.0235
0.1356	0.1026	0.1470	0.1305	0.0530	0.0019	0.0600	0.0088
0.1406	0.0684	0.1510	0.0893	0.0630	5.3314×10^{-4}	0.0704	0.0029
0.1456	0.0439	0.1545	0.0587	0.0730	1.3463×10^{-4}	0.0812	8.3796×10^{-4}
0.1506	0.0277	0.1575	0.0378	0.0830	3.0755×10^{-5}	0.0919	2.1680×10^{-4}
0.1556	0.0168	0.1596	0.0232	0.0930	6.3828×10^{-6}	0.1022	5.0027×10^{-5}
0.1606	0.0042	0.1616	0.0058	0.1030	2.8085×10^{-6}	0.1072	2.3080×10^{-5}

Table 3.6: Operating characteristics at time points $t = 458, \ldots, 460$ for the Astrazeneca series.

$u = Q_{0.39} = 19$				$u = Q_{0.50} = 25$			
t=458				t=458			
$P(C_{t,2} \mid y_t, D_t) = 0.1596$				$P(C_{t,2} \mid y_t, D_t) = 0.0563$			
k	α_2	$P(C_{t,2} \mid A_{t,2})$	$P(A_{t,2} \mid C_{t,2})$	k	α_2	$P(C_{t,2} \mid A_{t,2})$	$P(A_{t,2} \mid C_{t,2})$
0.1345	0.4380	0.1454	0.4732	0.0202	0.3200	0.0288	0.4556
0.1370	0.3195	0.1486	0.3530	0.0302	0.0857	0.0407	0.1727
0.1395	0.3195	0.1486	0.3530	0.0402	0.0360	0.0494	0.0879
0.1420	0.2818	0.1498	0.3138	0.0502	0.0142	0.0585	0.0412
0.1445	0.1984	0.1529	0.2255	0.0602	0.0050	0.0684	0.0170
0.1470	0.1814	0.1536	0.2071	0.0702	0.0016	0.0788	0.0062
0.1495	0.1249	0.1564	0.1451	0.0802	4.5795×10^{-4}	0.0893	0.0020
0.1520	0.0841	0.1585	0.0991	0.0902	1.1877×10^{-4}	0.0996	5.8510×10^{-4}
0.1545	0.0801	0.1588	0.0945	0.1002	2.8006×10^{-5}	0.1092	1.5132×10^{-4}
0.1570	0.0505	0.1603	0.0601	0.1102	6.0272×10^{-6}	0.1177	3.5094×10^{-5}
0.1595	0.0310	0.1611	0.0371	0.1202	1.1878×10^{-6}	0.1243	7.3016×10^{-6}
t=459				t=459			
$P(C_{t,2} \mid y_t, D_t) = 0.1404$				$P(C_{t,2} \mid y_t, D_t) = 0.0215$			
k	α_2	$P(C_{t,2} \mid A_{t,2})$	$P(A_{t,2} \mid C_{t,2})$	k	α_2	$P(C_{t,2} \mid A_{t,2})$	$P(A_{t,2} \mid C_{t,2})$
0.1485	0.5085	0.1548	0.5300	0.0291	0.3379	0.0394	0.4576
0.1495	0.4482	0.1556	0.4697	0.0391	0.1046	0.0519	0.1867
0.1505	0.3793	0.1567	0.4002	0.0491	0.0438	0.0619	0.0932
0.1515	0.3267	0.1576	0.3467	0.0591	0.0181	0.0716	0.0445
0.1525	0.3267	0.1576	0.3467	0.0691	0.0068	0.0817	0.0190
0.1535	0.3267	0.1576	0.3467	0.0791	0.0040	0.0868	0.0118
0.1545	0.2318	0.1591	0.2484	0.0891	0.0013	0.0969	0.0042
0.1555	0.2318	0.1591	0.2484	0.0991	3.6816×10^{-4}	0.1065	0.0013
0.1565	0.1995	0.1596	0.2144	0.1091	9.7163×10^{-5}	0.1151	3.8480×10^{-4}
0.1575	0.1995	0.1596	0.2144	0.1191	2.3413×10^{-5}	0.1224	9.8558×10^{-5}
0.1585	0.1346	0.1606	0.1456				
0.1595	0.1141	0.1609	0.1236				
0.1605	0.0704	0.1614	0.0765				
0.1615	0.0232	0.1617	0.0252				
t=460				t=460			
$P(C_{t,2} \mid y_t, D_t) = 0.1617$				$P(C_{t,2} \mid y_t, D_t) = 0.0469$			
k	α_2	$P(C_{t,2} \mid A_{t,2})$	$P(A_{t,2} \mid C_{t,2})$	k	α_2	$P(C_{t,2} \mid A_{t,2})$	$P(A_{t,2} \mid C_{t,2})$
0.1094	0.3529	0.1259	0.4063	0.0113	0.3077	0.0171	0.4660
0.1144	0.3077	0.1279	0.3599	0.0213	0.0555	0.0285	0.1401
0.1194	0.2148	0.1330	0.2611	0.0313	0.0136	0.0394	0.0473
0.1244	0.1451	0.1380	0.1831	0.0413	0.0046	0.0480	0.0195
0.1294	0.1260	0.1399	0.1611	0.0513	0.0014	0.0576	0.0070
0.1344	0.0851	0.1446	0.1126	0.0613	3.7183×10^{-4}	0.0679	0.0022
0.1394	0.0555	0.1491	0.0757	0.0713	9.0236×10^{-5}	0.0787	6.2842×10^{-4}
0.1444	0.0357	0.1528	0.0499	0.0813	1.9805×10^{-5}	0.0896	1.5689×10^{-4}
0.1494	0.0223	0.1561	0.0318	0.0913	3.9483×10^{-6}	0.1000	3.4939×10^{-5}
0.1544	0.0134	0.1587	0.0194	0.1013	1.7025×10^{-6}	0.1051	1.5827×10^{-5}
0.1594	0.0073	0.1604	0.0107	0.1113	2.9600×10^{-7}	0.1142	2.9909×10^{-6}

Table 3.7: Operating characteristics at time points $t = 450, \ldots, 453$ for the Glaxosmithkline series.

$u = Q_{0.39} = 13$				$u = Q_{0.50} = 18$			
t=450 $P(C_{t,2}) = 0.1883$				t=450 $P(C_{t,2}) = 0.0416$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1725	0.4099	0.1809	0.4298	0.0294	0.3495	0.0366	0.4345
0.1755	0.3495	0.1822	0.3691	0.0444	0.0509	0.0509	0.0893
0.1785	0.2075	0.1861	0.2238	0.0594	0.0051	0.0687	0.0120
0.1815	0.1347	0.1890	0.1476	0.0744	6.2032×10^{-4}	0.0834	0.0018
0.1845	0.1289	0.1893	0.1414	0.0894	1.2793×10^{-4}	0.0936	4.0718×10^{-4}
0.1875	0.0819	0.1919	0.0911	0.1044	9.4528×10^{-6}	0.1096	3.5205×10^{-5}
0.1905	0.0508	0.1941	0.0572	0.1194	5.3616×10^{-7}	0.1252	2.2811×10^{-6}
0.1935	0.0303	0.1958	0.0344	0.1344	2.3816×10^{-8}	0.1401	1.1341×10^{-7}
0.1965	0.0094	0.1978	0.0107	0.1494	8.4309×10^{-10}	0.1535	4.3979×10^{-9}
t=451 $P(C_{t,2}) = 0.1603$				t=451 $P(C_{t,2}) = 0.0222$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1688	0.4056	0.1779	0.4273	0.0272	0.3423	0.0342	0.4299
0.1718	0.3423	0.1794	0.3638	0.0422	0.0465	0.0491	0.0839
0.1748	0.1984	0.1838	0.2160	0.0572	0.0045	0.0659	0.0109
0.1778	0.1843	0.1843	0.2012	0.0722	5.2734×10^{-4}	0.0804	0.0016
0.1808	0.1206	0.1874	0.1339	0.0872	4.5577×10^{-5}	0.0958	1.6039×10^{-4}
0.1838	0.0756	0.1903	0.0852	0.1022	7.6165×10^{-6}	0.1065	2.9786×10^{-5}
0.1868	0.0465	0.1928	0.0531	0.1172	4.1818×10^{-7}	0.1221	1.8758×10^{-6}
0.1898	0.0465	0.1928	0.0531	0.1322	1.7978×10^{-8}	0.1373	9.0700×10^{-8}
0.1928	0.0275	0.1948	0.0317	0.1472	6.1587×10^{-10}	0.1510	3.4150×10^{-9}
0.1958	0.0084	0.1974	0.0099	0.1622	4.9408×10^{-12}	0.1662	3.0161×10^{-11}
t=452 $P(C_{t,2}) = 0.1949$				t=452 $P(C_{t,2}) = 0.0513$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1364	0.3840	0.1487	0.4186	0.0143	0.3193	0.0190	0.4227
0.1424	0.2763	0.1528	0.3094	0.0293	0.0120	0.0357	0.0298
0.1484	0.1657	0.1587	0.1929	0.0443	7.1593×10^{-4}	0.0504	0.0025
0.1544	0.1039	0.1640	0.1249	0.0593	5.8834×10^{-5}	0.0635	2.6065×10^{-4}
0.1604	0.0644	0.1689	0.0798	0.0743	3.5853×10^{-6}	0.0782	1.9553×10^{-5}
0.1664	0.0384	0.1736	0.0489	0.0893	1.6612×10^{-7}	0.0937	1.0853×10^{-6}
0.1724	0.0219	0.1779	0.0286	0.1043	5.9800×10^{-9}	0.1098	4.5790×10^{-8}
0.1784	0.0063	0.1854	0.0085	0.1193	1.7041×10^{-10}	0.1255	1.4916×10^{-9}
0.1844	0.0032	0.1886	0.0044	0.1343	3.9074×10^{-12}	0.1405	3.8277×10^{-11}
0.1904	7.1587×10^{-4}	0.1940	0.0010	0.1493	7.3124×10^{-14}	0.1539	7.8455×10^{-13}
0.1964	1.3876×10^{-4}	0.1972	2.0059×10^{-4}	0.1643	2.7066×10^{-16}	0.1683	3.1760×10^{-15}
t=453 $P(C_{t,2}) = 0.1896$				t=453 $P(C_{t,2}) = 0.0955$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1789	0.4200	0.1859	0.4365	0.0339	0.3650	0.0413	0.4449
0.1809	0.2842	0.1884	0.2995	0.0489	0.0601	0.0566	0.1004
0.1829	0.2477	0.1893	0.2622	0.0639	0.0121	0.0696	0.0248
0.1849	0.1618	0.1918	0.1735	0.0789	0.0017	0.0842	0.0043
0.1869	0.1506	0.1922	0.1618	0.0939	1.8007×10^{-4}	0.0997	5.2957×10^{-4}
0.1889	0.0971	0.1943	0.1055	0.1089	1.4137×10^{-5}	0.1155	4.8156×10^{-5}
0.1909	0.0952	0.1944	0.1035	0.1239	8.5245×10^{-7}	0.1309	3.2928×10^{-6}
0.1929	0.0598	0.1960	0.0656	0.1389	4.0269×10^{-8}	0.1452	1.7250×10^{-7}
0.1949	0.0358	0.1971	0.0394	0.1539	1.5163×10^{-9}	0.1580	7.0689×10^{-9}
0.1969	0.0197	0.1979	0.0218	0.1689	1.3806×10^{-11}	0.1713	6.9776×10^{-11}

Table 3.8: Operating characteristics at time points $t = 454, \ldots, 457$ for the Glaxosmithkline series.

$u = Q_{0.39} = 13$				$u = Q_{0.50} = 18$			
t=454				t=454			
$P(C_{t,2}) = 0.1982$				$P(C_{t,2}) = 0.0639$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1969	0.7558	0.1979	0.7594	0.0668	0.3766	0.0776	0.4378
0.1970	0.7558	0.1979	0.7594	0.0768	0.1419	0.0887	0.1884
0.1971	0.7558	0.1979	0.7594	0.0868	0.0589	0.0985	0.0868
0.1972	0.7558	0.1979	0.7594	0.0968	0.0220	0.1085	0.0357
0.1973	0.6452	0.1980	0.6487	0.1068	0.0127	0.1135	0.0217
0.1974	0.6452	0.1980	0.6487	0.1168	0.0071	0.1184	0.0127
0.1975	0.5482	0.1981	0.5513	0.1268	0.0010	0.1334	0.0021
0.1976	0.4897	0.1981	0.4927	0.1368	2.4113×10^{-4}	0.1427	5.1524×10^{-4}
0.1977	0.4897	0.1981	0.4927	0.1468	5.0216×10^{-5}	0.1514	1.1385×10^{-4}
0.1978	0.3786	0.1982	0.3811	0.1568	9.3408×10^{-6}	0.1595	2.2310×10^{-5}
0.1979	0.3786	0.1982	0.3811	0.1668	6.1357×10^{-7}	0.1694	1.5565×10^{-6}
0.1980	0.3786	0.1982	0.3811	0.1768	1.1486×10^{-8}	0.1789	3.0764×10^{-8}
0.1981	0.2687	0.1982	0.2705				
0.1982	0.1820	0.1983	0.1832				
t=455				t=455			
$P(C_{t,2}) = 0.1982$				$P(C_{t,2}) = 0.0645$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1878	0.5080	0.1921	0.5194	0.0428	0.3563	0.0519	0.4316
0.1888	0.3556	0.1938	0.3669	0.0528	0.1230	0.0619	0.1777
0.1898	0.3556	0.1938	0.3669	0.0628	0.0499	0.0705	0.0822
0.1908	0.3087	0.1943	0.3194	0.0728	0.0176	0.0799	0.0328
0.1918	0.2069	0.1958	0.2156	0.0828	0.0053	0.0897	0.0112
0.1928	0.2055	0.1958	0.2142	0.0928	0.0014	0.1000	0.0033
0.1938	0.1887	0.1960	0.1969	0.1028	3.2288×10^{-4}	0.1104	8.3267×10^{-4}
0.1948	0.1238	0.1970	0.1298	0.1128	6.5477×10^{-5}	0.1207	1.8461×10^{-4}
0.1958	0.1177	0.1971	0.1235	0.1228	1.1777×10^{-5}	0.1310	3.6019×10^{-5}
0.1968	0.0700	0.1978	0.0737	0.1328	1.8903×10^{-6}	0.1407	6.2086×10^{-6}
0.1978	0.0324	0.1982	0.0342	0.1428	2.7232×10^{-7}	0.1496	9.5105×10^{-7}
t=456				t=456			
$P(C_{t,2}) = 0.1827$				$P(C_{t,2}) = 0.1066$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1882	0.5088	0.1923	0.5200	0.0555	0.3597	0.0657	0.4254
0.1892	0.3565	0.1940	0.3675	0.0655	0.1115	0.0783	0.1572
0.1902	0.3558	0.1940	0.3668	0.0755	0.0444	0.0879	0.0702
0.1912	0.3100	0.1945	0.3204	0.0855	0.0268	0.0928	0.0448
0.1922	0.2081	0.1959	0.2166	0.0955	0.0088	0.1029	0.0163
0.1932	0.2067	0.1959	0.2152	0.1055	0.0025	0.1131	0.0051
0.1942	0.1901	0.1961	0.1981	0.1155	6.2622×10^{-4}	0.1233	0.0014
0.1952	0.1223	0.1972	0.1282	0.1255	1.3787×10^{-4}	0.1333	3.3099×10^{-4}
0.1962	0.1188	0.1972	0.1245	0.1355	2.6937×10^{-5}	0.1428	6.9250×10^{-5}
0.1972	0.0708	0.1979	0.0744	0.1455	4.6989×10^{-6}	0.1515	1.2822×10^{-5}
0.1982	0.0327	0.1982	0.0345	0.1555	7.3593×10^{-7}	0.1597	2.1158×10^{-6}
t=457				t=457			
$P(C_{t,2}) = 0.1831$				$P(C_{t,2}) = 0.1061$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1961	0.7248	0.1979	0.7317	0.0758	0.3933	0.0866	0.4491
0.1963	0.7248	0.1979	0.7317	0.0858	0.1707	0.0964	0.2169
0.1965	0.7248	0.1979	0.7317	0.0958	0.0715	0.1065	0.1004
0.1967	0.7248	0.1979	0.7317	0.1058	0.0279	0.1163	0.0428
0.1969	0.6845	0.1980	0.6913	0.1158	0.0095	0.1261	0.0157
0.1971	0.6845	0.1980	0.6913	0.1258	0.0028	0.1358	0.0050
0.1973	0.6067	0.1981	0.6129	0.1358	0.0015	0.1404	0.0027
0.1975	0.5498	0.1982	0.5557	0.1458	1.6950×10^{-4}	0.1535	3.4307×10^{-4}
0.1977	0.5498	0.1982	0.5557	0.1558	3.4913×10^{-5}	0.1612	7.4249×10^{-5}
0.1979	0.5498	0.1982	0.5557	0.1658	2.6603×10^{-6}	0.1708	5.9917×10^{-6}
0.1981	0.4761	0.1982	0.4814	0.1758	1.6104×10^{-7}	0.1779	3.7794×10^{-7}
0.1983	0.1964	0.1983	0.1987				

Table 3.9: Operating characteristics at time points $t = 458, \ldots, 460$ for the Glaxosmithkline series.

$u = Q_{0.39} = 13$				$u = Q_{0.50} = 18$			
t=458 $P(C_{t,2}) = 0.1884$				t=458 $P(C_{t,2}) = 0.0977$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1962	0.7517	0.1979	0.7584	0.0754	0.3925	0.0862	0.4485
0.1964	0.7256	0.1980	0.7324	0.0854	0.1697	0.0960	0.2160
0.1966	0.7256	0.1980	0.7324	0.0954	0.0709	0.1062	0.0998
0.1968	0.7256	0.1980	0.7324	0.1054	0.0276	0.1159	0.0424
0.1970	0.6852	0.1980	0.6917	0.1154	0.0093	0.1257	0.0156
0.1972	0.6852	0.1980	0.6917	0.1254	0.0028	0.1355	0.0050
0.1974	0.6075	0.1981	0.6136	0.1354	0.0014	0.1401	0.0027
0.1976	0.5504	0.1982	0.5561	0.1454	1.6640×10^{-4}	0.1532	3.3802×10^{-4}
0.1978	0.5504	0.1982	0.5561	0.1554	3.4203×10^{-5}	0.1610	7.3031×10^{-5}
0.1980	0.4764	0.1982	0.4814	0.1654	6.2986×10^{-6}	0.1676	1.4003×10^{-6}
0.1982	0.3864	0.1983	0.3905	0.1754	1.5678×10^{-7}	0.1778	3.6973×10^{-7}
t=459 $P(C_{t,2}) = 0.1960$				t=459 $P(C_{t,2}) = 0.0810$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1969	0.8451	0.1979	0.8491	0.0685	0.3796	0.0794	0.4398
0.1970	0.8451	0.1979	0.8491	0.0785	0.1459	0.0903	0.1921
0.1971	0.7429	0.1980	0.7468	0.0885	0.0612	0.1001	0.0894
0.1972	0.7429	0.1980	0.7468	0.0985	0.0231	0.1100	0.0370
0.1973	0.7429	0.1980	0.7468	0.1085	0.0134	0.1150	0.0225
0.1974	0.6834	0.1980	0.6872	0.1185	0.0041	0.1249	0.0075
0.1975	0.6661	0.1980	0.6699	0.1285	0.0011	0.1348	0.0022
0.1976	0.6661	0.1980	0.6699	0.1385	2.6091×10^{-4}	0.1440	5.4834×10^{-4}
0.1977	0.5656	0.1981	0.5690	0.1485	5.4834×10^{-5}	0.1527	1.2216×10^{-4}
0.1978	0.4682	0.1982	0.4711	0.1585	4.2826×10^{-6}	0.1641	1.0255×10^{-5}
0.1979	0.4389	0.1982	0.4418	0.1685	2.6415×10^{-7}	0.1731	6.6722×10^{-7}
0.1980	0.4389	0.1982	0.4418	0.1785	1.3076×10^{-8}	0.1793	3.4214×10^{-8}
0.1981	0.3466	0.1982	0.3490				
0.1982	0.2279	0.1983	0.2295				
t=460 $P(C_{t,2}) = 0.1018$				t=460 $P(C_{t,2}) = 0.1739$			
k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	k	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.1951	0.4857	0.1971	0.4908	0.0433	0.3572	0.0523	0.4322
0.1954	0.4857	0.1971	0.4908	0.0533	0.1243	0.0623	0.1789
0.1957	0.4857	0.1971	0.4908	0.0633	0.0506	0.0710	0.0830
0.1960	0.4138	0.1973	0.4186	0.0733	0.0179	0.0803	0.0332
0.1963	0.3963	0.1974	0.4010	0.0833	0.0054	0.0902	0.0113
0.1966	0.3154	0.1976	0.3195	0.0933	0.0014	0.1005	0.0033
0.1969	0.2615	0.1978	0.2652	0.1033	3.3114×10^{-4}	0.1109	8.4899×10^{-4}
0.1972	0.2615	0.1978	0.2652	0.1133	6.7361×10^{-5}	0.1212	1.8872×10^{-4}
0.1975	0.1986	0.1980	0.2016	0.1233	1.2153×10^{-5}	0.1314	3.6923×10^{-5}
0.1978	0.1325	0.1981	0.1346	0.1333	1.9569×10^{-6}	0.1411	6.3811×10^{-6}
0.1981	0.1098	0.1982	0.1116	0.1433	2.8279×10^{-7}	0.1500	9.8026×10^{-7}

Table 3.10: Astrazeneca Series: operating characteristics at different time points, with Criteria 1 and 2, considering $u = Q_{0.39} = 19$.

Criterion	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
1	450	0.0655	0.0855	$\{0,\ldots,6\}\cup\{22,\ldots,47\}$	0.0853	0.1037	0.1350
	451	0.0604	0.0804	$\{0,\ldots,5\}\cup\{21,\ldots,48\}$	0.1060	0.0944	0.1656
	452	0.0381	0.0581	$\{0,\ldots,3\}\cup\{20,\ldots,53\}$	0.0721	0.0715	0.1353
	453	0.0388	0.0588	$\{0,\ldots,3\}\cup\{20,\ldots,53\}$	0.0744	0.0721	0.1381
	454	0.0575	0.0775	$\{0,\ldots,5\}\cup\{21,\ldots,49\}$	0.0973	0.0924	0.1564
	455	0.1085	0.1185	$\{0,\ldots,11\}\cup\{23,\ldots,41\}$	0.2127	0.1323	0.2594
	456	0.1520	0.1560	$\{5,\ldots,15\}\cup\{25,\ldots,31\}$	0.3300	0.1592	0.3456
	457	0.1156	0.1256	$\{0,\ldots,11\}\cup\{23,\ldots,39\}$	0.2318	0.1372	0.2751
	458	0.1345	0.1420	$\{0,\ldots,13\}\cup\{24,\ldots,36\}$	0.2818	0.1498	0.3138
	459	0.1485	0.1525	$\{1,\ldots,15\}\cup\{25,\ldots,32\}$	0.3267	0.1576	0.3467
	460	0.1094	0.1194	$\{0,\ldots,11\}\cup\{23,\ldots,41\}$	0.2148	0.1330	0.2611
2	450	0.0655	0.1455	$\{30,\ldots,39\}$	9.2971×10^{-4}	0.1518	0.0022
	451	0.0604	0.1404	$\{29,\ldots,41\}$	0.0013	0.1455	0.0032
	452	0.0381	0.1181	$\{28,\ldots,45\}$	5.5679×10^{-4}	0.1283	0.0019
	453	0.0388	0.1188	$\{28,\ldots,45\}$	5.9269×10^{-4}	0.1287	0.0020
	454	0.0575	0.1375	$\{29,\ldots,41\}$	0.0011	0.1443	0.0028
	455	0.1085	0.1585	$\{30,\ldots,34\}$	0.0074	0.1603	0.0109
	456	0.1520	0.1610	$\{9,\ldots,12\}\cup\{28,29\}$	0.0738	0.1615	0.0784
	457	0.1156	0.1606	$\{31,32\}$	0.0042	0.1616	0.0058
	458	0.1345	0.1595	$\{0,\ldots,5\}\cup\{29,\ldots,32\}$	0.0310	0.1611	0.0371
	459	0.1485	0.1615	$\{8,9\}\cup\{29\}$	0.0232	0.1617	0.0252
	460	0.1094	0.1594	$\{30,\ldots,33\}$	0.0073	0.1604	0.0107

Table 3.11: Astrazeneca Series: operating characteristics at different time points, with Criteria 1 and 2, considering u=$Q_{0.50}$= 25.

Criterion	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
	450	0.0036	0.0136	$\{26,\ldots,82\}$	0.0113	0.0177	0.0549
	451	0.0031	0.0131	$\{26,\ldots,82\}$	0.0089	0.0166	0.0475
	452	0.0013	0.0113	$\{27,\ldots,85\}$	0.0012	0.0141	0.0124
	453	0.0014	0.0114	$\{27,\ldots,84\}$	0.0012	0.0143	0.0129
	454	0.0028	0.0128	$\{26,\ldots,82\}$	0.0077	0.0160	0.0436
1	455	0.0111	0.0311	$\{0\}\cup\{29,\ldots,73\}$	0.0131	0.0391	0.0463
	456	0.0325	0.0525	$\{0,\ldots,8\}\cup\{30,\ldots,65\}$	0.0534	0.0647	0.1063
	457	0.0130	0.0330	$\{0,1\}\cup\{29,\ldots,72\}$	0.0173	0.0416	0.0553
	458	0.0202	0.0402	$\{0,\ldots,4\}\cup\{29,\ldots,69\}$	0.0360	0.0494	0.0879
	459	0.0291	0.0491	$\{0,\ldots,7\}\cup\{30,\ldots,66\}$	0.0438	0.0619	0.0932
	460	0.0113	0.0313	$\{0\}\cup\{29,\ldots,73\}$	0.0136	0.0394	0.0473
	450	0.0036	0.0336	$\{32,\ldots,74\}$	2.2149×10^{-4}	0.0379	0.0023
	451	0.0031	0.0231	$\{30,\ldots,78\}$	6.7534×10^{-4}	0.0285	0.0062
	452	0.0013	0.0113	$\{27,\ldots,85\}$	0.0012	0.0141	0.0124
	453	0.0014	0.0214	$\{31,\ldots,80\}$	5.5180×10^{-5}	0.0256	0.0010
	454	0.0028	0.0228	$\{30,\ldots,78\}$	5.5741×10^{-4}	0.0277	0.0055
2	455	0.0111	0.0611	$\{35,\ldots,66\}$	3.5349×10^{-4}	0.0676	0.0022
	456	0.0325	0.1025	$\{39,\ldots,56\}$	5.2170×10^{-4}	0.1087	0.0017
	457	0.0130	0.0630	$\{35,\ldots,65\}$	$.53314\times10^{-4}$	0.0704	0.0029
	458	0.0202	0.0802	$\{37,\ldots,61\}$	4.5795×10^{-4}	0.0893	0.0020
	459	0.0291	0.0991	$\{39,\ldots,57\}$	3.6816×10^{-4}	0.1065	0.0013
	460	0.0113	0.0613	$\{35,\ldots,66\}$	3.7183×10^{-4}	0.0679	0.0022

Table 3.12: Glaxosmithkline Series: operating characteristics at different time points, with Criteria 1 and 2, considering $u = Q_{0.39} = 13$.

Criterion	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
1	450	0.1725	0.1755	$\{0,\ldots,8\}\cup\{16,\ldots,33\}$	0.3495	0.1822	0.3691
	451	0.1688	0.1718	$\{0,\ldots,8\}\cup\{16,\ldots,35\}$	0.3423	0.1794	0.3638
	452	0.1364	0.1424	$\{0,\ldots,6\}\cup\{15,\ldots,44\}$	0.2763	0.1528	0.3094
	453	0.1789	0.1789	$\{0,\ldots,9\}\cup\{16,\ldots,31\}$	0.4200	0.1859	0.4365
	454	0.1969	0.1979	$\{9,\ldots,12\}\cup\{17,18\}$	0.3786	0.1982	0.3811
	455	0.1878	0.1888	$\{0,\ldots,9\}\cup\{17,\ldots,27\}$	0.3556	0.1938	0.3669
	456	0.1882	0.1892	$\{0,\ldots,9\}\cup\{17,\ldots,27\}$	0.3565	0.1940	0.3675
	457	0.1961	0.1981	$\{13,\ldots,17\}$	0.4761	0.1982	0.4814
	458	0.1962	0.1982	$\{13,\ldots,16\}$	0.3864	0.1983	0.3905
	459	0.1969	0.1979	$\{10,\ldots,13\}\cup\{17,18\}$	0.4389	0.1982	0.4418
	460	0.1951	0.1963	$\{1,\ldots,10\}\cup\{17,\ldots,21\}$	0.3963	0.1974	0.4010
2	450	0.1725	0.1965	$\{23,\ldots,27\}$	0.0094	0.1978	0.0107
	451	0.1688	0.1958	$\{23,\ldots,28\}$	0.0084	0.1974	0.0099
	452	0.1364	0.1904	$\{25,\ldots,34\}$	7.1587×10^{-4}	0.1940	0.0010
	453	0.1789	0.1969	$\{22,\ldots,25\}$	0.0197	0.1979	0.0218
	454	0.1969	0.1982	$\{10,11\}\cup\{18\}$	0.1820	0.1983	0.1832
	455	0.1878	0.1978	$\{0,1\}\cup\{21,22\}$	0.0324	0.1982	0.0342
	456	0.1882	0.1982	$\{0\}\cup\{21,22\}$	0.0327	0.1982	0.0345
	457	0.1961	0.1983	$\{14\}\cup\{16\}$	0.1964	0.1983	0.1987
	458	0.1962	0.1982	$\{13,\ldots,16\}$	0.3864	0.1983	0.3905
	459	0.1969	0.1982	$\{11,12\}\cup\{17\}$	0.2279	0.1983	0.2295
	460	0.1951	0.1981	$\{4,\ldots,7\}\cup\{19,20\}$	0.1098	0.1982	0.1116

Table 3.13: Glaxosmithkline Series: operating characteristics at different time points, with Criteria 1 and 2, considering u=$Q_{0.50}$= 18.

Criterion	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
	450	0.0294	0.0444	$\{0\} \cup \{20, \ldots, 93\}$	0.0509	0.0509	0.0893
	451	0.0272	0.0422	$\{0\} \cup \{20, \ldots, 95\}$	0.0465	0.0491	0.0839
	452	0.0143	0.0293	$\{21, \ldots, 100\}$	0.0120	0.0357	0.0298
	453	0.0339	0.0489	$\{0, 1\} \cup \{20, \ldots, 91\}$	0.0601	0.0566	0.1004
	454	0.0668	0.0768	$\{0, \ldots, 6\} \cup \{20, \ldots, 76\}$	0.1419	0.0887	0.1884
1	455	0.0428	0.0528	$\{0, \ldots, 4\} \cup \{19, \ldots, 87\}$	0.1230	0.0619	0.1777
	456	0.0433	0.0533	$\{0, \ldots, 4\} \cup \{19, \ldots, 87\}$	0.1243	0.0623	0.1789
	457	0.0758	0.0858	$\{0, \ldots, 7\} \cup \{20, \ldots, 73\}$	0.1707	0.0964	0.2169
	458	0.0754	0.0854	$\{0, \ldots, 7\} \cup \{20, \ldots, 73\}$	0.1697	0.0960	0.2160
	459	0.0685	0.0785	$\{0, \ldots, 6\} \cup \{20, \ldots, 76\}$	0.1459	0.0903	0.1921
	460	0.0555	0.0655	$\{0, \ldots, 5\} \cup \{20, \ldots, 81\}$	0.1115	0.0783	0.1572
	450	0.0294	0.0744	$\{27, \ldots, 84\}$	6.2032×10^{-4}	0.0834	0.0018
	451	0.0272	0.0722	$\{27, \ldots, 85\}$	5.2734×10^{-4}	0.0804	0.0016
	452	0.0143	0.0443	$\{25, \ldots, 98\}$	7.1593×10^{-4}	0.0504	0.0025
	453	0.0339	0.0789	$\{26, \ldots, 82\}$	0.0017	0.0842	0.0043
	454	0.0668	0.1268	$\{29, \ldots, 65\}$	0.0010	0.1334	0.0021
2	455	0.0428	0.0928	$\{27, \ldots, 77\}$	0.0014	0.1000	0.0033
	456	0.0433	0.0933	$\{27, \ldots, 77\}$	0.0014	0.1005	0.0033
	457	0.0758	0.1358	$\{29, \ldots, 61\}$	0.0015	0.1404	0.0027
	458	0.0754	0.1354	$\{29, \ldots, 62\}$	0.0014	0.1401	0.0027
	459	0.0685	0.1285	$\{29, \ldots, 64\}$	0.0011	0.1348	0.0022
	460	0.0555	0.1155	$\{29, \ldots, 69\}$	6.2622×10^{-4}	0.1233	0.0014

Table 3.14: Results for the Astrazeneca series, with $u = Q_{0.39} = 19$. Percentages in parenthesis.

Time instant	Criterion	Alarms		Catastrophes	
		False	Total	Detected	Total
$t = 456$	1	21011 (0.6369)	32992	11981 (0.3755)	31906
	2	4381 (0.5886)	7443	3062 (0.0948)	32315
$t = 457$	1	17464 (0.7505)	23271	5807 (0.3105)	18705
	2	249 (0.5818)	428	179 (0.0095)	18761
$t = 458$	1	19618 (0.6958)	28193	8575 (0.3523)	24340
	2	1820 (0.5938)	3065	1245 (0.0504)	24713
$t = 459$	1	20963 (0.6449)	32508	11545 (0.3798)	30396
	2	1417 (0.5984)	2368	951 (0.0313)	30389
$t = 460$	1	16254 (0.7655)	21233	4979 (0.2914)	17089
	2	464 (0.6097)	761	297 (0.0170)	17433

Table 3.15: Results for the Astrazeneca series, with $u = Q_{0.50} = 25$. Percentages in parenthesis.

Time instant	Criterion	Alarms		Catastrophes	
		False	Total	Detected	Total
$t = 456$	1	4855 (0.9063)	5357	502 (0.1190)	4220
	2	38 (0.7917)	48	10 (0.0024)	4189
$t = 457$	1	1550 (0.9394)	1650	100 (0.0641)	1560
	2	43 (0.8431)	51	8 (0.0052)	1544
$t = 458$	1	3377 (0.9365)	3606	229 (0.0904)	2534
	2	48 (0.8136)	59	11 (0.0043)	2532
$t = 459$	1	3963 (0.9081)	4364	401 (0.1072)	3739
	2	30 (0.9091)	33	3 (0.0008)	3748
$t = 460$	1	1302 (0.9518)	1368	66 (0.0509)	1297
	2	44 (0.8627)	51	7 (0.0054)	1303

Table 3.16: Results for the Glaxosmithkline series, with $u = Q_{0.39} = 13$. Percentages in parenthesis.

Time instant	Criterion	Alarms		Catastrophes	
		False	Total	Detected	Total
$t = 454$	1	20607 (0.5452)	37794	17187 (0.3786)	45399
	2	9873 (0.5449)	18119	8246 (0.1819)	45340
$t = 457$	1	25776 (0.5397)	47761	21985 (0.4496)	48898
	2	10609 (0.5401)	19641	9032 (0.1848)	48867
$t = 458$	1	21062 (0.5429)	38795	17733 (0.3638)	48742
	2	21062 (0.5429)	38795	17733 (0.3638)	48742
$t = 459$	1	24280 (0.5499)	44152	19872 (0.4338)	45814
	2	12447 (0.5510)	22589	10142 (0.2198)	46145
$t = 460$	1	22415 (0.5663)	39583	17168 (0.4169)	41183
	2	5918 (0.5408)	10944	5026 (0.1226)	41006

Table 3.17: Results for the Glaxosmithkline series, with $u = Q_{0.50} = 18$. Percentages in parenthesis.

Time instant	Criterion	Alarms		Catastrophes	
		False	Total	Detected	Total
$t = 454$	1	12571 (0.8819)	14254	1683 (0.1988)	8464
	2	77 (0.8105)	95	18 (0.0022)	8120
$t = 457$	1	14895 (0.8718)	17085	2190 (0.2279)	9609
	2	125 (0.8013)	156	31 (0.0032)	9726
$t = 458$	1	14801 (0.8719)	16976	2175 (0.2263)	9611
	2	109 (0.7956)	137	28 (0.0030)	9430
$t = 459$	1	12749 (0.8775)	14529	1780 (0.2052)	8675
	2	98 (0.8305)	118	20 (0.0023)	8601
$t = 460$	1	9935 (0.9003)	11035	1100 (0.1678)	6554
	2	58 (0.8788)	66	8 (0.0012)	6521

Chapter 4

Conclusions and Future Directions of Research

This thesis focuses on the analysis of non-linear time series with emphasis on the application of optimal alarm systems and on the development of observation-driven models to address particular features commonly observed in time series of counts.

In Chapter 2, models with conditional heteroscedasticity were considered and the FIAPARCH(p, d, q) model was given particular attention. Estimation procedures were implemented following classical and Bayesian approaches. Under the classical perspective, only the Quasi-Maximum Likelihood Estimation procedure assuming conditional normality was used. Although QMLE standard errors obtained were slightly higher than standard deviations obtained under the Bayesian perspective, parameter estimates seem satisfactory. An optimal alarm system was constructed for the FIAPARCH$(1, d, 1)$ model and expressions for the alarm characteristics of the alarm system were given. Two criteria were tested and a choice could be made regarding the optimization of the operating characteristics, as we could conclude that, overall, Criterion 1 provided better estimates. The alarm system was tested in a frequencist perspective and probabilities of correct alarm and of detecting the event of around 17% and 21% were reached, respectively. The on-line alarm system was also implemented and the adaptive behaviour of the alarm system could be observed. After the simulation study an application was made to a particular real data series, the IBOVESPA returns data set, containing the daily returns of the São Paulo Stock Market during a fourteen year period. An optimal alarm system was constructed, taking as the event of interest the down-crossing of a particular level. Once again, the advantages of the on-line implementation where the past and present experiments are updated at each time point, became clear. The system adapts itself in order to produce a minimum number of false alarms allowing the probability of correct alarm to be near one. On the other hand, as few alarms were given, the detection probability never exceeded 30%.

In the second part of the work, we turned our attention to non-linear models used in the analysis of time series of counts. In Chapter 3, two fundamentally different goals were pursued. The implementation of an optimal alarm system was carried out for the INAPARCH(1,1) model and all the expressions for the alarm characteristics were obtained. An application was made concerning two integer-valued data series of the daily number of transactions in the Astrazeneca and Glaxosmithkline stocks. The event of interest was in both situations considered as the up-crossing of a fixed level u. Both percentiles 39 and 50 were considered as a fixed level. The designation of catastrophe may not be as adequate as in the application in Chapter 2, as these quantiles do not exactly represent rare events. Anyway, as these particular data series present many zero counts, the probability of catastrophe should be negligible if we were to consider higher percentiles as really rare events. As a consequence, the operating characteristics of the alarm system would not be satisfactory. This is an issue we would like to explore in future work considering the application of alarm systems to other real data time series exhibiting a significant number of zero counts.

Considering the behaviour of the alarm system above described, we would like to point out a few conclusions. Firstly, the adaptation of the criteria used in Chapter 2 to the integer-valued case in Chapter 3 seems satisfactory, as very reasonable operating characteristics were obtained. Overall, the first criterion used provided better estimates. The testing of the alarm system in the case were the theoretical operating characteristics were evaluated through frequency counts of the number of false alarms and undetected catastrophes, had a similar outcome to that of Chapter 2: frequency estimates tend to overestimate the theoretical values, considering both time series, both fixed levels $u = Q_{0.39}$ and $u = Q_{0.50}$ and both operating characteristics, the probability of correct alarm and the probability of detection. Better operating characteristics were obtained for the up-crossing of the 39^{th} percentile, for both time series, and once again, we suspect this behaviour is related to the nature of the particular time series used. Further developments should help to shed some light on these matters.

The second aforementioned goal pursued in Chapter 3 relates to the modelling of time series of counts. The presence of asymmetric overdispersion in data had never been addressed for integer-valued time series. As an innovative contribution of this thesis in this field of study, we propose the INAPARCH model as an integer-valued counterpart of the APARCH model, a well known model in the analysis of real-valued financial time series. The establishment of stationarity and ergodicity of the INAPARCH process is addressed by employing Markov chain theory and the concept of $\tau-$weak dependence by Dedecker and Prieur (2004). Sufficient conditions for the ergodicity of the INAPARCH(1,1) were obtained making use of the previous concepts. Nevertheless, an important issue that we expect to develop in further research is the establishment of necessary instead of sufficient conditions. A less restrictive parameter space should be useful in order to address wider practical applications.

The conditional Maximum Likelihood Estimation method was successfully applied to parameter estimation of the proposed model. Necessary calculations and asymptotic theory were straightforwardly developed for the INAPARCH(1,1) model. A simulation study was included in Chapter 3 in order to illustrate the methodology. The results deserve a few comments, the first of which is the reinforcement of the necessity to develop necessary conditions for ergodicity: five different parameter sets were considered in the simulations and the last one was intentionally chosen in order to be outside the admissible parameter range. Empirical results of preliminary data analysis pointed towards stability of the underlying process, for all cases considered. Also in the application section, Section 3.6, in which the conditional MLE procedure was applied to estimate two time series of tick-by-tick data generated from stock transactions, the estimated δ parameter fails the $\delta \geq 2$ condition for both data series. This did not come as a surprise as the same condition was violated in the original paper by Ding et al. (1993), where the real-valued APARCH model was adjusted to the Standard & Poor 500 stock market daily closing price index and the $\delta \geq 2$ condition was just a sufficient condition for the process to be covariance stationary. Although the results obtained with the conditional MLE method were satisfactory, it may be worthwhile to consider other estimation procedures in the future. For instance, it is known that the Autoregressive Conditional Poisson specification can also be estimated by a Bayesian posterior analysis using the Gibbs sampling scheme proposed by Bauwens and Lubrano (1998) for GARCH-type models. Also, the Quasi-Maximum Likelihood Estimation procedure, used in Chapter 2 for the real-valued FIAPARCH(p, d, q) model could be implemented for the integer-valued INAPARCH(p, q) case, considering the Poisson QML estimator and eventually result in enhanced robustness. As Wooldridge (1999) points out, specifying a distribution that depends on both conditional mean parameters and additional parameters, and then maximizing the likelihood with respect to all parameters, generally produces inconsistent estimates of the conditional mean parameters if some aspects of the true distribution are misspecified. This misleading inference could be avoided by allowing the density in the log-likelihood function to be different from the actual conditional density of the count variable, Y_t.

A final remark related to the application in Section 3.6, and already discussed therein, remains to be done about the estimated value of the γ parameter: it was negative for both real data time series, leading us to the conclusion that positive shocks should have stronger impact on overdispersion than negative ones. The existence of the leverage effect is not supported by the data. In the paper by Tse (1998), the conditional heteroscedasticity of the yen-dollar exchange rate was modelled, amongst others, through the APARCH and the FIAPARCH representations for the volatility. It was found that the asymmetry parameter, γ, though having positive estimates, could not be considered statistically significant. This results were afterwards contradicted by Tsui and Ho (2004), who found evidence of asymmetric volatility in some data series of daily returns of currencies measured against the dollar or the yen (with both positive and negative estimates for the γ parameter) being the asymmetric effects more significant for currencies measured against the yen. As future work we expect to apply the INAPARCH(p, q) in the analysis of a few more real data financial integer-valued time series. Different conclusions related to the presence of asymmetric overdispersion relative to the conditional mean of the process are expected to be drawn.

Appendix A

Additional Results for Conditional MLE

A.1 Estimation

For illustration purposes, the third derivative of $\ell_t(\boldsymbol{\theta})$ in order to ω is presented next. Other third order derivatives are easily obtained and the conclusion that all third derivatives are bounded by a sequence that converges in probability follows straightforwardly.

$$
\begin{aligned}
\frac{\partial^3 \ell_t(\boldsymbol{\theta})}{\partial \omega^3} &= \left(\frac{1-\delta}{\lambda_t^{\delta+1}} + \frac{(\delta+1)y_t}{\lambda_t^{\delta+2}}\right) \times \\
&\times \left(1+\delta(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \omega}\right)\left(\frac{\partial \lambda_t}{\partial \omega}\right)^2 + \\
&+ \left(\frac{\delta-1}{\lambda_t^{\delta}} - \frac{\delta y_t}{\lambda_t^{\delta+1}}\right)\left\{2(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial^2 \lambda_{t-1}}{\partial \omega^2}\frac{\partial \lambda_t}{\partial \omega} + \right. \\
&+ 2\left(\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2} + \beta(\delta-1)\lambda_{t-1}^{\delta-2}\right)\left(\frac{\partial \lambda_{t-1}}{\partial \omega}\right)^2\frac{\partial \lambda_t}{\partial \omega} + \\
&+ \left.\left(\frac{1}{\delta} + (\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial \lambda_{t-1}}{\partial \omega}\right)\frac{\partial^2 \lambda_t}{\partial \omega^2}\right\} +
\end{aligned}
$$

$$
\begin{aligned}
+ & \left(\frac{y_t}{\lambda_t^{\delta}} - \frac{1}{\lambda_t^{\delta-1}}\right)\Bigg\{(\alpha(I_{t-1}+\gamma)g_{t-1}^{\delta-1} + \beta\lambda_{t-1}^{\delta-1})\frac{\partial^3\lambda_{t-1}}{\partial\omega^3} + \\
+ & \ 3\left(\alpha(\delta-1)(I_{t-1}+\gamma)^2 g_{t-1}^{\delta-2} + \beta(\delta-1)\lambda_{t-1}^{\delta-2}\right)\frac{\partial\lambda_{t-1}}{\partial\omega}\frac{\partial^2\lambda_{t-1}}{\partial\omega^2} + \\
+ & \left(\alpha(\delta-1)(\delta-2)(I_{t-1}+\gamma)^3 g_{t-1}^{\delta-3} + \beta(\delta-1)(\delta-2)\lambda_{t-1}^{\delta-3}\right)\left(\frac{\partial\lambda_{t-1}}{\partial\omega}\right)^3\Bigg\}.
\end{aligned}
$$

A.2 Conditional ML estimation of the model parameters

The simulation study carried contemplated five different combinations for $\boldsymbol{\theta}$, displayed in Table 3.1, Chapter 3. For each set of parameters, time series of length 500 with 300 independent replicates from the INAPARCH$(1,1)$ model were simulated. A sample path and its corresponding sample ACF is presented in Figure 3.1, Chapter 3, for the combination of parameters C2. The remaining cases are presented in Figures A.1, A.2, A.3 and A.4.

Figure A.5 represents the boxplots of the bias of the conditional maximum likelihood estimates for combination of parameters C1, C3, C4 and C5, respectively, from top to bottom. In a similar way to Figure 3.2, Chapter 3, the α parameter seems to be correctly estimated, having a very small bias. On the other hand, the variability in the estimates obtained for the δ parameter is very high. In the C5 combination of parameters case, not only the δ parameter is difficult to estimate but also the ω parameter shows very high variability.

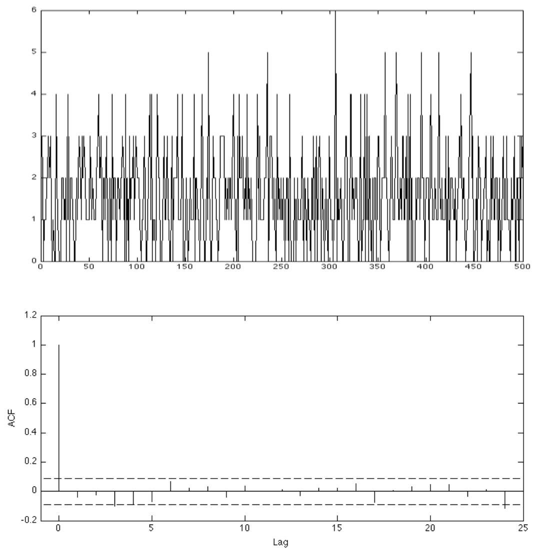


Figure A.1: Sample path for the INAPARCH$(1,1)$ process. Combination of parameters C1 (top) and its corresponding autocorrelation function (bottom).

A.3 Log-likelihood analysis

For C2 and C4, 300 samples were simulated considering values of δ varying from 2.0 to 3.0 (i.e., six different situations for each case). Preliminary data analysis was done with the construction of boxplots and histograms. Boxplots and histograms for Case 2 are presented in figures A.6, A.7, A.8 and A.9. Boxplots and histograms for Case 4 are presented in figures A.10, A.11, A.12 and A.13.

The log-likelihood was studied in the following manner: for each set of 300 samples the log-likelihood was calculated, varying the δ parameter in the range 2.0 to 3.0. Results are presented in Table A.1 for Case 4. For this Case 4, only the first value of the δ parameter results in a parameter set that lies inside the contractivity condition region. Nevertheless, same behaviour

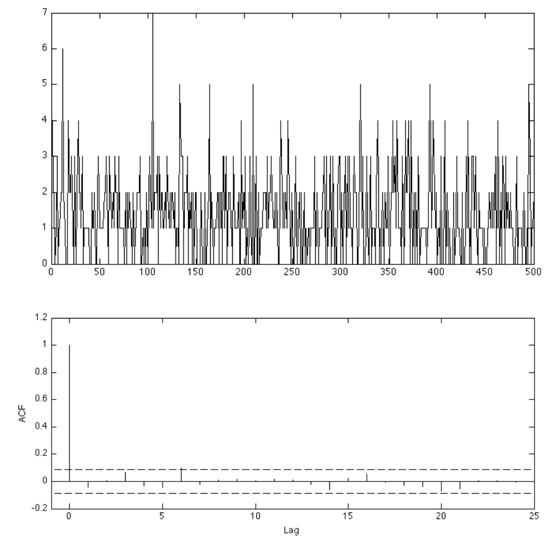


Figure A.2: Sample path for the INAPARCH(1,1) process. Combination of parameters C3 (top) and its corresponding autocorrelation function (bottom).

was observed for both Case 4 and Case 2, presented in Chapter 3, Section 3.4: the δ value for which the calculated log-likelihood was maximum was exactly what was expected for both cases and all 6 different situations.

Table A.1: Maximum likelihood estimation results for Case 4.

Samples simulated with	Log-likelihood for varying δ					
$\boldsymbol{\theta}$=$(2.30, 0.03, 0.68, 0.06, \delta)$	2.0	2.2	2.4	2.6	2.8	3.0
$\delta = 2.00$	-419.3641	-420.0895	-421.7561	-423.8736	-426.1726	-428.5032
$\delta = 2.20$	-435.4349	-434.6769	-435.1181	-436.2077	-437.6332	-439.2132
$\delta = 2.40$	-447.7643	-445.8791	-445.3893	-445.6977	-446.4591	-447.4683
$\delta = 2.60$	-457.7094	-454.9324	-453.7080	-453.4023	-453.6436	-454.2076
$\delta = 2.80$	-467.5063	-463.8594	-461.9165	-461.0084	-460.7380	-460.8626
$\delta = 3.00$	-473.2921	-469.0898	-466.6888	-465.3968	-464.8004	-464.6450

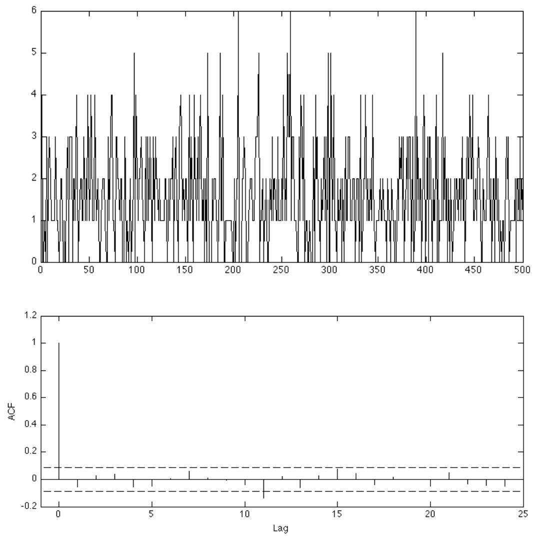


Figure A.3: Sample path for the INAPARCH$(1, 1)$ process. Combination of parameters C4 (top) and its corresponding autocorrelation function (bottom).

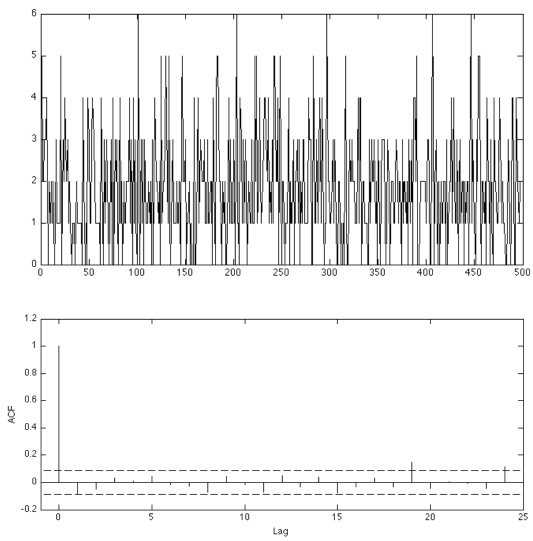


Figure A.4: Sample path for the INAPARCH$(1, 1)$ process. Combination of parameters C5 (top) and its corresponding autocorrelation function (bottom).

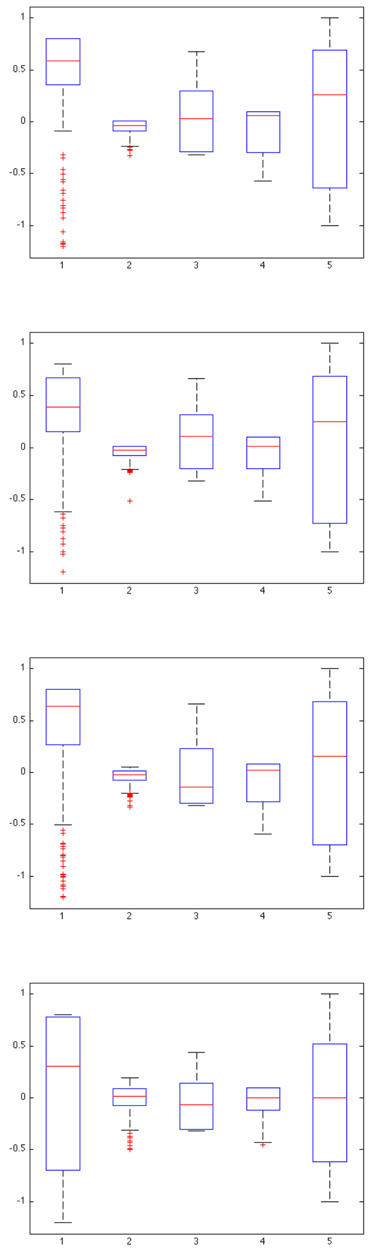


Figure A.5: Bias of conditional ML estimates, for combination of parameters C1, C3, C4 and C5, respectively, from top to bottom.

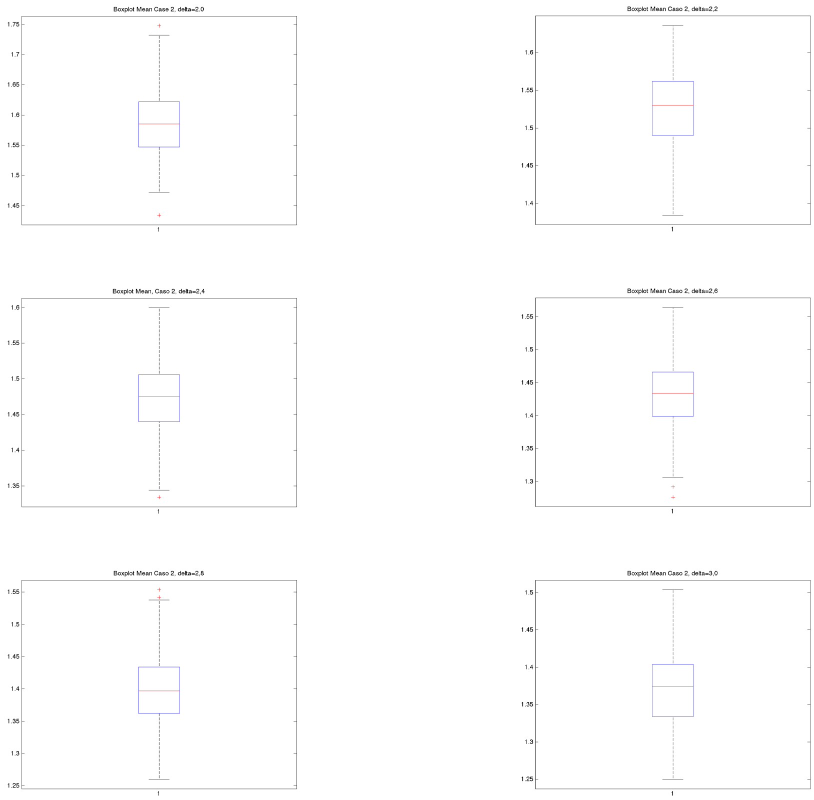


Figure A.6: Boxplots of the means, with δ varying from 2.0 to 3.0, for combination of parameters C2.

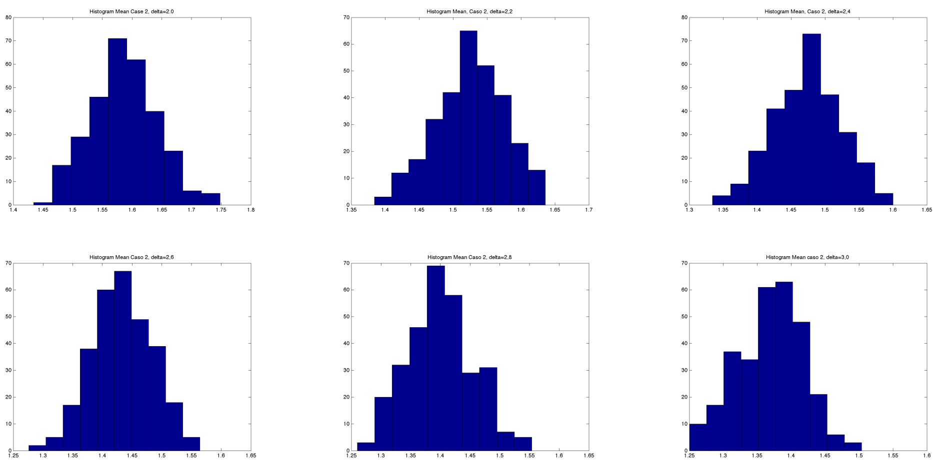


Figure A.7: Histograms of the means, with δ varying from 2.0 to 3.0, for combination of parameters C2.

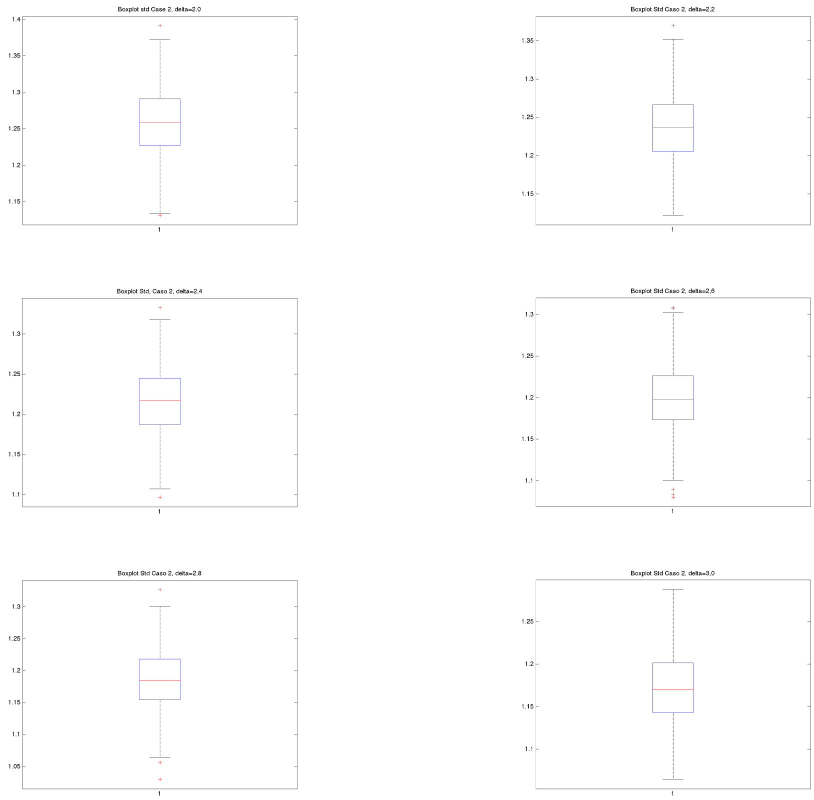


Figure A.8: Boxplots of the standard deviations, with δ varying from 2.0 to 3.0, for combination of parameters C2.

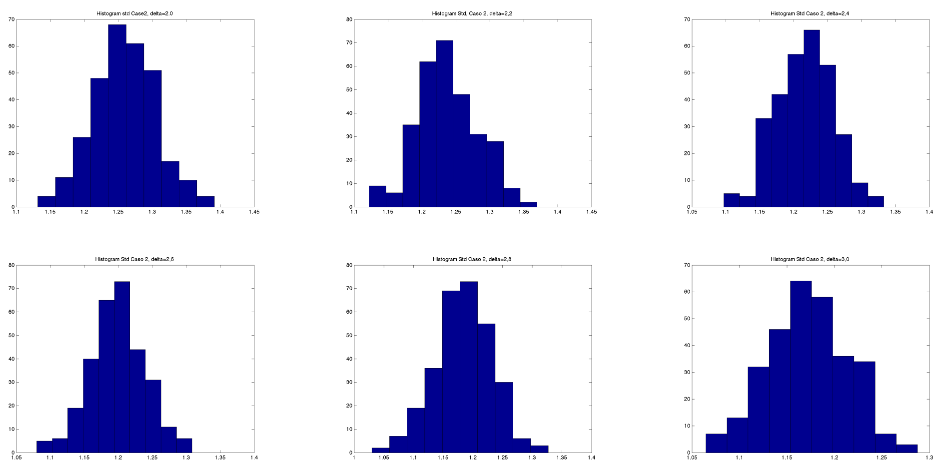


Figure A.9: Histograms of the standard deviations, with δ varying from 2.0 to 3.0, for combination of parameters C2.

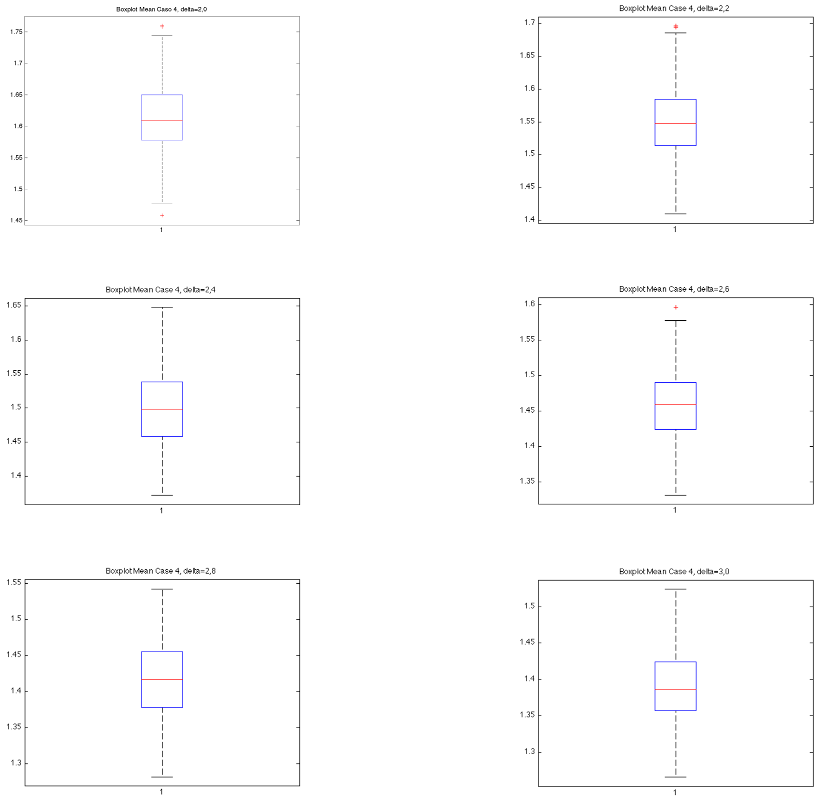


Figure A.10: Boxplots of the means, with δ varying from 2.0 to 3.0, for combination of parameters C4.

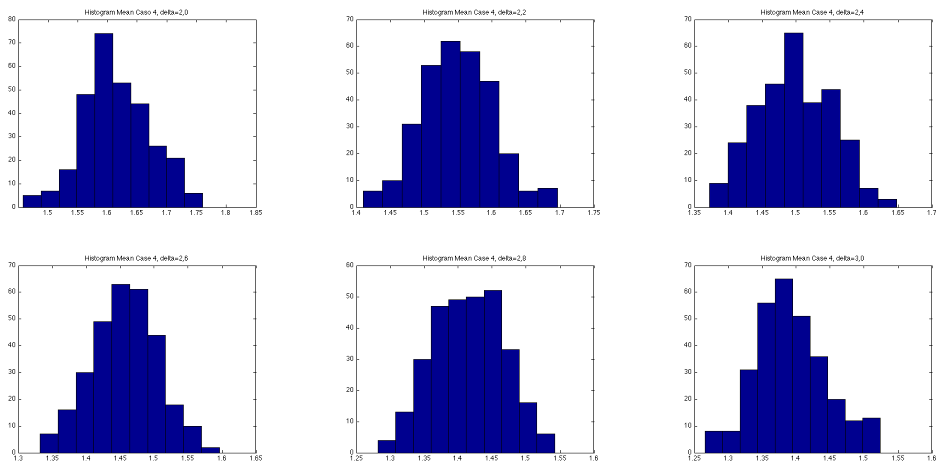


Figure A.11: Histograms of the means, with δ varying from 2.0 to 3.0, for combination of parameters C4.

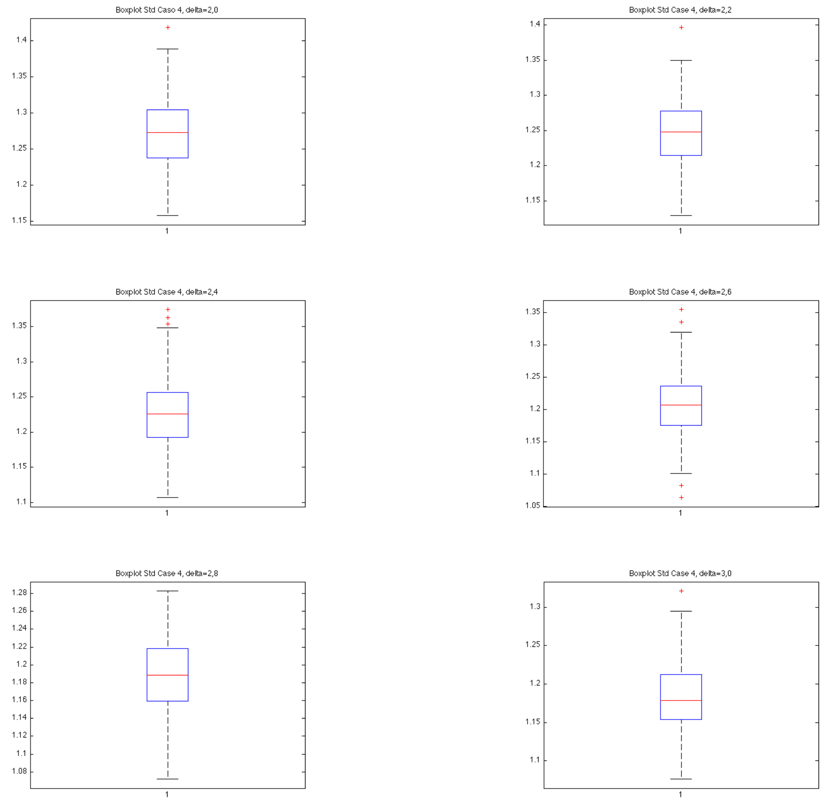


Figure A.12: Boxplots of the standard deviations, with δ varying from 2.0 to 3.0, for combination of parameters C4.

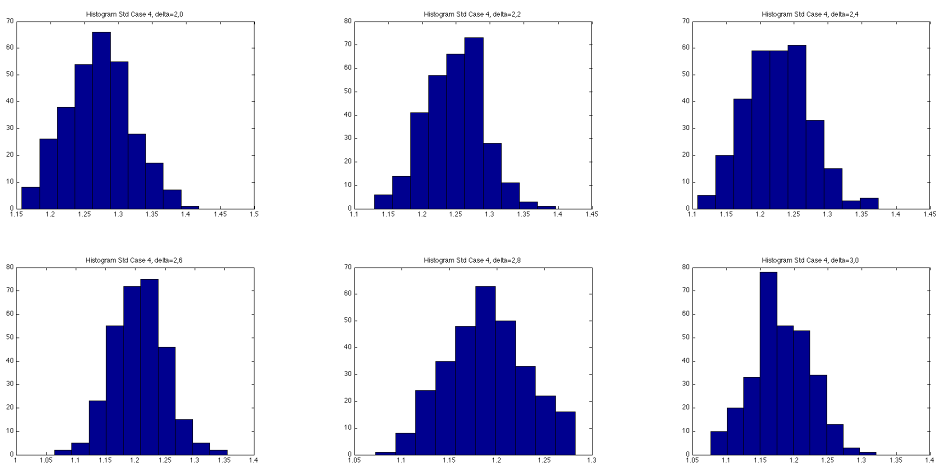


Figure A.13: Histograms of the standard deviations, with δ varying from 2.0 to 3.0, for combination of parameters C4.

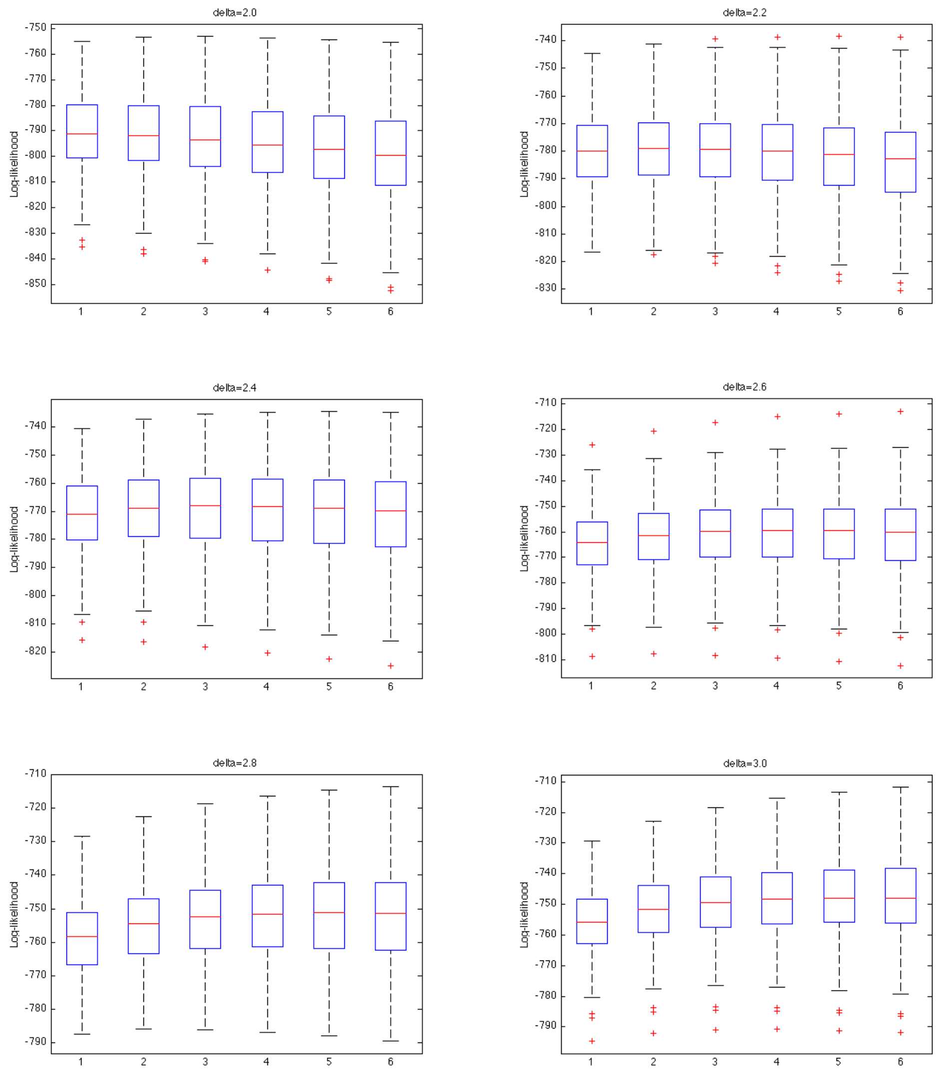


Figure A.14: Log-likelihood for varying δ, Case 4.

Appendix B

Matlab Code

B.1 Programs related to Chapter 2

Obtaining a sample from the FIAPARCH process.

```
clear all
n=2500;
m=n/2;
a0=0.40;   %(= ω)
a1=0.10;   %(= φ)
g1=0.68;   %(= γ)
b1=0.28;   %(= β)
delta=1.27;   %(= δ)
df=0.30;   %(= d)
z=normrnd(0,1,n,1);
y1(1,1)=normrnd(0,0.1)
h1(1,1)=(y1'*y1)
for i=2:n
    h1(i,1)=a0+a1*((abs(y1(i-1,1))-g1*y1(i-1,1))∧delta)+b1*h1(i-1,1);
    y1(i,1)=z(i,1)*h1(i,1)∧(1/delta);
end
```

```
for i=1:n
    B(i,1)=(abs(y1(i,1))-g1*y1(i,1))∧delta;
end
Kzero=sum(B)/n
lambda(1,1) =df;
aux(1,1) =a1-b1+df;
for i =2:n
soma=0;
    for j=2:n/2
        aux(j,1)=aux(j-1,1)*(j-1-df)/j;
        lambda(j,1)=b1*lambda(j-1,1)+[((j-1-df)/j)-a1]*aux(j-1,1);
        if i-j<=0
            soma=soma+lambda(j,1)*Kzero;
        else
            soma=soma+lambda(j,1)*B(i-j,1);
        end
    end
        somavar(i,1)=soma;
    h1(i,1)=(a0/(1-b1))+lambda(1,1)*B(i-1,1)+soma;
    y1(i,1)=(h1(i,1)∧(1/delta))*z(i,1);
    B(i,1)=(abs(y1(i,1))-g1*y1(i,1))∧delta;
end
save 'FIAPARCHsample'
```

Obtaining the volatility from a real data series considering the FIAPARCH model.

```
clear all
load 'residuos100ar10'
y1=residuos100AR10;
h1(1)=y1(1,1)'*y1(1,1)
```

```
n=length(y1)
a0=0.3903;
a1=0.0957;
g1=0.6782;
b1=0.2794;
delta=1.2744;
df=0.2952;
for i=1:n
    B(i,1)=(abs(y1(i,1))-g1*y1(i,1))∧delta;
end
Kzero=sum(B)/n
lambda(1,1)=a1-b1+df;
aux(1,1)=df;
for i=2:n
    soma=0;
    for j=2:n/2
        aux(j,1)=aux(j-1,1)*(j-1-df)/j;
        lambda(j,1)=b1*lambda(j-1,1)+[((j-1-df)/j)-a1]*aux(j-1,1);
        if i-j<=0
            soma=soma+lambda(j,1)*Kzero;
        else
            soma=soma+lambda(j,1)*B(i-j,1);
        end
    end
    somavar(i,1)=soma;
    h1(i,1)=(a0/(1-b1))+lambda(1,1)*B(i-1,1)+soma;
    z(i,1)=y1(i,1)/(h1(i,1)∧(1/delta));
end
save 'IBOsampleML'
```

Auxiliary function for the calculation of the volatility, considering the FIAPARCH model.

```
function s=fsigmaF(Banterior,somatorio)
a0=0.40;
a1=0.10;
g1=0.68;
b1=0.28;
delta=1.27;
df=0.30;
%Same function, considering an APARCH model
% e=y1(i-1,1)    previous epsilon value
% sa=h1(i-1,1)    previous sigma value
% s=(a0+a1*((abs(e)-g1*e)∧delta)+b1*sa)∧(1/delta);
lambda(1,1)=a1-b1+df;
s=(a0/(1-b1)+lambda(1,1)*Banterior+somatorio)∧(1/delta);
```

Calculating $P(C_{t,j}|\mathbf{x_2}, D_t, \theta)$, $P(C_{t,j}|D_t, \theta)$ and the alarm region (D, T and R, respectively).

```
load 'amostraFIAPARCHparametrosIbovespaML'
percentil95=3.136;
percentil90=2.293;
mu=percentil90
ti=2000
for t=1:5
    ti+t
    e1=B(ti+t-1,1);
    s1=somavar(ti+t,1);
    e2=B(ti+t,1);
    s2=somavar(ti+t+1,1);
    e3=B(ti+t+1,1);
```

```
    s3=somavar(ti+t+2,1);
    funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
    funtriplo=@(x,y,z) exp(-x.∧2/(2.*(fsigmaF(e1,s1).∧2))
-y.∧2./(2.*(fsigmaF(e2,s2)∧2))-z.∧2/(2.*(fsigmaF(e3,s3)∧2)))
./(((2*pi)∧(3/2))*fsigmaF(e1,s1)*fsigmaF(e2,s2)*fsigmaF(e3,s3));
    D=dblquad(funduplo,-100000,mu,mu,100000);
    T=triplequad(funtriplo,-100000,100000,-100000,mu,mu,100000);
    et=[-100:100];
    L=length(et);
    k=T
    e1=B(ti+t-1,1);
    s1=somavar(ti+t,1);
    s=0;
    for j=3:(ti+t+2)/2
        s=s+lambda(j,1)*B(ti+t+2-j);
    end
    for i=1:L
        e2=(abs(et(i))-g1*et(i))∧delta;
        s2=somavar(ti+t+1,1);
        h1tmais1=(a0/(1-b1))+lambda(1,1)*e2+s2;
        y1tmais1=(h1tmais1∧(1/delta))*z(ti+t+1,1);
        e3=(abs(y1tmais1)-g1*y1tmais1)∧delta;
        s3=lambda(2,1)*e2+s;
        funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
        P(i,t)=dblquad(funduplo,-100000,mu,mu,100000);
        if P(i,t)>k
            R(i,t)=et(i);
```

```
        else
            R(i,t)=999;
        end
    end
end
R
```

Calculating $P(C_{t,j}|\mathbf{x_2}, D_t, \theta)$, $P(C_{t,j}|D_t, \theta)$ and the alarm region for the application with the IBOVESPA data series (D, T and R, respectively).

```
load 'amostraIBO'
percentil25=-1.219; mu=percentil25
ti=3450
for t=1:10
    ti+t
    e1=B(ti+t-1,1);
    s1=somavar(ti+t,1);
    e2=B(ti+t,1);
    s2=somavar(ti+t+1,1);
    e3=B(ti+t+1,1);
    s3=somavar(ti+t+2,1);
    funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
    funtriplo=@(x,y,z) exp(-x.∧2/(2.*(fsigmaF(e1,s1).∧2))
-y.∧2./(2.*(fsigmaF(e2,s2)∧2))-z.∧2/(2.*(fsigmaF(e3,s3)∧2)))
./(((2*pi)∧(3/2))*fsigmaF(e1,s1)*fsigmaF(e2,s2)*fsigmaF(e3,s3));
    D=dblquad(funduplo,mu,100000,-100000,mu);
    T=triplequad(funtriplo,-100000,100000,mu,100000,-100000,mu);
    et=[-100:100];
    L=length(et);
```

```
    k=T
    e1=B(ti+t-1,1);
    s1=somavar(ti+t,1);
    s=0;
    for j=3:(ti+t+2)/2
        s=s+lambda(j,1)*B(ti+t+2-j);
    end
    for i=1:L
        e2=(abs(et(i))-g1*et(i))∧delta;
        s2=somavar(ti+t+1,1);
        h1tmais1=(a0/(1-b1))+lambda(1,1)*e2+s2;
        y1tmais1=(h1tmais1∧(1/delta))*z(ti+t+1,1);
        e3=(abs(y1tmais1)-g1*y1tmais1)∧delta;
        s3=lambda(2,1)*e2+s;
        funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
        P(i,t)=dblquad(funduplo,mu,100000,-100000,mu);
        if P(i,t)>k
            R(i,t)=et(i);
        else
            R(i,t)=999;
        end
    end
end
R
```

Varying alarm region with varying k.

```
load 'amostraFIAPARCHparametrosIbovespaML'
percentil90=2.293;
percentil95=3.136;
```

```
mu=percentil90
t=2050
e1=B(t-1,1);
s1=somavar(t,1);
e2=B(t,1);
s2=somavar(t+1,1);
e3=B(t+1,1);
s3=somavar(t+2,1);
funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3))
funtriplo=@(x,y,z) exp(-x.∧2/(2.*(fsigmaF(e1,s1).∧2))
-y.∧2./(2.*(fsigmaF(e2,s2)∧2))-z.∧2/(2.*(fsigmaF(e3,s3)∧2)))
./(((2*pi)∧(3/2))*fsigmaF(e1,s1)*fsigmaF(e2,s2)*fsigmaF(e3,s3))
D=dblquad(funduplo,-100000,mu,mu,100000);
T=triplequad(funtriplo,-100000,100000,-100000,mu,mu,100000);
et=[-10:0.1:20];
L=length(et);
P=zeros(L,8);
R=zeros(L,8);
k=T
e1=B(t-1,1);
s1=somavar(t,1);
s=0;
for j=3:(t+2)/2
    s=s+lambda(j,1)*B(t+2-j);
end
cont=0;
while k<T+0.04
    cont=cont+1
```

```
    for i=1:L
        e2=(abs(et(i))-g1*et(i))^delta;
        s2=somavar(t+1,1);
        h1tmais1=(a0/(1-b1))+lambda(1,1)*e2+s2;
        y1tmais1=(h1tmais1^(1/delta))*z(t+1,1);
        e3=(abs(y1tmais1)-g1*y1tmais1)^delta;
        s3=lambda(2,1)*e2+s;
        funduplo=@(x,y) exp(-x.^2./(2.*fsigmaF(e2,s2).^2)
-y.^2./(2.*fsigmaF(e3,s3).^2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
        P(i,cont)=dblquad(funduplo,-100000,mu,mu,100000);
        if P(i,cont)>k
            R(i,cont)=et(i);
        else
            R(i,cont)=999;
        end
    end
    k=k+0.005
end
R
```

Varying alarm region with varying k, for the application with the IBOVESPA data series.

```
load 'amostraIBOcorrigido'
percentil25=-1.219
mu=percentil25
t=3452
e1=B(t-1,1);
s1=somavar(t,1);
e2=B(t,1);
s2=somavar(t+1,1);
```

```
e3=B(t+1,1);
s3=somavar(t+2,1);
funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3))
funtriplo=@(x,y,z) exp(-x.∧2/(2.*(fsigmaF(e1,s1).∧2))
-y.∧2./(2.*(fsigmaF(e2,s2)∧2))-z.∧2/(2.*(fsigmaF(e3,s3)∧2)))
./(((2*pi)∧(3/2))*fsigmaF(e1,s1)*fsigmaF(e2,s2)*fsigmaF(e3,s3))
D=dblquad(funduplo,mu,100000,-100000,mu);
T=triplequad(funtriplo,-100000,100000,mu,100000,-100000,mu);
et=[-10:0.5:50];
L=length(et);
P=zeros(L,10);
R=zeros(L,10);
k=T
e1=B(t-1,1);
s1=somavar(t,1);
s=0;
for j=3:(t+2)/2
    s=s+lambda(j,1)*B(t+2-j);
end
cont=0;
while k<T+0.045
    cont=cont+1
    for i=1:L
        e2=(abs(et(i))-g1*et(i))∧delta;
        s2=somavar(t+1,1);
        h1tmais1=(a0/(1-b1))+lambda(1,1)*e2+s2;
        y1tmais1=(h1tmais1∧(1/delta))*z(t+1,1);
        e3=(abs(y1tmais1)-g1*y1tmais1)∧delta;
```

```
        s3=lambda(2,1)*e2+s;
        funduplo=@(x,y) exp(-x.∧2./(2.*fsigmaF(e2,s2).∧2)
-y.∧2./(2.*fsigmaF(e3,s3).∧2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
        P(i,cont)=dblquad(funduplo,mu,100000,-100000,mu);
        if P(i,cont)>k
            R(i,cont)=et(i);
        else
            R(i,cont)=999;
        end
    end
    k=k+0.01
end
R
```

Calculating the operating characteristics of the alarm system.

```
load 'amostraFIAPARCHparametrosIbovespaML'
percentil90=2.293;
percentil95=3.136;
mu=percentil90
t=2050
% Alarm Region:
RAinf1=-100000;
RAsup1=-1.3
RAinf2=6.7
RAsup2=100000;
e1=B(t-1,1);
s1=somavar(t,1);
e2=B(t,1);
s2=somavar(t+1,1);
e3=B(t+1,1);
```

```
s3=somavar(t+2,1);
fun=@(x) exp(-x.^2./(2.*fsigmaF(e1,s1).^2))./(sqrt(2*pi).*(fsigmaF(e1,s1)));
% Size of the alarm region
Tamanho=quad(fun,RAinf1,RAsup1)+quad(fun,RAinf2,RAsup2)
funD=@(x,y) exp(-x.*x./(2.*fsigmaF(e2,s2).^2)
-y.*y./(2.*fsigmaF(e3,s3).^2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
funT=@(x,y,z) exp(-x.^2/(2.*(fsigmaF(e1,s1)^2))
-y.^2/(2.*(fsigmaF(e2,s2)^2))-z.^2/(2.*(fsigmaF(e3,s3)^2)))
/(((2*pi)^(3/2))*fsigmaF(e1,s1)*fsigmaF(e2,s2)*fsigmaF(e3,s3));
D=dblquad(funD,-100000,mu,mu,100000)
T=triplequad(funT,-100000,100000,-100000,mu,mu,100000)
% Probability of the Alarm being Correct
PAC=(triplequad(funT,RAinf1,RAsup1,-100000,mu,mu,100000)+
triplequad(funT,RAinf2,RAsup2,-100000,mu,mu,100000))
/(quad(fun,RAinf1,RAsup1)+quad(fun,RAinf2,RAsup2))
% Probability of Detecting the Event
PD=(triplequad(funT,RAinf1,RAsup1,-100000,mu,mu,100000)
+triplequad(funT,RAinf2,RAsup2,-100000,mu,mu,100000))
/triplequad(funT,-100000,100000,-100000,mu,mu,100000)
```

Calculating the operating characteristics of the alarm system, for the application with the IBOVESPA data series.

```
load amostraIBOcorrigido
percentil25=-1.219
mu=percentil25
t=3516
% Alarm Region
RAinf1=-100000;
RAsup1=-8
RAinf2=41.7
```

```
RAsup2=100000;
e1=B(t-1,1);
s1=somavar(t,1);
e2=B(t,1);
s2=somavar(t+1,1);
e3=B(t+1,1);
s3=somavar(t+2,1);
fun=@(x) exp(-x.^2./(2.*fsigmaF(e1,s1).^2))./(sqrt(2*pi).*(fsigmaF(e1,s1)));
% Size of the Alarm Region (Size)
Size=quad(fun,RAinf1,RAsup1)+quad(fun,RAinf2,RAsup2)
funD=@(x,y) exp(-x.*x./(2.*fsigmaF(e2,s2).^2)
-y.*y./(2.*fsigmaF(e3,s3).^2))./(2*pi.*fsigmaF(e2,s2)*fsigmaF(e3,s3));
funT=@(x,y,z) exp(-x.^2/(2.*(fsigmaF(e1,s1)^2))
-y.^2/(2.*(fsigmaF(e2,s2)^2))-z.^2/(2.*(fsigmaF(e3,s3)^2)))
/(((2*pi)^(3/2))*fsigmaF(e1,s1)*fsigmaF(e2,s2)*fsigmaF(e3,s3));
D=dblquad(funD,mu,100000,-100000,mu)
T=triplequad(funT,-100000,100000,mu,100000,-100000,mu)
% Probability of the Alarm being Correct (PAC)
PAC=(triplequad(funT,RAinf1,RAsup1,mu,100000,-100000,mu)
+triplequad(funT,RAinf2,RAsup2,mu,100000,-100000,mu))
/(quad(fun,RAinf1,RAsup1)+quad(fun,RAinf2,RAsup2))
% Probability of Detecting the Event (PD)
PD=(triplequad(funT,RAinf1,RAsup1,mu,100000,-100000,mu)
+triplequad(funT,RAinf2,RAsup2,mu,100000,-100000,mu))
/triplequad(funT,-100000,100000,mu,100000,-100000,mu)
```

B.2 Programs related to Chapter 3

Obtaining a sample from the INAPARCH process.

```
clear all
n=500 % n=size
rep=300 % rep=number of samples
omega=2.3
alpha1=0.03
gamma1=0.68
beta1=0.06
delta=2.0
condunica=2^delta*(2*alpha1*delta+0.5*beta1)
lambda0=3
for j=1:rep
    x(1,j)=poissrnd(lambda0);
    lambda(1,j)=lambda0;
    for i=2:n
        lambda(i,j)=(omega+alpha1*(abs(x(i-1,j)-lambda(i-1,j))-gamma1*
*(x(i-1,j)-lambda(i-1,j)))^delta+beta1*lambda(i-1,j)^delta)^(1/delta);
        x(i,j)=poissrnd(lambda(i,j));
    end
end
figure
plot(x(:,1),'b')
legend('case 2')
save 'INAPARCHsampleCase2'
```

Function for the calculation of λ_t (necessary in estimation procedures).

```
function y=lambdatdelta(theta,amX,lambdai,n)
```

```
omega=theta(1);
alpha1=theta(2);
gamma1=theta(3);
beta1=theta(4);
delta=theta(5);
lambdatdelta(1)=lambdai;
for i=2:n
    lambdatdelta(i)=(omega+alpha1*(abs(amX(i-1)-lambdatdelta(i-1))
-gamma1*(amX(i-1)-lambdatdelta(i-1)))∧delta+beta1*lambdatdelta(i-1)∧
∧delta)∧(1/delta);
end y=lambdatdelta(n);
```

ML Estimation of INAPARCH samples.

```
clear all
load amostraINAPARCHcaso3
size=size(x);
n=size(:,2);
global amostra
x0=[2 0.1 0.6 0.1 2.5]    % parameter initialization
lb=[1.5 0.001 -0.999 0.001 2];    % lower bounds for the estimates
ub=[3.5 0.999 0.999 0.999 4];    % upper bounds for the estimates
options=optimset('GradObj','on');
for amostra=1:n
    phat=mle(x(:,amostra),'nloglf',@nllfdeltateste,'start',x0,'lowerbound',
lb,'upperbound',ub,'optimfun','fmincon')
    xlist(amostra,:)=phat;
    save 'xlist'
    amostra
end
xlist
```

Auxiliary function for the ML Estimation of INAPARCH samples.

```
function [Q, gradQ]=nllfteste(theta,x,cens,freq)
global amostra
omega=theta(1);
alpha1=theta(2);
gamma1=theta(3);
beta1=theta(4);
delta=theta(5);
load amostraINAPARCHcaso3
amX=x(:,amostra);
amlambda(1)=3;
n=size(amX,1);
Q=0;
for i=1:n
    Q=Q+amX(i)*log(lambdatdelta(theta,amX,amlambda(1),i))-
-lambdatdelta(theta,amX,amlambda(1),i)-log(factorial(amX(i)));
end
Q=-Q;
```

Calculating log-likelihood with varying δ.

```
clear all
load 'amostraINAPARCH300caso4delta30'
%theta=[2.30 0.03 0.68 0.06 1.80];    Case 2
theta=[2.30 0.05 0.68 0.08 1.80];    % Case 4
for amostra=1:300
    amlambda=lambda(:,amostra);
    amX=x(:,amostra);
    n=size(amX,1);
    for i=1:6
```

```
        theta(5)=theta(5)+0.2;
        Q=0;
        for j=1:n
            Q=Q+amX(j)*log(lambdatdelta(theta,amX,amlambda(1),j))-
-lambdatdelta(theta,amX,amlambda(1),j)-log(factorial(amX(j)));
        end
        L(i,amostra)=Q;
    end
    theta(5)=1.80;
end
L
save ('logverosim300caso4delta30','L')
```

ML Estimation of real data time series, based on the INAPARCH model.

```
clear all
load DADOS
x0=[2 0.2 -0.6 0.2 2.0]      % parameter initialization
lb=[0.001 0.001 -0.999 0.001 0.001];      % lower bounds for the estimates
ub=[10 0.999 0.999 0.999 10];      % upper bounds for the estimates
options=optimset('GradObj','on');
phat=mle(GSK,'nloglf',@nllfdeltatestedados,'start',x0,'lowerbound',
lb,'upperbound',ub,'optimfun','fmincon');
xlist=phat
```

Auxiliary function for the ML Estimation of real data time series, based on the INAPARCH model.

```
function [Q, grad_Q]=nllfdeltatestedados(theta,x,cens,freq)
omega=theta(1);
alpha1=theta(2);
gamma1=theta(3);
```

```
beta1=theta(4);
delta=theta(5);
load DADOS
amX=GSK;
amlambda(1)=18.8563;
%amlambda(1)=30.8283;
n=size(amX,1);
Q=0;
for i=1:n
    Q=Q+amX(i)*log(lambdatdelta(theta,amX,amlambda(1),i))-
-lambdatdelta(theta,amX,amlambda(1),i)-log(factorial(amX(i)));
end
Q=-Q;
```

Calculating $P(C_{t,j}|\mathbf{y_2}, D_t, \theta)$, $P(C_{t,j}|D_t, \theta)$ and the alarm region for the application with the Astrazeneca or the Glaxosmithkline data series (P1, P2 and R, respectively).

```
clear all
load DADOS
thetaGSK=[0.3781 0.1392 -0.3269 0.8791 0.9826];
thetaAZN21Set=[2.4862 0.2824 -0.2787 0.7501 1.0598];
% theta=thetaGSK;
theta=thetaAZN21Set;
omega=theta(1);
alfa=theta(2);
gama=theta(3);
beta=theta(4);
delta=theta(5);
funlambda=@(y,l)(omega+alfa*(abs(y-l)-gama*(y-l))∧delta+beta*l∧delta)
∧(1/delta);
```

```
mu=25
% amX=GSK;
amX=AZN21Set;
% lambda1GSK=18.8563;
lambda1AZN21Set=30.8283;
% lambdatdelt(1)=lambda1GSK;
lambdatdelt(1)=lambda1AZN21Set;
for i=2:501
    lambdatdelt(i)=(omega+alfa*(abs(amX(i-1)-lambdatdelt(i-1))-
-gama*(amX(i-1)-lambdatdelt(i-1)))^delta+beta*lambdatdelt(i-1)^delta)
^(1/delta);
end
ti=450
for t=1:11
    ti+t-1
    yt=amX(ti+t-1);
    lbt=funlambda(amX(ti+t-2),lambdatdelt(ti+t-2));
% Calculating P(C_{t,j}|y_2, D_t, θ), P1
    t1=0;
    for i=0:mu
        ytmais1=i;
        lbtmais1=funlambda(yt,lbt);
        t2=0;
        for j=mu+1:150
            ytmais2=j;
            lbtmais2=funlambda(i,lbtmais1);
            t2=t2+exp(-lbtmais2)*lbtmais2^ytmais2/factorial(ytmais2);
        end
        t1=t1+(exp(-lbtmais1)*lbtmais1^ytmais1/factorial(ytmais1))*t2;
```

```
    end
    P1=t1
% Calculating P(C_{t,j}|D_t,θ), P2
    t1=0;
    for i=0:150
        yt=i;
        lbt=funlambda(amX(ti+t-2),lambdatdelt(ti+t-2));
        t2=0;
        for j=0:mu
            ytmais1=j;
            lbtmais1=funlambda(i,lbt);
            t3=0;
            for k=mu+1:150
                ytmais2=k;
                lbtmais2=funlambda(j,lbtmais1);
                t3=t3+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
            end
            t2=t2+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*t3;
        end
        t1=t1+(exp(-lbt)*lbt∧yt/factorial(yt))*t2;
    end
    P2=t1
% Calculating the Alarm Region, R
    et=[0:1:150];
    L=length(et);
    kt2=P2;
    for i=1:L
        yt=et(i);
        lbt=funlambda(amX(ti+t-2),lambdatdelt(ti+t-2));
```

```
    s1=0;
    for j=0:mu
        ytmais1=j;
        lbtmais1=funlambda(yt,lbt);
        s2=0;
        for k=mu+1:150
            ytmais2=k;
            lbtmais2=funlambda(ytmais1,lbtmais1);
            s2=s2+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
        end
        s1=s1+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*s2;
    end
    P(i,t)=s1;
    if P(i,t)>kt2
        R(i,t)=et(i);
    else
        R(i,t)=999;
    end
  end
end
R
```

Varying alarm region with varying k, for the application with Astrazeneca and Glaxosmithkline data series.

```
clear all
load DADOS
thetaGSK=[0.3781 0.1392 -0.3269 0.8791 0.9826];
thetaAZN21Set=[2.4862 0.2824 -0.2787 0.7501 1.0598];
%theta=thetaGSK;
theta=thetaAZN21Set;
```

```
omega=theta(1);
alfa=theta(2);
gama=theta(3);
beta=theta(4);
delta=theta(5);
funlambda=@(y,l)(omega+alfa*(abs(y-l)-gama*(y-l))∧delta+beta*l∧delta)
∧(1/delta);
%amX=GSK;
amX=AZN21Set;
%lambda1GSK=18.8563;
lambda1AZN21Set=30.8283;
%lambdatdelt(1)=lambda1GSK;
lambdatdelt(1)=lambda1AZN21Set;
for i=2:501
    lambdatdelt(i)=(omega+alfa*(abs(amX(i-1)-lambdatdelt(i-1))-
-gama*(amX(i-1)-lambdatdelt(i-1)))∧delta+beta*lambdatdelt(i-1)∧delta)
∧(1/delta);
end
mu=19
t=460
yt=amX(t);
lbt=funlambda(amX(t-1),lambdatdelt(t-1));
s1=0;
for i=0:mu
    ytmais1=i;
    lbtmais1=funlambda(yt,lbt);
    s2=0;
    for j=mu+1:100
        ytmais2=j;
```

```
        lbtmais2=funlambda(i,lbtmais1);
        s2=s2+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
    end
    s1=s1+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*s2;
end
P1=s1
t1=0;
for i=0:100
    yt=i;
    lbt=funlambda(amX(t-1),lambdatdelt(t-1));
    t2=0;
    for j=0:mu
        ytmais1=j;
        lbtmais1=funlambda(i,lbt);
        t3=0;
        for k=mu+1:100
            ytmais2=k;
            lbtmais2=funlambda(j,lbtmais1);
            t3=t3+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
        end
        t2=t2+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*t3;
    end
    t1=t1+(exp(-lbt)*lbt∧yt/factorial(yt))*t2;
end
P2=t1
et=[0:1:100];
L=length(et);
P=zeros(L,12);
R=zeros(L,12);
```

```
kt2=P2
contador=0;
while kt2<P2+0.08
    contador=contador+1
    for i=1:L
        yt=et(i);
        lbt=funlambda(amX(t-1),lambdatdelt(t-1));
        s1=0;
        for j=0:mu
            ytmais1=j;
            lbtmais1=funlambda(yt,lbt);
            s2=0;
            for k=mu+1:100
                ytmais2=k;
                lbtmais2=funlambda(ytmais1,lbtmais1);
                s2=s2+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
            end
            s1=s1+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*s2;
        end
        P(i,contador)=s1;
        if P(i,contador)>kt2
            R(i,contador)=et(i);
        else
            R(i,contador)=999;
        end
    end
    kt2=kt2+0.005
end
R
```

Calculating the operating characteristics of the alarm system, for the application with the Astrazeneca and Glaxosmithkline data series.

```
clear all
load DADOS
thetaGSK=[0.3781 0.1392 -0.3269 0.8791 0.9826];
thetaAZN21Set=[2.4862 0.2824 -0.2787 0.7501 1.0598];
%theta=thetaGSK;
theta=thetaAZN21Set;
omega=theta(1);
alfa=theta(2);
gama=theta(3);
beta=theta(4);
delta=theta(5);
funlambda=@(y,l)(omega+alfa*(abs(y-l)-gama*(y-l))∧delta+beta*l∧delta)
∧(1/delta);
%amX=GSK;
amX=AZN21Set;
%lambda1GSK=18.8563;
lambda1AZN21Set=30.8283;
%lambdatdelt(1)=lambda1GSK;
lambdatdelt(1)=lambda1AZN21Set;
for i=2:501
    lambdatdelt(i)=(omega+alfa*(abs(amX(i-1)-lambdatdelt(i-1))-
-gama*(amX(i-1)-lambdatdelt(i-1)))∧delta+beta*lambdatdelt(i-1)∧delta)
∧(1/delta);
end
mu=19
t=460
```

```
% Alarm Region:
RAinf1=0;
RAsup1=0;
RAinf2=29;
RAsup2=35;
% Size of the Alarm Region
s1=0;
s2=0;
for i=RAinf1:RAsup1
    yt=i;
    lbt=funlambda(amX(t-1),lambdatdelt(t-1));
    s1=s1+exp(-lbt)*lbt^yt/factorial(yt);
end
for i=RAinf2:RAsup2
    yt=i;
    lbt=funlambda(amX(t-1),lambdatdelt(t-1));
    s2=s2+exp(-lbt)*lbt^yt/factorial(yt);
end
Size=s1+s2
% Probability of the Alarm being Correct (PAC)
t1=0;
v1=0;
for i=RAinf1:RAsup1
    yt=i;
    lbt=funlambda(amX(t-1),lambdatdelt(t-1));
    t2=0;
    for j=0:mu
        ytmais1=j;
        lbtmais1=funlambda(i,lbt);
```

```
        t3=0;
        for k=mu+1:100
            ytmais2=k;
            lbtmais2=funlambda(j,lbtmais1);
            t3=t3+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
        end
        t2=t2+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*t3;
    end
    t1=t1+(exp(-lbt)*lbt∧yt/factorial(yt))*t2;
end
for i=RAinf2:RAsup2
    yt=i;
    lbt=funlambda(amX(t-1),lambdatdelt(t-1));
    v2=0;
    for j=0:mu
        ytmais1=j;
        lbtmais1=funlambda(i,lbt);
        v3=0;
        for k=mu+1:100
            ytmais2=k;
            lbtmais2=funlambda(j,lbtmais1);
            v3=v3+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
        end
        v2=v2+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*v3;
    end
    v1=v1+(exp(-lbt)*lbt∧yt/factorial(yt))*v2;
end
numerator=t1+v1;
PAC=numerator/Size
```

```
% Probability of Detecting the Event (PD)
u1=0;
for i=0:100
    yt=i;
    lbt=funlambda(amX(t-1),lambdatdelt(t-1));
    u2=0;
    for j=0:mu
        ytmais1=j;
        lbtmais1=funlambda(i,lbt);
        u3=0;
        for k=mu+1:100
            ytmais2=k;
            lbtmais2=funlambda(j,lbtmais1);
            u3=u3+exp(-lbtmais2)*lbtmais2∧ytmais2/factorial(ytmais2);
        end
        u2=u2+(exp(-lbtmais1)*lbtmais1∧ytmais1/factorial(ytmais1))*u3;
    end
    u1=u1+(exp(-lbt)*lbt∧yt/factorial(yt))*u2;
end
P2=u1
PD=numerator/P2
```

Frequency estimation of the operating characteristics of the alarm system, for the application with the Astrazeneca and Glaxosmithkline data series.

```
clear all
load DADOS
thetaGSK=[0.3781 0.1392 -0.3269 0.8791 0.9826];
thetaAZN21Set=[2.4862 0.2824 -0.2787 0.7501 1.0598];
theta=thetaGSK;
```

```
%theta=thetaAZN21Set;
omega=theta(1);
alfa=theta(2);
gama=theta(3);
beta=theta(4);
delta=theta(5);
funlambda=@(y,l)(omega+alfa*(abs(y-l)-gama*(y-l))∧delta+beta*l∧delta)
∧(1/delta);
amX=GSK;
%amX=AZN21Set;
lambda1GSK=18.8563;
%lambda1AZN21Set=30.8283;
lambdatdelt(1)=lambda1GSK;
%lambdatdelt(1)=lambda1AZN21Set;
for i=2:501
   lambdatdelt(i)=(omega+alfa*(abs(amX(i-1)-lambdatdelt(i-1))-
-gama*(amX(i-1)-lambdatdelt(i-1)))∧delta+beta*lambdatdelt(i-1)∧delta)
∧(1/delta);
end
mu=18
t=460
% Alarm Region RAinf1=0
RAsup1=5
RAinf2=20
RAsup2=81
max=100000    % Number of iterations
alarms=0;    % Number of alarms
catastrophes=0;    % Number of catastrophes
correctalarms=0;    % Number of correct alarms
```

```
for i=1:max
    lbt(i)=funlambda(amX(t-1),lambdatdelt(t-1));
    yt(i)=poissrnd(lbt(i));
    lbtmais1(i)=funlambda(yt(i),lbt(i));
    ytmais1(i)=poissrnd(lbtmais1(i));
    lbtmais2(i)=funlambda(ytmais1(i),lbtmais1(i));
    ytmais2(i)=poissrnd(lbtmais2(i));
% Counting the number of alarms
    if yt(i)>=RAinf1
        if yt(i)<=RAsup1
            alarms=alarms+1;
        end
    end
    if yt(i)>=RAinf2
        if yt(i)<=RAsup2
            alarms=alarms+1;
        end
    end
% Counting the number of catastrophes
    if ytmais2(i)>mu
        catastrophes=catastrophes+1;
    end
% Counting the number of correct alarms
    if yt(i)>=RAinf1
        if yt(i)<=RAsup1
            if ytmais2(i)>mu
                correctalarms=correctalarms+1;
            end
        end
```

```
    end
    if yt(i)>=RAinf2
        if yt(i)<=RAsup2
            if ytmais2(i)>mu
                correctalarms=correctalarms+1;
            end
        end
    end
end
alarms
PAlarm=alarms/max
catastrophes
PCatastrophe=catastrophes/max
correctalarms
% Probability of the Alarm being Correct (PAC)
PAC=correctalarms/alarms
% Probability of Detecting the Event (PD)
PD=correctalarms/catastrophes
```

Appendix C

Abbreviations and Notation

$\gamma_{n,Y}$	sample ACVF
$\rho_{n,Y}$	sample ACF
ACF	AutoCorrelation Function
ACP	Autoregressive Conditional Poisson
ACVF	AutoCoVariance Function
APARCH	Asymmetric Power ARCH
ARCH	AutoRegressive Conditional Heteroscedasticity
ARMA	AutoRegressive Moving Average
AR	AutoRegressive
BINMA	Bivariate INMA
BOVESPA	São Paulo Stock Exchange (BOlsa de Valores do Estado de São PAulo)
CLT	Central Limit Theorem
DARMA	Discrete ARMA
DAR	Discrete AutoRegressive
DSC	Discrete Self Decomposable
DSINAR	Doubly Stochastic INAR
EVT	Extreme Value Theory
EWMA	Exponentially Weighted Moving Average

FIAPARCH	Fractionally Integrated Asymmetric Power ARCH
FIGARCH	Fractionally Integrated GARCH
FINGARCH	Functional INGARCH
GARCH	Generalized ARCH
GLARMA	Generalized Linear ARMA
GLM	Generalized Linear Models
GP	Generalized Poisson
IBOVESPA	BOVESPA Index
IGARCH	Integrated GARCH
INARMA	INteger-valued ARMA
INAR	INteger-valued AutoRegressive
INGARCH	INteger-valued GARCH
INMA	INteger-valued Moving Average
MA	Moving Average
MC	Monte Carlo
MLE	Maximum Likelihood Estimation
ML	Maximum Likelihood
MTD	Mixture Transition Distribution
NYSE	New York Stock Exchange
PM_{10}	Particulate Matter of aerodynamic diameter 10 μm or less
QINAR	Quasi-binomial INAR
QMLE	Quasi-Maximum Likelihood Estimation
RCINAR	Random Coefficient INAR
RUP	Reinforced Urn Processes
SAM	Stochastic Autoregressive Mean
SRE	Stochastic Recurrence Equation
SV	Stochastic Volatility
i.i.d.	independent identically distributed
r.v.('s)	random variable(s)

References

M. Al-Osh and A. Alzaid. Integer-valued moving average (INMA) process. *Statistical Papers*, 29(1):281–300, 1988.

M. A. Al-Osh and E.-E. A. Aly. First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics - Theory and Methods*, 21(9):2483–2492, 1992.

M. A. Al-Osh and A. A. Alzaid. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.

E.-E. A. Aly and N. Bouzar. Explicit stationary distributions for some Galton-Watson processes with immigration. *Communications in Statistics. Stochastic Models*, 10(2):499–517, 1994.

E.-E. A. Aly and N. Bouzar. Stationary solutions for integer-valued autoregressive processes. *International Journal of Mathematics and Mathematical Sciences*, 2005(1):1–18, 2005.

A. Alzaid and M. Al-Osh. First-order integer-valued autoregressive (INAR(1)) process: Distributional and regression properties. *Statistica Neerlandica*, 42(1):53–61, 1988.

A. Alzaid and M. Al-Osh. Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics*, 45(2):223–232, 1993.

A. A. Alzaid and M. Al-Osh. An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27(2):314–324, 1990.

M. Amaral-Turkman and K. Turkman. Optimal alarm systems for autoregressive processes: A Bayesian approach. *Computational Statistics & Data Analysis*, 10(3):307 – 314, 1990.

P. Amatyakul. The relationship between trading volume and jump. Processes in financial markets. *Duke Journal of Economics*, XXII, 2010.

R. Ambagaspitiya and N. Balaskrishnan. On the compound generalized Poisson distributions. *Astin Bulletin*, 24(2):255–263, 1994.

E. Amerio, P. Muliere, and P. Secchi. Reinforced urn processes for modeling credit default distributions. *International Journal of Theoretical and Applied Finance*, 7(4):407–423, 2004.

M. Antunes, M. Amaral-Turkman, and K. Turkman. A Bayesian approach to event prediction intervals. *Journal of Time Series Analysis*, 24:631–646, 2003.

R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30, 1996.

O. E. Barndorff-Nielsen. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37, 2004.

O. E. Barndorff-Nielsen and N. Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30, 2006.

L. Bauwens and M. Lubrano. Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal*, 1(Conference):C23–C46, 1998.

S.-I. Beckman. Oc-functions for a predictor-based alarm system. Master's thesis, Lund Institute of Technology, Sweden, 1987.

S.-I. Beckman, J. Holst, and G. Lindgren. Alarm characteristics for a flood warning system with deterministic components. *Journal of Time Series Analysis*, 11(1):1–18, 1990.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

T. Bollerslev and H. O. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184, 1996.

T. Bollerslev and J. M. Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172, 1992.

P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1-2):115–127, 1992.

A. L. Braga, A. Zanobetti, and J. Schwartz. The time course of weather-related death. *Epidemiology*, 12(6):662–667, 2001.

P. T. Brandt and J. T. Williams. A linear Poisson autoregressive model: The Poisson AR(p) model. *Political Analysis*, 9(2):164–184, 2001.

K. Brännäs and A. Hall. Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry*, 17(3):277–291, 2001.

K. Brännäs and J. Hellström. Generalized integer-valued autoregression. *Econometric Reviews*, 20(4):425–443, 2001.

K. Brännäs and J. Nordström. An integer-valued time series model for hotels that accounts for constrained capacity. *Studies in Nonlinear Dynamics & Econometrics*, 8(4):1–11, 2004.

K. Brännäs and J. Nordström. Tourist accommodation effects of festivals. *Tourism Economics*, 12(2):291–302, 2006.

K. Brännäs and A. M. M. S. Quoreshi. Integer-valued moving average modelling of the number of transactions in stocks. *Applied Financial Economics*, 20(18):1429–1440, 2010.

K. Brännäs, J. Hellström, and J. Nordström. A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting*, 18(1):19–30, 2002.

P. Brockwell and R. Davis. *Time Series: Theory and Methods, 2nd Ed.* Springer, New York, 1991.

P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting.* Springer, New York, 1996.

P. Bulla. *Application of reinforced urn processes to survival analysis.* PhD thesis, Bocconi University, Italy, 2005.

A. Cameron and P. Trivedi. *Regression Analysis of Count Data.* Cambridge University Press, Cambridge, 1998.

M. J. Campbell, S. A. Julious, C. K. Peterson, and A. Tobias. Atmospheric pressure and sudden infant death syndrome in Cook County, Chicago. *Paediatric and Perinatal Epidemiology*, 15(3):287–289, 2001.

G. C. Cawley, G. J. Janacek, M. R. Haylock, and S. R. Dorling. Predictive uncertainty in environmental modelling. *Neural Networks*, 20(4):537–549, 2007.

K. S. Chan and J. Ledolter. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90(429):242–252, 1995.

M.-H. Chiang and C.-H. Wang. Intradaily relationship between information revelation and trading duration under market trends: The evidence of MSCI Taiwan stock index futures. *Applied Economics Letters*, 11(8):495–501, 2004.

P. Cirillo and J. Hüsler. Extreme shock models: An alternative perspective. *Statistics & Probability Letters*, 81(1):25–30, 2011.

P. Cirillo, J. Hüsler, and P. Muliere. A nonparametric urn-based approach to interacting failing systems with an application to credit risk modeling. *International Journal of Theoretical and Applied Finance (IJTAF)*, 13(8): 1223–1240, 2010.

P. Cirillo, J. Hüsler, and P. Muliere. Alarm systems and catastrophes from a diverse point of view. *Methodology and Computing in Applied Probability*, 15(4):821–839, 2013.

C. Conrad and B. R. Haag. Inequality constraints in the fractionally integrated GARCH model. *Journal of Financial Econometrics*, 4(3):413–449, 2006.

C. Conrad, M. Karanasos, and N. Zeng. Multivariate fractionally integrated APARCH modeling of stock market volatility: A multi-country study. *Journal of Empirical Finance*, 18(1):147 – 159, 2011.

P. Consul. *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker, New York, 1989.

P. C. Consul and S. P. Mittal. A new urn model with predetermined strategy. *Biometrische Zeitschrift*, 17(2):67–75, 1975.

R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

C. Costa, M. G. Scotto, and I. Pereira. Optimal alarm systems for FIAPARCH processes. *REVSTAT - Statistcal Journal*, 8:37 – 55, 2010.

D. R. Cox, G. Gudmundsson, G. Lindgren, L. Bondesson, E. Harsaae, P. Laake, K. Juselius, and S. L. Lauritzen. Statistical analysis of time series: Some recent developments (with discussion and reply). *Scandinavian Journal of Statistics*, 8(2):93–115, 1981.

M. E. da Silva and V. L. Oliveira. Difference equations for the higher-order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis*, 25(3):317–333, 2004.

M. Dacorogna, R. Gencay, U. Muller, R. Olsen, and P. O.V. *An Introduction to High-Frequency Finance*. Academic Press, San Diego, 2001.

S. Das and M. Kratz. On devising various alarm systems for insurance companies. IIM Bangalore Working Paper 322 and ESSEC Working Paper 10008, ESSEC Business School and Indian Institute of Management, 2010.

S. Das and M. Kratz. Alarm system for insurance companies: A strategy for capital allocation. *Insurance: Mathematics and Economics*, 51(1):53 – 65, 2012.

R. Davis and T. Mikosch. The sample autocorrelations of financial time series models. In W. Fitzgerald, R. Smith, A. Walden, and P. Young, editors, *Nonlinear and Nonstationary Signal Processing*, pages 247–274. Cambridge University Press, Cambridge, 2001.

R. Davis, W. Dunsmuir, and Y. Wang. *Modelling Time Series of Count Data, in Subir Ghosh (Ed.), Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri, pags. 63–116*. Marcel Dekker, New York, 1999.

R. A. Davis, W. T. M. Dunsmuir, and S. B. Streett. Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790, 2003.

R. A. Davis, W. T. Dunsmuir, and S. B. Streett. Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability*, 7(2):149–159, 2005.

J. de Maré. Optimal prediction of catastrophes with applications to Gaussian processes. *The Annals of Probability*, 8(4):841–850, 1980.

J. Dedecker and C. Prieur. Coupling for τ- dependent sequences and applications. *Journal of Theoretical Probability*, 17:861–885, 2004.

P. Diaconis and D. Freedman. De Finetti's theorem for Markov chains. *The Annals of Probability*, 8(1):115–130, 1980.

Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83 – 106, 1993.

J.-P. Dion, G. Gauthier, and A. Latour. Branching processes with immigration and integer-valued time series. *Serdica Mathematical Journal*, 21(2): 123–136, 1995.

P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313 – 342, 1999.

P. Doukhan and O. Wintenberger. Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications*, 118(11):1997 –2013, 2008.

P. Doukhan, K. Fokianos, and D. Tjøstheim. On weak dependence conditions for Poisson autoregressions. *Statistics & Probability Letters*, 82(5):942 – 948, 2012.

J.-G. Du and Y. Li. The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12(2):129–142, 1991.

A.-L. Dutot, J. Rynkiewicz, F. E. Steiner, and J. Rude. A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling & Software*, 22(9):1261–1269, 2007.

F. Eggenberger and G. Pólya. Über die statistik verketteter vorgänge. *ZAMM - Journal of Applied Mathematics and Mechanics - Zeitschrift für Angewandte Mathematik und Mechanik*, 3(4):279–289, 1923.

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

R. F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50, 1986.

R. F. Engle and G. González-Rivera. Semiparametric ARCH models. *Journal of Business & Economic Statistics*, 9(4):345–359, 1991.

R. F. Engle and J. R. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162, 1998.

L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd Ed.* Springer, New York, 2001.

R. Ferland, A. Latour, and D. Oraichi. Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942, 2006.

R. B. Flemming, J. Bohte, and B. D. Wood. One voice among many: The supreme court's influence on attentiveness to issues in the United States, 1947-92. *American Journal of Political Science*, 41(4):1224–1250, 1997.

B. J. Fogarty and J. E. Monogan. Modeling time-series count data: The unique challenges facing political communication. Under revise & resubmit at *Social Science Research*, 2013.

K. Fokianos. Some recent progress in count time series. *Statistics*, 45(1): 49–58, 2011.

K. Fokianos and D. Tjøstheim. Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64(6):1205–1225, 2012.

K. Fokianos, A. Rahbek, and T. Dag. Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439, 2009.

J. Franke. Weak dependence of functional INGARCH processes. Technical Report 126, Technische Universität Kaiserslautern, 2010.

J. Franke and T. Subba Rao. Multivariate first order integer valued autoregressions. Technical report, Mathematics Department, UMIST, 1995.

R. Freeland. *Statistical Analysis of Discrete Time Series with Applications to the Analysis of Workers Compensation Claims Data*. PhD thesis, University of British Columbia, Canada, 1998.

R. K. Freeland and B. P. M. McCabe. Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, 25(5): 701–722, 2004.

A.-M. Fuertes, M. Izzeldin, and E. Kalotychou. On forecasting daily stock volatility: The role of intraday information and market conditions. *International Journal of Forecasting*, 25(2):259 –281, 2009.

H. Furstenberg and H. Kesten. Products of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–469, 1960.

A. R. Gallant, P. E. Rossi, and G. Tauchen. Stock prices and volume. *The Review of Financial Studies*, 5(2):199–242, 1992.

A. Garcia-Ferrer and R. A. Queralt. A note on forecasting international tourism demand in Spain. *International Journal of Forecasting*, 13(4): 539–549, 1997.

V. V. Gavrishchaka and S. B. Ganguli. Volatility forecasting from multi-scale and high-dimensional market data. *Neurocomputing*, 55(1-2):285–305, 2003.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

J. Geweke. Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, 5(1):57–61, 1986.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics, J.M. Bernardo, J.O. Berger, A.P. Dawin and A.F.M. Smith, (Eds.)*, pages 169–193. University Press, 1992.

P. Giot, S. Laurent, and M. Petitjean. Trading activity, realized volatility and jumps. *Journal of Empirical Finance*, 17(1):168–175, 2010.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

D. Gomes. *Processos Autoregressivos de Coeficientes Aleatórios na Modelação de Dados de Contagem.* PhD thesis, Universidade de Évora, Portugal, 2005.

D. Gomes and L. Canto e Castro. Generalized integer-valued random coefficient for a first order structure autoregressive (RCINAR) process. *Journal of Statistical Planning and Inference*, 139(12):4088–4097, 2009.

H. Grage, J. Holst, G. Lindgren, and M. Saklak. Level crossing prediction with neural networks. *Methodology and Computing in Applied Probability*, 12(4):623–645, 2010.

A. Guillou, M. Kratz, and Y. Le Strat. An Extreme Value Theory approach for the early detection of time clusters with application to the surveillance of Salmonella. *ArXiv e-prints*, 2010. URL `http://arxiv.org/pdf/1003.4466`.

A. Gut and J. Hüsler. Extreme shock models. *Extremes*, 2(3):295–307, 1999.

A. Heinen. Modelling time series count data: An autoregressive conditional Poisson model. CORE Discussion Papers 2003062, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2003.

M. L. Higgins and A. K. Bera. A class of nonlinear ARCH models. *International Economic Review*, 33(1):137–158, 1992.

P. A. Jacobs and P. A. W. Lewis. Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(1):94–105, 1978a.

P. A. Jacobs and P. A. W. Lewis. Discrete time series generated by mixtures II: Asymptotic properties. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(2):pp. 222–228, 1978b.

P. A. Jacobs and P. A. W. Lewis. Discrete time series generated by mixtures III: Autoregressive processes (DAR(p)). Technical Report NPS55-78-022, Naval Postgraduate School, Monterey, California, 1978c.

H. Joe. Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3):664–677, 1996.

P. Johansson. Speed limitation and motorway casualties: A time series count data regression approach. *Accident Analysis & Prevention*, 28(1):73–87, 1996.

R. Jung and R. Liesenfeld. Estimating time series models of count data using efficient importance sampling. *Allgemeines Statistisches Archiv*, 85: 387–407, 2001.

R. C. Jung and A. Tremayne. Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis*, 95(1):59–91, 2011.

R. C. Jung and A. R. Tremayne. Binomial thinning models for integer time series. *Statistical Modelling*, 6(2):81–96, 2006.

R. C. Jung, M. Kukuk, and R. Liesenfeld. Time series of count data: Modeling, estimation and diagnostics. *Computational Statistics & Data Analysis*, 51(4):2350–2364, 2006.

J. M. Karpoff. The relation between price changes and trading volume: A survey. *The Journal of Financial and Quantitative Analysis*, 22(1):109–126, 1987.

K. Katsouyanni, G. Touloumi, E. Samoli, A. Gryparis, A. Le Tertre, M. Y., and G. Rossi. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA 2 project. *Epidemiology*, 12(5):521–531, 2001.

B. Kedem and K. Fokianos. *Regression Models for Time Series Analysis*. Wiley, New York, 2002.

T. Kerr. False alarm and correct detection probabilities over a time interval for restricted classes of failure detection algorithms. *IEEE Transactions on Information Theory*, 28(4):619–631, 1982.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973.

H.-Y. Kim and Y. Park. A non-stationary integer-valued autoregressive model. *Statistical Papers*, 49(3):485–502, 2008.

G. Koop and L. Tole. Measuring the health effects of air pollution: To what extent can we really say that people are dying from bad air? *Journal of Environmental Economics and Management*, 47(1):30–54, 2004.

A. Y. C. Kuk and Y. W. Cheng. The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation*, 59(3):233–250, 1997.

D. Lambert and C. Liu. Adaptive thresholds: Monitoring streams of network counts. *Journal of the American Statistical Association*, 101(473):78–88, 2006.

A. Latour. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19(4): 439–455, 1998.

Y. Lee and J. A. Nelder. Conditional and marginal models: Another view. *Statistical Science*, 19(2):219–228, 2004.

R. Liesenfeld, I. Nolte, and W. Pohlmeier. Modelling financial transaction price movements: A dynamic integer count data model. *Empirical Economics*, 30(4):795–825, 2006.

G. Lindgren. Prediction from a random time point. *The Annals of Probability*, 3(3):412–423, 1975a.

G. Lindgren. Prediction of catastrophes and high level crossings. *Bulletin of the International Statistical Institute: Proc. of the 40th Session Warsaw*, 46(Issue 2):225–240, 1975b.

G. Lindgren. Model processes in nonlinear prediction with application to detection and alarm. *The Annals of Probability*, 8(4):775–792, 1980.

G. Lindgren. Optimal prediction of level crossings in Gaussian processes and sequences. *The Annals of Probability*, 13(3):804–824, 1985.

J.-C. Liu. Stationarity of a family of GARCH processes. *Econometrics Journal*, 12(3):436–446, 2009.

I. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall, London, 1997.

B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36:394–419, 1963.

R. A. Martin. A state-space approach to optimal level-crossing prediction for linear Gaussian processes. *IEEE Transactions on Information Theory*, 56:5083–5096, 2010.

R. A. Martin. Extreme value analysis of optimal level-crossing prediction for linear Gaussian processes. *Journal of Time Series Analysis*, 33(4): 583–607, 2012.

B. McCabe and G. Martin. Bayesian predictions of low count time series. *International Journal of Forecasting*, 21(2):315–330, 2005.

E. McKenzie. Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650, 1985.

E. McKenzie. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 18(3):679–705, 1986.

E. McKenzie. Innovation distributions for gamma and negative binomial autoregressions. *Scandinavian Journal of Statistics*, 14(1):79–85, 1987.

E. McKenzie. Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20(4):822–835, 1988.

E. McKenzie. Ch. 16. discrete variate time series. In D. Shanbhag and C. Rao, editors, *Stochastic Processes: Modelling and Simulation*, volume 21 of *Handbook of Statistics*, pages 573–606. Elsevier, 2003.

S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, 1994.

M. Mezzetti, P. Muliere, and P. Bulla. An application of reinforced urn processes to determining maximum tolerated dose. *Statistics & Probability Letters*, 77(7):740–747, 2007.

T. Mikosch. *Modeling Dependence and Tails of Financial Time Series, in B. Finkenstadt and H. Rootzén, (Eds.), Extreme Values in Finance, Telecommunications, and the Environment, pags.185–286*. Chapman & Hall/CRC, Boca Raton, 2003.

T. Mikosch and C. Stărică. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics*, 28(5):1427–1451, 2000.

M. Monteiro. *Contributos para a Análise Estatística de Séries de Contagem*. PhD thesis, Universidade de Aveiro, Portugal, 2010.

M. Monteiro, I. Pereira, and M. G. Scotto. Optimal alarm systems for count processes. *Communications in Statistics - Theory and Methods*, 37(19): 3054–3076, 2008.

P. A. Morettin. *Econometria Financeira, Um Curso em Séries Temporais Financeiras*. Editora Blucher, São Paulo, 2008.

P. Muliere, P. Secchi, and S. Walker. Urn schemes and reinforced random walks. *Stochastic Processes and their Applications*, 88(1):59–78, 2000.

B. J. Murteira, D. A. Müller, and K. F. Turkman. *Análise de Sucessões Cronológicas*. McGraw-Hill, Lisboa, 1993.

P. Neal and T. Subba Rao. MCMC for integer-valued ARMA processes. *Journal of Time Series Analysis*, 28(1):92–110, 2007.

D. B. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(3):318–334, 1990.

H. Neumann, Michael. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, 17(4):1268–1284, 2011.

J. Nordström. Tourism satellite account for Sweden 1992 - 93. *Tourism economics: The business and finance of tourism and recreation*, 2(1):13–42, 1996.

G. Nunnari. An improved back propagation algorithm to predict episodes of poor air quality. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 10(2):132–139, 2006.

S. Pantula. Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, 5(1):71–74, 1986.

J. S. Peake and M. Eshbaugh-Soha. The agenda-setting impact of major presidential TV addresses. *Political Communication*, 25(2):113–137, 2008.

G. G. S. Pegram. An autoregressive model for multilag Markov chains. *Journal of Applied Probability*, 17(2):350–362, 1980.

R. Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007.

F. Pokropp, W. Seidel, A. Begun, M. Heidenreich, and K. Sever. Control charts for the number of children injured in traffic accidents. In H.-J. Lenz and P.-T. Wilrich, editors, *Frontiers in Statistical Quality Control 8*, pages 151–171. Physica-Verlag HD, 2006.

S. Quoreshi. Bivariate time series modelling of financial count data. *Communications in Statistics - Theory and Methods*, 35(7):1343–1358, 2006.

A. E. Raftery. A model for high-order Markov chains. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):528–539, 1985a.

A. E. Raftery. A new model for discrete-valued time series: Autocorrelations and extensions. *Rassegna di Metodi Statistici ed Applicazioni*, 3-4 (Cagliari, 1983/84):149–162, 1985b.

S. Ramos, A. Amaral-Turkman, and M. Antunes. Bayesian classification for bivariate normal gene expression. *Computational Statistics & Data Analysis*, 54(8):2012–2020, 2010.

M. M. Ristić, A. S. Nastić, and A. V. Miletić Ilić. A geometric time series model with dependent Bernoulli counting series. *Journal of Time Series Analysis*, 34(4):466–476, 2013.

T. H. Rydberg and N. Shephard. Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1 (1):2–25, 2003.

J. Schwartz, R. Levin, and K. Hodge. Drinking water turbidity and pediatric hospital use for gastrointestinal illness in Philadelphia. *Epidemiology*, 8 (6):615–620, 1997.

G. Schwert. Stock volatility and the crash of '87. *Review of Financial Studies*, 3(1):77–102, 1990.

L. R. Shenton. *Quasibinomial Distributions*. John Wiley & Sons, Inc., 1986.

N. Shephard. Generalized linear autoregressions. Economics Papers 8, Economics Group, Nuffield College, University of Oxford, Feb. 1995.

J. Siven and J. Lins. Gain/loss asymmetry in time series of individual stock prices and its relationship to the leverage effect. *ArXiv e-prints*, 2009. URL `http://arxiv.org/pdf/0911.4679`.

R. L. Smith, J. M. Davis, J. Sacks, P. Speckman, and P. Styer. Regression models for air pollution and daily mortality: Analysis of data from Birmingham, Alabama. *Environmetrics*, 11(6):719–743, 2000.

F. W. Steutel and K. Van Harn. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7(5):893–899, 1979.

A. Svenson and J. Holst. Prediction of high water levels in the Baltic. *ISTATISTIK, Journal of the Turkish Statistical Association*, 1(3), 1998.

A. Svensson. *Event Prediction and Bootstrap in Time Series.* PhD thesis, Lund Institute of Technology, Sweden, 1998.

A. Svensson and J. Holst. Optimal prediction of level-crossing in piecewise linear Gaussian processes with changing catastrophe level. Technical report, Department of Mathematical Statistics, Lund Institute of Technology, 1997.

A. Svensson, J. Holst, R. Lindquist, and G. Lindgren. Optimal prediction of catastrophes in autoregressive moving average processes. *Journal of Time Series Analysis*, 17(5):511–531, 1996.

T. Sáfadi and I. Pereira. Bayesian analysis of FIAPARCH model: An application to São Paulo stock market. *International Journal of Statistics & Economics*, 5(A10):49–63, 2010.

J. Taylor, Stephen. *Modelling financial time series.* John Wiley & Sons, New York, 1986.

S. J. Taylor. Conjectured models for trends in financial prices, tests and forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 143 (3):338–362, 1980.

T. Teräsvirta. An introduction to univariate GARCH models. In T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen, editors, *Handbook of Financial Time Series*, pages 17–42. Springer Berlin Heidelberg, 2009.

L. C. Thomas. A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2):149–172, 2000.

D. Tjøstheim. Some recent theory for autoregressive count time series. *TEST*, 21(3):413–438, 2012.

A. Tobías and M. G. Scotto. Prediction of extreme ozone levels in Barcelona, Spain. *Environmental Monitoring and Assesment*, 100(1-3):23–32, 2005.

G. Touloumi, R. Atkinson, A. L. Tertre, E. Samoli, J. Schwartz, C. Schindler, J. M. Vonk, G. Rossi, M. Saez, D. Rabszenko, and K. Katsouyanni. Analysis of health outcome time series data in epidemiological studies. *Environmetrics*, 15(2):101–117, 2004.

Y. K. Tse. The conditional heteroscedasticity of the yen-dollar exchange rate. *Journal of Applied Econometrics*, 13(1):49–55, 1998.

A. K. Tsui and K.-Y. Ho. Conditional heteroscedasticity of exchange rates: Further results based on the fractionally integrated approach. *Journal of Applied Econometrics*, 19(5):637–642, 2004.

J. D. Ura. The supreme court and issue attention: The case of homosexuality. *Political Communication*, 26(4):430–446, 2009.

I. D. Vrontos, P. Dellaportas, and D. N. Politis. Full Bayesian inference for GARCH and EGARCH models. *Journal of Business & Economic Statistics*, 18(2):187–198, 2000.

T. Wang and Z. Huang. The relationship between volatility and trading volume in the Chinese stock market: A volatility decomposition perspective. *Annals of Economics and Finance*, 13(1):211–236, 2012.

L. R. Weatherford and S. E. Kimes. A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, 19 (3):401–415, 2003.

C. H. Weiß. Controlling correlated processes with binomial marginals. Preprint 277, Mathematische Institute der Julius-Maximilians-Universität Würzburg, 2007.

C. H. Weiß. Thinning operations for modelling time series of counts - a survey. *AStA Advances in Statistical Analysis*, 92(3):319–341, 2008a.

C. H. Weiß. Serial dependence and regression of Poisson INARMA models. *Journal of Statistical Planning and Inference*, 138(10):2975–2990, 2008b.

C. H. Weiß. Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, 18(4):507–519, 2009.

J. M. Wooldridge. Chapter 8. Quasi-likelihood methods for count data. In M. H. Pesaran and P. Schmidt, editors, *Handbook of Applied Economics Volume II: Microeconomics*. Blackwell Publishing, 1999.

N. Ye, J. Giordano, and J. Feldman. A process control approach to cyber attack detection. *Communications of the ACM*, 44(8):76–82, 2001.

W. Yin. An empirical research on China's stock market's volume-volatility relationship. *IMF World Economic Outlook*, 3:66–79, 2010.

J.-M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955, 1994.

S. L. Zeger. A regression model for time series of counts. *Biometrika*, 75(4): 621–629, 1988.

S. L. Zeger and B. Qaqish. Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, 44:1019–1031, 1988.

H. Zheng, I. V. Basawa, and S. Datta. Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis*, 27(3):411–440, 2006.

H. Zheng, I. V. Basawa, and S. Datta. First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137(1):212–229, 2007.

R. Zhu and H. Joe. A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models*, 19(2):235–254, 2003.