Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

METODOLOGIA STATISTICA PER LA RICERCA SCIENTIFICA

Ciclo XXVI

Settore Concorsuale di afferenza: 13/A5

Settore Scientifico disciplinare: SECS-P/05

ECONOMETRICS OF DEFAULT RISK

Presentata da: Arianna Agosto

Coordinatore Dottorato

Prof. Angela Montanari

Tutor

Prof. Giuseppe Cavaliere

Co-tutor

Prof. Anders Rahbek

Esame finale anno 2012/2013

Contents

Abstract

This thesis is the result of a project aimed at the study of a crucial topic in finance: default risk, whose measurement and modelling have achieved increasing relevance in recent years. We investigate the main issues related to the default phenomenon, under both a methodological and empirical perspective. The topics of default predictability and correlation are treated with a constant attention to the modelling solutions and reviewing critically the literature. From the methodological point of view, our analysis results in the proposal of a new class of models, called Poisson Autoregression with Exogenous Covariates (PARX). The PARX models, including both autoregressive end exogenous components, are able to capture the dynamics of default count time series, characterized by persistence of shocks and slowly decaying autocorrelation.

Application of different PARX models to the monthly default counts of US industrial firms in the period 1982-2011 allows an empirical insight of the defaults dynamics and supports the identification of the main default predictors at an aggregate level.

Acknowledgements

I am grateful to my supervisor Prof. Giuseppe Cavaliere for precious advice and for all I learned from him.

I express my sincere gratitude to my co-tutor Prof. Anders Rahbek for supporting my ideas and for the great experience in Copenhagen.

Thanks to all my research group for useful suggestions and comments. I am particularly grateful to Dr. Luca De Angelis for all his support.

I would like to thank Pablo Barbagallo from Moody's Corporation.

A special thank to Lucia for all the moments we shared in our PhD experience.

I am grateful to Dr. Enrico Moretto, who believes in me more than I do.

Many thanks to my family for teaching me to never give up and, last but not least, to Rocco for all his love and support.

Chapter 1

Introduction to Default Risk

This chapter explains how default risk can be defined and measured, motivating the importance of deriving models for its analysis and prediction. After giving a technical definition of the default event, we illustrate the main empirical evidences in the corporate default phenomenon as well as two crucial topics related to their interpretation - default predictability and correlation between corporate defaults. The structure and the motivation of the thesis work is then presented and connected to the economic and financial issues introduced.

1.1 Default risk: definition and measurement

Default risk is defined as the risk of loss from a counterparty failure to repay the owed amount in terms of either principal or interests of a loan. Default is considered as the most serious event related to credit risk, the last referring to the more comprehensive case of a change in the current value of a credit exposure due to an expected variation of the borrower solvency.

Banks and financial groups are highly involved in both corporate and retail default risk and are required to adopt methodologies for quantifying such risk and thereby determining the amount of capital necessary to support their business and to protect themselves against volatility in the level of losses. The default risk management is included in the Basel II regulation for the stability of the international banking system and comprises both general economic capital requirements and internal rating procedures. A key aspect in default risk management is the measurement of the Probability of Default, i.e. the probability that, following the definition given by the Bank of International Settlements, with regard to a particular obligor either or both of the two events have taken place:

- the bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realising securities (if held)
- the obligor is past due more than 90 days on any material credit obligation to the banking group.

There are two main approaches to default risk modelling: the *structural* and the *reduced form* approach. The first considers default as an endogenously determined event which can be predicted by the economic and financial conditions of the company, reflected in its balance sheet data and market value. Therefore, structural models study the evolution of structural firm variables such as the assets and debt values in order to determine the probability and the timing of bankruptcy, thus explicitly relating default to the first time the assets fall below a certain level - the *default barrier* - as an endogenously determined event. This approach was introduced by the seminal work of Merton (1974), which first relies on the option pricing theory for deriving the probability that the assets fall below the outstanding value of debt. Merton model is based on treating the assets of a firm as a call option held by the stockholders, whose price - the (known) market value - implies the probability of default. This approach is then extended by abandoning some irrealistic assumptions, such as the existence of a fixed default barrier given by the nominal total value of debt. Black and Cox (1976) introduce a time-varying threshold defined as a fraction of the nominal value of liabilities, as it is done by Leland (1994), which also considers the fiscal aspects of bakruptcy decision. Leland and Toft (1996) first evaluate the effects of the presence of coupons and of short-term debt roll-over. A recent development by Agosto and Moretto (2012) determines the curvature parameter of the nonconstant default barrier by using firm-specific balance sheet and market data. *Moody's KMV*, the proprietary model used by the rating agency Moody's for determining the probability of default, is the most famous application of a structural model and is based on the extension of the Merton model developed by Kealhofer, McQuown and Vasicek in 1989.

In contrast to the structural approach, reduced-form models consider default as an exogenously determined process and use immediately available market and credit data - mainly forward rates, rating and price of the issued bonds - rather than modelling the asset value dynamics. Jarrow and Turnbull (1995) and its development Jarrow, Lando and Turnbull (1997), for example, define a model which explicitly incorporates credit rating information into debt instruments pricing and can also be used for risk management purposes as it allows to derive the probabilities of solvency implied by credit spreads. An important class of reduced-form models is that of the so called *intensity* models. They consider the default time as the stochastic first jump time of a count process - Poisson in many cases - whose intensity is a function of latent or observable variables. Their link to probability of default modelling is clear if one thinks that the limit of the intensity of a count process, for a time period approximating zero, is the probability of observing one event. The popularity of intensity models has increased in recent years, as they allow for many econometric applications based on the estimation of default intensity through risk factors and business failure predictors. This approach is followed, for example, by Duffie and Singleton (1999) and Lando (1998) and, as we shall explain, can be effectively used for considering relevant aspects such as dependence between corporate defaults.

Looking at the empirical measures of default risk, the data typically used in risk management and published in rating agencies and financial institutions reports are:

- *default rate*: it is the most widely used measure of the default phenomenon incidence, being defined as the number of defaulting companies in a certain time period divided by the total number of debt issuers in the same period. An alternative definition, that we do not consider here, is the value-weighted default rate, which considers the incidence of defaults in terms of money loss;

- *default count*: it is the number of failures in a certain time period (typically a month). As we shall see, there are several reasons motivating the counting approach to default risk modelling;
- firm-specific measures, such as *distance-to-default*: this is a volatility-adjusted measure calculated and periodically published by Moody's, resulting from the application of the above mentioned KMV model. Following Crosbie and Bohn (2002), it can be defined as "the number of asset value's standard deviations between the market asset value and the default point".

Most of the works presented in the following are focused on default rates or counts modelling and often use "ready-available" measures of firm-specific risk such as distance-to-default.

1.2 The Default Clustering

Looking at the corporate defaults phenomenon under an aggregate perspective, the most relevant aspect is the strong empirical evidence that corporate defaults cluster in time: both default rates and counts show very high peaks, followed by periods of low incidence. This is clear from Figure 1.1, showing the time series of US default rates and counts among Moody's rated industrial firms from 1982 to 2011.

The potentially strong impact of the default clusters on the investors and financial institutions risk has increased the interest of the financial and econometric literature in the two main issues related to the presence of default peaks: default *predictability* and default *correlation*.

First, a central objective in risk management is finding macroeconomic variables and financial indicators that are able to predict the peaks in the number of defaults, in support of financial vigilance and central banks decisions. There are indeed many empirical studies analyzing the strong time variation of default frequencies and linking it to macroeconomic variables and business cycle indicators. This is done, amongst others, by Shumway (2001) and Duffie et al. (2007).

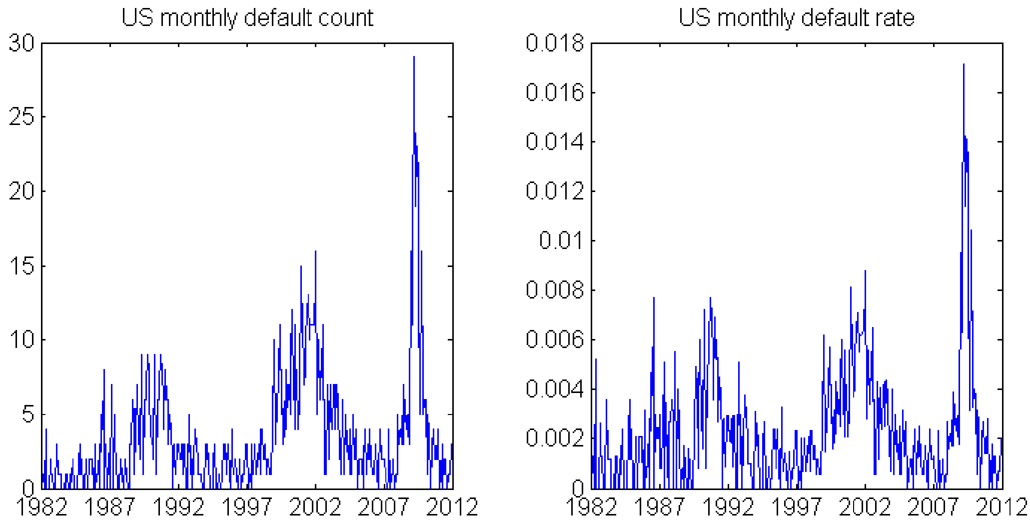

Figure 1.1: (a) Monthly default count of Moody's rated industrial firms from January 1982 to December 2011. (b) Monthly default rate of Moody's rated industrial firms from January 1982 to December 2011.

The interpretation of default clustering is also connected to the issue of correlation, as a high number of defaults in a short period could also be caused by commercial and financial links between the companies. The study of correlation between corporate defaults is an essential tool of credit risk management at portfolio level and its importance has increased in recent years for several reasons. First, banks minimum capital requirements in the Basel II approach are function, among the other things, of the borrowers joint default probability, measured by asset correlation. Second, there has been a large growth of financial instruments like Collateralized Debt Obligations, whose cash flows depend explicitly on default frequency at portfolio level. Furthermore, the evaluation of default probability at the level of an individual security is not able to give an adequate explanation of credit risk spreads, whose dynamics are influenced by commonality in corporate solvency.

The default clustering phenomenon has given rise to a debate about its possible explanation. An important question is whether cross-firm default correlation associated with observable macroeconomic and financial factors affecting corporate solvency is sufficient to explain the observed degree of default clustering or it is possible to document *contagion effects* by which one firm's default increases the likelihood of other firms defaulting. The "cascade" effect which seems to be generated by defaults could spread by means of contractual relationships (customer-supplier or borrower-creditor, for example) or through an "informational" channel, that means a change in the agents expectations of corporate solvency. An increased uncertainty on the credit market leading to a worsening in funding conditions, like credit crunch or higher interest rates, can indeed influence the risk perception. Furthermore, the default clusters could be linked to the systematic (aggregate) risk generated by common macroeconomic and financial risk factors affecting firm solvency: this case is usually excluded from the most strict definition of contagion, that refers instead to between-firms effects on default timing. The works we are going to present in the following chapter are related to default prediction and correlation, investigated through models for aggregate or firm-specific data on default events.

1.3 Motivation and overview

The aim of this work is to study how default risk can be measured and modelled. We contribute to the existing literature by defining, studying and applying a count time series model for the number of corporate defaults, providing a good in- and out-of-sample forecasting of default counts in an extended group of debt issuers.

Our model specification results from the analysis of the stylized facts of corporate default count time series presented in this chapter. First of all, as it often happens for rare events, the default phenomenon is characterized by overdispersion: the variance of the number of events is much higher than its mean, leading to series showing both peaks ("clusters") and periods of low incidence. Moreover, the default count time series are characterized by a slowly decreasing autocorrelation function, which is a typical feature of long-memory processes.

We start, in Chapter 2, with a review of the main econometric and financial models for default risk, with a final focus on intensity models applied to count time series of corporate defaults.

We then present, in Chapter 3, the main models for count data used in econometrics, which rely on the theory of Generalized Linear Models. For several reasons related to the empirical evidences in corporate default count time series, we focus on conditional Poisson models, taking Poisson Autoregression by Fokianos, Rahbek and Tjøstheim (2009) as our main reference. This model (reviewed in Section 3.6) is based on the definition of the count process as a sequence of Poisson drawings which are independent conditional on the past count history. The time-varying intensity (i.e. the expected number of events at time t) is specified as a linear function of lagged counts and intensities. This approach shares some similarities with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) approach for volatility (Bollerslev, 1986). The idea - which can be considered as the first part of our contribution - is that of modelling default clustering in a similar way to the models for volatility clustering, through an autoregressive model which also gives a measure of "persistence" of the series. The dependence of the process (i.e. the number of defaults, in our case) on its past history can indeed explain its long memory and allows to study it under the perspective of shocks persistence. Poisson Autoregression - differently from the traditional Poisson model - also allows for overdispersion.

The consideration that the expected number of defaults is probably influenced by the macroeconomic and financial context in which corporate firms operate has led us to the idea of extending Poisson Autoregression by including exogenous covariates. Thus, in Chapter 4, we present our methodological contribution, developing a class of Poisson intensity AutoRegressions with eXogeneous covariates (PARX) models that can be used for modelling and forecasting time series of counts. We analyze the time series properties and the conditions for stationarity and develop the asymptotic theory for this new model. This way we provide a flexible framework for analyzing dependence of default intensity on both the past number of default events and other relevant financial variables. It is also interesting to consider the impact of including a lagged covariate process on the estimated persistence.

In Chapter 5, we present an extended empirical study of US corporate defaults, based on the application of alternative PARX models. We consider the monthly default counts of US Moody's rated corporate firms: the rating agency Moody's provides monthly and annual reports showing default rates and counts and also offers some instruments for looking more analitically through the data. One of these services is the Credit Risk Calculator, which allows to create customized reports and get data on defaults and rating transitions for specific sectors in a given geographical area. We use a dataset which covers the period from January 1982 to December 2011 and consists in the monthly default counts of US Moody's rated corporate firms classified as "broad industrial", that means it excludes banking, financial and insurance companies as well as public utility and transportation activities. As we will see in the review part, the use of data on industrial firms is common in the corporate defaults analyses. We consider the impact on default intensity of several covariate processes, such as business cycle indicators, production indexes and rating downgrades. For analyzing the link between the financial and the credit market we also include a measure of realized volatility of returns. Realized volatility is expected to summarize the level of uncertainty during periods of financial turmoil when corporate defaults are more likely to cluster and we show that it is significantly and positively associated with the number of defaults.

Chapter 2

Econometric modelling of Default Risk

The two main issues related to the corporate default phenomenon - default predictability and correlation - are now analyzed through an overview of the existing financial and econometric literature of credit risk modelling, with a special focus on the models for default intensity, defined as the expected number of bankruptcies in a given period. These models often include macroeconomic and financial explanatory variables, in the aim of finding both common and firm-specific risk factors for solvency and default predictors. Furthermore, the count modelling framework allows extensions easing the analysis of dependence between default events.

2.1 Default prediction

The most obvious default predictor for a single firm is represented by its business and financial conditions, which can be summarized by balance sheet data such as leverage and net profit measures. This approach is natural in the above mentioned structural models, which are based on the study of the firm's asset evolution, but also characterizes a variety of statistical methods for credit risk measurement, such as *credit scoring*. It is due, for example, to Altman (1968) the development of a multiple discriminant statistical methodology applied to bankruptcy prediction through a set of financial and economic ratios which are shown to successfully discriminate between failing and nonfailing firms. The discriminant function includes variables such as working capital on total assets ratio, market on book value ratio and the sales amount. It is clear that this represents a microeconomic approach which seems not to be suitable when analyzing the default likelihood of large dimension or listed companies, which are expected to be more involved with the overall financial and macroeconomic scenarios.

Recently, there is a growing interest in the specification of models explaining the number or the frequency of corporate defaults with a set of exogenous covariates. An example can be found in Giesecke et al. (2011). They focus on modelling the *default rate*, which is one of the most used measure of the default phenomenon incidence, being defined as the number of defaulting companies in a certain time period divided by the total number of debt issuers in the same period, and periodically published in rating agencies reports. Their empirical analysis considers a large dataset of monthly default rates of US industrial firms, spanning the 1866-2008 period, and is based on the application of a regime-switching model, in the aim of examining the extent to which default rates can be predicted by financial and macroeconomic variables. The econometric specification is the following:

$$D_t = \alpha_t + \beta_k X_{k,t-1} + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.N(0, \sigma^2) \tag{2.1}$$

where $\boldsymbol{X}_{t-1}$ is a k-vector of exogenous explanatory variables and the β_k terms are the corresponding slope coefficients. The intercept term follows a three-state Markov chain taking values α_1, α_2 and α_3 - corresponding to "low", "medium" and "high" default regime respectively - and the π_{ij} probability of transition from state i to state j is the (i, j)-th entry of a transition matrix. Following Hamilton (2005), the model is estimated by a maximum likelihood algorithm based on the recursive updating of the probability $\xi_{i,t}$ of being in state i at time t, the recursion expression being:

$$\xi_{i,t} = \frac{\sum_{i=1}^{3} \pi_{ij} \xi_{i,t-1} \eta_{jt}}{\sum_{i=1}^{3} \sum_{j=1}^{3} \pi_{ij} \xi_{i,t-1} \eta_{jt}} \tag{2.2}$$

with conditional likelihood function η_{jt} given by

$$\eta_{jt} = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-\left(D_t - \alpha_{jt} - \sum_{k=1}^{N} \beta_k X_{k,t-1}\right)^2}{2\pi\sigma^2} \right) \tag{2.3}$$

Among the regressors the authors include both business cycle variables, such as GDP and Industrial Production (IP) growth, and financial covariates (stock returns, change in returns volatility and change in credit spread), as well as the lagged default rate itself. Several covariates, like the change in returns volatility and returns themselves, turn out to be significant in explaining default rates dynamics, while others, such as the growth in Industrial Production and the change in credit spreads, have a low explanatory power. An interesting point - which seems to be not deeply investigated in the paper - is the high value of the lagged default rate coefficient, highlighting the relevance of the autoregressive components in default rate evolution. The maximum likelihood estimate of the time-varying intercept α goes from a minimum of 0.007 in the "low" regime to a value of 0.111 under the worst scenario, so it is in general quite low. The "Dot-Com bubble" of 2001-2002, for instance, corresponds to a high default regime, although its severity is not comparable to other crisis periods such as the Great Depression. Other empirical studies which try to find a connection between the business cycle and the default rates are, amongst others, Kavvathas (2001) and Koopman and Lucas (2005).

A missing element in this kind of approach is the absence of firm-specific variables, which are instead present in other, even previous, works, like Duffie et al. (2007). This article provides maximum likelihood estimators of multi-period conditional probabilities of corporate default incorporating the dynamics of both firm-specific and macroeconomic variables. The empirical analysis is again based on a dataset of defaults among Moody's rated US industrial firms. With regard to the modelling framework, a Cox regression model for counting processes is used: this approach is shared by some of the works related to the analysis of default correlation presented in Section 2.2, so it will be described in detail later. The individual firm covariates considered in Duffie et al. (2007) are the previously defined *distance to default* and the firm trailing stock return, while the overall regressors are the trailing

S&P 500 returns and the three-month Treasury bill rate. It is quite surprising - and also recalls the results of Giesecke et al. (2011) - the lack of significance of other variables, such as credit spreads and GDP growth, that are instead expected to be relevant in default prediction.

2.1.1 The role of rating

When talking about default predictability, an analysis of the role of *credit rating* information cannot be avoided. Rating is, indeed, the main result of the evaluation of a company solvency, made by specialized agencies. The rating information is synthetical and categorical, two features that summarize the potential advantage of this kind of evaluation and explain the wide use of rating in support of pricing and investment decisions. Furthermore, rating agencies methodologies should rely on statistical and econometric models, thus giving a quantitative judgement which is reasonably thought to be objective. However, in the last years some well-known cases like that of Lehman Brothers, whose collapse was not preceeded by any “in time” rating downgrade - Standard & Poor’s maintained the investment-grade rating of “A” and Moody’s downgraded Lehman only one business day before the bankruptcy announcement - has given rise to a burning debate about the possible mistakes in rating evaluation and whether other aspects than a rational and documented quantitative analysis influence the action of rating agencies. Beyond the often unproductive and simplistic discussions trying to mark rating as “good” or “bad”, the question arising in a proper econometric analysis is whether the current rating of a firm is a good predictor of its default probability. There is a double link between rating and the probability of default (henceforth PD). First of all, “default” is one of the classes characterizing the rating scale: class “D” is present in the classification used by all the main rating agencies, such as Fitch, Moody’s and Standard & Poor’s. In the long-term rating assignment, the companies in the “default class” are those that have already failed to repay all or some of their obligations, even in case bankruptcy has not been officially declared yet; in the short-term rating scale, class “D” corresponds to an effective state of insolvency. Secondly, rating agencies periodical material establishes a correspondance between rating classes and PD, based on historical default rates of firms with different rating scores. As an example, we briefly describe the Moody's approach to rating attribution: the output of its proprietary (KMV) model - based on the application of Merton's option pricing formulas in order to derive the market value of assets and its volatility from the market value of equity (firm stocks) - is the so-called Expected Default Frequency (EDF). Figure 2.1 gives a graphical representation of EDF, as the probability that the firm assets fall below a certain threshold over a given time horizon, typically one year or more, based on the hypothesis of log-normal dynamics of the asset value which is typical of Black and Scholes modelling framework.

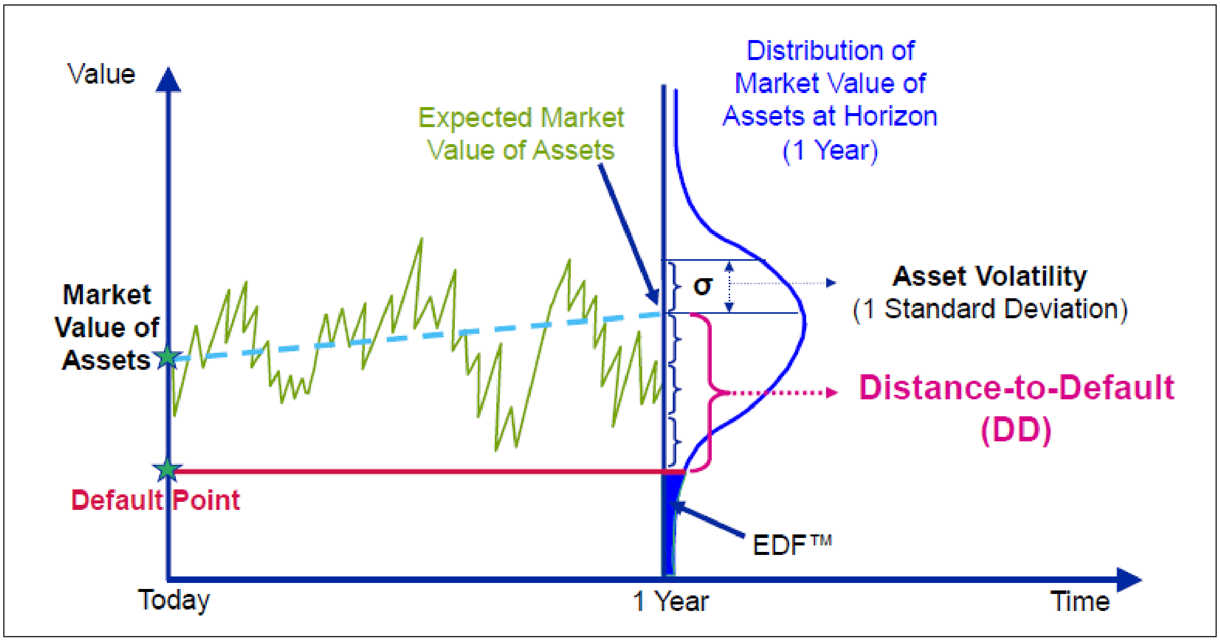

Figure 2.1 Illustration of EDF determined by Moody's KMV. Source: Moody's.

To each interval of EDF, Moody's associates a class of what the agency itself defines as *implied rating* and declares to be a relevant component of the overall rating, the latter also including qualitative and discretionary considerations. Thus, implied rating represents the link between rating and PD.

The econometric analysis of rating is mainly based on the modelling of rating history, that is the changes in a firm rating. This is also motivated by the fact that a kind of information widely used in the risk management of financial institutions is given by the *rating transition matrices*, both historical and forecasted. The general framework of the models for rating, chacterizing, among the others, Jarrow, Lando and Turnbull (1997), is the following. A Markov chain is defined on a finite space of states:

$$S = \{1, 2, ..., k\} \tag{2.4}$$

Each state corresponds to a different rating class, so that the k-th state is the default category, hence we may write, following, as an example, Moody's classification,

$$S = \{AAA, AA, ..., D\}$$

It is assumed that the Markovian process describing rating evolution is homogenous, i.e. its transition matrix does not change in time. The transition matrix Q for (2.4) is defined as follows:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & & \cdots & q_{1,k} \\ q_{2,1} & q_{2,2} & & & q_{2,k} \\ \vdots & & \ddots & & \vdots \\ q_{k-1,1} & q_{k-1,2} & \cdots & q_{k-1,k-1} & q_{k-1,k} \\ 0 & 0 & & \cdots & 1 \end{pmatrix}$$

where the generic entry $q_{i,j}$ is the probability that a company belonging to rating class i in t will have rating j in $t+1$. It is trivial that the following must be verified for $i = 1, ..., k$:

$$\begin{aligned} q_{i,j} &\geq 0 \\ q_{i,i} &= 1 - \sum_{\substack{j=1 \\ j \neq i}}^{k} q_{i,j} \quad \text{for } i = 1, ..., k \end{aligned}$$

Note that the last row corresponds to the obvious assumption that default is an absorbing state, i.e. it is not possible to move from state k to another. The assumed homogeneity implies that matrix, say, $Q(t,T)$, containing the probabilities $q(t,T)$ of being in state i at time t and in j at T, is obtained by simply multiplying it for itself:

$$Q(t,T) = Q^{T-t}$$

The transition probabilities are, in general, given by historical data on average rating changes rates. Another possibility is that of deriving "risk-neutral" transition probabilities by multiplying Q by a matrix containing credit risk premiums estimated from empirical credit spreads.

In this framework, the PD by time T calculated in t is defined as

$$PD(t,T) = 1 - q_{i,k}(t,T) \tag{2.5}$$

This approch is simple but operationally appealing. Lando and Skødeberg (2002) revisit it by introducing a corrected transition matrix that takes into account the rating changes occured between t and T, as ignoring them can lead to underestimate the probability of downgrade. A more complex intensity-based model for rating transitions has been, instead, proposed by Koopman et al. (2008).

With regard to the investigation of the predictive power of rating information through empirical analyses, a common strategy used in econometric works is that of analyzing how much the current rating of a firm really incorporates the stage of the business cycle and the risk profile of its sector, by studying dependence of the published rating transition probabilities on a set of indicators. Nickell et al. (2000) find that business cycle effects have a strong impact on rating especially for low-grade issuers, while Behar and Nagpal (2001) argue that the current rating of a firm seems not to incorporate much of the influence of the macroeconomic context on the default rates.

2.2 Default correlation and Contagion

When modelling the rate or the number of defaults, one of the main objective is finding macroeconomic variables and financial indicators able to predict the peaks in the number of defaults, in support of financial vigilance and central banks decisions. Another crucial topic a great part of the literature focuses on is default correlation: are corporate defaults independent rare events or are there connections between them? First, there are several works supporting the hypothesis of default correlation with empirical analyses. For example, Das et al. (2006) document default correlation -

derived as correlation between individual default probabilities in an intensity-based setting - in various economic sectors and emphasize that correlation effects are time-varying. They further claim that it is possible to distinguish between two "default regimes": a *high* regime characterized by a higher correlation and a *low* regime in which correlation is modest. Another important aspect is the already mentioned possibility of *contagion* effects by which one firm's default directly increases the likelihood of other firms defaulting, generating the "default cascade" effect which seems to characterize the crisis periods. Some examples of contagion models include Davis and Lo (2001), Jarrow and Yu (2001) and Azizpour and Giesecke (2008*a*). These models share the assumption that the default event of one firm directly triggers the default of other firms or causes their default probabilities to increase. A missing element in this kind of modeling is testing the hypothesis of conditional independence between default events, which are probably subject to a common source of randomness due to the mutual exposure to common risk factors. The test of the doubly stochastic assumption, i.e. the assumption that defaults are independent after conditioning on common factors, has been introduced in two recent works about contagion, Das et al. (2007) and Lando and Nielsen (2010), the latter reviewed in the following. Both examine whether default events in an intensity-based setting can be considered conditionally independent testing whether bankruptcy count behaves as a standard Poisson process. This means to verify in an intensity-based setting the *doubly stochastic* assumption, under which the default events are dependent only on exogenous variables.

A distinct class of models for contagion is that of the so-called *frailty* models. They aim at individuating latent (unobservable) factors acting as an additional channel for the spread of defaults. As stated in Azizpour and Giesecke (2008*b*), in frailty models default clustering is indeed explained by three kinds of factors:

- *observable common factors*: changes and shocks in the macroeconomic and financial context;
- *frailty factors*: unobservable common factors affecting corporate solvency;

- *contagion*: the direct negative impact that the default event has on other companies. This can be due to contractual relationships linking firms to each other, but also to the "informational" aspect, as bankruptcy announcements increase the market uncertainty and cause a decrease in the value of stocks portfolio of both industrial and banking firms, with important consequences on credit supply and companies financial conditions. The effects of default announcements are also treated in Lang and Stulz (1992).

In this class of models, including, among others, Duffie et al. (2009), Azizpour et al. (2010) and Koopman et al. (2011), both frailty and contagion effects are analyzed with *self-exciting* point processes. These are characterized by the specification of the conditional instantaneous default intensity of a counting process, that is of the infitesimal rate at which events are expected to occur around a certain time, allowing for dependence on the timing of previous events. The major reference for this approach is the self-exciting process defined by Hawkes (1971).

A different specification of conditional default intensity can be found in Focardi and Fabozzi (2005) and Chou (2012): both use the Autoregressive Conditional Duration (ACD) model introduced by Engle and Russell (1998). In the ACD model, the expectation of the duration, i.e. of the interval between two arrival times, conditional on the past is first specified and the conditional intensity is expressed as the product of a baseline hazard rate - as in the tradition of proportional hazard models for survival data - and a function of the expected duration.

2.3 The study of default correlation through count models

The economic and financial relevance of the default phenomenon, showing peaks of incidence like the sharp one in the crisis period of 2008-2010, has led to an increasing interest in modelling and forecasting time series of corporate default counts. Modeling time series of counts rather than the default rate is quite common and justified by the fact that the default rate denominator - the total number of borrowers in a certain economic sector or rating class - is usually known by the risk managers in a certain advance. It is also possible to note (see Figure 1.1, for instance) that the time series of default counts and default rates share a very similar trend.

2.3.1 Testing conditional independence of defaults

According to the *doubly stochastic* assumption, default events depend uniquely on exogenous variables, that means they are independent conditionally on common macroeconomic and financial factors. A method for testing this assumption is developed by Lando and Nielsen (2010), revisiting the method of *time change* test already used by Das et al. (2007), though reaching different results.

In Lando and Nielsen (2010), the default time of a firm is modelled through its stochastic default intensity. If the firm is alive at time t, then the conditional intensity at time t, i.e. the conditional mean default arrival rate for firm i satisfies

$$\lambda_{it} = \lim_{\Delta t \to 0} \frac{P(t < \tau_i \leq t + \Delta t \mid \tau_i \leq t, \mathcal{F}_t)}{\Delta t} \tag{2.6}$$

where τ_i is the default time for firm i. That means the probability of default within a small time period Δt after t is close to $\lambda_{it}\Delta t$, where λ_{it} depends on information available at time t as represented by $\mathcal{F}_t$.

The individual firm default intensity is then specified through a Cox regression:

$$\lambda_{it} = R_{it} \exp(\beta'_W W_t + \beta'_X X_{it}) \tag{2.7}$$

where W_t is the vector of covariates that are common to all companies whereas X_{it} contains firm-specific variables and R_{it} is a dummy variable which assumes value 1 if firm i is alive and observable at time t, zero otherwise. The crucial point is to determine the firm-specific and macroeconomic variables which are significant explanatory variables in the regression of default intensity.

The Cox regression model was introduced by Cox (1972) in a survival data setting and then extended to the general counting process framework by Andersen and Gill (1982). This approach arises from the Cox proportional hazard model, which is a semi-parametric proportional hazard model making no assumptions about the shape of the baseline hazard function $h(t)$ in the definition of the conditional intensity. The latter is in general expressed as:

$$h\left(t|\, X\right) = h(t) * \exp(\beta_1 X_1 + ... + \beta_p X_p)$$

The theory of Cox regression provides the partial log-likelihood to be maximized by standard techniques in order to draw inference about the parameters vector $\beta = (\beta_W, \beta_X)$:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \int_0^T \left(\beta_W' W_t + \beta_X' X_{it}\right) dN_i(t) \\ &\quad - \sum_{i=1}^{n} \int_0^T R_{it} \exp(\beta_W' W_t + \beta_X' X_{it}) 1_{(\tau_i \geqslant t)} dt \end{aligned} \tag{2.8}$$

where $N_i(t)$ is the one-jump process which jumps to 1 if firm i defaults at time t, n is the total numbers of firms and T is the terminal time point of the estimation.

The cumulative number of defaults among n firms is then defined as:

$$N(t) = \sum_{i=1}^{n} 1_{(\tau_i \leq t)}$$

The objective is to verify the assumption of orthogonality, i.e. that there are never exact simultaneous defaults. Under this assumption, the aggregate default intensity is the sum of the individual ones:

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t) 1_{(\tau_i \leq t)}$$

In order to execute the test, the cumulative default process has to be "time-scaled", that means the scale of time is replaced by the scale of intensity. This is done by defining the compensator

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

that allows to write the time-changed process as

$$J(t) = N(\Lambda^{-1}(t))$$

It is possible to show that $J(t)$ is a unit-rate Poisson process with jump times $V_i = \Lambda(\tau_{(i)})$ where $0 \leq \tau_{(1)} \leq \tau_{(2)} \leq ...$ are the ordered default times. As a consequence, the interarrival times V_1, $V_2 - V_1$,... are independent exponentially distributed variables and, for any $c > 0$, the jump times

$$Z_j = \sum_{i=1}^{n} 1_{]c(j-1),cj]} V_i$$

are independent Poisson variables of c intensity.

Testing orthogonality of defaults means splitting up the entire time period into intervals in which the cumulative integrated default intensity Λ increases by an integer c and verifying, by using several test statistics, if the default counts in each interval are independent and Poisson distributed with mean c. Note that the tested property is the independence of defaults conditional to observable common factors, with the aim of detecting an excess default clustering that is conceivable with the existence of contagion effects.

The data used by the authors are the monthly number of Moody's rated US corporate firms' defaults occured between 1982 and 2005.

With regard to covariates, W_t vector contains the following selection of macroeconomic variables:

- 1-year return on the S&P index
- 3-month US Treasury bill rate
- 1-year percentage change in the US industrial production, calculated from monthly data
- spread between the 10-year and the 1-year Treasury rate

while the firm-specific covariates entering vector X_{it} are:

- 1-year equity return
- 1-year Moody's distance-to-default

- quick ratio, calculated as the sum of cash, short-term investments and total receivables divided by the current liabilities
- log book asset value.

The results obtained in the paper by applying the time-change method and then using several test statistics - like the Fisher dispersion and the upper tail statistics - in order to test the Poisson assumption, lead to accept the hypothesis that default times are conditionally independent, that was rejected in Das et al. (2007). The authors claim that this is due to the use of a different set of explanatory variables and so that the contagion effects apparently revealed by the previous analysis are instead explained by missing covariates. They also argue that the time-change test is actually a misspecification test, as the hypothesis of correct intensity specification is satisfied by construction and that, furthermore, the doubly stochastic assumption is not needed for having orthogonality of default times. They find indeed no evidence of contagion by considering a different specification, that is the Hawkes *self-exciting* process

$$\lambda_{it} = R_{it}\left(\exp(\beta_W' W_t + \beta_X' X_{it}) + \int_0^t (\alpha_0 + \alpha_1 Y_s)\exp(-\alpha_2(t-s)dN_s + \delta)\right)$$

where Y_s is the log book asset value of the firm defaulting at time s. Model (2.13) explicitly includes a contagion effect through an affine function of Y so that larger firms' bankrupt has a higher impact on the individual default intensities. The exponential function makes the default impact decay exponentially with time, with α_2 measuring the time horizon of influence of a default on the overall intensity. Estimation can be carried out by partial maximum likelihood standard instruments (see, for example, Andersen et al., 1992).

In a recent extension of Lando and Nielsen (2010), Lando et al. (2013) replace the Cox multiplicative model with an additive default intensity, based on Aalen (1989) regression model, where the covariate effects act in an additive way on a baseline intensity. The authors claim that the advantage of this model is allowing for the introduction of time-varying effects without the need for estimation procedures more complex than the least squares methods. The focus moves from the test of the conditional independence hypothesis characterizing the previous paper to the search for predictive variables acting on default intensity with nonconstant magnitude. The results are partly different from those reached by the previous analysis: the time-varying effects of firm-specific variables like distance-to-default and short-to-long term debt are found significant, but none of the macroeconomic covariates - many of which already successfully employed in Lando and Nielsen (2010) - are. A problem in the interpretation of results is that some of the coefficients are negative, thus leading to negative default intensities, which is a nonsense from a technical point of view. With regard to this aspect, the authors claim that default intensity should be interpreted as a risk measure rather than an expected rate and that negative values could indicate that a firm is weakly involved in the risk of failure.

2.3.2 An Autoregressive Conditional Duration model of credit risk contagion

The use of self-exciting processes for representing the cascading phenomenon of bankruptcies was already present in another previous work, through a different specification. Focardi and Fabozzi (2005) propose indeed a self-exciting point process. The model belongs to the autoregressive conditional duration (ACD) family introduced by Engle and Russell (1998) and is based on the idea of modelling default clustering with econometric techniques that are the point process analogue of ARCH-GARCH models. Applying the ACD specification to the number of defaults, the default process in a time interval $(0, t)$ is defined as a sequence of default times $t_i, i = 1, 2, ...,$ with the related durations between defaults $\Delta t_i = (t_i + 1 - t_i)$. The model is specified in terms of the conditional densities of the durations, defining

$$E\left[\Delta t_i \mid \Delta t_{i-1}, ..., \Delta t_1\right] = \psi\left[\Delta t_{i-1}, ..., \Delta t_1, \theta\right] = \psi_i \tag{2.9}$$

and

$$\Delta t_i = \psi_i \varepsilon_t \tag{2.10}$$

where ε_t are i.i.d. variables and θ is a parameter.

It is then assumed that the expectation of the present duration is linearly determined by the last m durations between defaults and the last q expectations of durations:

$$\psi_i = \omega + \sum_{j=0}^{m} \alpha_j \Delta t_{i-j} + \sum_{j=0}^{q} \alpha_j \Delta t_{i-j} \tag{2.11}$$

This model is called an ACD(m, q) model.

The authors apply ACD models to simulated data of default durations in order to evaluate the impact of different expected durations on the value of a credit portfolio.

2.4 Concluding remarks

We have investigated how the econometric and financial literature has faced the modelling of default risk and the interpretation of the relative empirical results under the perspective of default predictability and correlation, also clarifying the origin and the issues of the current debate about contagion. The search for explanatory variables in the default rates and counts evolution has led to not always obvious results, because, for example, the link with business cycle indicators and macroeconomic variables does not appear so strong. We have also considered the discussion on the predictive power of rating and described some common approaches to the modelling of rating transitions. We have progressively focused on models which consider count processes for investigating the corporate defaults dynamics. Many of these models aim at analyzing default correlation. With regard to this topic, we claim that the idea of distinguishing between common factors and contagion, thus separating the systematic risk from other risk components, is worth being further investigated. An aspect which seems somewhat missing in the literature yet is that of the autoregressive components in the defaults dynamics, which could lead to interesting considerations about the persistence in the default phenomenon. It is, indeed, present in Focardi and Fabozzi (2005), but without considering the role of covariate processes, so giving a limited definition of contagion which does not take into account crucial aspects of credit and financial risk, and without presenting any application to real data. Our approach to default risk modelling, which we will present in Chapter 4, considers indeed both exogenous variables and autoregressive components and is applied to an empirical corporate default count time series in Chapter 5.

Chapter 3

Econometric modelling of Count Time Series

This chapter presents the main models for count time series. They are based on the theory of Generalized Linear Models for time series, that is reviewed in the first section. The aim of the next sections is to make a critical review, focused on the suitability of the presented models to explain some features commonly found in empirical count time series, such as overdispersion in the data. This is instrumental to the following of our work, which proposes a modelling framework for default count data, based on the extension of the Poisson autoregressive model introduced in the last section.

3.1 Generalized Linear Models for time series

It is well known that generalized linear models (GLM), introduced by Nelder and Wedderburn (1972), allow to extend ordinary linear regression to nonnormal data. Applying the theory of GLM to time series makes thus possible to handle very common processes like binary and count data, which are not normally distributed.

Before presenting the most important applications of GLM to the modelling of count data, it is important to present the concept of *partial likelihood*, introduced by Cox (1975). Partial likelihood is an useful tool when the observations are dependent and the covariates representing auxiliary information are also random and time dependent. In these situations the likelihood function is not readily available as the nonindependence prevents from deriving a simple factorization.

Consider a generic *response* time series $\{y_t\}$, $t = 1, ..., T$. If no other assumption is added, the joint density $f_\theta(y_1, ..., y_T)$, parametrized by vector $\boldsymbol{\theta}$, is defined as

$$f_{\boldsymbol{\theta}}(y_1, ...y_T) = f_{\boldsymbol{\theta}}(y_1) \prod_{t=2}^{T} f_{\boldsymbol{\theta}}(y_t \mid y_1, y_2, ..., y_{t-1}) \tag{3.1}$$

where the main difficulty is that, if no other assumption is made, the size of $\boldsymbol{\theta}$ increases as the series size T does. A more tractable likelihood function can be obtained by introducing limitations in conditional dependence such as Markovianity, according to which we could use, for example, the following factorization:

$$f_{\boldsymbol{\theta}}(y_1, ...y_T) = f_{\boldsymbol{\theta}}(y_1) \prod_{t=2}^{T} f_{\boldsymbol{\theta}}(y_t \mid y_{t-1}) \tag{3.2}$$

where inference regarding $\boldsymbol{\theta}$ can be based only on the product term, as the first factor is not dependent on T.

Then, consider the case where the response variable is observed jointly with some time-dependent random covariate X_t. Then the joint density of the X and Y observations can be written, using conditional probabilities, as:

$$f_{\boldsymbol{\theta}}(x_1, y_1, ..., x_T, y_T) = f_{\boldsymbol{\theta}}(y_1) \left[\prod_{t=2}^{T} f_{\boldsymbol{\theta}}(x_t \mid d_t) \right] \left[\prod_{t=2}^{T} f_{\boldsymbol{\theta}}(y_t \mid c_t) \right] \tag{3.3}$$

where $d_t = (y_1, x_1, ..., y_{t-1}, x_{t-1})$ and $c_t = (y_1, x_1, ..., y_{t-1}, x_{t-1}, x_t)$. The idea of Cox is to take into account only the second product of the right hand side of (3.3), which is a "partial" likelihood in the sense that it does not consider the conditional distribution of the covariate process X_t. Moreover, it does not specify the full joint distribution of the response and the covariate. Cox (1975) shows that the second product term in (3.3) can be used for inference, although it ignores a part of the information about $\boldsymbol{\theta}$.

The general definition of the partial likelihood (PL) relative to $\boldsymbol{\theta}$, $\mathcal{F}_{t-1}$ and the observations $Y_1, ..., Y_T$ applies this idea joint with that of limited conditional dependence mentioned above. Considering only what is known to the observer up to the present time allows for sequential conditional inference:

$$\mathrm{PL}(\boldsymbol{\theta};y_1,...,y_T)=\prod_{t=1}^{T}f_{\boldsymbol{\theta}}(y_t;\theta_t|\mathcal{F}_{t-1})=\prod_{t=1}^{T}f_{\boldsymbol{\theta}}(y_t;\boldsymbol{\theta}) \tag{3.4}$$

where $\mathcal{F}_{t-1}$ is the filtration generated by all is known to the observer by t and possibly including the information given by a random covariate process. Note that this definition simplifies to ordinary likelihood when there is no auxiliary information and the data are independent, while it becomes a conditional likelihood when a deterministic - i.e. known throughout the period of observation - covariate process is included. This formulation enables conditional inference for nonMarkovian processes where the response depends on autoregressive components and past values of covariates, as it does not require the full knowledge of the joint distribution of the response and the covariates.

The vector $\boldsymbol{\theta}$ maximizing equation (3.4) is called the *maximum partial likelihood estimator* (MPLE) and its theoretical properties have been studied by Wong (1986).

We now show how the theory of GLM and partial likelihood can be applied to time series (see Kedem and Fokianos, 2002 for a complete review).

Consider again the response series $\{y_t\}$, $t=1,...,T$ and include a p-dimensional vector of explanatory variables $\mathbf{x}_t=(x_{t,1},...,x_{t,p})'$. Then denote the σ-field generated by $y_{t-1},y_{t-2},...,\mathbf{x}_{t-1},\mathbf{x}_{t-2},...$ as

$$\mathcal{F}_{t-1}=\sigma\left\{y_{t-1},y_{t-2},...,\mathbf{x}_{t-1},\mathbf{x}_{t-2},...\right\}$$

where is often convenient to define $\mathbf{Z}_t=(y_t,\mathbf{x}_t)'$ which contains both the past values of the response and a set of covariates:

$$\mathcal{F}_{t-1}=\sigma\left\{\mathbf{Z}_{t-1},\mathbf{Z}_{t-2},...\right\}$$

The main feature of GLM for time series is the definition of the conditional expectation of y_t given the past of the process $\mathbf{Z}_t$:

$$\mu_t=E\left[y_t\mid\mathcal{F}_{t-1}\right] \tag{3.5}$$

It is worth to note that defining the expected value of y_t as a linear function of the covariates can lead to senseless results when the data are not normal. For instance, linear regression of μ_t on the covariates may lead to negative estimates of intensity when the response is Poisson distributed.

The GLM approach to time series can be stated in two steps:

1. *Random* component: the conditional distribution of the response given the past belongs to the exponential family of distributions, that is

$$f(y_t; \theta_t \mid \mathcal{F}_{t-1}) = \exp\{y_t\theta_t + b(\theta_t) + c(y_t)\} \tag{3.6}$$

   where θ_t is the *natural* (or *canonical*) parameter of the distribution.

   By setting $\prod_{t=1}^{T} f_{\boldsymbol{\theta}}(y_t; \theta_t \mid \mathcal{F}_{t-1}) = \prod_{t=1}^{T} f_{\boldsymbol{\theta}}(y_t; \theta_t)$, the latter product defines a partial likelihood in the sense of Cox (1975), as it is a nested sequence of conditioning history, not requiring the knowledge of the full likelihood.

2. *Systematic* component: there exists a monotone function $g(\cdot)$ such that

$$g(\mu_t) = \eta_t = \sum_{j=1}^{p} \beta_j Z_{(t-1)j} = \mathbf{Z}'_{t-1}\boldsymbol{\beta} \tag{3.7}$$

   where we call $g(\cdot)$ the *link* function, while we refer to η_t as the *linear predictor* of the model, and $\boldsymbol{\beta}$ is a vector of coefficients. It is quite common to include also X_t, i.e. the present value of x, in the covariate vector, if it is already known in $t-1$. It can happen, for instance, when x is a deterministic process or when y_t is a delayed output. We refer, then, to $g^{-1}(\cdot)$ as the *inverse link* function.

3.2 The Poisson Model

3.2.1 Model specification

When handling count data, a natural candidate is the Poisson distribution. If we assume that the conditional density of the response given the past, i.e. the available information up to time t, is that of a Poisson variable with mean λ_t, we get

$$f(y_t; \lambda_t \mid \mathcal{F}_{t-1}) = \frac{\exp(-\lambda_t)\lambda_t^{y_t}}{y_t!}, \quad t = 1, ..., T \tag{3.8}$$

In the Poisson model, the conditional expectation of the response is equal to its conditional variance:

$$E[y_t \mid \mathcal{F}_{t-1}] = Var[y_t \mid \mathcal{F}_{t-1}] = \lambda_t \tag{3.9}$$

Then we denote by $\{\mathbf{Z}_{t-1}\}$, $t = 1, ..., T$ a p-dimensional vector of covariates which may include past values of the response and other auxiliary information. A typical choice for $\mathbf{Z}_{t-1}$ is

$$\mathbf{Z}_{t-1} = (1, y_{t-1}, x_t)'$$

but it is also possible to consider interactions between the processes by defining, for instance, $\mathbf{Z}_{t-1} = (1, y_{t-1}, x_t, y_{t-1}x_t)'$.

Following the theory of of GLM and recalling (3.7), a suitable model is obtained by setting $\mu_t = \lambda_t$ and

$$g(\lambda_t) = \eta_t = \mathbf{Z}'_{t-1}\boldsymbol{\beta} \qquad t = 1, ..., T \tag{3.10}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters.

The most common model is that using the *canonical link* function, which is derived from the canonical form of the Poisson conditional density:

$$f(y_t; \lambda_t \mid \mathcal{F}_{t-1}) = \exp\{(y_t \ \log\lambda_t - \lambda_t) - \log\lambda_t!\}, \ t = 1, ..., T$$

where the natural parameter turns out to be $\log\lambda_t$.

Hence,

$$g(\lambda_t) = \log\lambda_t, \ t = 1, ..., T \tag{3.11}$$

is defined as the canonical link, while the inverse link function g^{-1} guarantees that $\lambda_t > 0$ for every t, as:

$$g^{-1}(\eta_t) = \exp(\eta_t), \quad t = 1, ..., T \tag{3.12}$$

The resulting definition of intensity

$$\lambda_t = \exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta}), \ t = 1, ..., T \tag{3.13}$$

characterizes the so-called *log-linear* model, which has been widely applied in econometrics since Hausman et al. (1984).

3.2.2 Inference

Consider first the estimation of the parameter vector $\boldsymbol{\beta}=(\beta_1, ..., \beta_p)$ for the general case of the Poisson model with $g(\lambda_t) = \mathbf{Z}'_{t-1}\boldsymbol{\beta}$. Recalling (3.4), the partial likelihood function is

$$\begin{aligned} \mathrm{PL}(\boldsymbol{\beta}) &= \prod_{t=1}^{T} f(y_t; \boldsymbol{\beta} \mid \mathcal{F}_{t-1}) \\ &= \prod_{t=1}^{T} \frac{\exp(-\lambda_t(\boldsymbol{\beta}))\lambda_t(\boldsymbol{\beta})^{y_t}}{y_t!} \end{aligned} \tag{3.14}$$

Hence, the partial log-likelihood is the following:

$$\begin{aligned} l(\boldsymbol{\beta}) &\equiv \log \mathrm{PL}(\boldsymbol{\beta}) \\ &= \sum_{t=1}^{T} y_t \log \lambda_t(\boldsymbol{\beta}) - \sum_{t=1}^{T} \lambda_t(\boldsymbol{\beta}) - \sum_{t=1}^{T} y_t \log y_t! \end{aligned} \tag{3.15}$$

The *partial score* function is then obtained by differentiating the log-likelihood:

$$\begin{aligned} \mathbf{S}_T(\boldsymbol{\beta}) &= \nabla l(\boldsymbol{\beta}) = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1}, ..., \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p} \right)' \\ &= \sum_{t=1}^{T} \mathbf{Z}_{t-1} \frac{\partial g^{-1}(\eta_t)}{\partial \eta_t} \frac{1}{\lambda_t(\boldsymbol{\beta})} (y_t - \lambda_t(\boldsymbol{\beta})) \end{aligned} \tag{3.16}$$

Then, the MPLE $\hat{\boldsymbol{\beta}}$ (see Wong, 1986) is obtained by solving the system

$$\mathbf{S}_T(\boldsymbol{\beta}) = \nabla l(\boldsymbol{\beta}) = \mathbf{0} \tag{3.17}$$

which has to be solved numerically, because is nonlinear. Besides the use of standard Newton-Raphson type algorithms, a possible method for solving (3.17) is the *Fisher scoring*, which is a modification of the Newton-Raphson algorithm where the observed information matrix is replaced by its conditional expectation, yielding some computational advantages. The application of the Fisher scoring method to the partial likelihood estimation of the Poisson model is presented in Kedem and Fokianos (2002).

Define first the observed information matrix as

$$\mathbf{H}_T(\boldsymbol{\beta}) = -\boldsymbol{\nabla}\boldsymbol{\nabla}' l(\boldsymbol{\beta}) \tag{3.18}$$

It admits the following decomposition:

$$\mathbf{H}_T(\boldsymbol{\beta}) = \mathbf{G}_T(\boldsymbol{\beta}) - \mathbf{R}_T(\boldsymbol{\beta}) \tag{3.19}$$

where $\mathbf{G}_T(\boldsymbol{\beta})$ is the *cumulative conditional information matrix*, which is defined as

$$\begin{aligned}\mathbf{G}_T(\boldsymbol{\beta}) &= \sum_{t=1}^{T} Cov\left[\mathbf{Z}_{t-1}\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t}\frac{1}{\lambda_t(\boldsymbol{\beta})}(y_t - \lambda_t(\boldsymbol{\beta})) \mid \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^{T} \mathbf{Z}_{t-1}\left(\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t}\right)^2 \frac{1}{\lambda_t(\boldsymbol{\beta})}\mathbf{Z}'_{t-1} \\ &= \mathbf{Z}'\mathbf{W}(\boldsymbol{\beta})\mathbf{Z}\end{aligned} \tag{3.20}$$

where $\mathbf{Z} = (\mathbf{Z}'_0, \mathbf{Z}'_1, ..., \mathbf{Z}'_{T-1})$ is a $T \times p$ matrix and $\mathbf{W}(\boldsymbol{\beta}) = diag(w_1, ..., w_T)$ with entries

$$w_t = \left(\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t}\right)^2 \frac{1}{\lambda_t(\boldsymbol{\beta})}, \quad t = 1, ..., T$$

and

$$\mathbf{R}_T(\boldsymbol{\beta}) = \sum_{t=1}^{T} \mathbf{Z}_{t-1} d_t(\boldsymbol{\beta}) \mathbf{Z}'_{t-1}(y_t - \lambda_t(\boldsymbol{\beta})) \tag{3.21}$$

with $d_t(\boldsymbol{\beta}) = [\partial^2 \log g^{-1}(\eta_t)/\partial \eta_t^2]$.

By substituting $\mathbf{H}_T$ with $\mathbf{G}_T$, if $\mathbf{G}_T^{-1}$ exists, the iterations take the form

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + \mathbf{G}_T^{-1}(\hat{\boldsymbol{\beta}}^{(k)})\mathbf{S}_T(\hat{\boldsymbol{\beta}}^{(k)}) \tag{3.22}$$

An interesting feature of the Fisher scoring is that it can be viewed as an *iterative reweighted least squares* (IRLS) method.

It should indeed be noted that equation (3.22) can be rewritten as

$$\mathbf{G}_T(\hat{\boldsymbol{\beta}}^{(k})\hat{\boldsymbol{\beta}}^{(k+1)} = \mathbf{G}_T(\hat{\boldsymbol{\beta}}^{(k)})\hat{\boldsymbol{\beta}}^{(k)} + \mathbf{S}_T(\hat{\boldsymbol{\beta}}^{(k)}) \tag{3.23}$$

where the right-hand side is a p-dimensional vector whose i-th element is

$$\begin{aligned}&\sum_{j=1}^{p}\left[\sum_{t=1}^{T}\frac{Z_{(t-1)j}Z_{(t-1)i}}{\sigma_t^2}\left(\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t}\right)^2\right]\hat{\beta}_j^{(k)} + \sum_{t=1}^{T}\frac{(y_t - \lambda_t)Z_{(t-1)i}}{\sigma_t^2}\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t} \\ =\ &\sum_{t=1}^{T} Z_{(t-1)i} w_t \left\{\eta_t + (y_t - \lambda_t)\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t}\right\}\end{aligned}$$

Thus, defining

$$\begin{aligned} q_t^{(k)} &= \sum_{t=1}^{T} Z_{(t-1)i}\hat{\beta}_j^{(k)} + (y_t - \lambda_t)\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t} \\ &= \eta_t(\hat{\boldsymbol{\beta}}^{(k)}) + (y_t - \lambda_t)\frac{\partial g^{-1}(\eta_t)}{\partial \eta_t} \end{aligned}$$

and, denoted by $\mathbf{q}^{(k)}$ the T-dimensional vector whose elements are the $q_t^{(k)}$, the right-hand side of (3.23) is equal to $\mathbf{Z}'\mathbf{W}(\boldsymbol{\beta}^{(k)})\mathbf{q}^{(k)}$. By applying (3.20) to the left side, (3.23) becomes

$$\mathbf{Z}'\mathbf{W}(\hat{\boldsymbol{\beta}}^{(k)})\mathbf{Z}\hat{\boldsymbol{\beta}}^{(k+1)} = \mathbf{Z}'\mathbf{W}(\hat{\boldsymbol{\beta}}^{(k)})\mathbf{q}^{(k)}$$

and the iteration simplifies to

$$\hat{\boldsymbol{\beta}}^{(k+1)} = (\mathbf{Z}'\mathbf{W}(\hat{\boldsymbol{\beta}}^{(k)})\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}(\hat{\boldsymbol{\beta}}^{(k)})\mathbf{q}^{(k)} \tag{3.24}$$

The limit for $k \to \infty$ of recursion (3.24) is the maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}$. In each iteration we can recognize the form of the weighted least squares with adjusted weight $\mathbf{W}(\boldsymbol{\beta}^{(k)})$ and the adjusted dependent variable $\mathbf{q}^{(k)}$. For initializing the recursions the conditional means can be replaced by the corresponding responses in order to get a first estimate of the weight matrix $\mathbf{W}$ and hence a starting point for $\boldsymbol{\beta}$.

When the canonical link is used, we have

$$\lambda_t(\boldsymbol{\beta}) = \exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta})$$

and several semplifications are possible. Indeed, for the log-linear model, equations (3.17) and (3.20) become

$$\mathbf{S}_T(\boldsymbol{\beta}) = \sum_{t=1}^{T} \mathbf{Z}_{t-1}(y_t - \lambda_t(\boldsymbol{\beta})) \tag{3.25}$$

and

$$\mathbf{G}_T(\boldsymbol{\beta}) = \sum_{t=1}^{T} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}\lambda_t(\boldsymbol{\beta}) \tag{3.26}$$

Moreover, as $d_t = 0$ in (3.21), $\mathbf{R}_T(\boldsymbol{\beta})$ vanishes and we get

$$\mathbf{H}_T(\boldsymbol{\beta}) = \mathbf{G}_T(\boldsymbol{\beta}) \tag{3.27}$$

thus for the log-linear model the Fisher scoring and Newton-Raphson methods coincide.

3.2.3 Asymptotic theory

In the general theory of GLM the following assumptions (see Fahrmeir and Kaufmann, 1985 for more details) allow to show consistency and asymptotic normality of the MPLE $\hat{\boldsymbol{\beta}}$:

Assumption 1 *The true parameter β belongs to an open set $\beta \subseteq R^p$.*

Assumption 2 *The covariate vector Z_{t-1} almost surely lies in a nonrandom compact subset Γ of R^p, such that $P\left[\sum_{t=1}^{T} \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}' > 0\right] = 1$. In addition, $Z_{t-1}'\beta$ lies almost surely in the domain H of the inverse link function g^{-1} for all $Z_{t-1} \in \Gamma$ and $\beta \in B$.*

Assumption 3 *The inverse link function g^{-1} is twice continuously differentiable and $|\partial h(\gamma)/\partial\gamma| \neq 0$.*

Assumption 4 *There is a probability measure ν on R^p such that $\int_{R^p} zz'\nu(dz)$ is positive definite, and such that, if the conditional distribution of Y_t belongs to the exponential family of distributions in canonical form and under (3.10), for Borel sets $A \subset R^p$,*

$$\frac{1}{T}\sum_{t=1}^{T} I_{[\mathbf{Z}_{t-1}\in A]} \xrightarrow{p} \nu(A)$$

as $T \to \infty$, at the true value of β.

Assumption 4 assures the existence of a $p \times p$ nonrandom *limiting information matrix*

$$\mathbf{G}(\boldsymbol{\beta}) = \int_{R^p} \mathbf{Z}\left(\frac{\partial g^{-1}(\eta)}{\partial\eta}\right)^2 \frac{1}{g^{-1}(\eta)}\mathbf{Z}'\nu(d\boldsymbol{z}) \tag{3.28}$$

with $\eta = \mathbf{Z}'\boldsymbol{\beta}$ such that

$$\frac{\mathbf{G}_T(\boldsymbol{\beta})}{T} \xrightarrow{p} \mathbf{G}(\boldsymbol{\beta}) \tag{3.29}$$

Once stated the above assumption, the following theorem, providing the asymptotic properties of the MPLE, can be presented.

Theorem 3.1 *For the Poisson model, as well as for the general case of GLM, it can be shown that, under assumptions 1-4, the maximum partial likelihood estimator is almost surely unique for all sufficiently large T and*

1. *The MPLE is consistent and asymptotically normal:*

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$$

and

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\beta}))$$

as $T \to \infty$.

2. *The following holds:*

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \frac{1}{\sqrt{T}}\mathbf{G}^{-1}(\boldsymbol{\beta})\mathbf{S}_T(\boldsymbol{\beta}) \xrightarrow{p} \mathbf{0}$$

as $T \to \infty$.

3.2.4 Hypothesis testing

Consider the test of the hypothesis

$$\mathrm{H}_0 : \mathbf{C}'\boldsymbol{\beta} = \boldsymbol{r}$$

where $\mathbf{C}$ is a known $p \times q$ matrix with full rank and $\boldsymbol{r}$ is a known q-dimensional column vector. Then denote by $\boldsymbol{\beta}_0$ the restricted maximum partial likelihood estimator under the null hypothesis.

The most commonly used test statistics for testing H_0 in the context of the Poisson model are:

- the *partial likelihood ratio* statistic

$$LR_T = 2\left\{\log \mathrm{PL}(\hat{\boldsymbol{\beta}}) - \log \mathrm{PL}(\boldsymbol{\beta}_0)\right\} \tag{3.30}$$

- the *Wald* statistic

$$W_T = (\mathbf{C}'\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'(\mathbf{C}'\mathbf{G}^{-1}(\boldsymbol{\beta}_0)\mathbf{C})^{-1}(\mathbf{C}'\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \tag{3.31}$$

- the *partial score* statistic

$$LM_T = \frac{1}{T}\mathbf{S}'_T(\tilde{\boldsymbol{\beta}})\mathbf{G}^{-1}(\tilde{\boldsymbol{\beta}})\mathbf{S}_T(\tilde{\boldsymbol{\beta}}) \tag{3.32}$$

Kedem and Fokianos (2002) prove the following theorem concerning the asymptotic distribution of the test statistics defined above.

Theorem 3.2 *Under the set of assumptions 1-4, the test statistics LR_T, W_T and LM_T are asymptotically equivalent. Furthermore, under* H_0*, their asymptotic distribution is chi-square with r degrees of freedom.*

3.2.5 Goodness of fit

In the context of Poisson regression for count time series, several definitions of residuals can be employed (see Cox and Snell, 1968).

- The *raw* residual is the difference between the response and its conditional expectation:

$$\hat{r}_t = y_t - \lambda_t(\hat{\boldsymbol{\beta}}), \quad t = 1, ..., T \tag{3.33}$$

- The *Pearson* residual is the standardized version of the raw residual, taking into account that the variance of Y_t is not constant:

$$\hat{e}_t = \frac{y_t - \lambda_t(\hat{\boldsymbol{\beta}})}{\sqrt{\lambda_t(\hat{\boldsymbol{\beta}})}}, \quad t = 1, ..., T \tag{3.34}$$

- The *deviance* residual

$$\hat{d}_t = sign(y_t - \lambda_t(\hat{\boldsymbol{\beta}}))\sqrt{l_t(y_t) - l_t(\lambda_t(\hat{\boldsymbol{\beta}}))} \tag{3.35}$$

can be viewed as the t-th contribute to the model *deviance*.

The notion of deviance is based on a likelihood comparison between the *full* (or *saturated*) model and the estimated model. The full model is that where λ_t is estimated directly from the data $y_1, ..., y_T$ disregarding $\boldsymbol{\beta}$, thus it has as many parameters as observations, as in this case the maximum partial likelihood of λ_t is y_t. The estimated model includes $p < T$ parameters instead. Since $l(\mathbf{y}; \mathbf{y}) \geq l(\hat{\lambda}_t; \mathbf{y})$, the *deviance statistic*

$$D = 2\left\{l(\mathbf{y}; \mathbf{y}) - l(\hat{\lambda}_t; \mathbf{y})\right\} \tag{3.36}$$

where $l(\mathbf{y};\mathbf{y}) = \sum_{t=1}^{T} y_t$ has been suggested as a measure of the model overall goodness of fit. Lower positive values correspond to a better fitted model. The deviance statistic has been shown to be have an approximate χ^2_{T-p} distributions under certain conditions (see Mc Cullagh, 1986).

In many generalized linear model, including the Poisson, Pearson residuals are known to be skewed and fat tails. It can be indeed convenient to use a normalizing transformation so that they are more likely to achieve approximate normality under the correct model, like the *Anscombe* residuals. In McCullagh and Nelder (1983) these are defined as:

$$\hat{a}_t = \frac{\frac{3}{2}y^{2/3} - \hat{\lambda}_t^{2/3}}{\hat{\lambda}_t^{1/6}} \tag{3.37}$$

Autocorrelation of Pearson residuals

The large sample properties of the MPLE stated by Theorem 3.1 imply that $\hat{e}_t$ is a consistent estimator of $e_t = \frac{y_t - \lambda_t(\boldsymbol{\beta})}{\sqrt{\lambda_t(\boldsymbol{\beta})}}$, so that the autocorrelation of the e_t's at lag k $\rho_e(k)$ can be consistently estimated by

$$\hat{\rho}_e(k) = \frac{1}{T}\sum_{t=k+1}^{T} \hat{e}_t\hat{e}_{t-k} \tag{3.38}$$

Li (1991) has proved the following theorem relative to the asymptotic distribution of the autocorrelation vector.

Theorem 3.3 *Under the correct model, the vector*

$$\frac{1}{\sqrt{T}}\hat{\boldsymbol{\rho}}_e = \left(\frac{\hat{\rho}_e(1)}{\sqrt{T}}, \frac{\hat{\rho}_e(2)}{\sqrt{T}}, ..., \frac{\hat{\rho}_e(m)}{\sqrt{T}}\right)$$

for some $m > 0$ *is asymptotically normally distributed with mean* $\mathbf{0}$ *and some diagonal limiting covariance matrix* (see Li, 1991 for details).

Testing the "whiteness" of Pearson residuals is used in many applications for goodness of fit analysis, as they should be a white noise, i.e. a sequence of uncorrelated random variables with mean 0 and finite variance, under the correct model (see Kedem and Fokianos, 2002). Plots of the sample autocorrelation function of Pearson residuals with confidence bands at $\pm 1.96/\sqrt{T}$ are commonly used for goodness of fit evaluation.

3.2.6 Model selection

In GLM for count time series, selection among competing models can be based on the traditional information criteria. The Akaike Information Criterion (AIC) introduced by Akaike (1974), in the partial likelihood estimation context is a function of the partial log-likelihood and the number of parameters:

$$\mathrm{AIC}(p) = -2 \log \mathrm{PL}(\hat{\boldsymbol{\beta}}) + 2p \tag{3.39}$$

The model with the number of parameters p which minimizes (3.39) is preferred.

The so-called Bayesian information criterion (BIC), following Schwarz (1978) is defined as

$$\mathrm{BIC}(p) = -2 \log \mathrm{PL}(\hat{\boldsymbol{\beta}}) + p \log T \tag{3.40}$$

3.3 The doubly-truncated Poisson model

The traditional Poisson model can be generalized, as in Fokianos (2001), by assuming that the conditional distribution of the response is *doubly truncated Poisson*. Let $\{Y_t\}$, $t = 1, ..., T$ be a time series of counts and suppose to obmit the values below a known fixed constant $c1$ and exceeding another known fixed constant $c2$, with $c1 < c2$. Then the doubly truncated Poisson conditional density is

$$f(y_t; \lambda_t; c1, c2 \mid \mathcal{F}_{t-1}) = \frac{\exp(-\lambda_t)\lambda_t^{y_t}}{y_t!\psi(c1, c2, \lambda_t)}, \quad t = 1, ..., T \tag{3.41}$$

where the function ψ is defined as

$$\psi(c1, c2, \lambda_t) = \begin{cases} \displaystyle\sum_{y=c1}^{c2} \frac{\lambda_t^y}{y!} & \text{if } 0 \leqslant c1 < c2 \\ \psi(0, c2, \lambda_t) & \text{otherwise} \end{cases}$$

and clearly $\psi(0, \infty, \lambda_t) = \exp(\lambda_t)$ leads to the common Poisson model. This generalization turns out to be useful for modelling truncated count data.

An often used specification is that obtained by setting $c1 = 1$ and $c2 = \infty$. In this case (3.41) becomes:

$$f(y_t; \lambda_t; 1, \infty \mid \mathcal{F}_{t-1}) = \frac{\lambda_t^{y_t}}{y_t!(\exp(-\lambda_t) - 1)}, \quad t = 1, ..., T$$

It should be noted that, differently from the traditional Poisson model, for the truncated Poisson model the conditional mean is not equal to the conditional variance, as

$$E^{tr}\left[y_t; c1, c2 \mid \mathcal{F}_{t-1}\right] = \lambda_t \frac{\psi(c1-1, c2-1, \lambda_t)}{\psi(c1, c2, \lambda_t)}$$

while

$$\begin{aligned} Var^{tr}\left[y_t; c1, c2 \mid \mathcal{F}_{t-1}\right] &= \frac{1}{\psi^2(c1, c2, \lambda_t)}\{\lambda_t^2 \psi(c1-2, c2-2, \lambda_t) \\ &\quad \cdot \psi(c1, c2, \lambda_t) + \lambda_t \psi(c1-1, c2-1, \lambda_t) \\ &\quad \cdot \left[\psi(c1, c2, \lambda_t) - \lambda_t \psi(c1-1, c2-1, \lambda_t)\right]\} \end{aligned}$$

As can be noticed from (3.41), the doubly truncated Poisson distribution belongs to the exponential family of distributions, hence its canonical link is the logarithm and the inverse link is the exponential. Therefore, we obtain again the log-linear model

$$\lambda_t \exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta})$$

and inference is based on maximization of the log-likelihood function derived by (3.41).

3.4 The Zeger-Qaqish model

Zeger and Qaqish (1988) define the following multiplicative model:

$$\begin{aligned} \mu_t(\boldsymbol{\beta}) &= \exp(\beta_0 + \beta_1 x_t + \beta_2 \log(\tilde{y}_{t-1})) \\ &= \exp(\beta_0 + \beta_1 x_t)\tilde{y}_{t-1}^{\beta_2}, \qquad t = 1, ..., T \end{aligned} \tag{3.42}$$

and no distributional assumption for the response y_t is specified. It is clear that, when $\beta_2 < 0$, there is an inverse relationship between $\tilde{y}_{t-1}$ and $\mu_t(\boldsymbol{\beta})$, while the conditional mean grows with $\tilde{y}_{t-1}$ when $\beta_2 > 0$. Observe that, when $\beta_2 < 0$, (3.42) reduces to a log-linear model.

In this formulation $\mathbf{Z}_{t-1} = (1, X_t, \log(\tilde{y}_{t-1}))'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, while $\tilde{y}_{t-1}$ is defined either as

$$\tilde{y}_{t-1} = \max(c, y_{t-1}), \quad 0 < c < 1$$

or

$$\tilde{y}_{t-1} = y_{t-1} + c, \quad c > 0$$

so that $y_{t-1} = 0$ is not an absorbing state.

Equation (3.42) defines the first conditional moment. With respect to the conditional variance it is assumed:

$$Var[y_t \mid \mathcal{F}_{t-1}] = \phi V(\mu_t) \tag{3.43}$$

where $V(\cdot)$ is a known variance function defining the relationship between the conditional mean and the conditional variance, and φ is an unknown dispersion parameter. The so-called *working variance* $\phi V(\mu_t)$ allows to accomodate some features found in the data. For example, the variance model $\phi\mu_t$, with $\varphi > 1$, may hold for count data where the conditional variance exceeds the conditional mean. As can be seen, in this model the assumptions on the response distribution concern only the first and second conditional moments.

A possible extension of (3.42) is the following multiplicative error model:

$$\mu_t(\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_t) \left(\frac{\tilde{y}_{t-1}}{\exp(\beta_0 + \beta_1 x_{t-1})} \right)^{\beta_2} \qquad t = 1, ..., T$$

which can be generalized by considering, as in Kedem and Fokianos (2002), the following model:

$$\mu_t(\boldsymbol{\beta}) = \exp\left[\mathbf{x}_t'\boldsymbol{\gamma} + \sum_{i=1}^{q} \theta_i(\log \tilde{y}_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\gamma})\right] \qquad t = 1, ..., T \tag{3.44}$$

where $\boldsymbol{\beta} = (\boldsymbol{\gamma}', \theta_1, ..., \theta_q)'$ is an $s + q$-dimensional parameter vector and $\{\mathbf{x}_t\}$ is an s-dimensional covariate vector of covariates. Note that when $s = 2$, $q = 1$, $\boldsymbol{\gamma} = (\beta_0, \beta_1)$, $\mathbf{x}_t = (1, x_t)'$ and $\theta_1 = \beta_2$, (3.44) reduces to (3.42).

Turning to the theory of inference for the Zeger-Qaqish model (3.49), we consider the case where c is known. In this case, the estimation of the parameter vector $\boldsymbol{\beta}$ can be carried out by using the *quasi-score* function:

$$\mathbf{S}_T(\boldsymbol{\beta}) = \sum_{t=1}^{T} \mathbf{Z}_{t-1} \frac{\delta\mu_t}{\delta\eta_t} \frac{(y_t - \mu_t(\boldsymbol{\beta}))}{\phi V(\mu_t(\boldsymbol{\beta}))} \tag{3.45}$$

which resembles the score function (3.16), except that the true conditional variance is replaced by the working variance.

According to the theory of quasi-partial maximum likelihood estimation for GLM (see Wedderburn, 1974), the estimator $\hat{\boldsymbol{\beta}}_q$ is consistent and asymptotically normal:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\beta})\mathbf{G}_1(\boldsymbol{\beta})\mathbf{G}^{-1}(\boldsymbol{\beta}))$$

where $\mathbf{G}(\boldsymbol{\beta})$ and $\mathbf{G}_1(\boldsymbol{\beta})$ are the following matrices:

$$\mathbf{G}_T(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_{t-1} \left(\frac{\delta\mu_t}{\delta\eta_t}\right)^2 \frac{1}{\phi V(\mu_t(\boldsymbol{\beta}))} \mathbf{Z}'_{t-1} \xrightarrow{p} \mathbf{G}(\boldsymbol{\beta})$$

and

$$\mathbf{G}_1(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_{t-1} \left(\frac{\delta\mu_t}{\delta\eta_t}\right)^2 \frac{\sigma_t^2(\boldsymbol{\beta})}{\phi^2 V^2(\mu_t(\boldsymbol{\beta}))} \mathbf{Z}'_{t-1} \xrightarrow{p} \mathbf{G}_1(\boldsymbol{\beta})$$

where $\sigma_t^2(\boldsymbol{\beta})$ denotes the true conditional variance. In practice, the covariance matrix of $\hat{\boldsymbol{\beta}}_q$ is estimated by replacing the parameters $\varphi, \boldsymbol{\beta}, \sigma_t^2(\boldsymbol{\beta})$ by their respective estimates. The true conditional variance $\sigma_t^2(\boldsymbol{\beta})$ is replaced by $\left(y_t - \mu_t(\hat{\boldsymbol{\beta}}_q)\right)^2$. The dispersion parameter ϕ can be estimated by

$$\hat{\phi} = \frac{1}{T-s} \sum_{t=1}^{T} \hat{e}_t$$

where $\hat{e}_t$ is the Pearson residual at time t:

$$\hat{e}_t = \frac{y_t - \mu_t(\hat{\boldsymbol{\beta}}_q)}{\sqrt{V(\mu_t(\hat{\boldsymbol{\beta}}_q)}}$$

3.5 Overdispersion and negative binomial regression

The equality of mean and variance characterizing the Poisson model makes it non-suitable when the data show *overdispersion*, i.e. the response variance is higher than the mean. We will show in the following that the introduction of lagged values of the response among the regressors for λ_t allows the unconditional variance to be higher than the unconditional mean, differently from the traditional Poisson model with only exogenous regressors. However, in general, when modelling count data the problem of overdispersion should be addressed. Several *post-hoc* tests - i.e. performed after modelling the data - have been proposed in order to detect overdispersion. One of them is the *Pearson statistic*, defined as the sum of squared Pearson residuals:

$$\chi^2 = \sum_{t=1}^{T} \frac{\left(y_t - \lambda_t(\hat{\boldsymbol{\beta}})\right)^2}{\lambda_t(\hat{\boldsymbol{\beta}})} \tag{3.46}$$

Its distribution was studied, among the others, by McCullagh (1986) and McCullagh and Nelder (1989). Under suitable regularity conditions, its distribution converges to a chi-square with $T - p$ degrees of freedom.

A distribution which is known to fit overdispersed count data is the *negative binomial*. If the conditional density of a time series given the past is that of a negative binomial variable with parameters p_t and r , its distributional law is

$$f(y_t; p_t, r \mid \mathcal{F}_{t-1}) = \binom{y_t + r - 1}{r - 1} p_t^r (1 - p_t)^{y_t}, \quad t = 1, ..., T \tag{3.47}$$

where p_t is the probability that an event occurs in t while r is the *scale* parameter and its inverse $1/r$ is known as the *overdispersion* parameter. The conditional mean $E\left[Y_t \mid \mathcal{F}_{t-1}\right] = \mu = \frac{r(1-p_t)}{p_t}$ is lower than the conditional variance $Var\left[Y_t \mid \mathcal{F}_{t-1}\right] = \frac{r(1-p_t)}{p_t^2}$.

The systematic component of the GLM in the negative binomial case, linking p_t, and thus the expected conditional value, to a set of covariates $\mathbf{Z}$, can be defined, as in Davis and Wu (2009), through the following logit model:

$$-\log\left(\frac{p_t}{1-p_t}\right) = \exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta}) \tag{3.48}$$

yielding

$$\mu = r \exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta}) \tag{3.49}$$

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ maximizes the partial log-likelihood function

$$l(\boldsymbol{\beta}) \equiv \log \mathrm{PL}(\boldsymbol{\beta}) = -r\sum_{t=1}^{T}\log(1+\exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta})) - \sum_{t=1}^{T} Y_t \log(1+\exp(\mathbf{Z}'_{t-1}\boldsymbol{\beta})) + \log\prod_{t=1}^{T}\binom{y_t + r - 1}{r-1} \tag{3.50}$$

Several optimization algorithms have been proposed by Hilbe (2007).

As we said, negative binomial is often used as an alternative to the Poisson model. For testing the Poisson model against the negative binomial distribution, a commonly used test statistic is that characterizing the *Z test*, which Lee (1986) defines as follows:

$$Z = \frac{\sum_{t=1}^{T}\left(y_t - \lambda_t(\hat{\boldsymbol{\beta}})\right)^2 - \lambda_t(\hat{\boldsymbol{\beta}})}{\sqrt{2}\sum_{t=1}^{T}\lambda_t(\hat{\boldsymbol{\beta}})} \tag{3.51}$$

and is shown to have asymptotic standard normal distribution. As the probability limit of the numerator is shown to be positive under the alternative hypothesis that the negative binomial distribution is preferable, a one-sided test is convenient. In particular, the Poisson specification is rejected in favour of the negative binomial with a level of significance α if

$$\sum_{t=1}^{T}\left(y_t - \lambda_t(\hat{\boldsymbol{\beta}})\right)^2 - \lambda_t(\hat{\boldsymbol{\beta}}) > \boldsymbol{c}_\alpha\sqrt{2}\sum_{t=1}^{T}\lambda_t(\hat{\boldsymbol{\beta}})$$

where $\boldsymbol{c}_\alpha$ is the critical value.

3.6 Poisson Autoregression

Fokianos, Rahbek and Tjøstheim (2009), henceforth FRT (2009), study a particular Poisson time series model, characterized by a linear autoregressive intensity and allowing to fit data showing a very slowly decreasing dependence. This model was already existing in literature and shown to fit some financial count data satisfactorily, but FRT (2009) is the first work to study ergodicity and develop the asymptotic theory, which is crucial for likelihood inference.

3.6.1 Model specification

FRT (2009) study the properties of the following Poisson model:

$$\begin{aligned} y_t|\mathcal{F}_{t-1}^{Y,\lambda} &\sim Pois(\lambda_t) \\ \lambda_t &= \omega + \alpha y_{t-1} + \beta\lambda_{t-1} \qquad t \geqslant 1 \end{aligned} \tag{3.52}$$

where the parameters ω, α and β are assumed to be positive. In addition, λ_0 and y_0 are assumed to be fixed.

By introducing for each time point t a "scaled" Poisson process $N_t(\cdot)$ of unit intensity, it is possible to rephrase (3.52) so that the response is defined explicitly as a function of the conditional mean:

$$\begin{aligned} y_t &= N_t(\lambda_t) \\ \lambda_t &= \omega + \alpha y_{t-1} + \beta\lambda_{t-1} \qquad t \geqslant 1 \end{aligned} \tag{3.53}$$

where y_t is then equal to the number of events of $N_t(\cdot)$ in the time interval $[0, \lambda_t]$. The rephrased model (3.53) is found to be more convenient when proving the asymptotic normality of the parameter estimates. Furthermore, expressing y_t as a function of conditional mean - which in the Poisson model is equal the conditional variance - recalls the first defining equation in the GARCH model. It is interesting to note that the sum $(\alpha + \beta)$ can be considered as a measure of persistence in intensity, just as the sum of the ARCH and GARCH parameters in the GARCH model can be read as a measure of persistence in volatility.

Both (3.52) and (3.53) refer to the theory of generalized linear model (GLM) for count time series. Here the random component is the Poisson distribution, as the unobserved process λ_t can be expressed as a function of the past values of the observed process y_t after recursive substitution.

The peculiarities of this approach are mainly two. First, it is characterized by a noncanonical link function - the identity - while, as we have seen, the traditional Poisson model uses the log-linear specification. The other contribution is the introduction of an autoregressive feedback mechanism in $\{\lambda_t\}$, while in the tradition of GLM the intensity is function of a vector of covariates, possibly including the lagged value of the response. This aspect makes the model able to capture a strong persistence with a small number of parameters.

As said before, although FRT (2009) is the first work studying ergodicity of (3.53), that is critical in developing the asymptotic theory, this model was already been considered in the econometric literature. It belongs indeed to the class of observation-driven models for time series of counts studied, among the others, by Zeger and Qaqish (1988) and, more recently, by Davis et al. (2003) and Heinen (2003). The latter defines, in particular, an Autoregressive Conditional Poisson model (ACP), which is a more general form of 3.53 including several lags of counts and intensity. A strong motivation for the analysis of this class of models is that is shown to well approximate some common financial count time series, such as the number of trades in a short time interval (Rydberg and Shephard, 2000 and Streett, 2000).

In particular, Ferland et al. (2006) define model (3.53) explicitly as an an integer-valued GARCH(1,1), i.e. an INGARCH(1,1), and show that Y_t is stationary provided that $0 \leqslant \alpha + \beta < 1$. In particular,

$$E[y_t] = E[\lambda_t] = \mu = \omega/(1 - \alpha - \beta)$$

They further show that all the moments are finite if and only if $0 \leqslant \alpha + \beta < 1$.

Turning to the second moments, as

$$Var[y_t] = \mu \left(1 + \frac{\alpha^2}{1 - (\alpha + \beta)^2}\right)$$

it is immediate to conclude that $Var[Y_t] \geqslant E[Y_t]$, with equality when $\alpha = 0$. Thus, including the past values of the response in the evolution of intensity leads to over-dispersion, a feature often found in real count data.

3.6.2 Ergodicity results

A crucial point in the analysis of this model is to prove the geometric ergodicity of the joint process (y_t, λ_t), where y_t is the observed component, while the intensity process is latent. The notion of geometric ergodicity for a Markov chain process can be summarized as follows. First, the concept of φ-irreducibility has to be introduced. Consider the homogenous Markov chain Z_t defined on a σ-field $\mathcal{M}$ on A, where $P^t(z, B) = P(Z_t \in B \mid Z_0 = z)$ is the probability of moving from $z \in \mathcal{A}$ to the set $B \in \mathcal{M}$ in t steps. The Markov chain (Z_t) is said *ϕ-irreducible* if, for some nontrivial σ-finite measure ϕ on $(A, \mathcal{M})$,

$$\forall B \in \mathcal{M} \quad \phi(B) > 0 \Rightarrow \forall x \in A, \exists t > 0, \quad P^t(z, B) > 0$$

If a ϕ-irreducible Markov chain is positive recurrent (see Meyn and Tweedie, 1996), then there exists a (unique) invariant distribution, that is a probability measure π such that

$$\forall B \in \mathcal{B} \ \ \pi(B) = \int P(z, B)\pi(dz)$$

Finally, (Z_t) is said to be *geometrically ergodic* if there exists a $\rho \in (0, 1)$ such that

$$\forall x \in A \quad \rho^{-t} \left\| P^t(z, \cdot) - \pi \right\| \to 0 \quad \text{as } t \to +\infty$$

Thus, geometric ergodicity states convergence to the invariant distribution.

FRT (2009) succeed in proving geometric ergodicity of (y_t, λ_t) by using an approximated (perturbed) model and proving that is geometrically ergodic under some restrictions on the parameter space. Then, they show that the perturbed model can be made arbitrarily close to the unperturbed one, allowing to extend the results to the latter.

The perturbed model is defined as:

$$\begin{aligned} y_t^m &= N_t(\lambda_t^m) \\ \lambda_t^m &= \omega + \alpha y_{t-1}^m + \beta \lambda_{t-1}^m + \varepsilon_{t,m} \end{aligned} \tag{3.54}$$

where λ_0^m and y_0^m are fixed and

$$\begin{aligned} \varepsilon_{t,m} &= c_m I\left\{y_{t-1}^m = 1\right\} u_t, \\ c_m &> 0 \\ c_m &\longrightarrow 0 \text{ as } m \longrightarrow \infty \end{aligned}$$

where $I\{\cdot\}$ is the indicator function and $\{U_t\}$ is a sequence of i.i.d. uniform random variables on $(0, 1)$ such that $\{U_t\}$ and $\{N_t\}$ are independent. The introduction of

$\{U_t\}$ enables to establish $\phi-$irreducibility, where ϕ is the Lebesgue measure with support $[k, \infty)$ for some $k \geq \lambda^*$, with $\lambda^* = \omega/(1-\beta)$ solution of $\lambda = \omega + \beta\lambda$. The proof that the point λ^* is reachable, and so that $\{\lambda_t\}$ is open set irreducible on $[\lambda^*, \infty)$, provided that $\beta < 1$, is instead given (see FRT, 2009 for details) without using any perturbation.

The following lemma allows to complete the proof of ergodicity of (3.53), establishing that the perturbed model can be made arbitrarily close to the unperturbed one.

Lemma 3.1 *With* (y_t, λ_t) *and* (y_t^m, λ_t^m) *defined by (3.53) and (3.54) respectively, if* $0 \leq \alpha + \beta \leq 1$, *then the following statements hold:*

1. $|E(\lambda_t^m - \lambda_t)| = |E(y_t^m - y_t)| \leq \delta_{1,m}$
2. $E(\lambda_t^m - \lambda_t)^2 \leq \delta_{2,m}$
3. $E(y_t^m - y_t)^2 \leq \delta_{3,m}$

and $\delta_{i,m} \longrightarrow 0$ *as* $m \longrightarrow \infty$ *for* $i = 1, 2, 3$. *Furthermore, with* m *sufficiently large,* $|\lambda_t^m - \lambda_t| \leq \delta$ *and* $|y_t^m - y_t| \leq \delta$ *for any* $\delta > 0$ *almost surely.*

3.6.3 Estimation of parameters

Denoting by $\boldsymbol{\theta}$ the three-dimensional vector of unknown parameters, i.e. $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$, the conditional likelihood function for $\boldsymbol{\theta}$ based on (3.52) in terms of the observations $y_1, ..., y_T$ given the starting values λ_0, y_0 is the following:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{\exp(-\lambda_t(\boldsymbol{\theta}))\lambda_t^{y_t}(\boldsymbol{\theta})}{y_t!} \tag{3.55}$$

where $\lambda_t(\boldsymbol{\theta}) = \omega + \alpha y_{t-1}(\boldsymbol{\theta}) + \beta\lambda_{t-1}$, while, denoting the true parameter vector by $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0)'$, we can write $\lambda_t = \lambda_t(\boldsymbol{\theta}_0)$.

Thus the conditional log-likelihood function is given, up to a constant, by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} (y_t \log \lambda_t(\boldsymbol{\theta}) - \lambda_t(\boldsymbol{\theta})) \tag{3.56}$$

while the score function is

$$\mathbf{S}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1 \right) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{3.57}$$

where $\partial\lambda_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is a three-dimensional vector with components

$$\begin{aligned}\frac{\partial\lambda_t}{\partial\omega} &= 1+\beta\frac{\partial\lambda_{t-1}}{\partial\omega}, \qquad \frac{\partial\lambda_t}{\partial\alpha} = \lambda_{t-1}+\beta\frac{\partial\lambda_{t-1}}{\partial\alpha},\\ \frac{\partial\lambda_t}{\partial\beta} &= y_{t-1}+\beta\frac{\partial\lambda_{t-1}}{\partial\beta}\end{aligned}$$

The solution of $\mathbf{S}_T(\boldsymbol{\theta}) = \mathbf{0}$ yields the conditional maximum likelihood estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$.

The Hessian matrix is then obtained by further differentiation of the score equations (3.57):

$$\begin{aligned}\mathbf{H}_T(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\\ &= \sum_{t=1}^{T}\frac{y_t}{\lambda_t^2(\boldsymbol{\theta})}\left(\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)'\\ &\quad -\sum_{t=1}^{T}\left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})}-1\right)\frac{\partial^2\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\end{aligned} \tag{3.58}$$

In order to study the asymptotic properties of the maximum likelihood estimator for the unperturbed model which are presented in the following, it is again helpful to use the ergodic properties of the perturbed model, whose likelihood function, based on the Poisson assumption and the independence of U_t from (y_t^m, λ_t^m), is defined as

$$L^m(\boldsymbol{\theta}) = \prod_{t=1}^{T}\frac{\exp(-\lambda_t^m(\boldsymbol{\theta}))(\lambda_t^m(\boldsymbol{\theta}))^{y_t^m}}{y_t^m!}\prod_{t=1}^{T} f_u(U_t)$$

where f_u denotes the uniform density. Note that, as $L^m(\boldsymbol{\theta})$ and $L(\boldsymbol{\theta})$ has the same form, then $\mathbf{S}_T^m(\boldsymbol{\theta})$ and $\mathbf{H}_T^m(\boldsymbol{\theta})$ are the counterpart of $\mathbf{S}_T(\boldsymbol{\theta})$ and $\mathbf{H}_T(\boldsymbol{\theta})$, where (y_t, λ_t) are replaced by (y_t^m, λ_t^m).

3.6.4 Asymptotic theory

FRT (2009) prove that the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal by first showing these properties for $\hat{\boldsymbol{\theta}}^m$. For proving consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}^m$ they take advantage of the fact that the log-likelihood function is three times differentiable, which allows to apply Lemma 1 of Jensen and Rahbek (2004). The latter states consistency and asymptotic normality of the maximum likelihood estimator for the traditional GARCH(1,1) model when some assumptions on parameters are relaxed. It is then shown that the score function, the information matrix and the third derivatives of the perturbed likelihood tend to the corresponding quantities of the unperturbed likelihood function. This allows to use proposition 6.3.9 of Brockwell and Davis (1991), stating convergence in distribution of a random vector when some conditions are satisfied.

Before formulating the theorem stating the main result, it is necessary to define the lower and upper values of each component of $\boldsymbol{\theta}$, $\omega_L < \omega_0 < \omega_U$, $\alpha_L < \alpha_0 < \alpha_U < 1$, and $\beta_L < \beta_0 < \beta_U$:

$$\begin{aligned} O(\boldsymbol{\theta}_0) &= \{\boldsymbol{\theta}|0 < \omega_L \leq \omega \leq \omega_U, \\ \boldsymbol{\theta}|0 &< \omega_L \leq \omega \leq \omega_U, \\ 0 &< \alpha_L \leq \alpha \leq \alpha_U < 1 \text{ and} \\ 0 &< \beta_L \leq \beta \leq \beta_U \} \end{aligned}$$

The following theorem states the properties of consistency and asymptotically normality of the maximum likelihood estimator, under a stationarity condition.

Theorem 3.3 *Under model (3.53), assuming that at the true value* $\boldsymbol{\theta}_0, 0 < \alpha_0 + \beta_0 < 1$, *there exists a fixed open neighborhood* $O = O(\boldsymbol{\theta}_0)$ *of* $\boldsymbol{\theta}_0$ *such that with probability tending to* 1, *as* $T \to \infty$, *the log-likelihood function has a unique maximum point* $\hat{\boldsymbol{\theta}}$ *and, furthermore,* $\hat{\boldsymbol{\theta}}$ *is consistent and asymptotically normal:*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\theta}))$$

where the conditional information matrix $\mathbf{G}(\boldsymbol{\theta})$ *is defined as*

$$\mathbf{G}(\boldsymbol{\theta}) = E\left[\frac{1}{\lambda_t(\boldsymbol{\theta})}\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)'\right] \tag{3.59}$$

and can be consistently estimated by

$$\begin{aligned} \mathbf{G}_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} Var\left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \mid \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^{T} \frac{1}{\lambda_t(\boldsymbol{\theta})}\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)' \end{aligned} \tag{3.60}$$

The standard errors of parameter estimates can be obtained from matrix $\mathbf{G}_T(\boldsymbol{\theta})$.

3.7 Concluding remarks

We have reviewed the main models for time series of counts used in econometrics. They belong to the class of GLM and their estimation relies on partial likelihood theory. We have deeply analyzed one of the most used count model, which is the Poisson with log-linear intensity. Then we have introduced a recently developed Poisson model: Poisson Autoregression by Fokianos, Rahbek and Tjøstheim (FRT, 2009). This model defines intensity as a linear function of its own past values and the past number of events and is able to capture the overdispersion and the strong persistence characterizing many count data. As these features are also found in the corporate default count time series, we can think to Poisson Autoregression as an useful tool for the count time series analysis of the default phenomenon.

Chapter 4

A new Poisson Autoregressive model with Exogenous Covariates

We have concluded the previous chapter by presenting Poisson Autoregression by Fokianos, Rahbek and Tjøstheim [FRT] (2009) and explaining its potential advantages in modelling overdispersed and long-memory count data, which are features found in the corporate default counts that will be the object of our empirical study in Chapter 5. Though, this formulation does not consider the role of covariate processes in the intensity dynamics, i.e. in the distribution of the number of events. We claim that including exogenous predictors in the conditional mean specification can enrich the analysis of count time series and also improve the in- and out-of-sample forecasting performance, especially when applying the model to empirical time series strongly linked to the financial and economic context. In this chapter we then propose and develop a class of Poisson intensity AutoRegressions with eXogenous covariates (PARX) models. Extending the theory developed by FRT (2009) allowing for covariate processes requires a strong theoretical effort which is a relevant part of our methodological contribution. First, we provide results on the time series properties of PARX models, including conditions for stationarity and existence of moments. We then provide an asymptotic theory for the maximum-likelihood estimators of the parameters entering the model, allowing inference and forecasting.

4.1 Related literature

The PARX model is related to a recent literature on GARCH models augmented by additional covariates with the aim of improving the volatility forecasting performance. In many cases the lagged squared returns offer just a weak signal about the level of volatility and, as a consequence, the approximation provided by standard GARCH models is poor when volatility changes rapidly to a new level. Realized volatility measures calculated from high-frequency financial data and introduced in the literature by seminal works such as Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002) can be useful to improve the approximation of these models. These measures are found indeed to approximate the level of volatility very well. The first models including realized volatility measures in the GARCH equation are the so-called GARCH-X models estimated by Engle (2002), but are quite incomplete as they do not explain the variation in the realized measures. More complete models are those introduced by Engle and Gallo (2006) and the HEAVY model of Shephard and Sheppard (2010), both specifying multiple latent volatility processes, and the Realized GARCH model of Hansen et al. (2012), which combines a GARCH structure for the daily returns with an integrated model for realized measures of volatility. More generally, there are several works presenting empirical analyses where the time-varying volatility is explained by past returns and volatilities together with additional covariates, typically the volume of transactions as a proxy of the flow of information reaching the market (see, for example, Lamoureux and Lastrapes, 1990 and Gallo and Pacini, 2000). An econometric analysis of ARCH and GARCH models including exogenous covariates can be found in Han and Park (2008) and Han and Kristensen (2013). The PARX shares the same motivation and modelling approach of the presented literature, except that the variable of interest in our case is the time-varying Poisson intensity.

4.2 Specification of PARX models

Consider the Poisson model for the counts y_t, conditional on past intensity and counts, denoted by λ_{t-m} and y_{t-m}, for $m \geq 1$, respectively, as well as past values of an explanatory variable x_t:

$$y_t \mid \mathcal{F}_{t-1} \sim Pois(\lambda_t) \tag{4.1}$$

where $\mathcal{F}_{t-1} = \sigma(y_{t-m},\ \lambda_{t-m},\ x_{t-m}, m \geq 1)$ and λ_t is the, potentially time-varying, Poisson intensity. Following FRT (2009), equation (4.1) can be rewritten in terms of an i.i.d. sequence $N_t(\cdot)$ of Poisson processes with unit-intensity

$$y_t = N_t(\lambda_t) \tag{4.2}$$

The time-varying intensity is specified in terms of the linear link function considered in FRT (2009), here augmented by an exogenous covariate $x_t \in \mathbb{R}$ entering the intensity through a known function $f : \mathbb{R} \to \mathbb{R}^+$:

$$\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_i \lambda_{t-i} + \gamma f(x_{t-1}) \tag{4.3}$$

The parameters of interest are given by $\omega > 0$, and $\alpha_1, ..., \alpha_p$, $\beta_1, ..., \beta_q$ and $\gamma \geq 0$. It is easy to observe that, when $\gamma = 0$, the model reduces to the Poisson Autoregression in FRT (2009). Also note that we define a more general specification, allowing for p lags of the response and q lags of the intensity. We can then use the notation PARX(p, q) in an analogous way as GARCH(p, q) identifies a GARCH models where p lags of the returns and q lags of the volatility are included. The presence of the lagged covariate value rather than the value at time t allows the definition of a conditional intensity that is known at time t given the information available up to time $t - 1$.

In order to carry out multi-step ahead forecasting, we close the model by imposing a Markov-structure on the covariate,

$$x_t = g(x_{t-1}, \varepsilon_t; \eta) \tag{4.4}$$

for some function $g(x, \varepsilon; \eta)$ which is known up to parameter η and where ε_t is an i.i.d. error term. We will assume that $\{\varepsilon_t\}$ and $\{N_t(\cdot)\}$ are mutually independent so that there is no feedback effect from y_t to x_t.

4.3 Time series properties

We here provide sufficient conditions for a PARX process to be stationary and ergodic with polynomial moments of a given order[1]. The analysis is carried out by applying recent results on so-called *weak dependence* developed in Doukhan and Wintenberger (2008). The notion of weak dependence allows to prove the existence of a strictly stationary solution for a large variety of time series models called *chains with infinite memory*, defined by the equation

$$X_t = F(X_{t-1}, X_{t-2}, ..., ; \xi_t) \quad \text{a.s. for } t \in \mathcal{T}$$

where F takes values in a Banach space and ξ_t constitutes an i.i.d. sequence (see Doukhan and Wintenberger, 2008 for details). These models can be seen as a natural extension either of linear models or Markov models. While weak dependence is a slightly weaker concept than the geometric ergodicity used in FRT (2009), it does imply that a strong law of large numbers as well as a central limit theory, both used for the results on econometric inference shown in the following, apply.

Specifically, we make the following assumptions:

Assumption 1 $|f(x) - f(\tilde{x})| \leq L \|x - \tilde{x}\|$, for some $L > 0$ and for every pair of points $x, \tilde{x} \in \mathbb{R}$.

Assumption 2 $E\left[\|g(x; \varepsilon_t) - g(\tilde{x}; \varepsilon_t)\|^s\right] \leq \rho \|x - \tilde{x}\|^s$ for some $\rho < 1$, $s \geq 1$ and for every pair of points $x, \tilde{x} \in \mathbb{R}$, and $E\left[\|g(0; \varepsilon_t)\|^s\right] < \infty$.

Assumption 3 $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$.

Assumption 4 $(\varepsilon'_t, N_t(\cdot))$ are i.i.d.

A few remarks on these assumptions are needed.

First, Assumption 1 states that f satisfies the Lipschitz condition. This assumption will be weakened in the following in order to gain flexibility in the choice of the function f.

Assumption 2 concerns, instead, a function g defining the structure of the covariate process and requires it to be L_s-Lipschitz for all values of x. This is a key

[1] All theorems and lemmas are proved in Appendix A.

assumption when proving stationarity of many popular time series models, including the linear autoregressive ones.

Assumption 3 implies that the function $L(y,\lambda) = \omega + \sum_{i=1}^{p} \alpha_i y_i + \sum_{i=1}^{q} \beta_i \lambda_i$ is Lipschitz. This assumption is imposed in Doukhan and Wintenberger (2008) for applying the weak dependence theory and it is identical to the condition imposed in FRT (2009) for the Poisson autoregressive model.

Finally, Assumption 4 rules out dependence in the two error terms driving the model. It could be weakened, still satisfying the conditions of Doukhan and Wintenberger (2008), by allowing the two joint innovation terms to be Markov processes. This would accomodate "leverage intensity-effects" if $\{\varepsilon_t\}$ and $\{N_t(\cdot)\}$ are negatively correlated. Though, for our purpose here we maintain Assumption 4.

In the following we provide a theorem stating the existence of a stationary solution for process y_t under the assumptions defined above. Before stating it, we briefly present the theory of weak dependence developed by Doukhan and Wintenberger (2008). They use the notion of weak dependence introduced by Dedecker and Prieur (2004) and defined as follows.

Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-subalgebra of $\mathcal{C}$ and Z a generic random variable with values in A. Assume that $\|Z\|_1 < \infty$, where $\|\cdot\|_m$ denotes the L^m norm, i.e. $\|Z\|_m^m = E\|Z\|^m$ for $m \geq 1$, and define the coefficient τ as

$$\tau(\mathcal{M}, Z) = \left\| \sup \left\{ \left| \int f(z) \mathbb{P}_{X|\mathcal{M}}(dz) - \int f(z) \mathbb{P}_X(dz) \right| \quad \text{with } f \in \Lambda_1(A) \right\} \right\|_1$$

An easy way to bound this coefficient is based on a coupling argument:

$$\tau(\mathcal{M}, Z) \leq \|Z - W\|_1$$

for any W with the same distribution as Z and independent of $\mathcal{M}$. Under certain conditions on the probability space $(\Omega, \mathcal{C}, \mathbb{P})$ (see Dedecker and Prieur, 2004), then there exists a Z^* such that $\|Z - Z^*\|_1$ and, using the definition of τ, the dependence between the past of the sequence $(Z_t)_{t\in\mathcal{T}}$ and its future k-tuples may be assessed. Consider the norm $\|z - w\| = \|z_1 - w_1\| + ... + \|z_k - w_k\|$ on A^k, set $\mathcal{M}_p = \sigma(Z_t, t \leq p)$ and define

$$\begin{aligned}\tau_k(r) &= \max_{1\le l\le k}\frac{1}{l}\sup\left\{\tau(\mathcal{M}_p,(Z_{j1},...,Z_{jl}))\ \text{ with } p+r\le j_1,...,j_l\right\},\\ \tau_\infty(r) &= \sup_{k>0}\tau_k(r)\end{aligned}$$

The time series $(Z_t)_{t\in\mathcal{T}}$ is said τ-weakly dependent when its coefficients $\tau_\infty(r)$ tend to 0 as r tends to infinity. The notion of geometric ergodicity (see 3.6.2) is stronger and refers to the rate of convergence of the Markov chain transition probabilities to the invariant distribution. It requires the ϕ-irreducibility of the Markov chain and in FRT (2009) is shown for an approximated (perturbated) Poisson Autoregressive model.

Theorem 4.1 *Under Assumptions 1-4 there exists a τ-weakly dependent stationary and ergodic solution $X_t^*=(y_t^*,\lambda_t^*,x_t^*)$ with $E\left[\|X_t^*\|^s\right]<\infty$ and $\tau(r)=\max\left(\sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i),\rho\right)$.*

The above theorem complements the results of FRT (2009). Note that here we provide sufficient conditions for weak dependence of the actual model, not an approximated version. On the other hand, we do not show the stronger property of geometric ergodicity.

Given the existence of a stationary distribution, it can easily be shown that

$$E[y_t]=E[\lambda_t]=\mu=\frac{\omega+\gamma E\left[f(x_{t-1})\right]}{1-\ \sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i)}$$

and furthermore $Var\left[y_t\right]\ge E[y_t]$. Thus, by including past values of the response and covariates in the evolution of the intensity, the PARX model generates overdispersion, which is a prominent feature in many count time series.

An important consequence of Theorem 4.1 is that, using again the results of Doukhan and Wintenberger (2008), if Assumptions 1-4 are satisfied then the (strong) law of large numbers (LLN) applies to any function $h(\cdot)$ of $X_t=(y_t,\lambda_t,x_t)$ provided $E\left[\|h(X_t^*)\|\right]<\infty$. As a lemma we note that the same applies independently of the choice of initial values (y_0,λ_0,x_0), that is:

Lemma 4.1 *If $X_t=F(X_{t-1},\xi_t)$ with ξ_t i.i.d. and X_t τ-weakly dependent, then $\frac{1}{T}\sum_{t=1}^{T}h(X_t)\overset{a.s.}{\to}E\left[h(X_t^*)\right]$ provided that $E\left[\|h(X_t)\|\right]<\infty$.*

Note that no role is played by the initial values in what stated above.

Also observe that when ε_t is an i.i.d.$(0, \sigma^2)$ sequence and $E\left[h^2(X_t^*)\right] < \infty$, it follows by Lemma 4.1 and a CLT for martingales (see Brown, 1971) that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} h(X_t)\ \varepsilon_t \xrightarrow{d} N(0, \sigma^2 E\left[h^2(X_t^*)\right]) \tag{4.5}$$

It is worth remarking that the Lipschitz condition in Assumption 1 rules out some unbounded transformations $f(x)$ of x_t, such as $f(x) = \exp(x)$.

In order to handle such situations we introduce a truncated model:

$$\lambda_t^c = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \lambda_{t-i}^c + \gamma f(x_{t-1}) I\left\{\|x_{t-1}\| \leq c\right\} \tag{4.6}$$

for some cut-off point $c > 0$.

We can then relax Assumption 1 allowing $f(x)$ to be locally Lipschitz in the following sense:

Assumption 1' *For all* $c > 0$, *there exists some* $L_c < \infty$ *such that*

$$|f(x) - f(\tilde{x})| \leq L\|x - \tilde{x}\|, \quad \|x\|, \|\tilde{x}\| \leq c$$

By replacing Assumption 1 with Assumption 1', we now obtain, by identical arguments as in the proof of Theorem 4.1, that the truncated process has a weakly dependent stationary and ergodic solution. Though this approach recalls the approximated GARCH-type Poisson process introduced in FRT (2009), the reasoning is different. In FRT (2009) an approximated process was needed to establish geometric ergodicity of the Poisson process, while here we introduce the truncated process in order to handle the practice - often used in literature - of introducing non-log realized volatility measures as exogenous covariate. Note that, as $c \to \infty$, the truncated process approximates the untruncated one $(c = +\infty)$ in the following sense:

Lemma 4.2 *Under Assumptions 1'- 4 together with* $E\left[f(x_t^*)\right] < \infty$,

$$\begin{aligned} \left|E\left[\lambda_t^c - \lambda_t\right]\right| &= \left|E\left[y_t^c - y_t\right]\right| \leq \delta_1(c), \\ E\left[\lambda_t^c - \lambda_t\right]^2 &\leq \delta_2(c), \quad E\left[y_t^c - y_t\right]^2 \leq \delta_3(c) \end{aligned}$$

where $\delta_k(c) \to 0$ as $c \to \infty$, $k = 1, 2, 3$.

The above result is akin to Lemma 2.1 in FRT (2009). The additional assumption of $E[f(x_t^*)]$ being finite needs to be verified on a case-by-case basis. For example, with $f(x) = exp(x)$, then this holds if x_t^* has a Gaussian distribution, or some other distribution for which the moment generating function, or Laplace transform, is well-defined.

4.4 Maximum likelihood estimation

Denote by $\boldsymbol{\theta} = (\omega, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \in \mathbb{R}^{p+q+2}$, where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)'$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)'$ the set of unknown parameters entering the PARX model in (4.2)-(4.3). The conditional log-likelihood function in terms of observations $y_1, ..., y_T$, given the initial values $(\lambda_0, \lambda_{-1}, ..., \lambda_{-q+1}, y_0, y_{-1}, ..., y_{-p+1})$, takes the form

$$L_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \qquad \text{where } l_t(\boldsymbol{\theta}) = \; y_t \log \lambda_t(\boldsymbol{\theta}) - \lambda_t(\boldsymbol{\theta}) \tag{4.7}$$

where we have left out a constant term and

$$\lambda_t(\boldsymbol{\theta}) = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \lambda_{t-i}(\boldsymbol{\theta}) + \gamma f(x_{t-1})$$

The maximum likelihood estimator is then computed as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta \in \Theta} L_T(\boldsymbol{\theta}) \tag{4.8}$$

where $\Theta \subset \mathbb{R}^{p+q+2}$ is the parameter space.

We now impose the following conditions on the parameters:

Assumption 5 *Assume that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{p+q+2}$, with Θ compact and $\boldsymbol{\theta}_0 \in int\Theta$. Moreover, for all $\boldsymbol{\theta} = (\omega, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \in \Theta$, $\beta_i \leq \beta_U < 1/q$ for $i = 1, 2, ..., q$ and $\omega \geq \omega_L > 0$.*

Under this assumption together with the ones used to establish stationarity of the model, we obtain the following asymptotic result for the maximum likelihood estimator:

Theorem 4.2 *Under Assumptions 1-5, $\hat{\boldsymbol{\theta}}$ is consistent and*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, G^{-1}) \qquad G = -E\left[\left. \frac{\delta^2 l_t(\boldsymbol{\theta})}{\delta \boldsymbol{\theta} \delta \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right] \tag{4.9}$$

An important remark is the following. If the distribution of y_t is misspecified, thus there is an error term in the definition of intensity, but it still holds that $E[y_t] = \lambda_t$, we expect the asymptotic properties of the maximum likelihood estimator to remain correct except that the asymptotic variance now takes the sandwich form $G^{-1}\Omega G^{-1}$ where

$$\Omega = E\left[\left. \frac{\delta l_t(\boldsymbol{\theta})}{\delta \boldsymbol{\theta}} \frac{\delta l_t(\boldsymbol{\theta})}{\delta \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right]$$

See Gourieroux et al. (2004) for an analysis of Quasi-Maximum Likelihood Estimation (QMLE) of Poisson models.

Theorem 4.2 generalizes the result of FRT (2009) to allow for estimation of parameters associated with additional regressors in the specification of λ_t. By combining the arguments in FRT (2009) with Lemma 4.2, the asymptotic result can be extended to allow f to be locally Lipschitz (see Assumption 1').

More precisely, we define the likelihood quantities for the approximated, or truncated, model as

$$L_T^c(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t^c(\boldsymbol{\theta}), \qquad \text{where } l_t^c(\boldsymbol{\theta}) = \; y_t^c \log \lambda_t^c(\boldsymbol{\theta}) - \lambda_t^c(\boldsymbol{\theta}) \tag{4.10}$$

It immediately follows that the results of Theorem 4.2 holds for the QMLE $\hat{\boldsymbol{\theta}}^c$ of $L_T^c(\boldsymbol{\theta})$. However, as the approximated likelihood function can be made arbitrarily close to the true likelihood as $c \to \infty$, one can show that we can replace Assumption 1 in Theorem 4.2 by Assumption 1':

Theorem 4.3 *Under Assumptions 1', 2-5 and $E\left[f(x_t^*)\right] < \infty$,* then $\hat{\boldsymbol{\theta}}$ *is consistent and*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, G^{-1}) \qquad G = -E\left[\left. \frac{\delta^2 l_t(\boldsymbol{\theta})}{\delta \boldsymbol{\theta} \delta \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right] \tag{4.11}$$

With the above theorem we have generalized the asymptotic results by allowing the assumptions on function f to be relaxed.

4.5 Forecasting

The PARX model can be used to generate forecasts of both the intensity, λ_t, and the number of events, y_t. It is important to remark that, for multi-step forecasting, we also need to estimate the model for x_t as given in (4.4). Given that x_t is exogenous, we can estimate the parameters entering equation (4.4) independently of $\boldsymbol{\theta}$. If no model is available for x_t, only one-step ahead forecasts are possible. In the following, we treat the parameters entering the model as known for notational ease. In practice, the unknown parameters are simply replaced by their estimates. Forecasting of Poisson autoregressive processes is similar to forecasting of GARCH processes (see, e.g., Hansen et al, 2012, Section 6.2) since it proceeds in two steps. First, a forecast of the time-varying parameter - the variance in the case of GARCH, the intensity in the case of PARX - is obtained; then, this is substituted into the conditional distribution of the observed process y_t.

Consider the forecasting of λ_t. A natural one-step ahead forecast is

$$\lambda_{T+1 \mid T} = \omega + \sum_{i=1}^{p} \alpha_i y_{T+1-i} + \sum_{i=1}^{q} \beta_i \lambda_{T+1-i} + \gamma f(x_T) \tag{4.12}$$

More generally, a multi-step ahead forecast of the distribution of y_{T+h}, for some $h > 1$, takes the form

$$F_{T+h \mid T}(y) = F\left(y \mid \lambda_{T+h \mid T}\right)$$

where $\lambda_{T+h \mid T}$ is the final output of the following recursion:

$$\lambda_{T+h \mid T} = \omega + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)\lambda_{T+k-i \mid T} + \gamma f(x_{T+k-i \mid T}), \qquad k = 1, ..., h \tag{4.13}$$

where the initial value $\lambda_{T+1 \mid T}$ derives from (4.12) and $x_{T+k \mid T}$, $k = 1, ..., h-1$, is obtained from some forecast procedure based on (4.4). For example, if the model for x_t is an AR, the natural forecast is

$$y_{T+h \mid T} := E\left[y_{T+h} \mid \mathcal{F}_t\right] = \lambda_{T+h \mid T}$$

together with the $1-\alpha$ confidence interval (as implied by the forecast distribution) for some $\alpha \in (0,1)$. The symmetric $1-\alpha$ confidence interval takes the form

$$CI_{1-\alpha} = \left[Q\left(\alpha/2 | \lambda_{T+h \mid T}\right), Q\left(1-\alpha/2 | \lambda_{T+h \mid T}\right)\right]$$

where $p \longmapsto Q(p|\lambda)$ denotes the quantile function of a Poisson distribution with intensity λ. The quantile function is available in standard statistical software packages, such as Matlab. The forecasting results can be used to evaluate competing PARX models, e.g. based on different choices of covariates. A number of different tests have been proposed in the literature for comparing forecasting models. One can either use forecast evaluation methods based on point forecast, $y_{T+h\,|\,T}$, as proposed in, among others, Christoffersen and Diebold (1997). Alternatively, the evaluation of the forecast distribution can be made by using the so-called *scoring rules* (Diebold et al., 1998). These take as starting point some loss function $S(P, y)$ whose arguments are the probability forecast, P, and the future realization, y. For instance, the log-score, $S(P, y) = \log P(y)$ can be used for ranking probability forecast methods by comparing their average scores. A test based on the scoring rules is the likelihood ratio test studied by Amisano and Giacomini (2007). Suppose we have two competing PARX models with corresponding intensity forecasts $\lambda^{(1)}_{T+h\,|\,T}$ and $\lambda^{(2)}_{T+h\,|\,T}$. We then define the corresponding log-likelihood functions given the actual outcome in period $T + h$,

$$\lambda^{(k)}_{T+h\,|\,T} = y_{T+h} \log \lambda^{(k)}_{T+h\,|\,T} - \lambda^{(k)}_{T+h\,|\,T}\ , \qquad k = 1, 2$$

and compare the two forecasting models in terms of the Kullback-Leibler distance across $k \geq 1$ realizations and corresponding forecasts

$$LR = \frac{1}{k+1} \sum_{T=m}^{m+k} \left\{ \lambda^{(1)}_{T+h\,|\,T} - \lambda^{(2)}_{T+h\,|\,T} \right\}$$

where $m \geq 1$ is the "training sample size" with $\{y_t, x_t : t = 1, ..., m\}$ being used to obtain the parameter estimates. If $LR > 0$ (< 0) we prefer the first (second) model. Amisano and Giacomini (2007) show that LR follows a normal distribution as $k \to \infty$.

4.6 Finite-sample simulations

In this section we present a simulation study with the aim of evaluating the performance of MLE for PARX models. We consider the results of simulations from

PARX models with different covariate processes, mainly distinguishing between long-memory and short-memory processes. The objective is indeed to show not only the satisfactory performance of the estimation algorithm, but also the flexibility of PARX in terms of choice of the covariates.

4.6.1 Simulation design

This experiment[2] is focused on the finite-sample behaviour of MLE for PARX models. We evaluate the parameter estimates for different sample sizes, in order to verify not only the accuracy but also the convergence to the asymptotic Gaussian distribution. In particular, our study is organized as follows. We simulate and fit the PARX(1,1) model

$$y_t \mid \mathcal{F}_{t-1} \sim Pois(\lambda_t)$$

$$\lambda_t = \omega + \alpha_1 y_{t-1} + \beta_1 \lambda_{t-1} + \gamma \exp(x_{t-1})$$

Though here our Monte Carlo experiment is shown for a PARX(1,1) model only, the results are very similar if more lags of the response and intensity are included. We choose the exponential function as the positive function f for including the generated exogenous covariate in the model (see Equation 4.3). This allows to evaluate the parameter estimates when the Lipschitz condition on f is relaxed, allowing for unbounded transformation to be employed (see assumption A'). The exponential transformation will also be used in our empirical study.

We examine different cases, based on alternative choices of the function $g(x, \varepsilon; \eta)$ in

$$X_t = g(x_{t-1}, \varepsilon_t; \eta)$$

The cases included in our simulation design are the following:

- **Case 1**: stationary AR(1) covariate

  $x_t = \varphi x_{t-1} + \varepsilon_t$

  $\phi = 0.50$

[2] We use Matlab for writing the data generation and estimation code.

- **Case 2:** MA(1) covariate

  $x_t = \theta x_{t-1} + \varepsilon_t$

  $\theta = 0.50$

- **Case 3**: ARFIMA (0,0.25,0) covariate

  $\Delta_j^d x_t = x_{t-1} + \varepsilon_t$

  $d = 0.25$ where, using the backward shift operator L, $\Delta_j^d = (1-L)^d = \sum_{k=0}^{j} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}$, with $\Gamma(\cdot)$ denoting the gamma function and j denoting the truncation order of the theoretical infinite sum $\Delta^d = (1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}$.

In each case the innovation process $\{\varepsilon_t\}$ is chosen to be i.i.d. normal with variance σ^2 such that the variance of the covariate model is 1 and thus facilitating comparisons. In all cases the initial values are set to $x_0 = 0$. Note that the choice of a fractional differencing order $d = 0.25$ for the fractional white noise satisfies the stationarity condition for autoregressive fractionally integrated processes $|d| < 0.50$, so that Assumption 2 on the Lipschitz condition is not violated.

For each case we consider four alternative scenarios for the data-generating parameter values, changing the value of the sum of the persistence parameters $\alpha_1 + \beta_1$:

- Scenario 1: null coefficient of intensity:

  $\omega = 0.10$, $\alpha_1 = 0.30$, $\beta_1 = 0.00$, $\gamma = 0.50$

- Scenario 2 - “low” persistence:

  $\omega = 0.10$, $\alpha_1 = 0.30$, $\beta_1 = 0.20$, $\gamma = 0.50$

- Scenario 3 - “high” persistence with the coefficient of the response larger than the coefficient of intensity:

  $\omega = 0.10$, $\alpha_1 = 0.70$, $\beta_1 = 0.25$, $\gamma = 0.50$

- Scenario 4 - "high" persistence with the coefficient of intensity larger than the coefficient of the response:

  $\omega = 0.1$, $\alpha_1 = 0.25$, $\beta_1 = 0.70$, $\gamma = 0.50$

The first scenario is comparable to an ARCH model as only the lagged response is included. Note that none of the presented scenarios violates the condition of stationarity $\sum_{i=1}^{\max(p,q)}(\alpha_i + \beta_i) < 1$ (Assumption 3) that we have imposed when developing the asymptotic theory.

For all scenarios we simulate for sample sizes $T \in \{100, 250, 500, 1000\}$ with 1000 replications. We also include small sample sizes for providing insights into the quality of the estimates for short length count time series which are commonly modeled in many empirical applications.

4.6.2 Results

As discussed above, our study of the MLE performance in finite samples concerns both the accuracy and the speed of convergence to normality. In Tables 4.1 to 4.6, the mean of the parameter estimates (obtained averaging out the results from all the replications) is reported in the fourth column, while the fifth shows the root mean square error (RMSE) of the estimates. The sixth and the seventh column report the skewness and the kurtosis of the estimates distribution. We also perform a Kolmogorov-Smirnov test on the estimates for testing against the standard normal distribution and report the corresponding p-value in the last column. In what follows, we comment the results obtained for the cases with AR/MA (short-memory) covariates and long-memory covariates separately.

Results for the short-memory covariates

In Tables 4.1 to 4.4, we show the results for the case where short-memory processes are included in the intensity specification. We consider a stationary AR(1) and a stationary MA(1), thus two short-memory processes characterized by a different rate of decrease of the autocorrelation function. The results are very similar. In both cases the estimate precision is fully satisfactory for a sample size of 500. We can also note a relevant improvement moving from $T = 100$ to $T = 250$. The best results are obtained in the first and second (low persistence) scenarios (see Tables 4.1 to 4.4). The "worst" scenario appears to be the third, i.e. when the value of persistence is close to one and the coefficient of the response α_1 is higher than the coefficient of intensity β_1. Moreover, even in this case, the approximation improves quicky as the sample size increases. The less accurate estimate is that of the constant (ω) parameter. Convergence to normality is evident in both cases and for all the scenarios considered, as normality is never rejected at a 5% significance level when the sample size is at least 500.

Results for the long-memory covariates

Case 3 considers the inclusion of a fractionally integrated process (Tables 4.5 to 4.6). ARFIMA processes are weakly stationary if the condition $|d| < 0.50$ (as in our experiment) is satisfied, but have slowly-decaying autocorrelations compared to the exponential rate of decay typical of ARMA models. Considering this case separately is then convenient. The results do not show substancial differences with respect to the previously examined case of AR/MA covariates. Again, the approximation is satisfactory, except for the constant parameter in Scenario 3, which substantially improves for a sample size of 1000, though. Convergence to normality is confirmed, as the only rejection for sample sizes larger than 250 concerns the constant parameter in Scenarios 3 and 4 (see Tables 4.5 to 4.6).

Table 4.1: Results of simulations for PARX(1,1) with stationary AR(1) covariate. Scenario 1: null coefficient of intensity. Scenario 2: "low" persistence.

		Scenario 1						Scenario 2					
Sample size	Parameter	True	Mean	RMSE	Skewness	Kurtosis	KS p-value	True	Mean	RMSE	Skewness	Kurtosis	KS p-value
T=100	ω	0.10	0.09	0.16	0.23	3.65	0.36	0.10	0.10	0.18	0.35	3.82	0.01
	α_1	0.30	0.28	0.13	0.10	3.85	0.32	0.30	0.27	0.11	-0.05	3.38	0.97
	β_1	0.00	0.02	0.15	0.15	3.85	0.31	0.20	0.22	0.14	0.07	4.14	0.34
	γ	0.50	0.51	0.07	0.06	3.31	0.85	0.50	0.51	0.07	0.35	3.26	0.32
T=250	ω	0.10	0.09	0.07	-0.11	3.28	0.85	0.10	0.10	0.08	0.33	3.64	0.19
	α_1	0.30	0.30	0.07	0.05	3.27	0.87	0.30	0.29	0.07	0.04	2.87	0.99
	β_1	0.00	0.00	0.08	-0.02	3.13	0.93	0.20	0.21	0.08	-0.08	2.93	0.63
	γ	0.50	0.50	0.04	0.18	3.01	0.49	0.50	0.50	0.04	0.01	2.97	0.92
T=500	ω	0.10	0.10	0.05	0.15	2.94	0.66	0.10	0.10	0.05	0.27	3.07	0.35
	α_1	0.30	0.30	0.04	0.15	3.64	0.33	0.30	0.30	0.04	-0.17	3.17	0.87
	β_1	0.00	0.00	0.05	-0.13	3.26	0.17	0.20	0.20	0.05	0.13	3.16	0.16
	γ	0.50	0.50	0.02	0.06	3.02	0.34	0.50	0.50	0.02	0.00	3.06	0.75
T=1000	ω	0.10	0.10	0.03	0.21	3.18	0.38	0.10	0.10	0.04	0.25	3.08	0.42
	α_1	0.30	0.30	0.03	-0.09	3.05	0.52	0.30	0.30	0.03	0.02	2.92	0.61
	β_1	0.00	0.00	0.03	0.11	3.04	0.98	0.20	0.20	0.03	-0.03	2.95	0.71
	γ	0.50	0.50	0.02	-0.01	3.23	0.74	0.50	0.50	0.02	0.13	2.81	0.32

Table 4.2: Results of simulations for PARX(1,1) with stationary AR(1) covariate. Scenario 3: "high" persistence due to high coefficient of the response. Scenario 4: "high" persistence due to high coefficient of intensity.

		Scenario 3						Scenario 4					
Sample size	Parameter	True	Mean	RMSE	Skewness	Kurtosis	KS p-value	True	Mean	RMSE	Skewness	Kurtosis	KS p-value
T=100	ω	0.10	0.20	0.48	0.41	4.20	0.01	0.10	0.15	0.30	0.56	3.94	0.07
	α_1	0.70	0.67	0.12	-0.46	5.21	0.12	0.25	0.18	0.15	-0.92	3.50	0.00
	β_1	0.25	0.27	0.12	0.48	5.41	0.18	0.70	0.77	0.15	0.89	3.46	0.00
	γ	0.50	0.52	0.13	0.09	3.22	0.86	0.50	0.51	0.11	0.04	2.94	0.84
T=250	ω	0.10	0.22	0.36	0.61	4.01	0.00	0.10	0.13	0.21	0.36	3.30	0.13
	α_1	0.70	0.69	0.06	0.16	3.14	0.26	0.25	0.23	0.06	-0.20	3.91	0.72
	β_1	0.25	0.26	0.06	-0.15	3.07	0.45	0.70	0.72	0.06	0.20	3.92	0.64
	γ	0.50	0.50	0.07	0.19	3.97	0.75	0.50	0.50	0.05	0.05	3.20	0.81
T=500	ω	0.10	0.15	0.23	0.57	3.53	0.00	0.10	0.11	0.13	0.30	2.88	0.21
	α_1	0.70	0.70	0.05	0.12	3.12	0.99	0.25	0.24	0.04	-0.04	2.79	0.86
	β_1	0.25	0.25	0.05	-0.07	3.16	0.99	0.70	0.71	0.04	0.01	2.85	0.96
	γ	0.50	0.50	0.03	0.08	2.92	0.94	0.50	0.50	0.02	0.05	2.87	0.95
T=1000	ω	0.10	0.13	0.17	0.46	3.59	0.05	0.10	0.10	0.10	0.30	3.20	0.24
	α_1	0.70	0.70	0.03	-0.01	3.01	0.71	0.25	0.24	0.02	0.03	2.88	0.79
	β_1	0.25	0.25	0.03	0.05	3.02	0.73	0.70	0.71	0.02	-0.02	2.89	0.81
	γ	0.50	0.50	0.03	-0.02	2.78	0.99	0.50	0.50	0.02	0.07	2.91	0.99

Table 4.3: Results of simulations for PARX(1,1) with MA(1) covariate. Scenario 1: null coefficient of intensity. Scenario 2: "low" persistence.

		Scenario 1						Scenario 2					
Sample size	Parameter	True	Mean	RMSE	Skewness	Kurtosis	KS p-value	True	Mean	RMSE	Skewness	Kurtosis	KS p-value
T=100	ω	0.10	0.10	0.19	0.37	5.74	0.01	0.10	0.13	0.19	0.93	4.78	0.00
	α_1	0.30	0.28	0.12	0.15	3.24	0.51	0.30	0.27	0.14	-0.29	3.94	0.37
	β_1	0.00	0.00	0.20	0.07	3.80	0.16	0.20	0.20	0.24	0.21	4.21	0.03
	γ	0.50	0.52	0.11	0.26	3.24	0.70	0.50	0.51	0.16	0.25	2.89	0.29
T=250	ω	0.10	0.10	0.10	0.28	3.21	0.03	0.10	0.10	0.12	0.49	3.48	0.16
	α_1	0.30	0.29	0.07	-0.07	3.03	0.81	0.30	0.29	0.07	-0.06	3.17	1.00
	β_1	0.00	0.00	0.10	-0.04	3.19	0.55	0.20	0.21	0.10	-0.14	3.41	0.72
	γ	0.50	0.50	0.06	0.11	2.85	0.86	0.50	0.51	0.07	0.12	2.77	0.69
T=500	ω	0.10	0.11	0.08	0.15	2.95	0.83	0.10	0.11	0.09	0.50	3.94	0.05
	α_1	0.30	0.30	0.05	0.06	3.63	0.34	0.30	0.30	0.05	0.05	2.81	0.96
	β_1	0.00	0.00	0.07	0.07	3.01	0.71	0.20	0.20	0.08	-0.05	3.30	0.54
	γ	0.50	0.50	0.05	0.01	2.92	0.72	0.50	0.50	0.05	0.04	3.11	0.40
T=1000	ω	0.10	0.10	0.05	0.22	3.33	0.29	0.10	0.10	0.07	0.35	3.31	0.40
	α_1	0.30	0.30	0.03	-0.04	2.94	0.99	0.30	0.30	0.03	-0.12	3.43	0.40
	β_1	0.00	0.00	0.05	0.15	2.94	0.73	0.20	0.20	0.05	0.08	3.32	0.57
	γ	0.50	0.50	0.03	0.17	3.27	0.93	0.50	0.50	0.04	-0.05	3.00	0.88

Table 4.4: Results of simulations for PARX(1,1) with MA(1) covariate. Scenario 3: "high" persistence due to high coefficient of the response. Scenario 4: "high" persistence due to high coefficient of intensity.

		Scenario 3						Scenario 4					
Sample size	Parameter	True	Mean	RMSE	Skewness	Kurtosis	KS p-value	True	Mean	RMSE	Skewness	Kurtosis	KS p-value
T=100	ω	0.10	0.30	0.60	0.35	4.93	0.00	0.10	0.16	0.34	0.55	3.96	0.02
	α_1	0.70	0.67	0.11	-0.03	3.07	0.25	0.25	0.17	0.16	-0.80	3.62	0.00
	β_1	0.25	0.26	0.12	-0.08	3.38	0.92	0.70	0.77	0.16	0.74	3.48	0.00
	γ	0.50	0.51	0.27	0.25	3.19	0.68	0.50	0.52	0.19	0.04	3.29	0.71
T=250	ω	0.10	0.24	0.35	0.82	5.00	0.01	0.10	0.17	0.33	0.48	3.54	0.02
	α_1	0.70	0.69	0.07	0.08	3.19	1.00	0.25	0.23	0.06	-0.22	2.82	0.38
	β_1	0.25	0.25	0.07	-0.09	3.26	0.96	0.70	0.71	0.06	0.15	2.97	0.35
	γ	0.50	0.50	0.15	0.11	3.33	0.98	0.50	0.51	0.13	0.25	2.86	0.16
T=500	ω	0.10	0.18	0.25	0.65	4.08	0.02	0.10	0.17	0.26	0.73	4.35	0.10
	α_1	0.70	0.70	0.04	0.05	3.26	0.97	0.25	0.24	0.04	0.15	3.09	0.79
	β_1	0.25	0.25	0.05	-0.02	3.32	0.97	0.70	0.70	0.04	-0.22	3.31	0.49
	γ	0.50	0.50	0.10	-0.01	3.06	1.00	0.50	0.51	0.08	0.01	3.24	0.99
T=1000	ω	0.10	0.15	0.17	0.33	3.40	0.36	0.10	0.13	0.14	0.39	2.99	0.03
	α_1	0.70	0.70	0.03	0.01	2.99	0.84	0.25	0.25	0.02	0.08	2.79	0.95
	β_1	0.25	0.25	0.03	-0.07	2.98	0.56	0.70	0.70	0.03	-0.12	2.85	0.90
	γ	0.50	0.50	0.09	0.13	3.19	0.91	0.50	0.50	0.05	-0.04	3.08	0.98

Table 4.5: Results of simulations for PARX(1,1) with ARFIMA(0,0.25,0) covariate. Scenario 1: null coefficient of intensity. Scenario 2: "low" persistence.

		Scenario 1						Scenario 2					
Sample size	Parameter	True	Mean	RMSE	Skewness	Kurtosis	KS p-value	True	Mean	RMSE	Skewness	Kurtosis	KS p-value
T=100	ω	0.10	0.12	0.20	0.74	5.30	0.00	0.10	0.11	0.18	0.81	4.21	0.00
	α_1	0.30	0.29	0.13	-0.05	3.80	0.47	0.30	0.27	0.13	-0.18	3.55	0.43
	β_1	0.00	-0.01	0.23	0.00	4.42	0.16	0.20	0.21	0.19	0.02	4.05	0.31
	γ	0.50	0.51	0.13	0.17	3.21	0.50	0.50	0.51	0.12	0.17	3.11	0.32
T=250	ω	0.10	0.10	0.09	0.39	3.71	0.14	0.10	0.12	0.12	0.47	3.49	0.08
	α_1	0.30	0.30	0.07	0.17	3.48	0.70	0.30	0.29	0.07	0.01	3.02	0.57
	β_1	0.00	0.00	0.10	-0.24	3.66	0.33	0.20	0.20	0.10	-0.15	3.15	0.81
	γ	0.50	0.50	0.06	0.26	3.06	0.39	0.50	0.50	0.07	0.25	2.95	0.85
T=500	ω	0.10	0.10	0.07	0.30	3.44	0.22	0.10	0.10	0.07	0.30	3.13	0.54
	α_1	0.30	0.30	0.05	-0.05	2.94	0.95	0.30	0.30	0.05	-0.10	2.96	0.96
	β_1	0.00	0.00	0.07	0.03	3.13	1.00	0.20	0.20	0.07	0.12	3.02	0.90
	γ	0.50	0.50	0.04	-0.01	2.98	0.59	0.50	0.50	0.05	0.02	2.79	0.97
T=1000	ω	0.10	0.10	0.05	0.09	2.93	0.73	0.10	0.10	0.05	0.40	3.68	0.14
	α_1	0.30	0.30	0.03	-0.05	3.04	0.81	0.30	0.30	0.03	-0.15	2.83	0.56
	β_1	0.00	0.00	0.05	0.03	3.13	0.82	0.20	0.20	0.05	0.07	3.03	0.80
	γ	0.50	0.50	0.03	0.09	3.11	0.74	0.50	0.50	0.03	0.08	2.79	0.43

Table 4.6: Results of simulations for PARX(1,1) with ARFIMA(0,0.25,0) covariate. Scenario 3: "high" persistence due to high coefficient of the response. Scenario 4:"high" persistence due to high coefficient of intensity.

		Scenario 3						Scenario 4					
Sample size	Parameter	True	Mean	RMSE	Skewness	Kurtosis	KS p-value	True	Mean	RMSE	Skewness	Kurtosis	KS p-value
T=100	ω	0.10	0.29	0.57	0.73	5.33	0.01	0.10	0.16	0.30	0.63	3.96	0.02
	α_1	0.70	0.66	0.12	-0.20	3.44	0.74	0.25	0.17	0.16	-0.88	3.34	0.00
	β_1	0.25	0.27	0.12	0.14	3.33	0.61	0.70	0.78	0.16	0.80	3.30	0.00
	γ	0.50	0.52	0.15	0.15	2.79	0.69	0.50	0.51	0.14	-0.02	3.14	0.81
T=250	ω	0.10	0.29	0.44	0.76	3.95	0.00	0.10	0.18	0.25	1.15	5.93	0.00
	α_1	0.70	0.69	0.07	-0.22	3.37	0.59	0.25	0.23	0.05	-0.25	4.35	0.58
	β_1	0.25	0.25	0.07	0.19	3.45	0.29	0.70	0.71	0.06	0.09	3.98	0.84
	γ	0.50	0.50	0.14	0.11	3.20	0.45	0.50	0.51	0.14	0.33	3.47	0.30
T=500	ω	0.10	0.19	0.27	0.67	3.64	0.00	0.10	0.13	0.14	0.71	3.81	0.00
	α_1	0.70	0.70	0.04	-0.09	2.88	0.34	0.25	0.24	0.04	0.06	2.83	0.47
	β_1	0.25	0.25	0.05	0.03	2.83	0.76	0.70	0.71	0.04	-0.09	2.81	0.29
	γ	0.50	0.50	0.09	0.06	2.76	0.85	0.50	0.51	0.07	0.15	3.02	0.46
T=1000	ω	0.10	0.15	0.19	0.40	3.23	0.11	0.10	0.12	0.11	0.37	3.28	0.02
	α_1	0.70	0.70	0.03	0.10	3.28	0.19	0.25	0.24	0.02	0.03	2.97	0.95
	β_1	0.25	0.25	0.03	-0.04	3.18	0.31	0.70	0.70	0.03	-0.06	3.12	0.97
	γ	0.50	0.50	0.07	0.17	2.92	0.36	0.50	0.51	0.05	0.05	2.94	0.77

4.7 Concluding remarks

In this chapter we have defined and studied the properties of Poisson Autoregressions with Exogenous Covariates (PARX). Specifically, we have developed both the asymptotic and estimation theory, in addition to establishing the conditions for stationarity and ergodicity of the defined process. We have also considered how forecasting can be carried out and evaluated in our framework. In the last section we have conducted a simulation study of different PARX models, i.e. including different covariates. The results show a good performance of MLE and very little differences among the alternative PARX models considered. In the empirical analysis discussed in the next chapter, we will show that the PARX model is extremely useful for investigating the corporate defaults phenomenon.

Chapter 5

Empirical study of Corporate Default Counts

So far we have presented default risk and the main measures and models for analyzing it (see Chapter 1 and 2). We have presented and discussed the literature of default correlation as well as several studies investigating the phenomenon - which is central in risk management - of the default peaks predictability. We have reviewed regression models including variables which may explain the incidence of corporate defaults phenomenon, in terms of either default rates or counts. We have progressively focused on models for default counts, encouraged by the fact that the same clusters shown in the default rates time series are also evident in the time series of bankruptcy counts. Furthermore, as previously said, the main point in default rate prediction is forecasting the number of defaulting issuers by a certain time horizon. The predicted default *intensity* - the expected number of defaults - can be an easy and immediate instrument in bank risk management communications. The count models typically used for rare events like the Poisson, presented in Chapter 3 together with other count time series models, seem to be suitable. Our idea of using Poisson models with both autoregressive components and exogenous regressors for capturing the default clustering has led to the definition of a new model called Poisson Autoregression with Exogenous Covariates (PARX). How Poisson Autoregressions and PARX models perform when handling actual corporate default data and how the results of their application should be intepreted are the research questions we address in this chapter.

5.1 Overview of the approach

We investigate the corporate default dynamics through a count time series approach including autoregressive components and exogenous variables, sharing some similarities with the generalized autoregressive models for conditional volatility. Our analysis of corporate defaults dynamics is made under an *aggregate* perspective, which does not take into account firm-specific conditions determining the individual probability of default of a company. This study tries indeed to measure an overall default risk concerning debt issuers of considerable relevance in terms of dimension, because we consider defaults among rated, thus in most cases listed, firms. The default intensity of high dimension firms is expected to be linked to *common risk factors* arising from the financial and macroeconomic context, as well as possible *contagion* effects. We claim that this approach can give a useful measure of the general tendency in the corporate default dynamics, providing a measure of “systematic” default risk which can support the traditional analysis of individual firm solvency conditions.

5.2 Corporate default counts data

The time series of corporate default counts we analyze here refers to the monthly number of bankruptcies among Moody’s rated United States firms in the period going from January 1982 to December 2011. The default count dataset is one of the risk monitoring instruments provided by Moody’s Credit Risk Calculator (CRC), which allows to download historical default rates and counts in the form of customized reports, with many options in terms of time interval length and economic sectors. We choose to focus our study on the *industrial* sector: this means to include all the firms covering nonfinancial activities and exclude banking, financial and insurance companies. This choice is quite common in the study of corporate default counts (see, for instance, Das et al., 2007, Lando and Nielsen, 2010 and Lando et al. 2013) and motivated by the convenience of considering the real and financial economy default events separately, at least in the first place. Other categories typically excluded are the public utilities and transportation activities, because of their peculiar management structure, often linked to the public sector.

More generally, the choice of using US data is motivated by the good quality and organization of the default data material, at least from the 1980s. The Bankruptcy Reform Act of 1978, amending the Bankruptcy Act of 1898, is the first complete expression of the US default law, trying to give protection to the creditors as well as the chance to the borrowers to reorganize their activity. With this act, the default legislation becomes uniform in all the federal states. The Bankruptcy Reform Act of 1978 continues to serve as the federal law that governs the bankruptcy cases today, and again a strong emphasis is given to business reorganization (see Skeel, 2001 for a history of the US bankruptcy law). However, in the US as in many European countries, during the period from World War II through the 1970s, bankruptcy was a nearly exceptional event. With the exception of Northeastern railroads, there were not many notable business failures in the U.S. in that time. During the 1970s, there were only two corporate bankruptcies of prominence: Penn Central Transportation Corporation in 1970 and W.T. Grant Company in 1975. It is interesting that the failure of Penn Central and Northeastern railroads is often cited as the first documented case of contagion, as the major case of the railroads default was the missed payment of obligations by Penn Central. Both Das et al. (2007) and Lando and Nielsen (2010) cite the Penn Central case in their empirical analyses. The small number of defaults before the 1980s explains our choice of using January 1982 as the starting period of our empirical analysis.

Some first considerations about the time series of corporate default counts in US in the last thirty years can be made by inspecting a simple plot of our data, shown in Figure 5.1.

The first evidence from Figure 5.1 is that the data show the peaks typically found in corporate default counts time series and also referred to as "default clusters". The long memory of the series is evident from the slowly decaying autocorrelation function (see Figure 5.2).

Looking more in detail at the peak periods and trying to connect them with

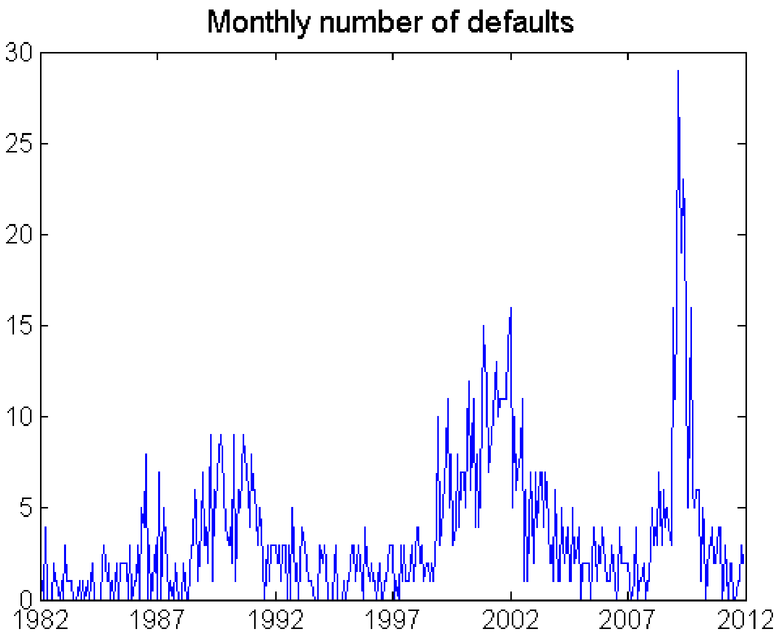

Figure 5.1: Monthly default counts of US Moody's rated industrial firms from January 1982 to December 2011.

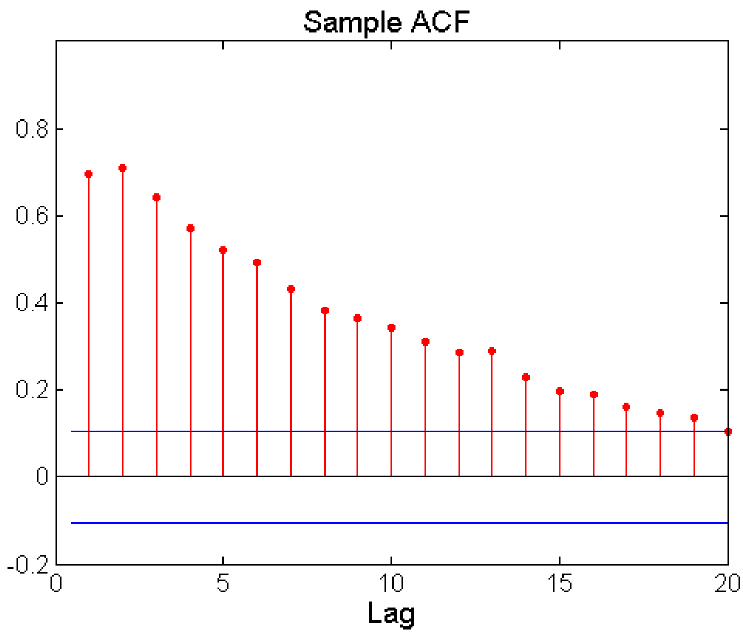

Figure 5.2: Autocorrelation function of the monthly default counts.

the financial crises, during the 1980s and early 1990s many bankruptcies took place. Many well-known companies filed for bankruptcy, mainly encouraged by reorganization opportunities. These Include LTV, Eastern Airlines, Texaco, Continental Airlines, Allied Stores, Federated Department Stores, Greyhound, Maxwell Communication and Olympia & York. Indeed, also the financial sector lived years of trouble between the 1980s and the 1990s, like the well-known "savings and loan" crisis. The financial crisis did not involve the banking sector only, as the 1987 market crash showed. The second peak in our series appears in the 1999-2002 period and, again, this is not surprising: in the years 2000-2001 a strong financial crisis took place, starting from the so-called "Dot-com" (or "Tech") bubble, causing the recession of 2001 and 2002. After a period of stability from 2003 to 2007, a new peak characterizes the final part of our sample, from 2008 to 2010, starting from the financial sector with the subprime crisis of 2007 and spreading to the real, as a global and systemic crisis, in the following years.

It is interesting to compare the default count time series to macroeconomic indicators such as the monthly Leading Index published by the Federal Reserve. The Leading Index includes the Coincident Index and a set of variables that "lead" the economy: the state-level housing permits, the state initial unemployment insurance claims, the delivery times from the Institute for Supply Management (ISM) manufacturing survey, the interest rate spread between the 10-year Treasury bond and the 3-month Treasury bill.

Looking at Figure 5.3, the low level in the late 1980s and earlier 1990s as well as in 2000-2002 confirms the previous analysis, and again the last crisis turns out to be the most dramatic period. Another relevant index, explicitly signalling the phases of the business cycle, is the recession indicator released by the National Bureau of Economic Research (NBER): the NBER recession indicator is a time series which consists in dummy variables that distinguish the periods of expansion and recession, where a value of 1 indicates a recessionary period, while a value of 0 signals an expansionary one. The shaded areas created by the recession dates in Figure 5.4 confirm the previous identification of three turbolence periods (1982-1991, 2000-2002, 2008-2010). In our analysis we shall also consider the connection between the

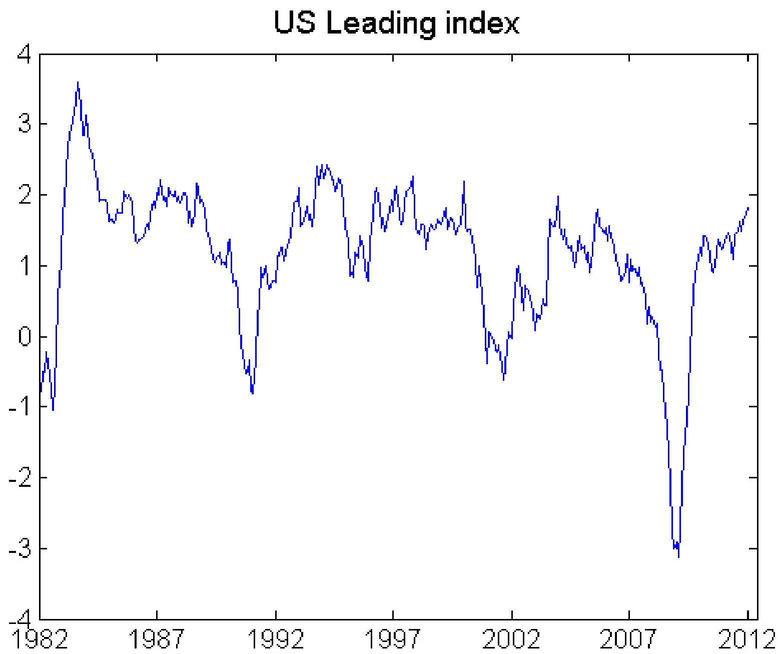

Figure 5.3: Monthly Leading Index from January 1982 to December 2011.

business cycle and the number of corporate defaults.

Based on the previous considerations, in Table 5.1 we show some descriptive statistics of the data in different subsamples of our dataset, which includes a total of 360 observations. In particular, we distinguish the three clusters of the late 1980s and early 1990s, the first 2000s and 2007-2010 respectively. In addition to the mean, the standard deviation and the median we also report the variance, underlying that all the considered subsamples present data overdispersion.

It is interesting to note that the effects on defaults of the crisis spread in 2000

Table 5.1: Descriptive statistics of the default count data.

Sample	Mean	Std. Dev.	Variance	Median
first cluster: 1986-1991	3.54	3.54	7.50	3
second cluster: 2000-2003	7.69	3.79	14.83	7
third cluster: 2007-2010	5.96	6.65	44.17	4
whole dataset	3.51	3.95	15.57	2

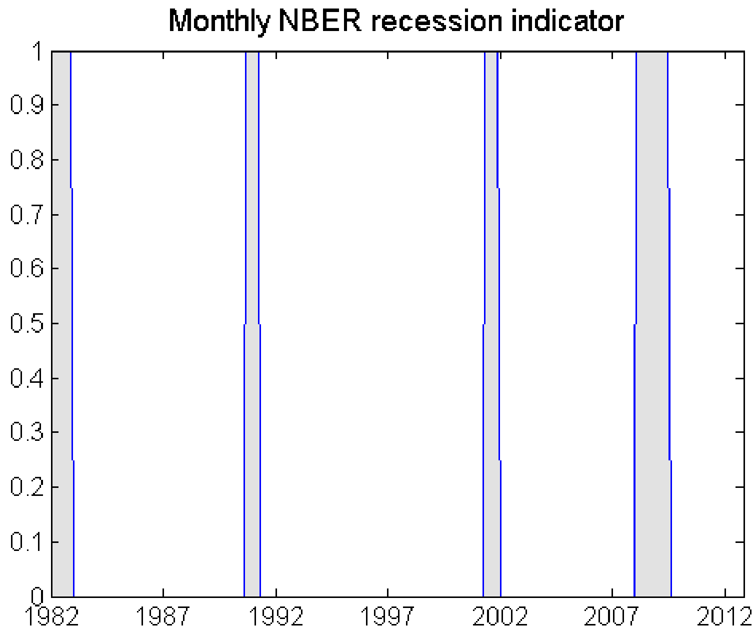

Figure 5.4: Monthly NBER recession indicator from January 1982 to December 2011.

are the most severe in terms of average number of defaults. In the last financial and economic crisis period the most relevant aspect is instead the variance, as the number of defaults explodes and decreases quickly, while the previous clusters are more lasting in time.

5.3 Choice of the covariates

Our empirical study concerns the time series analysis and modelling of the number of corporate defaults and also aims at measuring the impact of the macroeconomic and financial context on the defaults phenomenon. This needs some reflections about the variables to be considered, that are expected to be *common factors* for corporate solvency conditions and thus to be predictive of the default clusters. This section complements the previous - describing the default counts dataset which will be our response time series - by presenting the other data included in our study and motivating our choices. The covariates presented in the following can be divided into two groups:

- *financial and credit market* variables
- *production and macroeconomic* indicators

All the variables are included using monthly frequency data.

5.3.1 Financial market variables

The performance of the financial market influences both firms returns on financial investments, thus their profitability, and their funding capability, two aspects which strongly affect the liquidity and solvency conditions. Not only the stock market, but also the monetary market, which includes short-term financial instruments such as Treasury Bills, deposits and short-lived mortgages, is part of the financial market and a relevant part of the credit market, where the companies raise funds. With respect to funding, important variables are those expressing its cost, thus the interest rates and the relations between different interest rates, i.e. the *credit spreads*, which are widely used for deriving the implied differences in risk. The market is not the only evaluator of the corporate debt issuers, which are subject to the risk to become insolvent, but also to that of being downgraded by the rating agencies. Based on the above considerations, the financial and credit market variables we consider here are a measure of realized volatility of returns, the spread between the Moody's Baa rated corporate bonds yield and the 10-year Treasury rate and the number of Moody's downgrades.

Realized Volatility of returns

Our choice of using a measure of volatility of the stock returns rather than the returns themselves is motivated by the features of the corporate defaults time series, whose dynamics are mostly driven by *variance*. Indeed, as expected for rare events, the mean number of defaults is low and the level often comes back to zero. It is interesting to investigate the link between the financial market and the corporate defaults dynamics, which is expected to be strong in the crisis periods. Realized volatility deserves a special insight for several reasons. First, as for each of the covariates we include in PARX models, it is important to analyze its time series properties and verify whether the assumptions on its dynamics (see in particular Assumption 2 in Chapter 4) are satisfied. Furthermore, estimating a model for the covariate processes allows multi-step ahead forecasting (see Section 4.5). Recalling Section 4.1, the traditional realized volatility measures rely on the theory of a series of seminal papers by Andersen, Bollerslev, Diebold and Labys (2001), Andersen, Bollerslev, Diebold and Ebens (2001), and Barndorff-Nielsen and Shephard (2002), showing that the daily integrated variance, i.e. the integral of the instantaneous variance over the one-day interval, can be approximated to an arbitrary precision using the sum of intraday squared returns. Furthermore, other works such as Andersen, Bollerslev, Diebold, and Labys (2003) show that direct time series modelling of realized volatility strongly outperforms both the GARCH and stochastic volatility models.

Our approach refers to this theory, even though is not really high-frequency: we construct a proxy of *monthly* realized volatility by using the *daily* returns. Monthly volatility proxies of this kind can be found, for example, in French, Schwert and Stambaugh (1987) and Schwert (1989). According to this approach we define the following measure for the S&P 500 monthy realized volatility:

$$RV_t = \sum_{i=1}^{n_t} r_{i,t}^2 \tag{5.1}$$

where $r_{i,t}$ is the i-th daily return on the S&P 500 index in month t and n_t is the number of trading days in month t.

The high values of skewness (9.02) and kurtosis (100.26) of our proxy of realized variance indicate that it is far from being normally distributed. Nonnormality is pointed out in empirical works based on realized volatility measures from high frequency data, such as Martens et al. (2009). Realized volatility time series usually show high variance and peaks, recalling the sharp spikes of infinite variance processes that have often been used for modelling the stock market prices (see, for example, Fama, 1965). The logarithmic transformation of our monthly realized volatility (see Figure 5.5 (a)) is more suitable for standard time series modelling, because the variance is lower and there are no outlier observations. The high and slowly decaying autocorrelation (see Figure 5.5 (b)) suggests the use of long memory processes such

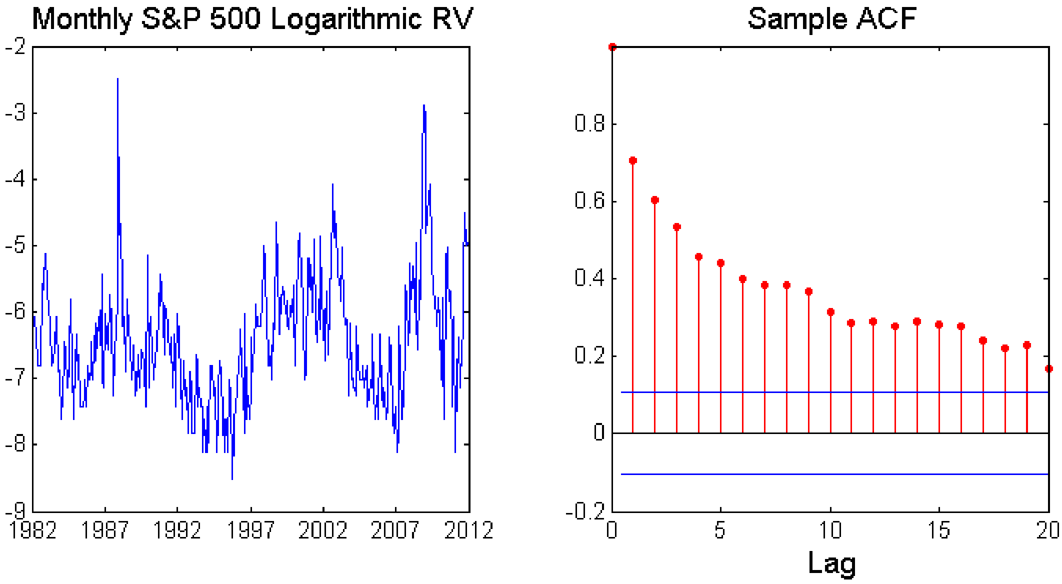

Figure 5.5: (a) Logarithm of S&P 500 monthly realized volatility. (b) Autocorrelation function of logarithmic realized volatility.

as ARFIMA. The long memory of realized volatility is a crucial point in some recent works on this topic - such as Andersen, Bollerslev and Diebold (2007) and Corsi (2009) - and put in doubt that the needed stationarity condition is satisfied. However, the same works claim that the long memory is "apparent" in the sense that the persistence in realized volatility series can be effectively captured by a special class of autoregressive models, which include different autoregressive parts corresponding to volatility components realized over different time horizons. These models are called Heterogeneous Autoregressive model of Realized Volatility (HAR-RV).

Corsi (2009) defines a HAR model for daily realized volatility calculated from intraday data by considering three volatility components corresponding to time horizons of one day ($1d$), one week ($1w$) and one month ($1m$). These "heterogeneous" lags can be interpreted as taking into account financial returns variability with respect to different investment time horizons. The specification proposed by the author for the daily realized volatility is the following:

$$RV_t^{(d)} = c + \beta^{(d)} RV_{t-1d}^{(d)} + \beta^{(w)} RV_{t-1d}^{(w)} + \beta^{(m)} RV_{t-1d}^{(m)} + \varepsilon_t \tag{5.2}$$

where $RV_t^{(d)} = \sqrt{\sum_{i=0}^{n_t} r_{i,t}^2}$ and n_t number of available intraday squared returns while

$RV_{t-1}^{(w)}$ and $\beta^{(m)}RV_{t-1}^{(m)}$ denote the weekly and monthly realized volatility respectively, computed as:

$$RV_t^{(w)} = \tfrac{1}{5}(RV_t^{(d)} + RV_{t-1d}^{(d)} + ... + RV_{t-4d}^{(d)})$$
$$RV_t^{(m)} = \tfrac{1}{22}(RV_t^{(m)} + RV_{t-1d}^{(m)} + ... + RV_{t-21d}^{(m)})$$

where the multiperiod volatilities are calculated as the simple averages of the daily ones during the period.

This model is shown to be able to reproduce the long memory of the empirical volatility. The model performance in terms of both in-sample and out-of-sample forecasting is comparable to that of fractionally integrated models and can be estimated more easily, since OLS can be employed.

Adapting this approach to our monthly realized volatility could be useful for carrying out multi-step ahead forecasting in a PARX model including this variable. A possible choice of the "heterogeneous" lags suitable for our monthly measure would be including the first lag of logarithmic realized volatility and the last half-year logarithmic realized volatility. The latter is computed as the simple average of the last six monthly logarithmic realized volatility. This yields the following model:

$$\log RV_t = c + \beta^{(1m)} \log RV_{t-1} + \beta^{(6m)} \log RV_{t-1}^{(6m)} + \varepsilon_t \tag{5.3}$$

where RV_t is defined in (5.1), while for the longer period component we have:

$$\log RV_t^{(6m)} = \frac{1}{6}(\log RV_t + \log RV_{t-1} + ... + \log RV_{t-5})$$

Following the notation of Corsi (2009), this specification corresponds to a HAR(2) model, because two volatility components are entered.

As an example, estimation of (5.3) for the logarithm of monthly realized volatility in the period from 1982 to 2011 yields the following model:

$$\log RV_t = \underset{(0.2711)}{-1.1030} + \underset{(0.0580)}{0.5543} \log RV_{t-1} + \underset{(0.0527)}{0.2733} \log RV_{t-1}^{(6m)}$$

which is a stationary autoregressive process.

Baa/10-year Treasury spread

The default risk premium, i.e. the risk premium the investors require for accepting the risk of corporate default, is often calculated as the difference between the yields

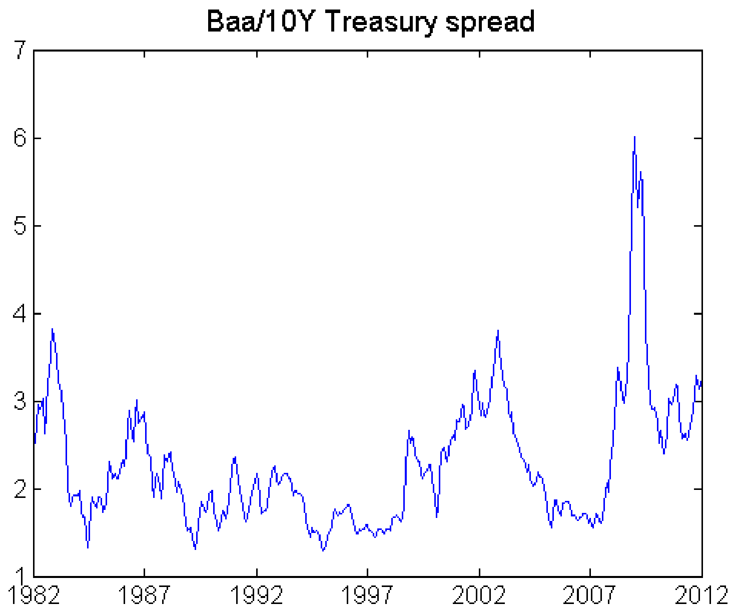

Figure 5.6: Monthly spread between Baa Moody's seasoned corporate bonds and 10-year Treasury yield.

on corporate bonds and the yields on government securities - mainly the Treasury bills - which are expected to be risk free. The spreads on Treasury rates can be considered as an implied default risk, which we expect to be positively correlated to default intensity. One of the most used is the Baa/10-year Treasury spread, i.e. the difference between the Moody's seasoned Baa corporate bond yield and the constant maturity 10-year Treasury rate. Our source for both rates is the FRED website[1], provided by the Federal Reserve Bank of St. Louis. Being a measure of the market perception of credit risk, the Baa/10-year spread is usually higher during recession periods, when the investors are worried of default risk even for upper-medium quality firms like the Baa rated. This is evident from Figure 5.6: look, for example, at the high peak in the last crisis period.

[1] http://research.stlouisfed.org/.

Number of downgrades

The monthly counts of defaults are not the only data we get from Moody's CRC, which also provides the monthly rating transition matrices, where each entry is the number of firms moving from a rating class to another (see 2.1.1 for a comprehensive analysis of rating and its modelling). As discussed before, the main role of rating is to give an objective evaluation of corporate solvency. Therefore, the number of firms which are downgraded, i.e. moved to a lower rating class, it is naturally expected to be predictive of an increased default probability. However, the capability of rating to be a default predictor is not so fair, and, as seen, also put under discussion by several econometric analyses, like, among the others, Blume et al. (1998) and Nickell et al. (2000). Thus we think that is important to measure whether and how much the number of downgrades can support the prediction of the number of defaults. At a first sight (see Figure 5.7), most of the downgrade peaks correspond to the recession periods and the default clusters, except for the first peak taking place in 1982, which is due to a credit rating refinement carried out and announced by Moody's, which modifies the classes number and assignment (see Tang, 2009).

5.3.2 Production and macroeconomic indicators

Change in Industrial Production Index

The Industrial Production Index is an economic indicator that measures the real output for all facilities located in the United States manufacturing, mining and utilities. It is compiled by the Federal Reserve System on a monthly basis in order to bring attention to short–term changes in the industrial production. As it measures the movements in the industrial output, it is expected to highlight the structural developments in the economy. Its change can be considered as an indicator of the growth in the industrial sector and is already used as a default intensity regressor in Lando and Nielsen (2010). The monthly percentage change in Industrial Production index (Figure 5.8) is computed as the logarithmic difference of the monthly Industrial production Index downloaded from the FRED website.

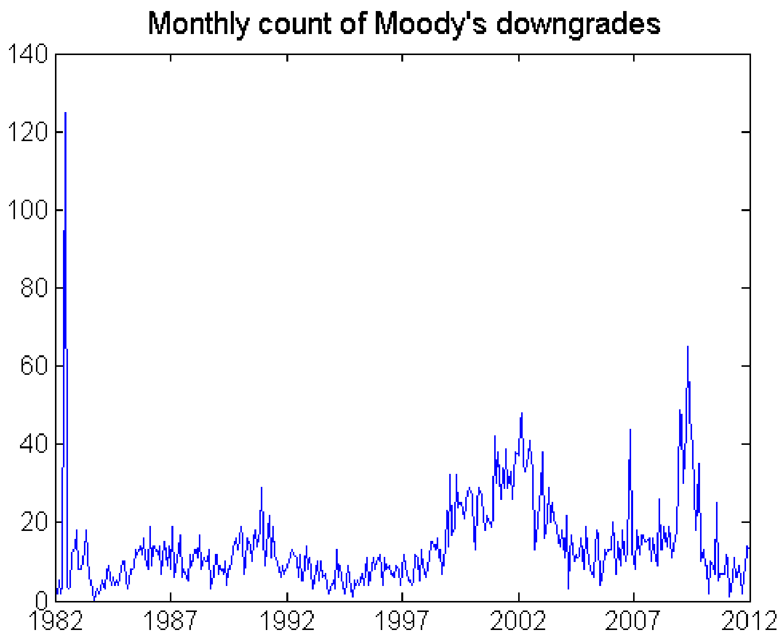

Figure 5.7: Monthly number of downgrades among industrial Moody's rated firms.

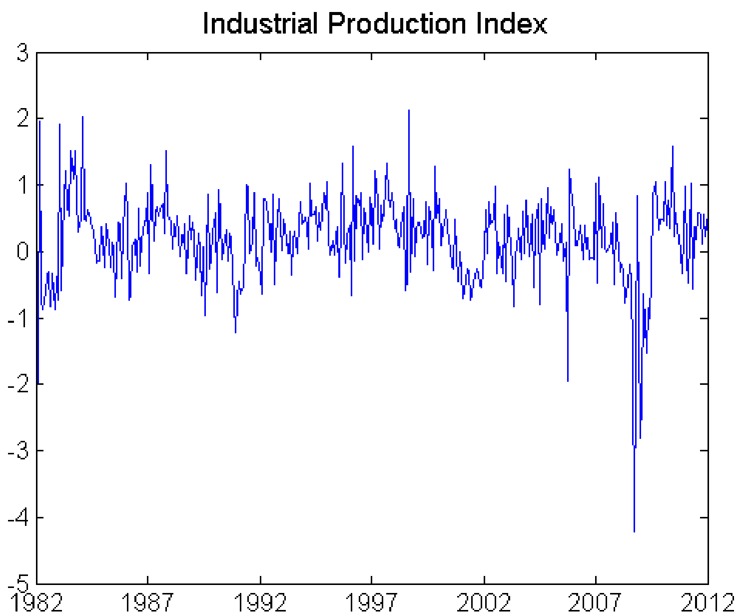

Figure 5.8: Monthly percentage change in Industrial Production Index.

Leading Index and NBER recession indicator

As our analysis of the default phenomenon is made under an aggregate perspective, we claim that the effect of business cycle on the default intensity has to be measured through overall indicators, representing the state of the economy - such as the Leading Index published by the Federal Reserve - or signalling the expansion and recession periods, as captured by the NBER recession index. They have been presented in Section 5.2. The data for both variables are downloaded by the FRED website.

For each financial and macroeconomic covariate described above, we perform an Augmented Dickey-Fuller (ADF) test, rejecting the null hypothesis of presence of unit roots in all the cases. All the variables introduced above can thus be employed in the following analysis, since they satisfy the Lipschitz condition (see Assumption 2 in Chapter 4). For realized volatility, the ADF test has been performed on the series in logarithms, whose properties we have previously investigated.

5.4 Poisson Autoregressive models for corporate default counts

The first objective of our analysis of corporate default counts dynamics is to evaluate whether the inclusion of exogenous variables can improve the prediction of the number of defaults. In particular, we consider alternative PARX models by including different covariates and compare the results. Furthermore we compare the PARX models with the Poisson Autoregression without exogenous regressors as proposed by FRT (2009) (PAR). We mainly focus on two aspects: first, we evaluate which of the chosen variables allow to explain the default intensity; second, we compute the value of the estimated persistence. As seen before, the latter allows to measure the persistence of shocks in the default counts process. We also aim at evaluating whether the inclusion of different covariates has a different impact on the estimated persistence: the magnitude of the autoregressive coefficients is expected to decline in the case one or more covariates explain most of the series long memory. This objective is thus similar to that of several empirical studies which consider the impact of covariates, such as the trading volume, in the GARCH specification (see, for instance, Lamoureux and Lastrapes, 1990 and Gallo and Pacini, 2000) and evaluate their effect on the ARCH and GARCH parameter estimates. In our context, the financial and macroeconomic variables explaining the default intensity can be considered as common factors influencing the solvency conditions of all companies.

As seen before, in PARX models negative covariates are handled by transforming them through a positive function f, which can be chosen case by case, as long as the Lipschitz condition stated in Assumption 1' of Chapter 4 is satisfied. The specification which generalizes (4.3) by including an n-dimensional vector of covariates is the following:

$$\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \lambda_{t-i} + \sum_{i=1}^{n} \gamma_i f_i(x_{i,t-1}) \tag{5.4}$$

where $\omega > 0$, $\alpha_1, \alpha_2, \beta_1, \gamma_i \geq 0$, $f : \mathbb{R} \rightarrow \mathbb{R}^+$.

According to the choice motivated in the previous section, the covariates included are the following:

- S&P 500 realized volatility (RV) (see Section 5.3.1 for details on its computation)
- Baa Moody's rated to 10-year Treasury bill spread (BAA_TB)
- Moody's downgrade count (DG)
- NBER recession indicator ($NBER$)
- percentage change in Industrial Production Index (IP)
- Leading Index (LI)

Function f is simply the identity for covariates assuming only positive values, while we use the absolute value for transforming the two variables which assume also negative values, that are the percentage change in the Industrial Production Index (IP) and the value of Leading Index (LI). Both are also expected to be negatively correlated to default intensity. Then, for capturing the asymmetric effect of positive and negative values of these covariates, we introduce a dummy variable which is 1 when the value is lower than zero. This solution is analogous to that adopted in the GJR-GARCH model by Glosten et al. (1993), where a dummy variable is introduced for capturing the asymmetric effect of positive and negative lagged returns. According to Engle and Ng (1993), in the volatility modelling this approach outperforms other specifications that overcome the problem of nonnegativity, such as the EGARCH by Nelson (1991). As to realized volatility covariate, in the previous section we have analyzed its logarithmic transform, which is stationary according to the ADF test performed. Furthermore, as we have seen, our logarithmic realized volatility has similar properties to the realized volatility measures analyzed in literature, whose long memory can be effectively captured by stationary HAR processes (Corsi, 2009). Variable RV can then be considered as the exponential transformation of the logarithmic realized volatility, satisfying the model assumptions.

Preliminary model selection based on information criteria and likelihood ratio tests leads to choose $p = 2$ and $q = 1$, i.e. two lags of the response and one lag of intensity. Thus, the model including all the six covariates - nesting all the estimated models presented in the next section - is specified as

$$\begin{aligned}\lambda_t &= \omega + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 \lambda_{t-1} + \gamma_1 RV_{t-1} + \gamma_2 BAA\_TB_{t-1} + \gamma_3 DG_{t-1} \\ &\quad + \gamma_4 NBER_{t-1} + \gamma_5 \left|IP_{t-1}\right| + \gamma_6 I_{\{IP_{t-1}<0\}} \left|IP_{t-1}\right| + \gamma_7 \left|LI_{t-1}\right| + \gamma_8 I_{\{LI_{t-1}<0\}} \left|LI_{t-1}\right| \end{aligned} \tag{5.5}$$

5.4.1 Results

Table 5.2 shows the results obtained by estimating[2] nine different PARX models. The upper portion of Table 5.2 reports the parameter estimates (standard errors in brackets). The lower portion reports, for each model, two information criteria, i.e. the AIC (Akaike, 1974) and the BIC (Schwarz, 1978), and the p-value of the likelihood ratio (LR) test. The latter compares each estimated model with respect to the one which includes all the six covariates (*All* in Table 5.2), thus following a specific-to-general model selection approach. The second column reports the results for the *PAR* model, i.e. the model with no covariates. The columns from third to eighth in Table 5.2 report the results of estimation of models including one covariate

[2] We write in Matlab the optimization code for maximum likelihood estimation.

at the time. As explained above, for covariates IP and LI we also consider the effect of negative values separately, by introducing a dummy variable as in (5.5). The first evidence from our results is that the autoregressive components play the main role in the defaults dynamics. The estimated persistence is indeed not far from one in all the models. The number of defaults in the US economy shows a high persistence of shocks, supporting our proposal of a model able to capture long memory. But can exogenous covariates explain the strong autocorrelation, and then the clusters, of defaults? The first evidence is that several of the covariates we have considered are found significant in explaining default intensity when included one at the time. They are the S&P 500 index realized volatility, the Baa Moody's rated to 10-year Treasury spread, the number of Moody's downgrades and the NBER recession indicator[3]. First of all, we think that it is of particular interest that a financial variable as realized volatility accounts for a real economic issue as defaults of industrial firms. The inclusion of realized volatility is indeed new in default risk analysis. While the use of credit spreads like the Baa to 10-year Treasury Bill is quite common in default risk prediction - especially in the reduced-form models mentioned in Chapter 1, using a pricing approach to default risk measurement - the inclusion of the number of downgrades among the regressors of default counts is new as well. In fact, there are in literature several works focusing on the link between the rating transitions and the business cycle - like, among the others, Nickell et al. (2000) and Behar and Nagpal (2001) - but not estimating a direct relation between downgrades and defaults at an aggregate level. The significance of the NBER recession indicator highlights a connection between the business cycle and the defaults dynamics and confirms the idea of a relation between economic recession and default clusters. The effect of the macroeconomic context on default intensity is also captured by including the Industrial Production Index and the Leading Index. The asymmetric effect of the positive and negative values of variables IP and LI on default intensity is confirmed,

[3] All the mentioned covariates are found significant at a level of 5% or less, except for the number of downgrades, which is found significant at the 10% level.

as they are found significant only when assuming negative values[4]: both a decrease in Industrial Production and a decrease in the value of Leading Index result in a higher predicted level of default risk. According to the LR test, as well as information criteria, all the models including one covariate at the time are preferable to the *PAR* model, thus highlighting that covariates are needed to account for the default phenomenon. Among these PARX models, according to both the information criteria and the LR test, the best are *RV* and *LI*. Realized volatility of returns and negative values of Leading Index are indeed the only two significant covariates in the *All* model (5.5), including all the covariates. The result that the number of defaults is positively associated to the level of uncertainty shown by the financial market only one month before is of particular interest and could be effectively used for risk management operational purposes. Furthermore, the significance of Leading Index shows that the macroeconomic context is relevant in default prediction. This is not an obvious result, as the existence of a link between macroeconomic variables and corporate default phenomenon is not always supported by similar analyses in the econometric literature. While, for example, Keenan, Sobehart, and Hamilton (1999) and Helwege and Kleiman (1997) forecast aggregate US corporate default rates using various macroeconomic variables, including industrial production, interest rates and indicators for recession, in some recent works the estimated relation between the default rates and the business cycle is not so strong. In particular, the empirical results of both Duffie et al. (2009) and Giesecke et al. (2011) show a not significant role of production growth and Lando et al. (2013) find that, conditional on individual firm risk factors, no macroeconomic covariate is significant in explaining default intensity.

Looking now at the estimated persistence ($\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\beta}_1$) and comparing it between *PAR* and *All* models, we observe that the inclusion of covariates leads to a small decrease in the level of persistence (from 0.9155 to 0.8758), which is not significant. The large value of the estimated persistence and its substantial invariance when exogenous covariates are included indicates that the autoregressive parts of the model

[4]For models "*IP*" and "*LI*", as well as "*All*", we perform a restricted maximization of the log-likelihood function by constraining the coefficients to be positive.

explain most of the slowly decaying behaviour of the autocorrelation function characterizing the default dynamics (see Figure 5.2). However, finding significant variables in default count time series is of relevant interest in default risk evaluation and forecasting. An increase in the level of the identified risk factors can indeed be a "warning" for risk managers and, in general, default risk evaluators.

The final model we obtain on the basis of our model selection procedure is labelled *LMRV & LI*$^{(-)}$ in Table 5.2. Here we include both the S&P 500 realized volatility and the Leading Index - when taking negative values - in the model specification.

Table 5.2: Estimation results of different PARX models.

	PAR	RV	BAA_TB	DG	NBER	IP	LI	RV & LI$^{(-)}$	All
$\hat{\omega}$	0.3015	0.1690	0.1166	0.2065	0.2897	0.2023	0.2949	0.2324	0.2083
	(0.0832)	(0.0685)	(0.1621)	(0.0930)	(0.0816)	(0.0944)	(0.1465)	(0.0717)	(0.2081)
$\hat{\alpha}_1$	0.2409	0.1966	0.2273	0.2208	0.2280	0.2127	0.1927	0.1850	0.1801
	(0.0443)	(0.0447)	(0.0448)	(0.0448)	(0.0445)	(0.0451)	(0.0452)	(0.0450)	(0.0457)
$\hat{\alpha}_2$	0.2148	0.1796	0.2217	0.1976	0.2063	0.1453	0.1979	0.1878	0.1834
	(0.0667)	(0.0617)	(0.0660)	(0.0653)	(0.0657)	(0.0643)	(0.0635)	(0.0618)	(0.0633)
$\hat{\beta}_1$	0.4598	0.5263	0.4298	0.4547	0.4696	0.5520	0.4979	0.5177	0.5123
	(0.0755)	(0.0663)	(0.0797)	(0.0750)	(0.0746)	(0.0675)	(0.0724)	(0.0686)	(0.0723)
$\hat{\gamma}_1$		63.991						28.092	24.313
		(15.565)						(13.659)	(14.368)
$\hat{\gamma}_2$			0.2407						0.0000
			(0.0867)						(0.0951)
$\hat{\gamma}_3$				0.0171					0.0059
				(0.0090)					(0.0092)
$\hat{\gamma}_4$					0.4196				0.0000
					(0.1883)				(0.4656)
$\hat{\gamma}_5$						0.0000			0.0000
						(0.1423)			(0.1647)
$\hat{\gamma}_6$						0.6945			0.0000
						(0.2113)			(0.1843)
$\hat{\gamma}_7$							0.0000		0.0000
							(0.0644)		(0.0821)
$\hat{\gamma}_8$							0.9413	0.7297	0.7540
							(0.2245)	(0.1954)	(0.3189)
$\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\beta}_1$	0.9155	0.9026	0.8788	0.8731	0.9039	0.9100	0.8885	0.8905	0.8758
	(0.0267)	(0.0170)	(0.0169)	(0.0213)	(0.0168)	(0.0202)	(0.0223)	(0.0261)	(0.0241)
AIC	-1352.04	-1368.82	-1359.86	-1352.88	-1354.94	-1360.52	-1375.06	-1377.52	-1365.84
BIC	-1336.47	-1349.36	-1340.40	-1333.42	-1335.48	-1337.17	-1351.71	-1354.17	-1319.14
LR test (p-value)	0.0000	0.4424	0.0455	0.0047	0.0095	0.0982	0.9931	0.9964	

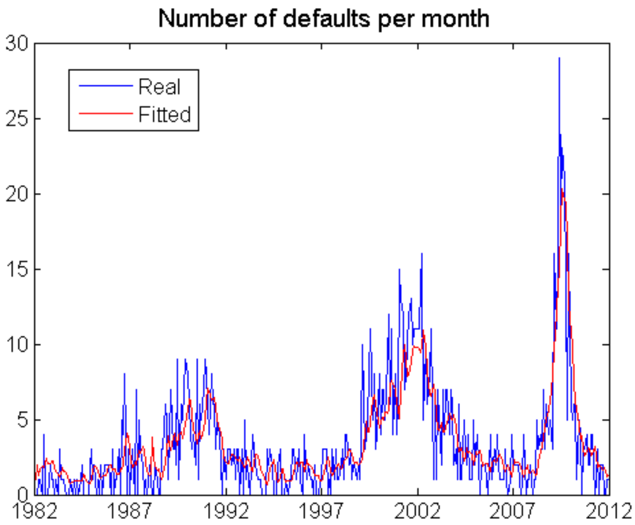

Figure 5.9: Observed and fitted monthly number of defaults from January 1982 to December 2011 for PARX model including logarithmic realized volatility and Leading Index.

5.4.2 Goodness of fit analysis

Overall, as can be seen from Figure 5.9, the model including realized volatility and Leading Index using the prediction $\hat{y}_t = \hat{\lambda}_t$ captures the default counts dynamics satisfactorily.

A commonly used diagnostic check for Poisson-type count models is to test the absence of autocorrelation in the Pearson residuals (see Section 3.2.5), which are the standardized version of the raw residual $y_t - \lambda_t(\hat{\boldsymbol{\theta}})$, taking into account that the conditional variance of y_t is not constant. In fact, the sequence of Pearson residuals estimates the sequence

$$e_t = \frac{y_t - \lambda_t}{\sqrt{\lambda_t}}, \quad t = 1, ..., T$$

which, as previously seen, is an uncorrelated process with mean zero and constant variance under the correct model. In addition, no significant serial correlation should be found in the sequence e_t^2 as well. As can be seen from Figure 5.10, the Pearson residuals of our final estimated model do not show significant autocorrelation at any lag, thus approximate a white noise satisfactorily. In order to check the adequacy of

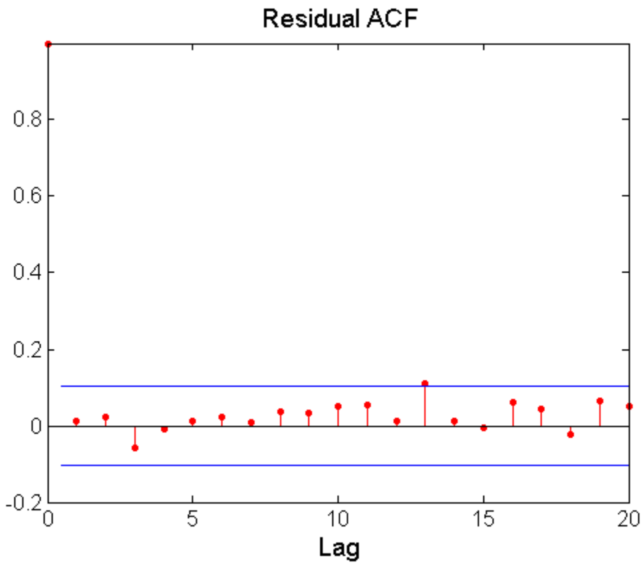

Figure 5.10: Autocorrelation function of Pearson residuals for PARX model including logarithmic realized volatility and Leading Index.

our model, following Jung et al. (2006) we perform a Ljung-Box test for the Pearson residuals and the squared Pearson residuals including 30 lags. The resulting p-values (0.661 and 0.373 respectively) indicate that the model successfully accounts for the dynamics of the first and second order moments of our default counts.

An important point concerning the PARX model goodness of fit analysis in the specific case of our empirical study should be considered: when applying the PARX model to the counts of defaults, the aim is to capture the default clusters and signal the periods where the default intensity, and thus the default risk, is higher. Then, the model performance is crucial when the number of observed events is relatively high. In this respect, Table 5.3 compares the empirical (second column) and estimated frequencies (third column) for different values of y_t. Each of the estimated frequencies is computed as the probability of observing a count falling in the range defined in the first column, under the estimated model.

In order to test the equality between theoretical and observed frequencies, we employ the test derived in the following and similar to the common test for equality of Bernoulli proportions. Suppose that we want to test the equality of the empirical and theoretical frequency of y_t values belonging to a subset A of $\mathbb{N} \cup \{0\} = \{0, 1, 2, ...\}$.

Count	Empirical frequency	Estimated frequency	p-value
$y_t = 0$	0.18	0.12	0.001
$0 < y_t \leq 5$	0.62	0.68	0.002
$y_t > 5$	0.21	0.19	0.384
$5 < y_t \leq 10$	0.14	0.14	0.741

Table 5.3: Empirical and estimated frequencies of default counts.

First define

$$Z_t = I\left(y_t \in A\right)$$

and

$$\pi_t = \Pr\left(Z_t = 1 | \mathcal{F}_{t-1}\right)$$

It can be noted that $E\left(Z_t - \pi_t | \mathcal{F}_{t-1}\right) = 0$, i.e. $Z_t - \pi_t$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$. The conditional variance of each $Z_t - \pi_t$ variable can be derived as follows:

$$\begin{aligned} V\left(Z_t - \pi_t | I_{t-1}\right) &= E\left(\left(Z_t - \pi_t\right)^2 | I_{t-1}\right) = E\left(Z_t^2 | I_{t-1}\right) + \pi_t^2 - 2\pi_t E\left(Z_t | I_{t-1}\right) \\ &= E\left(Z_t^2 | I_{t-1}\right) + \pi_t^2 - 2\pi_t^2 \\ &= E\left(Z_t^2 | I_{t-1}\right) - \pi_t^2 = \pi_t - \pi_t^2 = \pi_t\left(1 - \pi_t\right) \end{aligned}$$

Define now

$$S_T = \sum_{t=1}^{T}\left(Z_t - \pi_t\right)$$

As the sequence $\pi_t\left(1 - \pi_t\right)$ is a stationary and ergodic process, we have that the mean of the conditional variances is asymptotically constant:

$$V\left(\frac{S_T}{\sqrt{T}}\right) = \frac{1}{T}\sum_{t=1}^{T}\pi_t\left(1 - \pi_t\right) \xrightarrow{p} \sigma^2$$

This allows to apply the Martingale Central Limit Theorem (Brown, 1971) to S_T and state that

$$s_T = \frac{S_T}{\sqrt{\sum_{t=1}^{T}\pi_t\left(1 - \pi_t\right)}} \rightarrow_d N\left(0, 1\right)$$

A one-sided or two sided test can be constructed based on $N(0,1)$ critical values, replacing the unknown π_t's with their estimates

$$\hat{\pi}_t = \Pr\left(Z_t = 1|\lambda_t(\hat{\boldsymbol{\theta}})\right)$$

given by the model.

The last column of Table 5.3 shows the p-value of the two-sided test constructed as above for different A subsets.

As can be seen from Table 5.3, for values larger than 5 and for the subset $(5, 10]$, we accept the null hypothesis of equality between the empirical and theoretical proportion at the 5% significance level. It is a good result that the model correctly estimates the frequency of defaults when the relevance of the phenomenon becomes considerable. Prediction is indeed not crucial in periods of stability, when defaults are rare and isolated events. Equality is rejected when the number of defaults is null or very low.

Some considerations have to be made about the incidence of zero counts. Default of rated firms is a rare event, nearly exceptional in periods of economic expansion and financial stability. Thus, default count time series are characterized by a high number of zero observations. In our default counts dataset, there are 63 zeros on a total of 360 observations, corresponding to a proportion of 17.5%. In the PARX models, the distribution of the number of events conditional on its past and on the past value of a set of covariates is Poisson. The Poisson distribution does allow for zero observations. At each time t, the probability of having a zero count is given by $\exp(-\lambda_t)$, i.e. the probability corresponding to value 0 in a Poisson distribution of intensity λ_t. An aspect often investigated in Poisson regression models specification analysis is whether the incidence of zero counts is greater than expected for the Poisson distribution. In our application, the analysis of the incidence of zero counts should take into account two main points. First, the empirical frequency of zero counts has to be compared to that implied by the PARX model. Then, the relevance of a possible underestimate of the number of zeros has to be evaluated with respect to our specific case.

Figure 5.11 can give an idea of the relation between the observed zeros and the

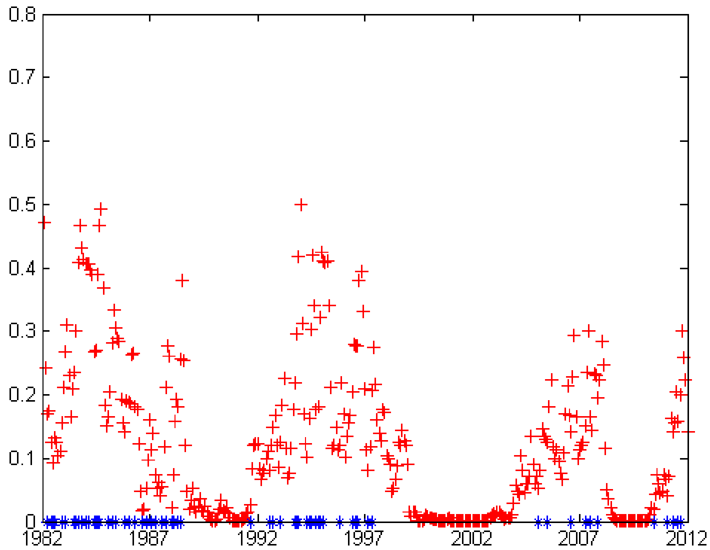

Figure 5.11: Empirical zero counts (asterisks) and probability of having a zero count under the estimated model (crosses).

probability of having a sampling zero under the model assumptions. There is a clear correspondance between the periods characterized by a higher number of zeros and the probability of having a sampling zero. The latter reaches values of more than 40% in the two most "zero-inflated" periods of 1982-1987 and 1994-1997. There is only one part of the series, around year 1987, showing an estimated frequency of less than 10% zeros when the empirical one is high. However, this period anticipates that of the last eighties financial crisis, characterized by a rapidly increasing number of defaults and corresponding to a decrease in the estimated zero counts probability.

A possible way of accounting for excess zeros in the Poisson models is to define mixture models such as those proposed and applied in the works of Mullahy (1986), Johnson, Kotz, and Kemp (1992) and Lambert (1992) and known as Zero-Inflated Poisson models (ZIP). In ZIP models, an extra-proportion of zeros is added to that implied by the Poisson distribution.The zeros from the Poisson distribution can be considered as sampling zeros, occurring by chance, while the others are structural zeros, not depending on the regressors dynamics. It is worth to note that in our application, considering an aggregate data of default incidence, the distinction between structural and sampling zeros is not so relevant. First of all, the occurrence of the single default is linked to the individual firm history and to occasional - and difficult to predict - individual events. Furthermore, the zero-inflated periods are those where the importance of default prediction is low.

5.5 Out-of-sample prediction

We perform a forecasting experiment for evaluating the PARX model out-of-sample performance. We focus, in particular, on the out-of-sample prediction in the period going from January 2008 to December 2011, corresponding to the last financial crisis and showing a sharp peak in the number of defaults. In particular, we perform a series of static one-step-ahead forecasts, updating the parameter estimates at each observation. The PARX model we consider includes the S&P 500 realized volatility and the negative values of the Leading Index, which is the preferable model according to the selection presented in the previous section. We also compare the results with those obtained with the PAR model, for evaluating whether the covariates included improve the prediction. Table 5.4 shows the results of both point (third and sixth column) and interval (columns fourth to fifth and sixth to seventh) estimate at each step, from $h = 1$ to $h = 48$, corresponding to the last observation in our dataset. Following Section 4.5, the point estimate of y_{T+h} is defined as

$$\hat{y}_{T+h|T+h-1} = \hat{\lambda}_{T+h|T+h-1}$$

while the 95% confidence interval for the estimate of $\hat{y}_{T+h|T+h-1}$ is given by

$$CI_{1-\alpha} = \left[Q\left(\alpha/2|\hat{\lambda}_{T+h|T+h-1}\right), Q\left(1-\alpha/2|\hat{\lambda}_{T+h|T+h-1}\right)\right]$$

where $\alpha = 0.05$. In Table 5.4, $Q\left(\alpha/2|\hat{\lambda}_{T+h\,|\,T}\right)$ and $Q\left(1-\alpha/2|\hat{\lambda}_{T+h\,|\,T}\right)$ are indicated as “min” and “max” respectively. We also report, as performance measures, the mean absolute error (MAE) and the root mean square error (RMSE). According to both indicators, the PARX model slightly outperforms the model without covariates. A comparison between the two models is also possible from Figure 5.12, plotting the actual number of defaults joint to the minimum (“min”) and maximum ("max") value of the forecast confidence interval for the PARX (first panel) and the PAR (second panel) model. Not surprisingly, in both cases the peak of March 2009, corresponding to an outlier in the default count time series, is out of the forecasting interval. There is indeed for both models a delay of three months in predicting the sharpest peak of the series. However, the PARX model predicts four defaults more than the PAR in the peak, thus considering the realized volatility - as a proxy of the financial market uncertainty- and the Leading Index - summarizing the macroeconomic context - allows to reduce the underestimate of the number of defaults in this cluster. Furthermore, the rapid increase of the default counts starting from November 2008 is captured better from the PARX model, whose predicted values increase more quickly than the number of defaults forecasted by the PAR. The high value of persistence, not far from one in all the estimates, and the consequent slow decrease of the autocorrelation lead the predicted series to decrease more slowly than the empirical series of default counts. Overall, the PARX model performs better than the PAR in capturing the default clustering.

		PARX			PAR		
h	y_{T+h}	$\hat{y}_{T+h\|T+h-1}$	min	max	$\hat{y}_{T+h\|T+h-1}$	min	max
1	5	1.094	0	4	1.081	0	3
2	3	1.786	0	5	1.779	0	5
3	4	2.340	0	6	2.337	0	6
4	3	2.703	0	6	2.705	0	6
5	7	2.958	0	7	2.919	0	7
6	3	3.635	0	8	3.602	0	8
7	6	3.953	1	8	3.893	1	8
8	4	4.230	1	9	4.108	1	8
9	5	4.501	1	9	4.287	1	9
10	4	4.590	1	9	4.334	1	9
11	3	4.589	1	9	4.321	1	9
12	16	6.077	2	11	4.059	1	8
13	11	8.912	4	15	6.046	2	11
14	16	11.066	5	18	8.144	3	14
15	29	13.230	7	21	10.178	4	17
16	19	17.216	10	26	15.170	8	23
17	23	19.602	11	29	18.103	10	27
18	21	20.121	12	29	18.963	11	28
19	14	20.290	12	30	19.600	11	29
20	5	18.369	10	27	17.767	10	26
21	16	14.690	8	23	13.650	7	21
22	6	12.512	6	20	11.867	6	19
23	5	11.062	5	18	10.786	5	18
24	6	8.255	3	14	7.840	3	14
25	6	6.705	2	12	6.470	2	12
26	1	5.970	2	11	6.012	2	11
27	5	4.731	1	9	4.617	1	9
28	4	3.799	1	8	3.788	1	8
29	0	3.926	1	8	4.116	1	9
30	3	3.161	0	7	3.063	0	7
31	3	2.546	0	6	2.387	0	6
32	4	2.801	0	6	2.790	0	6
33	2	3.076	0	7	3.205	0	7
34	2	3.055	0	7	3.138	0	7
35	4	2.599	0	6	2.652	0	6
36	4	2.723	0	6	2.917	0	7
37	0	3.212	0	7	3.479	0	8
38	1	2.660	0	6	2.762	0	6
39	3	1.832	0	5	1.822	0	5
40	0	1.964	0	5	2.070	0	5
41	2	1.876	0	5	1.912	0	5
42	2	1.595	0	4	1.656	0	5
43	0	1.849	0	5	1.971	0	5
44	0	1.613	0	4	1.634	0	5
45	1	1.185	0	4	1.049	0	3
46	1	1.298	0	4	1.018	0	3
47	3	1.465	0	4	1.224	0	4
48	2	1.933	0	5	1.802	0	5
MAE		2.543			2.840		
RMSE		4.119			4.613		

Table 5.4: Out-of-sample estimation results of PARX and PAR model.

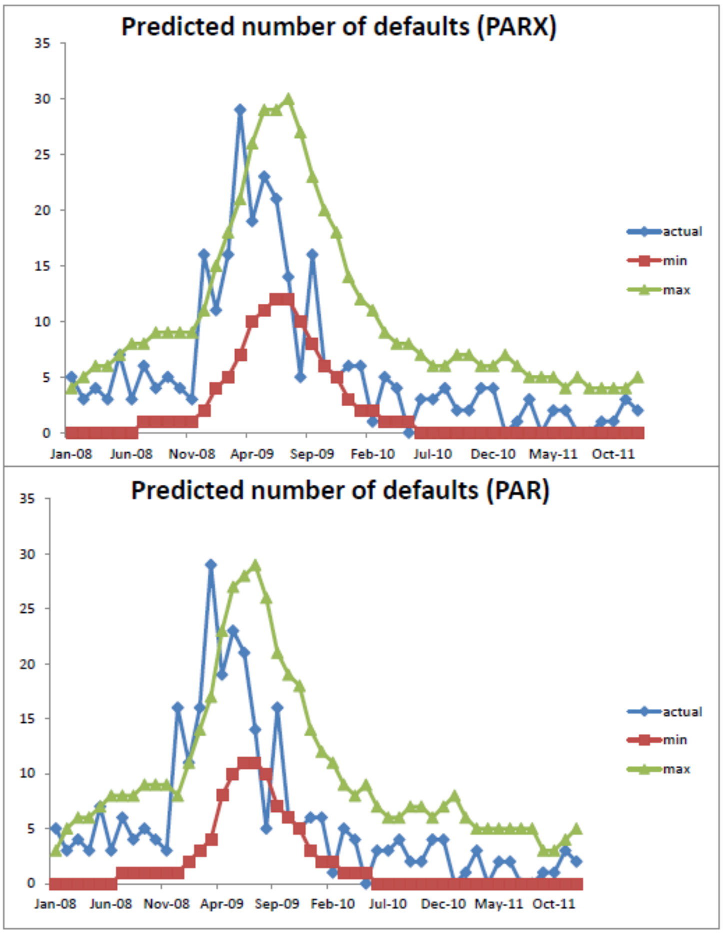

Figure 5.12: Actual and forecasted number of defaults with PARX (first panel) and PAR (second panel) model.

5.6 Concluding remarks

In this chapter we have presented an empirical analysis of the corporate default dynamics. Our study is based on the estimation of Poisson Autoregressive models for the monthly count of defaults among Moody's rated industrial firms in the period from January 1982 to December 2011. The objectives of our analysis is two-fold: first, we want to evaluate whether there are macroeconomic and financial variables which can be useful in default prediction; secondly, an important point is to consider the relevance of the autoregressive components, whose presence is an essential part of our modelling approach. We estimate both the Poisson Autoregression without covariates (PAR) and different PARX models, including macroeconomic and financial covariates. Our results show that all the PARX models estimated are preferable to the PAR. The more relevant covariates in explaining default intensity according to our results are a macroeconomic variable - the Leading Index released by the Federal Reserve - and a financial variable - the realized volatility of the S&P 500 returns. At our knowledge, this is the first work showing a positive association between the financial market uncertainty captured by the realized volatility and the number of corporate defaults. The link between the returns realized volatility and the defaults dynamics worths to be further investigated. Another aspect which should be further analyzed is the high persistence in the default intensity estimated by PARX models. The persistence of the shocks in the number of defaults could be caused by both persistence in the common default risk factors and contagion effects among firms. Overall, our results show that the PARX model including realized volatility and Leading Index fits the data satisfactorily and captures the default clustering. We have also performed a forecasting experiment in order to evaluate the PARX model out-of-sample performance during the 2008-2011 crisis period and reached quite satisfactory results, showing that including covariates improves the out-of-sample prediction of the default counts.

Chapter 6

Conclusions

We have developed this thesis work in the aim of studying the modelling of default risk, proposing a new modelling framework and highlighting the main factors influencing the corporate defaults dynamics.

We have started from the analysis of the stylized facts in corporate default counts and rates time series. The default phenomenon, as most rare events, is characterized by overdispersion - the variance of the number of events is much higher than its mean - leading to series showing both peaks ("clusters") and periods of low incidence. Moreover, the defaults time series are characterized by a slowly decreasing autocorrelation function, which is a typical feature of long-memory processes. In recent years, encouraged by the increasing relevance of the default phenomenon during the financial crisis started in 2008, the econometric and financial literature has shown a growing interest in default risk modelling. In particular, as seen in Chapter 2, in most works the topic of default predictability has been investigated by analyzing the link between the default clusters and the macroeconomic context. Another relevant aspect in default prediction is the role of rating, which we have analyzed both in the theoretical part of the thesis and in our empirical study. Several recent works - we have reviewed in details the approach of Das et al. (2006), Lando and Nielsen (2010), Lando et al. (2013) - have developed and applied models based on counting processes, where the modelled variable is the default intensity, i.e. the expected number of defaults in the time unit, typically a month. The use of counts eases the test of independence of default events conditional on common macroeconomic and financial factors. Comparing the distribution of the default counts to a Poisson distribution with constant intensity is the crucial feature of the cited works and has inspired our idea: modelling defaults with a *conditional* Poisson models with *time-varying intensity*, allowing for overdispersion and slowly decaying autocorrelation of the counts through the inclusion of autoregressive dynamics. We have, then, reviewed the recent literature of Autoregressive Conditional Poisson models (ACP), focusing on Poisson Autoregression by Fokianos, Rahbek and Tjøstheim (2009), which is the first work studying ergodicity of these models and providing the asymptotic theory, allowing for inference. Defining an autoregressive Poisson model for default counts, linking the expected number of default events on its past history, is the first part of our contribution. The inclusion of autoregressive components is also relevant in the analysis of correlation between corporate defaults, linked to the recent debate about the possible existence of default contagion effects.

The consideration that the expected number of defaults is probably influenced by the macroeconomic and financial context in which corporate firms operate has led us to the idea of extending Poisson Autoregression by Fokianos, Rahbek and Tjøstheim (2009) (PAR) by including exogenous covariates. This is our methodological contribution, developed in Chapter 4, where we have presented a class of Poisson intensity AutoRegressions with eXogeneous covariates (PARX) models that can be used for modelling and forecasting time series of counts. We have analyzed the time series properties and the conditions for stationarity for this new models, also developing the asymptotic theory. The PARX models provide a flexible framework for analyzing dependence of default intensity on both the past number of default events and other relevant variables. In Chapter 5 we have applied different Poisson Autoregressive models, presenting an extended empirical study of US corporate defaults based on Moody's monthly default count data. The time interval considered, going from January 1982 to December 2011, includes three clusters of defaults corresponding to three crisis periods: the last eighties financial markets crisis, the 2000-2001 information technology bubble and the financial and economic crisis started in 2008. We have proposed and motivated a selection of covariates which can potentially explain the default clusters and the strong autocorrelation in the number of defaults. An original feature is, in particular, the inclusion of a measure of intra-monthly realized volatility, computed from daily S&P 500 returns. Realized volatility is indeed expected to summarize the uncertainty on financial markets, characterizing the periods of financial turmoil when defaults are more likely to cluster. According to the results of our empirical analysis, the one-month lagged realized volatility of returns is the most relevant covariate in explaining default intensity, together with the one-month lagged Leading Index. The latter is a macroeconomic indicator provided by the Federal Reserve and including a set of variables expected to anticipate the US economic tendency. At our knowledge, ours is the first work showing a positive association between the financial market uncertainty captured by the realized volatility and the number of corporate defaults. Also the inclusion of the Leading Index is new and its significance highlights the predictive role of the business cycle, which previous works try to include using GDP and industrial production growth, not always found significant in explaining default frequencies. Overall, our results have shown that the PARX model including realized volatility and Leading Index fits the default count data satisfactorily and captures the default clustering. We have also performed a forecasting experiment in order to evaluate the PARX model out-of-sample performance during the 2008-2011 crisis period and reached quite satisfactory results, showing that including covariates improves the out-of-sample prediction of the default counts. However, the default counts dynamics are mainly led by the autoregressive components and show a high persistence of shocks, even when significant exogenous covariates are included. In this respect, the main consideration arising is that the modelling of the aggregate default intensity should be supported by the analysis of firm-specific, or, at least, sector-specific variables. Sector profit indexes, for example, could improve the default prediction, as solvency is strongly linked to the firms balance sheet data. Including less aggregate data in default risk analysis could also allow to identify the risk factors linked to correlation among the solvency conditions of different companies. The fact that the autoregressive components have a stronger role than the overall default risk factors in explaining the defaults dynamics is an interesting result. However, it is not sufficient to state that contagion effects explain the autocorrelation in the number of defaults, as long as the commercial and financial links among companies are not taken into account. Another important aspect to point out relative to the prominent role of the autoregressive part is that it should not discourage the search and the analysis of exogenous risk factors. Finding variables significantly associated to the number of defaults can indeed provide warning signals in default risk evaluation.

At the aggregate level, the default phenomenon is influenced by the financial and macroeconomic context, but, at the same time, has an effect on it. The most immediate example is that of the credit spreads - included in our empirical study - which reflect the level of default risk connected to financial positions. A higher default risk also affects the agents expectations, having an impact on the uncertainty captured by the financial returns volatility. When the number of defaults is high, also the companies investment decisions and the commercial links among firms are affected, with consequences on industrial production. These considerations suggest the relaxing of the covariate exogeneity assumption and, as a future development of our work, the definition of a multivariate model. Another aspect which should be further analyzed is the usefulness of the PARX models for defaults at the operational level: the relevance of a new model for default risk should be evaluated with respect to the actual needs in risk management practices. As an example, one of the main applications of the models for default risk concerns the pricing of corporate bonds. Measuring how much our estimated default intensity reflects in the market price of the financial instruments issued by rated companies could support the evaluation of the PARX models performance.

Appendix A

Proofs

Proof of Theorem 4.1

Define $\zeta := \max\left(\sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i), \rho\right) < 1$. Moreover, consider the norm given by $\|(x,\lambda)\|_w := w_x\|x\| + w_\lambda\|\lambda\|$, where $w_x, w_y > 0$ are chosen below. Next, with $\alpha = (\alpha_1, ..., \alpha_p)$ and $\beta = \left(\beta_1, ..., \beta_q\right)$, and, correspondingly, N of dimension p and λ of dimension q,

$$F(x,\lambda;\varepsilon,N) = (g(x;\varepsilon), \omega + \alpha N(\lambda) + \beta\lambda + \gamma f(x))', \tag{A.1}$$

consider, with $\bar{N}_t = (N_t, ..., N_{t-p})'$,

$$\begin{aligned}
&E\left[\left\|F\left(x,\lambda;\varepsilon_t,\bar{N}_t(\cdot)\right) - F\left(\tilde{x},\tilde{\lambda};\varepsilon_t,\bar{N}_t(\cdot)\right)\right\|_w\right] \\
&= w_x E\left[\|g(x;\varepsilon) - g(\tilde{x};\varepsilon)\|\right] + w_\lambda E\left[\left|\alpha\left\{\bar{N}_t(\lambda) - \bar{N}_t\left(\tilde{\lambda}\right)\right\} + \beta\left\{\lambda - \tilde{\lambda}\right\} + \gamma\{f(x) - f(\tilde{x})\}\right|\right] \\
&\leq w_x\rho^{1/s}\|x-\tilde{x}\| + w_\lambda\sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i)\left\|\lambda-\tilde{\lambda}\right\| + w_\lambda\gamma L\|x-\tilde{x}\| \\
&= \left[w_x\rho^{1/s} + w_\lambda\gamma L\right]\|x-\tilde{x}\| + w_\lambda\sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i)\left\|\lambda-\tilde{\lambda}\right\| \qquad \text{(A.2)}
\end{aligned}$$

If $\zeta = \sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i)$, then choose $w_\lambda = w_x\left(\zeta - \rho^{1/s}\right)/(\gamma L)$ such that,

$$E\left[\left\|F\left(x,\lambda;\varepsilon_t,\bar{N}_t(\cdot)\right) - F\left(\tilde{x},\tilde{\lambda};\varepsilon_t,\bar{N}_t(\cdot)\right)\right\|_w\right] \leq \zeta\left\|(x,\lambda) - \left(\tilde{x},\tilde{\lambda}\right)\right\|_w. \tag{A.3}$$

If $\zeta = \rho^{1/s}$, then choose,

$$w_x\rho^{1/s} + w_\lambda\gamma L = (1+\delta)\rho^{1/s}w_x,$$

or $w_\lambda = \delta\rho^{1/s}w_x/(\gamma L)$, for some small $\delta > 0$, such that $(1+\delta)\rho < 1$, and hence

$$E\left[\left\|F\left(x,\lambda;\varepsilon_t,\bar{N}_t(\cdot)\right)-F\left(\tilde{x},\tilde{\lambda};\varepsilon_t,\bar{N}_t(\cdot)\right)\right\|_w\right] \leq (1+\delta)\zeta\left\|(x,\lambda)-\left(\tilde{x},\tilde{\lambda}\right)\right\|_w$$

.

Finally, $E\left[\left\|F\left(0,0;\varepsilon_t,\bar{N}_t\right)\right\|_w\right] = w_x E\left[\|g(0;\varepsilon)\|\right] + w_\lambda(\gamma f(0)+\omega) < \infty$ by Assumption 4. Then the result holds by Corollary 3.1 in Doukhan and Wintenberger (2008).

That y_t is stationary is clear. Next, with $z_t := (x_t', \lambda_t)'$ consider

$$\begin{aligned} P\left((y_t,z_t)\in A\times B \mid \mathcal{M}_{y,t-p},\mathcal{M}_{z,t-p}\right) &= P\left(y_t\in A \mid z_t\in B,\, \mathcal{M}_{y,t-p},\mathcal{M}_{z,t-p}\right) \\ &\quad P\left(z_t\in B \mid \mathcal{M}_{y,t-p},\mathcal{M}_{z,t-p}\right), \end{aligned}$$

where $\mathcal{M}_{x,t-k} = \sigma(x_{t-k}, x_{t-k-1}, ...)$. Now by definition of the process,

$$P\left(y_t\in A \mid z_t\in B,\, \mathcal{M}_{y,t-p},\mathcal{M}_{z,t-p}\right) = P\left(y_t\in A \mid z_t\in B\right).$$

Next, using the Markov chain property of z_t,

$$P\left(z_t\in B \mid \mathcal{M}_{y,t-p},\mathcal{M}_{z,t-p}\right) = P\left(z_t\in B \mid \mathcal{M}_{z,t-p}\right),$$

where the right hand side by τ weak dependence of z_t converges to the marginal $P(z_t\in B)$ as $p\to\infty$. Hence so does $P\left((y_t,z_t)\in A\times B \mid \mathcal{M}_{y,t-p},\mathcal{M}_{z,t-p}\right)$ for any A, B and p, $p\to\infty$.

Now consider $E\left[|y_t^*|^s\right] = \sum_{j=0}^{s}\binom{s}{j}E[(\lambda_t^*)^j]$, where

$$E[\lambda_t^*] = \sum_{i=1}^{\max(p,q)}\left(\alpha_i+\beta_i\right)E\left(\lambda_t^*\right) + \gamma E f\left(x_{t-1}^*\right) + \omega$$

$$(\lambda_t^*)^s = \sum_{j=0}^{s}\binom{s}{j}\left(\alpha\bar{y}_{t-1}^* + \beta\bar{\lambda}_{t-1}^*\right)^j\left(\omega+\gamma f\left(x_{t-1}^*\right)\right)^{s-j},$$

with $\bar{y}_t = (y_t, ..., y_{t-p+1})'$ and $\bar{\lambda}_t = (\lambda_t, ..., \lambda_{t-q+1})'$.

Hence,

$$\begin{aligned} E[(\lambda_t^*)^s] &= \sum_{j=0}^{s}\binom{s}{j}E\left[\left(\alpha\bar{y}_{t-1}^*+\beta\bar{\lambda}_{t-1}^*\right)^j\left(\omega+\gamma f\left(x_{t-1}^*\right)\right)^{s-j}\right] \\ &= E\left[\left(\alpha\bar{y}_{t-1}^*+\beta\bar{\lambda}_{t-1}^*\right)^s + E\left(\omega+\gamma f\left(x_{t-1}^*\right)\right)^s\right] + E\left[r_{s-1}\left(\bar{y}_{t-1}^*,\bar{\lambda}_{t-1}^*, f\left(x_{t-1}^*\right)\right)\right], \end{aligned}$$

with $r_{s-1}(y, \lambda, z)$ an $(s-1)$-order polynomial in $\left(\bar{y}, \bar{\lambda}, z\right)$ and so $E\left[r_{s-1}(\cdot)\right] < \infty$ by induction assumption.

Moreover, $E\left[\left(\omega + \gamma f\left(x^*_{t-1}\right)\right)^s\right] < \infty$ by applying Doukhan and Wintenberger (2008) (Theorem 3.2) on x_t and applying Assumption 2, such that we are left with considering terms of the form,

$$
\begin{aligned}
& E\left[\left(\alpha_i y^*_{t-1-i} + \beta_i \lambda^*_{t-1-i}\right)^s\right] \\
&= \sum_{j=0}^{s} \binom{s}{j} \alpha_i^j \beta_i^{s-j} E\left[\left(y^*_{t-1-i}\right)^j \left(\lambda^*_{t-1-i}\right)^{s-j}\right] \\
&= \sum_{j=0}^{s} \binom{s}{j} \alpha_i^j \beta_i^{s-j} \sum_{k=0}^{j} \binom{j}{k} E\left[\left(\lambda^*_t\right)^{s+(k-j)}\right] \\
&= \sum_{j=0}^{s} \binom{s}{j} \alpha_i^j \beta_i^{s-j} E\left[\left(\lambda^*_t\right)^s\right] + C \\
&= \left(\alpha_i + \beta_i\right)^s E\left[\left(\lambda^*_t\right)^s\right] + C,
\end{aligned}
$$

as by induction assumption all $E\left[\left(\lambda^*_t\right)^k\right] < \infty$, for $k < s$. Collecting terms,

$$
E\left[\left(\lambda^*_t\right)^s\right] = \left[\sum_{i=1}^{\max(p,q)} \left(\alpha_i + \beta_i\right)\right]^s E\left[\left(\lambda^*_t\right)^s\right] + \tilde{C},
$$

which for $\sum_{i=1}^{\max(p,q)} \left(\alpha_i + \beta_i\right) < 1$ has a well-defined solution. □

Proof of Lemma 4.1

In terms of initial values, consider, next, a process $X_t = F\left(X_{t-1}, \varepsilon_t\right)$, where $\left\|F(x;\xi) - F(\tilde{x};\xi)\right\|_\phi \leq \rho \left\|x - \tilde{x}\right\|$, $|\rho| < 1$ and $\left\|g(0;\varepsilon)\right\|_\phi < \infty$, which is τ-weakly dependent. With X^*_t denoting the stationary solution and $X_0 = x$ fixed, we wish to show, for some $h \leq \phi$, $\frac{1}{T}\sum_{t=1} h\left(X_t\right) \overset{a.s.}{\rightarrow} E[h\left(X^*_t\right)]$. Now,

$$
\frac{1}{T}\sum_{t=1} h\left(X_t\right) = \frac{1}{T}\sum_{t=1} \left[h\left(X_t\right) - h\left(X^*_t\right)\right] + \frac{1}{T}\sum_{t=1} h\left(X^*_t\right),
$$

and

$$
\left|\frac{1}{T}\sum_{t=1} \left[h\left(X_t\right) - h\left(X^*_t\right)\right]\right| \leq \frac{1}{T}\sum_{t=1} \left|h\left(X_t\right) - h\left(X^*_t\right)\right|.
$$

Assume furthermore that $|h(x) - h(\tilde{x})| \leq L||x - \tilde{x}||$, $\phi(z) \geq z$, $z > 0$, then we find by repeated use of iterated expectations,

$$\begin{aligned} E\left[|h(X_t) - h(X_t^*)|\right] &= E\left[E\left[|h(X_t) - h(X_t^*)|\,|\,X_{t-1}, X_{t-1}^*\right]\right] \\ &\leq LE\left[E\left[\left\|g(X_{t-1}, \varepsilon_t) - g\left(X_{t-1}^*, \varepsilon_t\right)\right\|\,|\,X_{t-1}, X_{t-1}^*\right]\right] \\ &\leq L\rho E\left[E\left[\left\|X_{t-1} - X_{t-1}^*\right\|\,|\,X_{t-1}, X_{t-1}^*\right]\right] \\ &= L\rho E\left[\left\|X_{t-1} - X_{t-1}^*\right\|\right] \leq L\rho^t E\left[\|X_0 - X_0^*\|\right] \end{aligned}$$

Proof of Lemma 4.2

The proof mimics the proof of Lemma 2.1 in Fokianos, Rahbek and Tjøstheim (2009), where the case $p = q = 1$ is treated. Without loss of generality, set here $p = q$, such that, by definition,

$$\lambda_t^c - \lambda_t = \sum_{i=1}^{p} \left[\alpha_i \left(y_{t-i}^c - y_{t-i}\right) + \beta_i \left(\lambda_{t-i}^c - \lambda_{t-i}\right)\right] + \gamma e_t^c, \tag{A.4}$$

with $e_t^c := f(x_{t-1})\,\mathbb{I}(\|x_{t-1}\| \geq c)$. Hence $E[\lambda_t^c - \lambda_t] = \sum_{i=0}^{t-1}\left(\sum_{j=1}^{p}\left[\alpha_j + \beta_j\right]\right)^i E\left(e_{t-i}^c\right)$, and, as $\sum_{j=1}^{p}\left[\alpha_j + \beta_j\right] < 1$, $\left|E\left(e_{t-i}^c\right)\right| \leq \zeta_1(c)$ with $\zeta_1(c) \to 0$ as $c \to \infty$, the first result holds with $\delta_1(c) := \zeta_1(c) / \left(1 - \sum_{j=1}^{p}\left[\alpha_j + \beta_j\right]\right)$. Next,

$$\begin{aligned} E\left(\lambda_t^c - \lambda_t\right)^2 = &\sum_{i=1}^{p} \alpha_i^2 E\left(y_{t-i}^c - y_{t-i}\right)^2 + \beta_i^2 E\left(\lambda_{t-i}^c - \lambda_{t-i}\right)^2 + \gamma^2 E\left(e_t^c\right)^2 \\ &+ 2\sum_{i,j=1,i<j}^{p} \alpha_i\beta_j E\left(\lambda_{t-j}^c - \lambda_{t-j}\right)\left(y_{t-i}^c - y_{t-i}\right) \\ &+ 2\sum_{i=1}^{p} \alpha_i E\left[\left(\lambda_{t-i}^c - \lambda_{t-i}\right)\gamma e_t^c\right] + 2\sum_{i=1}^{p} \beta_i \gamma E\left[e_t^c\left(y_{t-i}^c - y_{t-i}\right)\right] \\ &+ 2\sum_{i,j=1,i<j}^{p} \alpha_i\alpha_j E\left(y_{t-j}^c - y_{t-j}\right)\left(y_{t-i}^c - y_{t-i}\right) \\ &+ 2\sum_{i,j=1,i<j}^{p} \beta_i\beta_j E\left(\lambda_{t-j}^c - \lambda_{t-j}\right)\left(\lambda_{t-i}^c - \lambda_{t-i}\right) \end{aligned}$$

With $\lambda_t^c \geq \lambda_t$, and $t \leq s$,

$$\begin{aligned}
&E\left[\left(\lambda_t^c-\lambda_t\right)\left(y_s^c-y_s\right)\right] \\
&=E\left[E\left(\left(\lambda_t^c-\lambda_t\right)\left(y_s^c-y_s\right) \mid \mathcal{F}_{s-1}\right)\right] \\
&=E\left[\left(\lambda_t^c-\lambda_t\right) E\left(N_s\left[\lambda_s, \lambda_s^c\right]\right)\right]=E\left(\lambda_t^c-\lambda_t\right)\left(\lambda_s^c-\lambda_s\right),
\end{aligned} \tag{A.5}$$

where $\mathcal{F}_{s-1}=\sigma\left(x_k, N_k,\ k \leq s-1\right)$ and $N_t\left[\lambda_t, \lambda_t^c\right]$ is the number of events in $\left[\lambda_t, \lambda_t^c\right]$ for the unit-intensity Poisson process N_t. Likewise for $\lambda_t \geq \lambda_t^c$. Also observe that, still for $t \leq s$,

$$\begin{aligned}
&E\left[\left(y_t^c-y_t\right)\left(y_s^c-y_s\right)\right]=E\left[E\left(\left(y_t^c-y_t\right)\left(y_s^c-y_s\right) \mid \mathcal{F}_{s-1}\right)\right] \\
&=E\left[\left(y_t^c-y_t\right) E\left(\left(y_s^c-y_s\right) \mid \mathcal{F}_{s-1}\right)\right]=E\left(y_t^c-y_t\right)\left(\lambda_s^c-\lambda_s\right),
\end{aligned} \tag{A.6}$$

For $t \geq s$, note that the recursion for $\left(\lambda_t^c-\lambda_t\right)$ above gives,

$$\begin{aligned}
\lambda_t^c-\lambda_t &=\sum_{i=1}^p\left[\alpha_i\left(y_{t-i}^c-y_{t-i}\right)+\beta_i\left(\lambda_{t-i}^c-\lambda_{t-i}\right)\right]+\gamma e_t^c \\
&=\sum_{i=1}^p \beta_i\left[\sum_{j=1}^p\left[\alpha_j\left(y_{t-i-j}^c-y_{t-i-j}\right)+\beta_j\left(\lambda_{t-i-j}^c-\lambda_{t-i-j}\right)\right]+\gamma e_{t-i}^c\right] \\
&+\sum_{i=1}^p\left[\alpha_i\left(y_{t-i}^c-y_{t-i}\right)+\gamma e_t^c\right] \\
&=\ldots \\
&=\sum_{j=1}^{t-s}\left(a_j\left(y_{t-j}^c-y_{t-j}\right)+g_j e_{t-j}\right)+\sum_{j=1}^p\left[c_j\left(\lambda_{s-j}^c-\lambda_{s-j}\right)+d_j e_s^c+h_j\left(y_{s-j}^c-y_{s-j}\right)\right].
\end{aligned} \tag{A.7}$$

Observe that a_j, g_j, c_j, d_j and h_j are all summable. Using this, we find,

$$\begin{aligned}
E\left[\left(\lambda_t^c-\lambda_t\right)\left(y_s^c-y_s\right)\right]=&\, E\left(\sum_{j=1}^{t-s}\left(a_j\left(y_{t-j}^c-y_{t-j}\right)+g_j e_{t-j}\right)\left(y_s^c-y_s\right)\right) \\
&+E\left(\sum_{j=1}^p\left[c_j\left(\lambda_{s-j}^c-\lambda_{s-j}\right)+d_j e_s^c+h_j\left(y_{s-j}^c-y_{s-j}\right)\right]\left(y_s^c-y_s\right)\right)
\end{aligned} \tag{A.8}$$

Collecting terms, one finds $E\left(\lambda_t^c-\lambda_t\right)^2$ is bounded by $C \sum_{j=1}^t \psi_j E\left(e_{t-j}^c\right)^2$ for some constant C, some ψ_i with $\sum_{i=1}^{\infty} \psi_i<\infty$ and which therefore tends to zero.

Finally, using again the properties of the Poisson process N_t, we find

$$E\left(y_t^c - y_t\right)^2 \leq E\left[\left(\lambda_t^c - \lambda_t\right)^2\right] + \left|E\left(\lambda_t^c - \lambda_t\right)\right| \leq E\left(\lambda_t^c - \lambda_t\right)^2 + \delta_1\left(c\right). \tag{A.9}$$

This completes the proof of Lemma 4.2.

Proof of Theorem 4.2

We provide the proofs for the case of $p = q = 1$ as the general case is complex in terms of notation. With $p = q = 1$,

$$\lambda_t\left(\theta\right) = \omega + \alpha y_{t-1} + \beta \lambda_{t-1}\left(\theta\right) + \gamma f\left(x_{t-1}\right).$$

The result is shown by verifying the conditions in Kristensen and Rahbek (2005, Lemma X).

Score

The score $S_T\left(\theta\right) = \partial L_T\left(\theta\right) / \left(\partial\theta\right)$ is given by

$$S_T\left(\theta\right) = \sum_{t=1}^{T} s_t\left(\theta\right), \quad \text{where } s_t\left(\theta\right) = \left(\frac{y_t}{\lambda_t\left(\theta\right)} - 1\right)\frac{\partial \lambda_t\left(\theta\right)}{\partial\theta}. \tag{A.10}$$

Here, with $\eta = \left(\omega, \alpha, \gamma\right)'$ and $v_t = \left(1, y_{t-1}, f\left(x_{t-1}\right)\right)'$

$$\frac{\partial \lambda_t\left(\theta\right)}{\partial \eta} = v_t + \beta \frac{\partial \lambda_{t-1}\left(\theta\right)}{\partial \eta} \tag{A.11}$$

$$\frac{\partial \lambda_t\left(\theta\right)}{\partial \beta} = \lambda_{t-1}\left(\theta\right) + \beta \frac{\partial \lambda_{t-1}\left(\theta\right)}{\partial \gamma} \tag{A.12}$$

In particular, with $\lambda_t = \lambda_t\left(\theta_0\right)$,

$$s_t\left(\theta_0\right) = \frac{\partial \lambda_t\left(\theta\right)}{\partial\theta}\xi_t, \quad \xi_t := \left(\frac{N_t\left(\lambda_t\right)}{\lambda_t} - 1\right). \tag{A.13}$$

and where $\dot{\lambda}_t = \partial \lambda_t\left(\theta\right) / \left(\partial\theta\right)_{\theta=\theta_0}$. This is a martingale difference sequence with respect to $\mathcal{F}_t = \mathcal{F}\left(y_{t-k}, x_{t-k}, \lambda_{t-k}, \quad k = 0, 1, 2, ...\right)$ as $E\left(\xi_t | \mathcal{F}_{t-1}\right) = 0$. It therefore follows by the CLT for martingales, see, e.g., Brown (1971), that $\sqrt{T} S_T\left(\theta_0\right) \to^d N\left(0, \Omega\right)$, where

$$\Omega = E\left[s_t\left(\theta_0\right) s_t\left(\theta_0\right)'\right],$$

if we can show that the quadratic variation converges, $\langle S_T(\theta_0)\rangle \to^P \Omega$. To this end, observe that $E\left[\xi_t^2|\mathcal{F}_{t-1}\right] = 1/\lambda_t < 1/\omega_0$. Thus,

$$\langle S_T(\theta_0)\rangle = \frac{1}{T}\sum_{t=1}^{T} E\left[s_t(\theta_0)\, s_t(\theta_0)'\,|\mathcal{F}_{t-1}\right] = \frac{1}{T}\sum_{t=1}^{T}\dot{\lambda}_t\dot{\lambda}_t'/\lambda_t, \tag{A.14}$$

where $\dot{\lambda}_t = \partial\lambda_t(\theta)/(\partial\theta)_{\theta=\theta_0}$. As $\dot{\lambda}_0 = 0$,

$$\dot{\lambda}_t = (v_t', \lambda_{t-1})' + \beta\dot{\lambda}_{t-1} = \sum_{i=1}^{t-1}\beta^i\left(v_{t-i}', \lambda_{t-1-i}\right)', \tag{A.15}$$

By the same arguments as in the proof of Theorem 4.1, it is easily checked that the augmented process $\tilde{X}_t := \left(X_t, \dot{\lambda}_t\right)$, with X_t defined in Theorem 4.1, is weakly dependent with second moment. Since $\lambda_t \geq \omega$, it therefore follows that $E\left[\dot{\lambda}_t^*\left(\dot{\lambda}_t^*\right)'/\lambda_t^*\right] < \infty$. Thus, we can employ Lemma 4.1 to obtain that $\frac{1}{T}\sum_{t=1}^{T}\dot{\lambda}_t\dot{\lambda}_t'/\lambda_t \to \Omega$.

Information

It is easily verified that

$$-\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} = \frac{y_t}{\lambda_t^2(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta}\frac{\partial\lambda_t(\theta)}{\partial\theta'} - \left(\frac{y_t}{\lambda_t(\theta)} - 1\right)\frac{\partial^2\lambda_t(\theta)}{\partial\theta\partial\theta'}, \tag{A.16}$$

where

$$\frac{\partial^2\lambda_t(\theta)}{\partial\eta\partial\beta} = \frac{\partial\lambda_{t-1}(\theta)}{\partial\eta} + \beta\frac{\partial^2\lambda_{t-1}(\theta)}{\partial\eta\partial\beta} = \sum_{i=1}^{t-1}\beta^i\frac{\partial\lambda_{t-i}(\theta)}{\partial\eta} \tag{A.17}$$

$$\frac{\partial^2\lambda_t(\theta)}{\partial\beta^2} = 2\frac{\partial\lambda_{t-1}(\theta)}{\partial\beta} + \beta\frac{\partial^2\lambda_{t-1}(\theta)}{\partial\beta^2} = 2\sum_{i=1}^{t-1}\beta^i\frac{\partial\lambda_{t-i}(\theta)}{\partial\beta} \tag{A.18}$$

$$\frac{\partial^2\lambda_t(\theta)}{\partial\eta^2} = \beta\frac{\partial^2\lambda_t(\theta)}{\partial\eta^2} = ... = 0 \tag{A.19}$$

In particular, the augmented process $\tilde{X}_t(\theta) := \left(X_t'(\theta), \dot{\lambda}_t(\theta), \ddot{\lambda}_t(\theta)\right)'$ can be shown to be weakly dependent with second moments for $\theta \in \Theta$. In particular, for all $\theta \in \Theta$,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} = \frac{1}{T}\sum_{t=1}^{T} h\left(\tilde{X}_t(\theta)\right) \to^P E\left[h\left(\tilde{X}_t^*(\theta)\right)\right], \quad h\left(\tilde{X}_t(\theta)\right) = \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}.$$

Moreover, $\theta \mapsto \partial^2 l_t(\theta)/(\partial\theta\partial\theta')$ is continuous and satisfies

$$\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right\| \leq \bar{D}\left(\tilde{X}_t(\bar{\theta})\right) := \frac{y_t}{\omega_L^2}\frac{\partial\lambda_t(\bar{\theta})}{\partial\theta}\frac{\partial\lambda_t(\bar{\theta})}{\partial\theta'} - \left(\frac{y_t}{\omega_L} - 1\right)\frac{\partial^2\lambda_t(\bar{\theta})}{\partial\theta\partial\theta'},$$

where $\bar{\theta} = (\omega_U, \alpha_U, \beta_U, \gamma_U)$ contains the maximum values of the individual parameters in Θ, with $E\left[\bar{D}\left(\tilde{X}_t^*(\theta)\right)\right] < \infty$. For example,

$$\frac{\partial \lambda_t(\theta)}{\partial \beta} = \lambda_{t-1}(\theta) + \beta \frac{\partial \lambda_{t-1}(\theta)}{\partial \gamma} \leq \sum_{i=0}^{t-1} \beta_U^i \lambda_{t-1-i}\left(\bar{\theta}\right) = \frac{\partial \lambda_t\left(\bar{\theta}\right)}{\partial \beta} \tag{A.20}$$

and

$$\frac{\partial^2 \lambda_t(\theta)}{\partial \beta^2} = 2\frac{\partial \lambda_{t-1}(\theta)}{\partial \beta} + \beta \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial \beta^2} \leq 2\sum_{i=0}^{t-1} \beta_U^i \dot{\lambda}_{t-1-i}\left(\bar{\theta}\right) = \frac{\partial^2 \lambda_t\left(\bar{\theta}\right)}{\partial \beta^2}. \tag{A.21}$$

It now follows by Lemma X in Kristensen and Rahbek (2005) that

$$\sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} - E\left[h\left(\tilde{X}_t^*(\theta)\right)\right] \right\| \xrightarrow{p} 0. \tag{A.22}$$

Proof of Theorem 4.3

The proof follows by noting that Lemmas 3.1-3.4 in FRT (2009) hold for our setting. The only difference is that the parameter vector θ include γ loading $f(x_{t-1})$. However, as $E[f(x_{t-1})] < \infty$, all the arguments remain identical as is easily seen upon inspection of the proofs of the lemmas in FRT (2009).

Bibliography

Aalen, O. O. (1989), "A model for non-parametric regression analysis of life times", in J. Rosinski, W. Klonecki, and A. Kozek (eds.), *Mathematical Statistics and Probability Theory*, vol. 2 of Lecture Notes in Statistics, pp. 1–25, Springer, New York.

Agosto, A., and Moretto, E. (2012), "Exploiting default probabilities in a structural model with nonconstant barrier", *Applied Financial Economics*, 22:8, 667-679.

Akaike, H. (1974), "A new look at the statistical model identification", *lEEE Transactions on Automatic Control*, AC-19, 716-723.

Amisano, G., and Giacomini, R. (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", *Journal of Business and Economic Statistics*, 25, 177-190.

Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1992), *Statistical Models Based on Counting Processes*, Springer-Verlag.

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study", *Annals of Statistics*, 10, 1100–1120.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007), "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility", *The Review of Economic and Statistics*, 89, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), "The distribution of realized exchange rate volatility", *Journal of the American Statistical Association*, 96, 42–55.

Azizpour, S., Giesecke, K., (2008a), "Premia for Correlated Default Risk. Department of Management Science and Engineering", Stanford University. Unpublished manuscript.

Azizpour, S., Giesecke, K., (2008b), "Self-exciting Corporate Defaults: Contagion vs. Frailty", Department of Management Science and Engineering, Stanford University. Unpublished manuscript.

Azizpour, S., Giesecke, K., (2010), "Azizpour, S., Giesecke, K., (2010), "Exploring the sources of default clustering", Department of Management Science and Engineering, Stanford University. Unpublished manuscript.

Barndorff-Nielsen, O., and Shephard, N., 2002, "Estimating quadratic variation using realized variance", *Journal of Applied Econometrics* 17, 457–477.

Behar, R., and Nagpal, K. (2001), "Dynamics of rating transition", Algo Research Quarterly, 4 (March/June), 71–92.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307–327.

Blume, M. E., Lim, F., and Craig, A. (1998), "The Declining Credit Quality of U.S. Corporate Debt: Myth or Reality?", *The Journal of Finance*, 53, 1389-1413.

Brockwell, P.J. and Davis, R. A. (1991), *Time Series: Data Analysis and Theory*, Springer, New York, 2nd edition.

Brown, B. M. (1971), "Martingale Central Limit Theorems", *The Annals of Mathematical Statistics*, 42, 59-66.

Christoffersen, P.F. and Diebold, F.X. (1997), "Optimal Prediction Under Asymmetric Loss," *Econometric Theory*, 13, 808-817.

Chou, H. (2012), "Using the autoregressive conditional duration model to analyse the process of default contagion", *Applied Financial Economics*, 22:13, 1111-1120.

Czado, C., Gneiting, T. and Held, L. (2009), "Predictive Model Assessment for Count Data," *Biometrics* 65, 1254–1261.

Cox, D. R. (1972), "Regression models and life-tables (with discussion)", *Journal of the Royal Statistical Society*, Series B, 34, 187-220.

Cox, D. R. (1975), "Partial likelihood", *Biometrika*, 62, 69-76.

Cox, D. R., and Snell, E. J. (1968), "A general definition of residuals", *Journal of the Royal Statistical Society*, Series B, 30, 248-275.

Corsi, F. (2009), "A Simple Approximate Long-Memory Model of Realized Volatility", *Journal of Financial Econometrics*, 7, 174–196.

Crosbie, P. J., and Bohn, J. (2002), "Modeling default risk", Technical report, KMV, LLC.

Das, S.R., Duffie, D., Kapadia, N., and Saita, L. (2007), "Common failings: How corporate defaults are correlated," *Journal of Finance* 62, 93–117.

Davis, M., and Lo, V. (2001), "Modeling default correlation in bond portfolios", in C. Alexander, ed., *Mastering Risk Volume 2: Applications*, Prentice Hall, pp. 141-151.

Davis, A. R., and Wu (2009), R., "A negative binomial model for time series of counts", *Biometrika*, 96, 735-749.

Dedecker, J. and Prieur, C. (2004), "Coupling for τ-dependent sequences and applications", *Journal of Theoretical Probability*, 17, 861–855.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998), "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, 39, 863-883.

Doukhan, P., and Wintenberger, O. (2008), "Weakly dependent chains with infinite memory", *Stochastic Processes and their Applications*, 118, 1997-2013.

Duffie, D., and Singleton, K. (1999), "Modeling Term Structure of Defaultable Bonds", *The Review of Financial Studies*, 12:4, 687-720.

Duffie, D., Saita, L., and Wang, K. (2007), "Multi-period corporate default prediction with stochastic covariates", *Journal of Financial Economics*, 83, 635-665.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009), "Frailty Correlated Default", *Journal of Finance*, 64, 2089-2123.

Engle, R. F. (2002), "New frontiers for ARCH models", *Journal of Applied Econometrics*, 17, 425–446.

Engle, R. F., and Gallo, G. M. (2006), "A multiple indicators model for volatility using intra-daily data", *Journal of Econometrics*, 131, 3-27.

Engle, R. F., and Ng, V. (1993), "Measuring and testing of the impact of news on volatility", *Journal of Finance*, 48, 1749-1778.

Engle, R. F., and Russell, J.R. (1998), "Autoregressive conditional duration: a new model for irregularly spaced transaction data", *Econometrica*, 66:5, 1127-62.

Fahrmeir, L., and Kaufmann, H. (1985), "Consistency and asymptotic normality of the maximum likelihood estimates in generalized linear models", *Annals of Statistics*, 13, 342-368.

Fama, E. F. (1965), "The Behavior of Stock-Market Prices", *The Journal of Business*, 38, 34-105.

Ferland, R., Latour, A., and Oraichi, D. (2006), "Integer-Valued GARCH Processes", *Journal of Time Series Analysis*, 27, 923–942.

Focardi, S.M., and Fabozzi, F.J. (2005), "An autoregressive conditional duration model of credit-risk contagion", *The Journal of Risk Finance*, 6, 208 - 225.

Fokianos, K. (2001), "Truncated Poisson regression for time series of counts", *Scandinavian Journal of Statistics*, 28, 645-659.

Fokianos, K., and Kedem, B. (2004), "Partial Likelihood Inference for Time Series Following Generalized Linear Models", *Journal of Time Series Analysis*, 25, 173–197.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009), "Poisson autoregression", *Journal of the American Statistical Association*, 104, 1430–1439.

French, K. R., Schwert, G. W., and Staumbaugh, R. F. (1987), *Journal of Financial Economics*, 19, 3-29.

Gallo, G. M., and Pacini, B. (2000), "The effects of trading activity on market volatility", *The European Journal of Finance* 6, 163–175.

Giesecke, K., Longstaff, F., Schaefer, S., and Strebulaev, I. (2011), "Corporate bond default risk: A 150-year perspective", *Journal of Financial Economics*, 102, 233-250.

Glosten. L. R.. Jagannathan. R.. and Runkle. D. (1993), "Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, 48, 1779-1802.

Gourieroux, C., Monfort, A. and Trognon, A. (1984), "Pseudo Maximum Likelihood Methods Theory", *Econometrica*, 52, 681-700.

Hamilton, J. (2005), Regime-Switching Models", *The New Palgrave Dictionary of Economics.*

Han, H., and Park, J.Y. (2008), "Time series properties of ARCH processes with persistent covariates", *Journal of Econometrics*, 146, 275–292.

Han, H., and Kristensen, D. (2013), "Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates," CeMMAP working papers CWP18/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Hansen, P.R., Huang, Z. and Shek, H.W. (2012) "Realized GARCH: A joint model for returns and realized measures of volatility," *Journal of Applied Econometrics,* 27, 877–906.

Hausman, A., Hall, B. H., and Griliches, Z. (1984) Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica*, 52, 909-938.

Hawkes, A.G., (1971), "Spectra of some self-exciting and mutually exciting point processes", *Biometrika*, 58, 83–90.

Heinen, A. (2003), "Modeling time series count data: An autoregressive conditional Poisson model", CORE Discussion Paper 2003/62, Center of Operations research and Econometrics, Université Catholique de Louvain.

Hilbe, J. M. (2007), *Negative binomial regression*, Cambridge University Press.

Jarrow, R.,and Turnbull, S. (1995), "Pricing options on Financial Securities Subject to Default Risk", *Journal of Finance*, 50, 53–86.

Jarrow, R., Lando, D., Turnbull, S. (1997), "A Markov model for the term structure of credit risk spreads", *Review of Financial Studies*, 481–523.

Jarrow, R. and Fan, Y. (2001), "Counterparty risk and the pricing of defaultable securities", *Journal of Finance*, 56, 555-576.

Jensen, S. T., and Rahbek, A. (2004), "Asymptotic Inference for Nonstationary GARCH", *Econometric Theory*, 20, 1203–1226.

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992), *Univariate Discrete Distributions*, second edition, John Wiley & Sons, Inc., New York.

Jung, R.C., Kukuk, M. and Liesenfeld, R. (2006), "Time series of count data: modeling, estimation and diagnostics", *Computational Statistics and Data Analysis*, 51, 2350-2364.

Kavvathas, D., "Estimating credit rating transition probabilities for corporate bonds", Working paper, University of Chicago.

Koopman, S.J., and Lucas, A. (2005), "Business and Default Cycle for Credit Risk", *Journal of Applied Econometrics*, 20: 311–323.

Koopman, S.J., Lucas, A., and Monteiro, A. (2008), "The multi-state latent factor intensity model for credit rating transitions", *Journal of Econometrics*, 142, 399-424.

Koopman, S.J., Lucas, A., and Schwaab, B., "Modeling frailty-correlated defaults using many macroeconomic covariates", *Journal of Econometrics*, 162, 312-325.

Kedem, B., and Fokianos, K. (2002), *Regression Models for Time Series Analysis*, Hoboken, NJ: Wiley.

Kristensen, D. and Rahbek, A. (2005), "Asymptotics of the QMLE for a Class of ARCH(q) Models", *Econometric Theory*, 21, 946–961.

Lambert, D. (1992), "Zero-inflated Poisson regression, with an application to defects in manufacturing", *Technometrics*, 34, 1-14.

Lamoureux, C. G., and Lastrapes, W. D. (1990), "Heteroskedasticity in stock return data: Volume versus GARCH effects", *Journal of Finance*, 45, 221–229.

Lando, D. (1998), "On Cox processes and credit risky securities, *Review of Derivatives Research*, 2, 99–120.

Lando, D., and Nielsen, M. (2010), "Correlation in corporate defaults: Contagion or conditional independence?", *Journal of Financial Intermediation*, 19, 355-372.

Lando, D., Medhat, M., Nielsen, M., and Nielsen, S. (2013), "Additive Intensity Regression Models in Corporate Default Analysis", *Journal of Financial Econometrics*, 11, 443–485.

Lando, D., and Skødeberg, T. M. (2002), "Analyzing rating transitions and rating drift with continuous observations", *Journal of Banking and Finance*, 26, 423-444.

Lang, L.H.P., Stulz, R.M., (1992), "Contagion and competitive intra-industry effects of bankruptcy announcements. An empirical analysis", *Journal of Financial Economics,* 32, 45–60.

Leland, H. E. (1994), "Corporate debt value, bond covenants, and the optimal capital structure", *Journal of Finance*, 49, 1213–52.

Leland, H. E. and Toft, K. B. (1996), "Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads", *Journal of Finance*, 60, 987–1019.

Li, W. K. (1991), "Testing model adequacy for some Markov regression models for time series", *Biometrika*, 78, 83-89.

Martens, M., van Dijk, D., de Pooter. M. (2004),"Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements", *International Journal of Forecasting*, 25, 282-303.

McCullagh, P. (1986), "The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data", *Journal of the American Statistical Association*, 81:393, 104-107.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, Chapman & Hall, New York.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, Chapman & Hall, London, 2nd edition.

Meitz, M., and Saikonnen, P. (2008), "Ergodicity,Mixing and Existence of Moments of a Class of Markov Models With Applications to GARCH and ACD Models", *Econometric Theory*, 24, 1291–1320.

Meyn, S. P., and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, London: Springer.

Merton, R. C. (1974), "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, 29, 49–70.

Mullahy, J. (1986), "Specification and testing of some modified count data models", *Journal of Econometrics*, 33, 341{365.

Nelder, J. A., and Wedderburn, R. W. M. (1972), "Generalized linear models", *Journal of the Royal Statistical Society*, Series A, 135:370-384.

Nelson. D. B. (1991). "Conditional Heteroskedasticity in Asset Pricing: A New Approach", *Econometrica*, 59, 347-370.

Nickell, P., Perraudin, W., and Varotto, S. (2000), "Stability of rating transitions", *Journal of Banking and Finance*, 24, 203-227.

Rydberg, T. H., and Shephard, N. (2000), "A Modeling Framework for the Prices and Times of Trades on the New York Stock Exchange," in *Nonlinear and Nonstationary Signal Processing*, eds. W. J. Fitzgerlad, R. L. Smith, A. T. Walden, and P. C. Young, Cambridge: Isaac Newton Institute and Cambridge University Press, pp. 217–246.

Shephard, N. and Sheppard, K. (2010), Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* 25, 197-231.

Shumway, T. (2001), Forecasting bankruptcy more efficiently: A simple hazard model, *Journal of Business*, 74, 101–124.

Schwarz, G. (1978), "Estimating the dimension of a model", *Annals of Statistics*, 6, 461-464.

Schwert, G. W. (1989), "Why Does Stock Market Volatility Change Over Time?", *The Journal of Finance*, 44, 1115-1153.

Skeel, D. A. (2001), "Debt's Dominion: A History of Bankruptcy Law in America", Princeton University Press.

Streett, S. (2000), "Some Observation Driven Models for Time Series of Counts," Ph.D. thesis, Colorado State University, Dept. of Statistics.

Tay, A.S and Wallis, K.F. (2000) "Density Forecasting: A Survey", *Journal of Forecasting*, 19, 235-254.

Tang, T. T. (2009), "Information asymmetry and firms' credit market access: Evidence from Moody's credit rating format refinement", *Journal of Financial Economics*, 93, 325-351.

Wedderburn, R. W. M. (1974), "Quasi-likelihood functions, generalized linear models and the Gaussian method", *Biometrika*, 61, 439-447.

Wong, W. H. (1986), "Theory of partial likelihood", *Annals of Statistics*, 14, 88-123.

Zeger, S. L., and Qaqish, B. (1988), "Markov Regression Models for Time Series: A Quasi-Likelihood Approach," *Biometrics*, 44, 1019–1031.