

Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)

**Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen
and Anders Rahbek**

CREATES Research Paper 2015-11

Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)*

Arianna Agosto[†]

Giuseppe Cavaliere[‡]

Dennis Kristensen[§]

Anders Rahbek[¶]

*We thank Bent J. Christensen, David Lando, Peter C.B. Phillips, Enrique Sentana, as well as participants at the 2014 (EC)² conference, the 6th Italian Congress of Econometrics and Empirical Economics (ICEEE 2015) and a seminar at Imperial College, for useful comments. The authors acknowledge support from Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Cavaliere and Rahbek thank the Italian Ministry of Education, University and Research (MIUR), PRIN project “Multivariate statistical models for risk assessment”, for financial support. Rahbek acknowledges research support by the Danish Council for Independent Research, Sapere Aude – DFF Advanced Grant (Grant no.: 12-124980). Kristensen acknowledges research support by the ESRC through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 and the European Research Council (grant no. ERC-2012-StG 312474). We also thank Moody’s Investors Service for providing us with data.

[†]Gruppo Bancario Credito Valtellinese, 23100 Sondrio, Italy. E-mail: agosto.arianna@creval.it.

[‡]Corresponding author. Department of Statistical Sciences, University of Bologna, I-40126 Bologna, Italy; Department of Economics, University of Copenhagen. E-mail: giuseppe.cavaliere@unibo.it.

[§]Department of Economics, University College London, WC1E 6BT, United Kingdom; Institute of Fiscal Studies; CREATES, University of Aarhus. E-mail: d.kristensen@ucl.ac.uk.

[¶]Department of Economics, University of Copenhagen, 1353 Copenhagen K, Denmark; CREATES, University of Aarhus. E-mail: anders.rahbek@econ.ku.dk.

Abstract: We develop a class of Poisson autoregressive models with additional covariates (PARX) that can be used to model and forecast time series of counts. We establish the time series properties of the models, including conditions for stationarity and existence of moments. These results are in turn used in the analysis of the asymptotic properties of the maximum-likelihood estimators of the models. The PARX class of models is used to analyse the time series properties of monthly corporate defaults in the US in the period 1982-2011 using financial and economic variables as exogeneous covariates. Results show that our model is able to capture the time series dynamics of corporate defaults well, including the well known default counts clustering found in data. Moreover, we find that while in general current defaults do indeed affect the probability of other firms defaulting in the future, in recent years economic and financial factors at the macro level are capable to explain a large portion of the correlation of US firms defaults over time.

Keywords: corporate defaults, count data, exogeneous covariates, Poisson autoregression, estimation.

JEL codes: C13, C22, C25, G33.

1 Introduction

There is a strong ongoing interest in modelling and forecasting time series of corporate defaults. A stylized fact of defaults is that they tend to cluster over time. The default clustering phenomenon has been explored in the financial literature, giving rise to a debate about its causes, with several works trying to distinguish between “contagion” effects and comovements in corporate solvency due to common macroeconomic and financial factors; see, for example, Das *et al.* (2007) and Lando and Nielsen (2010) who investigate the role of systematic risk in default correlation by using the monthly count of industrial firms’ bankruptcies.

We here propose a novel class of dynamic Poisson models for describing and forecasting corporate defaults, which we call Poisson AutoRegressions with eXogeneous covariates (PARX). PARX models are an extension of the Poisson autoregression in Fokianos, Rahbek and Tjøstheim (2009), which is here augmented by including – in addition to lagged intensity and counts – a set of exogenous covariates as predictors. They provide a flexible framework within which we are able to analyze dependence of default intensity on past default intensities as well as on other relevant financial and economic variables. These additional predictors are meant to summarize the level of uncertainty during periods of financial turmoil and/or economic downturns; that is, when corporate defaults are more likely to cluster together. We also consider the impact of auxiliary information on the estimates of persistence parameters which express the degree of dependence on the past history of the process.

Our modelling approach radically differs from those appearing in the available literature on corporate defaults dynamics. Most existing papers (see, e.g., Das et al., 2007; Lando and Nielsen, 2010) model corporate defaults in terms of a bivariate continuous-time model for the timing of default together with the firm’s debt outstanding at default. The timings are modelled using a proportional hazards model where a latent frailty risk factor is included to capture the over-all financial risk together with observed economic factors. These models tend to be difficult to implement due to the dynamic latent variables and require careful handling of macroeconomic factors since these are not observed at the precise times of defaults. An exception is Koopman, Lucas and Schwaab (2012) who model defaults using a binomial specification where, similar to the PARX model, the probability of default is a time-varying functions of underlying factors. Similar to the frailty models, their specification involve unobserved components which have to be integrated out in the estimation which is done using computationally burdensome Monte Carlo methods. In contrast, PARX models defaults through a dynamic Poisson process, can easily handle exogeneous covariates, and estimation and forecasting is straightforward to implement in standard software packages.

PARX provides new insights into the dynamics of corporate defaults among Moody’s rated US firms during the period 1982-2011. Various macroeconomic and financial variables,

meant to capture the state of the US economy and financial markets, are included to investigate whether corporate defaults are driven by economic fundamentals and/or contagion effects during this period. We find that important explanators of corporate defaults are the over-all volatility of the US stock market and the Leading Index of the US economy, but that there are also contagion effects present in the dynamics. A structural break analysis shows that these relationships are not stable over time though and the relative importance of the different factors have been changing over the sample period. Interestingly, we find that the contagion effects have been diminishing over time and that corporate defaults during the recent financial crisis were mostly driven by macroeconomic and financial fundamentals.

This paper also contributes to the literature on econometric and statistical analysis of Poisson autoregressions. First, we provide new results on the time series properties of PARX models, including conditions for stationarity and existence of moments. Second, we provide an asymptotic theory for the maximum likelihood estimators of the parameters entering the model. These results extend and complement the ones found in, among others, Rydberg and Shephard (2000), Streett (2000), Ferland et al. (2006) and Fokianos, Rahbek and Tjøstheim (2009). As an important tool in the econometric analysis is the concept of τ -weak dependence; this is a relatively new stability concept which proves to be simpler to verify for discrete-valued Markov chains compared to existing stability concepts such as geometric ergodicity.

PARX models are also related to a recent literature on GARCH models augmented by additional co-variates with the aim of improving the forecast performance. These models include GARCH-X models, the so-called HEAVY model as proposed by Shephard and Sheppard (2010), and the Realized GARCH model of Hansen et al (2012); see also Han and Kristensen (2014) for econometric analysis of such models. In these models, the time-varying volatility is explained by past returns, volatilities together with additional co-variates, usually a realized volatility measure. PARX share the same motivation and modelling approach, except that the variable of interest in our case is discrete and so the technical analysis and the applications are different.

The paper is organized as follows. In Section 2 we introduce the class of PARX models and discuss them in relation to existing models. Time series properties of the models are investigated in Section 3. Maximum-likelihood based inference is formally analyzed in Section 4. Specifically, our reference maximum likelihood estimator is discussed in Section 4.1, while its finite sample properties are studied in Section 4.2 by Monte Carlo simulations. Moreover, Section 4.3 illustrates how the estimated PARX specification can be used for forecasting purposes. Section 5 contains the empirical analysis of US default counts. Section 6 concludes. All auxiliary lemmas and mathematical proofs are contained in the Appendix.

2 Modelling Defaults with PARX

We here set up a general dynamic model for time series count data, motivated by the empirical application where we analyze the dynamics of US corporate defaults. Let $y_t \in \{0, 1, 2, \dots\}$, $t \geq 1$, be a time series of counts, such as the number of defaults in a given period, say, a month; we then wish to model the dynamics of this process both in terms of its own past, y_{t-m} , $m \geq 1$, but also in terms of additional covariates $x_t := (x_{1t}, x_{2t}, \dots, x_{d_x t})' \in \mathbb{R}^{d_x}$. In the empirical application these include relevant macroeconomic and financial factors such as realized volatility measures, recession indicators, and measures of economic activity and financial stability. We do so by modelling y_t as following a conditional Poisson distribution with time-varying intensity, λ_t , expressed as a function of past counts and factors, that is,

$$y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad (1)$$

where $\mathcal{F}_{t-1} := \mathcal{F} \{y_{t-m}, x_{t-m} : m \geq 1\}$ and $\text{Poisson}(\lambda)$ denotes a Poisson random variable with intensity parameter λ .

The time-varying intensity λ_t is specified as a linear function of past counts and past intensities as considered in Fokianos, Rahbek and Tjøstheim (2009) which is here augmented by the exogenous variables contained in x_t . Specifically, x_t enters the intensity through a function $f : \mathbb{R}^{d_x} \rightarrow [0, \infty)$,

$$\lambda_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^q \beta_i \lambda_{t-i} + f(x_{t-1}). \quad (2)$$

The parameters of interest are given by $\omega > 0$, α_i ($i = 1, 2, \dots, p$) and $\beta_i \geq 0$ ($i = 1, 2, \dots, q$), together with the additional parameters entering the function f . A possible specification of the function f , which will be extensively used in the empirical analysis of Section 5, is the following additive specification,

$$f(x) := \sum_{i=1}^{d_x} \gamma_i f_i(x_i), \quad (3)$$

where $f_i : \mathbb{R} \mapsto [0, \infty)$, $i = 1, \dots, d_x$, are known functions, while $\gamma := (\gamma_1, \dots, \gamma_{d_x})' \in [0, \infty)^{d_x}$ is a vector of unknown parameters. Observe that with $\gamma = 0$, the model reduces to the Poisson autoregression (PAR) considered in Fokianos, Rahbek and Tjøstheim (2009). However, in general, the inclusion of additional covariates x_t will improve on in- and out-of-sample performance of the model, as we demonstrate in the empirical analysis, and provide further insights into how economic factors influence defaults.

The above specification allows for flexible dynamics of the number of defaults in terms of past defaults, captured by $\sum_{i=1}^p \alpha_i y_{t-i}$, and exogenous factors, as described by $f(x_{t-1})$. The

component $\sum_{i=1}^q \beta_i \lambda_{t-i}$ is a parsimonious way of incorporating "long memory" of these two components in a fashion similar to GARCH models. For example, with $p = 1$ and $q = 1$,

$$\lambda_t = \frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} y_{t-i} + \sum_{i=1}^{\infty} \beta^{i-1} f(x_{t-i}),$$

where we have assumed that $\alpha + \beta = \alpha_1 + \beta_1 < 1$. Thus, when $\beta > 0$, the model allows for all past defaults to effect the current number of defaults, and for long-run effects of exogenous covariates. More generally, one can interpret the value of $\alpha_1 + \dots + \alpha_p$ as a measure of dynamic contagion effects since a large value of $\alpha_1 + \dots + \alpha_p$ implies that past defaults have a large impact on current default probabilities. In the extreme case, when $\alpha_1 + \dots + \alpha_p = 0$, the model implies conditional (on x_{t-1}) independence between current and past defaults; see Lando and Nielsen (2010) for a further discussion.

The PARX model for default counts has strong similarities with the GARCH class of processes with exogenous covariates, or GARCH-X; see Han and Kristensen (2014) and references therein. Specifically, GARCH-X specifications model the conditional volatility, say h_t , of a given return r_t as

$$h_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i} + f(x_{t-1}),$$

where x_t is a set of covariates. It is also worth noticing the connection to the so-called HEAVY model of Shephard and Sheppard (2010) as well as the realized GARCH model of Hansen et al (2012). These two models consider GARCH-type specifications where, in its simplest form, the volatility process h_t satisfies

$$h_t = \omega + \alpha x_{t-1} + \beta h_{t-1}; \tag{4}$$

here the exogenous variable x_{t-1} is usually a (realized) measure of past volatility obtained from high-frequency data. An important difference with respect to our PARX models (as well as to GARCH-X models) is that in (4) the exogenous variable does not enter as an additional regressor but instead it replaces the past squared return y_{t-1}^2 . However, the PARX specification in (2) reduces to an HEAVY-type specification when the parameters linking past counts to current intensity are set to zero; i.e., $\alpha_i = 0$, $i = 1, 2, \dots, p$. In this case, the PARX model has the simpler form

$$\lambda_t = \omega + f(x_{t-1}) + \sum_{i=1}^q \beta_i \lambda_{t-i}, \tag{5}$$

corresponding to the case of conditional independence discussed earlier. While parts of the structure of GARCH-X type models are similar to that of the PARX model, a crucial difference is that while the former class of models are designed to capture the evolution of the (conditional) variance of a continuously distributed variable, the latter are modelling the full distribution of a count process.

3 Properties of PARX processes

In this section we provide sufficient conditions for a PARX process to be stationary and ergodic with polynomial moments of a given order. This result will be used in the estimation theory; in particular, we use these time series properties to show that estimators of model parameters are normally distributed in large samples. This result in turn allows us to use standard tools for inference.

The analysis is carried out by applying results on so-called τ -weak dependence, henceforth weak dependence, recently developed in Doukhan and Wintenberger (2008). Weak dependence is a stability concept for Markov chains that implies stationarity and ergodicity and so establishes, amongst other things, a (uniform) law of large numbers [LLN] for the process. It is related to alternative concepts of stability and mixing of time series such as (geometric) ergodicity (see, for example, Fokianos, Rahbek and Tjøstheim, 2009) but is simpler to verify for discrete-valued data. Christou and Fokianos (2013) employed the same techniques in the analysis of a class of negative binomial time series models.

Weak dependence basically requires that the time series is a stochastic contraction. To establish this property for the PARX model, we first rewrite the Poisson model (1) in terms of an i.i.d. sequence $N_t(\cdot)$ of Poisson processes with unit-intensity,

$$y_t = N_t(\lambda_t). \quad (6)$$

Next, we complete the model by imposing a Markov-structure on the set of covariates; that is,

$$x_t = g(x_{t-1}, \varepsilon_t), \quad (7)$$

for some function $g(x, \varepsilon)$ and with ε_t being an i.i.d. error term. The above structure could be generalized to $x_t = g(x_{t-1}, \dots, x_{t-m}, \varepsilon_t)$ for some $m \geq 1$, thereby allowing for more flexible dynamics of the covariates included in the model. However, we maintain eq. (7) for simplicity in the following.

We then impose the following assumptions on the complete model:

Assumption 1 (Markov) The innovations ε_t and $N_t(\cdot)$ are i.i.d.

Assumption 2 (Exogenous stability) $E[\|g(x; \varepsilon_t) - g(\tilde{x}; \varepsilon_t)\|^s] \leq \rho \|x - \tilde{x}\|^s$, for some $\rho < 1$, and $E[\|g(0; \varepsilon_t)\|^s] < \infty$, for some $s \geq 1$.

Assumption 3 (PARX stability) (i) $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$ and (ii) $|f(x) - f(\tilde{x})| \leq L \|x - \tilde{x}\|$, for some $L > 0$.

Assumption 1 implies that (y_t, x_t) can be embedded in a Markov chain and so we can employ the theory of weak dependence. Assumption 2 imposes a stochastic contraction

condition on $g(x, \varepsilon)$ w.r.t. x which is satisfied for many popular time series models such as (stable) linear autoregressive ones. This assumption is used to show, as a first step, that x_t is weakly dependent. Finally, Assumption 3(i) implies that the function $L(y, \lambda) = \omega + \sum_{i=1}^p \alpha_i y_i + \sum_{i=1}^q \beta_i \lambda_i$, where $y = (y_1, \dots, y_p)$ and $\lambda = (\lambda_1, \dots, \lambda_q)$, is a contraction mapping with contraction coefficient $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)$. It is identical to the conditions imposed in Fokianos, Rahbek and Tjøstheim (2009) for the Poisson autoregressive model (without exogenous regressors and with $p = q = 1$) to be stationary. Assumption 3(ii) restricts how x_t can enter the Poisson intensity; it excludes certain functions, such as the exponential one. This assumption will, however, be weakened at the end of this section.

Together the three assumptions imply that the PARX model admits a stationary and weakly dependent solution, as shown in the following theorem.

Theorem 1 *Under Assumptions 1–3, there exists a weakly dependent stationary and ergodic solution, which we denote $X_t^* = (y_t^*, \lambda_t^*, x_t^*)'$, to eqs. (1)-(2) and (7) with $E[\|X_t^*\|^s] < \infty$ and $s \geq 1$ given in Assumption 2.*

The above theorem complements the results of Fokianos, Rahbek and Tjøstheim (2009), who derive sufficient conditions for an approximate Poisson Autoregression to be geometrically ergodic. We here allow for exogeneous variables to enter the model, and provide sufficient conditions for weak dependence directly for this extended model.

One particular consequence of the above theorem is that the expected long-run number of defaults equals

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

and furthermore, that $\text{Var}[y_t] > E[y_t]$. Thus, by including past values of the response as well as covariates in the evolution of intensity, PARX models generate overdispersion in the marginal distribution, a feature that is prominent in many count time series, including corporate defaults.

One further consequence of Theorem 1 is that it gives us access to the strong Law of Large Numbers (LLN) for stationary and ergodic processes, $T^{-1} \sum_{t=1}^T h(X_t^*) \xrightarrow{P} E[h(X_t^*)]$ for *any* function $h(\cdot)$ of $X_t = (y_t, \lambda_t, x_t)'$ provided $E[\|h(X_t^*)\|] < \infty$. In the asymptotic theory of the proposed estimators, the likelihood function is computed based on a set of fixed initial values for the Poisson intensity. In order to analyze the asymptotic behaviour of the likelihood function in this setting, we need to generalize the LLN result to hold for any solution with arbitrary initialization. This extension is stated in the following lemma:

Lemma 1 *Let $\{X_t\}$ be a process satisfying $X_t = F(X_{t-1}; \xi_t)$ with ξ_t i.i.d., $E[\|F(x; \xi_t) - F(\tilde{x}; \xi_t)\|^s] \leq \rho \|x - \tilde{x}\|$ and $E[\|F(0; \xi_t)\|^s] < \infty$ for some $s \geq 1$. For any function $h(x)$ satisfying (i) $\|h(x)\|^{1+\delta} \leq C(1 + \|x\|^s)$ for some $C, \delta > 0$ and (ii) for some $c > 0$, there exists $L_c > 0$*

so that $\|h(x) - h(\tilde{x})\| \leq L_c \|x - \tilde{x}\|$ for $\|x - \tilde{x}\| \leq c$, it then holds that $T^{-1} \sum_{t=1}^T h(X_t) \xrightarrow{P} E[h(X_t^*)]$.

Remark 1 Suppose that the assumptions of Lemma 1 are satisfied for some $s \geq 1$. Then for any sequence $\{u_t\}$ satisfying $E[u_t | \mathcal{F}_{t-1}] = 0$ and $E[u_t u_t' | \mathcal{F}_{t-1}] = \Sigma(X_{t-1})$ w.r.t. some filtration \mathcal{F}_t , where the conditional variance $\Sigma(x)$ satisfies $\|\Sigma(x)\| \leq C(1 + \|x\|^s)$, it holds that:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T u_t \xrightarrow{d} N(0, E[\Sigma(X_t^*)]). \quad (8)$$

This result follows readily from standard CLT for stationary martingale differences (see e.g. Brown, 1971). This CLT proves to be important for the asymptotic analysis of the maximum likelihood estimator provided in the next section. \square

We end this section by noticing that the Lipschitz condition in Assumption 1 rules out some unbounded transformations $f(x_t)$ of x_t , such as the specification in (3) with $f_i(x_i) = \exp(x_i)$ for some $1 \leq i \leq d_x$. Such situations can be handled by introducing a truncated model,

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

for some cut-off point $c > 0$, and with y_t^c the corresponding Poisson process. We can then relax $f(x)$ to be locally Lipschitz in the following sense:

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

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

By replacing Assumption 1 with Assumption 1' we now obtain, by identical arguments as in the proof of Theorem 1, that the truncated process has a weakly dependent stationary and ergodic solution. While this approach is similar to the approximation of Poisson AR process as used in Fokianos, Rahbek and Tjøstheim (2009), the reasoning here is different. In Fokianos, Rahbek and Tjøstheim (2009), an approximating process was needed in order to establish geometric ergodicity of the Poisson GARCH process, while here we introduce the truncated process in order to handle the often applied practice of introducing non-bounded or exponential transformations of the regressors in the model.

In the next Lemma we formally prove that, as $c \rightarrow \infty$, the truncated process approximates the untruncated one ($c = +\infty$).

Lemma 2 Under Assumptions 1'-3 together with $E[f(x_t^*)] < \infty$,

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

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

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

4 Estimation and Forecasting

In this section, we describe how the PARX model can be estimated and the estimated model used for forecasting. We provide an asymptotic theory for the estimated parameters allowing for statistical inference, and present the results of a simulation study investigating the finite-sample properties of the estimator.

4.1 Estimation

We consider the model for y_t as specified in (1)-(2) and (7), that is with conditional intensity given by

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

where $\theta = (\omega, \alpha, \beta, \gamma) \in \Theta \subset (0, \infty) \times [0, \infty)^{p+q+d_x}$, where $\alpha = (\alpha_1, \dots, \alpha_p)'$, $\beta = (\beta_1, \dots, \beta_q)'$, and $\gamma = (\gamma_1, \dots, \gamma_{d_x})'$. We let $\theta_0 = (\omega_0, \alpha_0, \beta_0, \gamma_0)$, where $\alpha_0 = (\alpha_{0,1}, \dots, \alpha_{0,p})'$, $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,q})'$, and $\gamma_0 = (\gamma_{0,1}, \dots, \gamma_{0,d_x})'$, denote the true, data-generating parameter value.

The conditional log-likelihood function of θ in terms of the observations $(y_1, x_0), \dots, (y_T, x_{T-1})$, given some initial values $\lambda_0, \lambda_{-1}, \dots, \lambda_{1-q}, y_0, \dots, y_{1-p}$ and x_0 , takes the form

$$L_T(\theta) = \sum_{t=1}^T l_t(\theta), \quad l_t(\theta) := y_t \log \lambda_t(\theta) - \lambda_t(\theta) \quad (9)$$

where we have left out any constant terms. The maximum likelihood estimator (MLE) is then computed as

$$\hat{\theta} := \arg \max_{\theta \in \Theta} L_T(\theta). \quad (10)$$

In order to analyze the large sample properties of $\hat{\theta}$, we impose the following conditions on the parameters and the exogenous regressors:

Assumption 4 Θ is compact and for all $\theta = (\omega, \alpha, \beta, \gamma) \in \Theta$, $\beta_i \leq \beta^U < 1/q$ for $i = 1, 2, \dots, q$ and $\omega \geq \omega_L > 0$, with ω_L and β^U being fixed.

Assumption 5 The polynomials $A(z) = \sum_{i=1}^p \alpha_{0,i} z^i$ and $B(z) = \sum_{i=1}^q \beta_{0,i} z^i$ have no common roots; for any $(a, b) \neq (0, 0)$, $\sum_{i=1}^{\max\{p,q\}} a_i y_{t-i} + \sum_{i=1}^{d_x} b_i f(x_{i,t})$ has a nondegenerate distribution.

Assumption 4 imposes weak restrictions on the parameter space; these are similar to the ones imposed in the analysis of estimators of GARCH models and rule out β 's greater than one (for which $\lambda_t(\theta)$ is explosive) and ω 's equal to zero. The latter is used to ensure that $\lambda_t(\theta)$ is bounded away from zero.

Assumption 5 is an identification condition which is similar to the one found for GARCH models with exogenous regressors: The first part is the standard condition found for GARCH models (see, e.g., Berkes et al, 2003), while the second part rules out that the exogeneous covariates are colinear with each other and the observed count process (see Han and Kristensen, 2014 for a similar condition).

Under this assumption, together with those used earlier to establish stationarity and existence of moments, we obtain the following asymptotic result for the MLE conditional on the initial values:

Theorem 2 *Under Assumptions 1–4 with $s \geq 1$, $\hat{\theta}$ is consistent. Furthermore, if $\theta_0 \in \text{int } \Theta$ and $s \geq 2$,*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1}(\theta_0)), \quad H(\theta) := -E \left[\frac{\partial^2 l_t^*(\theta)}{\partial \theta \partial \theta'} \right],$$

where $l_t^*(\theta)$ denotes the likelihood function evaluated at the stationary solution.

Remark 2 If the model is mis-specified, we expect the asymptotic properties of the MLE to remain correct except that θ_0 is now the pseudo-true value maximizing the pseudo-likelihood and the asymptotic variance takes the well-known sandwich form $H^{-1}(\theta_0) \Omega(\theta_0) H^{-1}(\theta_0)$, where

$$\Omega(\theta) = E \left[\frac{\partial l_t^*(\theta)}{\partial \theta} \frac{\partial l_t^*(\theta)}{\partial \theta'} \right];$$

see White (1982) and Gouriéroux, Monfort and Trognon (1984). □

Remark 3 The assumption $\theta_0 \in \text{int } \Theta$ rules out cases where some of the parameters are zero. We detail how this assumption can be relaxed at the end of this section. The requirement on s , as defined in Assumption 2, is used to ensure that the likelihood function has a well-defined limit and that the moments in the information matrix $H(\theta)$ exist. □

The above theorem generalizes the result of Fokianos, Rahbek and Tjøstheim (2009) to allow for estimation of parameters associated with additional regressors in the specification of λ_t . It is established under the assumption that f is globally Lipschitz as stated in Assumption 1. By combining the arguments in Fokianos, Rahbek and Tjøstheim (2009) with Lemma 2, the asymptotic result can be extended to allow f to be locally Lipschitz, see Assumption 1'. More precisely, define the likelihood quantities for the approximating, or truncated, model as

$$L_T^c(\theta) = \sum_{t=1}^T l_t^c(\theta), \quad \text{where } l_t^c(\theta) = y_t^c \log \lambda_t^c(\theta) - \lambda_t^c(\theta).$$

It immediately follows that the results of Theorem 2 holds for the QMLE of $L_T^c(\theta)$, $\hat{\theta}^c$ say. However, as the approximating likelihood function can be made arbitrarily close to the true likelihood as $c \rightarrow \infty$, one can show that we can replace Assumption 1 in Theorem 2 by Assumption 1'.

Theorem 3 *Under Assumptions 1', 2-5, and $E[f_i(x_{it}^*)] < \infty$, $i = 1, \dots, d_x$, the conclusions of Theorem 2 remain valid.*

It will often be of interest to investigate if some of the elements of θ are zero, as for example $\gamma_i = 0$ or $\alpha_i = 0$. In order to allow for this, where under the null the parameter vector θ is on the boundary of the parameter space Θ , we complement the results of Theorems 2-3. To do so, we apply the general theory of Andrews (1999), see also Demos and Sentana (1998) and Francq and Zakoian (2009) to obtain the following corollary where we state this explicitly for the case of testing one parameter equal to zero (more general cases of multiple parameters on the boundary can be handled as in Francq and Zakoian, 2009). Here, we denote the standard t statistic for the null hypothesis $H_0 : \theta_{i0} = 0$ with $t_i = \sqrt{T}\hat{\theta}_i/\hat{\sigma}_{ii}$, where $\hat{\sigma}_{ii}^2$ is a consistent estimator of the i -th diagonal element of H^{-1} as defined in Theorem 2. The likelihood ratio test for the same null hypothesis is denoted by LR_i .

Corollary 1 *Under Assumptions 1-5 and H_0 with $\theta_{j0} \neq 0$ for all $j \neq i$,*

$$t_i \xrightarrow{d} \max\{0, Z\}, \tag{11}$$

$$LR_i \xrightarrow{d} (\max\{0, Z\})^2, \tag{12}$$

where Z is standard normally distributed.

Remark 4 For a given significance level $\eta < 1/2$, the $(1 - \eta)$ quantile of the asymptotic distribution in (11) equals the $(1 - \eta)$ quantile of the standard normal distribution, see e.g. Andrews (2000, p.404). Hence, in this case standard normal critical values apply to the t -statistic. The same does not hold for the LR statistic, as for any $\eta \in (0, 1)$, the $(1 - \eta)$

quantile of the asymptotic distribution in (12) equals the $(1 - 2\eta)$ quantile of the $\chi^2(1)$ distribution, see e.g. Francq and Zakoian (2009). \square

4.2 Finite Sample Performance

In this section we present results from a small simulation study aimed at evaluating the finite-sample performance of the MLE presented in the previous section. We consider the PARX(1,1) model (1) with conditional intensity given by

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

The use of an exponential link function is motivated by the empirical application where this is employed for some of the covariates. We examine the performance under two different data generating processes (DGP's) for the covariate x_t :

DGP 1 x_t is a stationary autoregressive process, $x_t = \varphi x_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \text{i.i.d.} N(0, 1)$, initialized at $x_0 \sim \text{i.i.d.} N(0, 1/(1 - \varphi^2))$; the AR parameter is set to $\varphi = 1/2$.

DGP 2 x_t is a stationary fractionally integrated process, $\Delta_+^d x_t = \varepsilon_t$, where the operator Δ_+^d is given by $\Delta_+^d z_t := \Delta^d z_t \mathbb{I}(t \geq 1) = \sum_{i=0}^{t-1} \pi_i(-d) z_{t-i}$ with $\pi_i(v) = (i!)^{-1}(v(v+1)\dots(v+i-1))$ denoting the coefficients in the usual binomial expansion of $(1-z)^{-v}$; ε_t is i.i.d. $N(0, 1)$ and $d = 1/4$.

These two DGP's represent typical time series behavior found in the factors used in the empirical application. The first DGP satisfies the theoretical conditions used in the asymptotic analysis of the MLE, while the second one does not since it is not a Markov chain. However, DGP 2 remains stationary and so we expect that the theory remains valid for this DGP as well.

Since the distribution of y_t is not invariant to the scale of the covariate x_t , in each case x_t has been re-scaled by its unconditional variance. We report results for $\omega = 0.10$, $\alpha = 0.30$, $\gamma = 0.5$ and four alternative scenarios for β : $\beta = 0$ (no feedback from lagged intensity to current intensity), $\beta = 0.20$ (low persistence) and $\beta = 0.70$ (high persistence). In all cases considered, the model admits a stationary solution, see section 3. Finally, we consider samples of size $T \in \{100, 250, 500, 1000\}$. For each experiment, the number of Monte Carlo replications is set to $N = 1000$.

Results for the case of DGP 1 are presented in Table 1. For each parameter, the mean and root mean square error (RMSE) (over the $N = 1000$ Monte Carlo replications) of the corresponding estimator are reported. Furthermore, the p -value obtained from a Kolmogorov-Smirnov (KS) test for the hypothesis of $N(0, 1)$ distribution of each parameter estimator is reported.

The performance of the MLE for DGP 1 seems largely satisfactory for moderate and large sample sizes. For samples of $T \geq 250$ observations and for all scenarios considered, the hypothesis of $N(0, 1)$ distribution of $\hat{\theta}_i$ is never rejected at any conventional significance level. For samples of $T = 100$, the degree of persistence of the process (here captured by the β coefficient) seems to affect the distribution of the estimators. Specifically, while in the case of lowest persistence ($\beta = 0$) the hypothesis of $N(0, 1)$ distribution of $\hat{\theta}_i$ is never rejected, in the cases of stronger persistence ($\beta = 0.2$ and $\beta = 0.7$) normality is rejected for the estimator constant term ω (both when $\beta = 0.2$ and $\beta = 0.7$) and for the PAR parameters α and β (when $\beta = 0.7$). These deviations from normality, however, do not persist for larger sample sizes. Finally, it is worth noticing that the parameter which delivers the highest RMSE is the constant term, ω .

Next, consider the results for DGP 2 as presented in Table 2: Compared to DGP 1, x_t now has higher persistence. Despite this, for $T \geq 250$, with the only exception of the constant term ω , results do not show substantial differences relative to the ones for DGP 1; that is, the asymptotic $N(0, 1)$ approximation is largely satisfactory. In the case of high persistence ($\beta = 0.7$), normality of $\hat{\omega}$ is rejected at the 1% significance level even when $T = 1000$. This is consistent with the findings of Han and Kristensen (2014) for the GARCH-X model who also find that the intercept is less precisely estimated in the presence of persistent regressors.

[Table 1 and Table 2 about there]

4.3 Forecasting

Once the PARX model has been estimated, it can be used to forecast future number of defaults, y_t . Forecasting of Poisson autoregressive processes is similar to forecasting of GARCH processes (see, *e.g.*, Hansen et al, 2012, Sec. 6.2) in that it proceeds in two steps: First, a forecast of the time-varying parameter (variance in the case of GARCH, intensity in the case of PARX) is obtained. This is then substituted into the conditional distribution of the observed process y_t . Consider first the forecasting of λ_t . A natural one-step ahead forecast, given available information at time T and parameters θ , is

$$\lambda_{T+1|T}(\theta) = \omega + \sum_{i=1}^p \alpha_i y_{T+1-i} + \sum_{i=1}^q \beta_i \lambda_{T+1-i}(\theta) + \gamma f(x_T). \quad (13)$$

More generally, a multi-step ahead forecast of λ_{T+h} , for some $h \geq 1$, solves the following recursive scheme,

$$\lambda_{T+k|T}(\theta) = \omega + \sum_{i=1}^{\max\{p,q\}} \{\alpha_i + \beta_i\} \lambda_{T+k-i|T}(\theta) + \gamma f(x_{T+k-1|T}), \quad k = 1, \dots, h, \quad (14)$$

with initial value $\lambda_{T+1|T}(\theta)$ coming from eq. (13). Here, $x_{T+k|T}$, $k = 1, \dots, h-1$, is a forecast of x_{T+h} given available information at time T . This is similar to GARCH-X, HEAVY and realized GARCH models, that also take as input a forecasting model for exogenous covariates.

Once we have computed a forecast of the underlying intensity, $\lambda_{T+h|T}(\theta)$, this can in turn be used to generate a forecast distribution of y_{T+h} ,

$$P(y_{T+h} = y | \mathcal{F}_T) = \text{Poisson}(y | \lambda_{T+1|T}(\theta)), \quad y \in \{0, 1, 2, \dots\},$$

where $\text{Poisson}(y | \lambda) = \lambda^y \exp(-\lambda) / y!$ is the probability function of a Poisson distribution with intensity λ . This is related to the well-known concept of density forecasts (see Tay and Wallis, 2000, for a review) except that we are here working with a discrete-valued distribution. A simple way of representing the forecast distribution is by reporting a measure of central tendency, such as

$$y_{T+h|T} := E[y_{T+h} | \mathcal{F}_t] = \lambda_{T+h|T}(\theta),$$

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

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

where $p \mapsto Q(p | \lambda)$ denotes the quantile function of a Poisson distribution with intensity λ .

5 Empirical Analysis

The aim of this section is to provide an empirical analysis of US corporate default counts using PARX models. Specifically, by including exogenous regressors in the intensity specification and by testing whether they cause a significant decrease in the impact of past default counts, we are able to investigate to what extent autocorrelation (as well as clustering over time) in default counts depends on common (aggregate) risk factors. That is, testing the existence of autocorrelation in default counts after correcting for common risk factors can be viewed as testing the existence of *contagion* effects over time. The alternative hypothesis – *i.e.*, default counts, after correcting for the common risk factors, do not depend on the number of past defaults – is similar to the idea of conditional independence of default counts as discussed in the recent literature by Das *et al.* (2007), Lando and Nielsen (2010) and Lando *et al.* (2013), among others. With respect to this literature, however, by means of PARX models we are able to model explicitly the conditional dependence through the specification of the time-dependent intensity process.

The data set on defaults consist of monthly number of bankruptcies among Moody's rated industrial firms in the United States in the 1982–2011 period ($T = 360$ observations),

collected from Moody’s Credit Risk Calculator (CRC). Figure 1(a, b), which shows default counts and the corresponding autocorrelation function, reveals three important stylized facts of defaults: (i) the high temporal dependence in default counts; (ii) the existence of default clusters over time; (iii) overdispersion of the distribution of default counts (the average is 3.51 while the variance is 15.57). It will be shown later in this section that all these empirical properties are well explained using PARX specifications.

[Figure 1(a, b) about here]

The choice of covariates to be included in our PARX models is important, as they represent the common risk factors conditional on which defaults could be independent over time. Following Lando and Nielsen (2010) we consider the following financial, credit market and macroeconomic variables: Baa Moody’s rated to 10-year Treasury spread (SP), the number of Moody’s rating downgrades (DG), year-to-year change in Industrial Production Index (IP), Leading Index released by the Federal Reserve (LI), the recession indicator released by the National Bureau of Economic Research¹ (NB).² Moreover, in order to shed some light on the possible impact of uncertainty in the financial markets on the number of future defaults, we also consider realized volatility (RV) on the S&P 500. RV is computed as a proxy of the S&P 500 monthly realized volatility using daily squared returns (that is, $RV_t := \sum_{i=1}^{n_t} r_{i,t}^2$ with $r_{i,t}$ denoting the i -th daily return on the S&P 500 index in month t and n_t being the number of trading days in month t).

Since Industrial Production and Leading Index take on both negative and positive values, we decompose them into their negative and positive parts and let $IP^{(+)} := \mathbb{I}_{\{IP \geq 0\}} |IP|$, $IP^{(-)} := \mathbb{I}_{\{IP < 0\}} |IP|$ and similarly for LI . This gives us a total of eight candidate covariates.

5.1 Full-sample Analysis

We here provide an analysis for the full sample 1982-2011. Preliminary covariate and lag selection using all eight covariates suggests the following specification of default intensity,

$$\lambda_t = \omega + \sum_{i=1}^2 \alpha_i y_{t-i} + \beta \lambda_{t-1} + \gamma_1 RV_{t-1} + \gamma_2 SP_{t-1} + \gamma_3 DG_{t-1} + \gamma_4 NB_{t-1} + \gamma_5 IP_{t-1}^{(+)} + \gamma_6 LI_{t-1}^{(-)}, \quad (15)$$

¹This time series is released by the Federal Reserve Bank of St. Louis interpreting the Business Cycle Expansions and Contractions data provided by The National Bureau of Economic Research (NBER) at <http://www.nber.org/cycles/cyclesmain.html>. A value of 1 indicates a recessionary period, while a value of 0 denotes an expansionary period.

²Data are obtained from the FRED website, provided by the Federal Reserve Bank of St. Louis, <http://research.stlouisfed.org/>, except for the number of Moody’s rating downgrades, which we collect from Moody’s CRC.

which is a special case of model (1)-(2) where we set $p = 2$, $q = 1$ and $\beta = \beta_1$.

[Table 3 about there]

Table 3 shows the estimation results for the full PARX(2,1) model in (15), along with the PAR(2,1) model (*i.e.*, the model without covariates) and nested specifications based on subsets of the six included covariates. For each specification, we report parameter estimates together with corresponding t statistics, standard (AIC and BIC) information criteria and the LR statistic relative to the maintained PARX model. Among the various models considered, the preferred PARX model, in terms of the information criteria as well as by the LR tests, is the one only including realized volatility and the leading index.

To our knowledge, the link between realized volatility (reflecting uncertainty in financial markets) and defaults of industrial firms has not been documented earlier in the literature. Similarly, significance of the Leading Index highlights a clear link between macroeconomic factors and corporate defaults, which is not generally found using standard econometric techniques. For instance, recent empirical results of Duffie *et al.* (2009) and Giesecke *et al.* (2011) do not show a significant role of production growth while Lando *et al.* (2013) find that, conditional on individual firm risk factors, no macroeconomic covariate seems to explain significantly default intensity. However, once we control for the information contained in realized volatility and the negative component of the Leading Index, none of the other four covariates (NBER recession indicator, interest rate spread, and number of downgrades) are found to be relevant in predicting future defaults.

We analyze the existence of contagion effects by investigating whether by including covariates, past default counts have a smaller impact, *i.e.* a significant decrease in $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in a given model with covariates (PARX) relative to the corresponding one without covariates (PAR). As remarked previously, conditional independence over time would require that α_1 and α_2 are both zero,³ which would imply that conditional intensity can fully be explained by past covariates only. Indeed, the inclusion of covariates leads to a decrease in $\hat{\alpha}_1 + \hat{\alpha}_2$ for almost all the models considered. On the other hand, the null hypothesis $H_0 : \alpha_1 + \alpha_2 = 0$ is rejected for all specifications. Therefore, although part of the dependence over time in default counts can be explained by the set of covariates considered, even after correcting for such covariates a strong link between conditional intensity and past default counts remains. This result seems to indicate that – at least on the basis of the exogenous regressors considered – significant evidence of contagion effects over time is likely to occur.

³It is worth noticing that this approach is related to empirical studies aiming at measuring the impact of covariates, such as the trading volumes, on future volatility using GARCH models (see, for instance, Lamoureux and Lastrapes, 1990 and Gallo and Pacini, 2000).

We run a number of model (mis)specification tests on the selected model. First, to check in-sample fit, we plot in Figure 2 the actual default counts (y_t) together with the predicted defaults ($\hat{y}_t := \hat{\lambda}_t = \lambda_t(\hat{\theta})$) and its confidence bands (see Section 4.3). As can be seen from this figure, the model captures the default counts dynamics well. The associated generalized, or Pearson, residuals (see Gouriéroux *et al.*, 1987; Kedem and Fokianos, 2002) – formally defined as $\hat{e}_t = \hat{\lambda}_t^{-1/2}(y_t - \hat{\lambda}_t)$ ($t = 1, \dots, T$) and reported in Figure 3 – also appear to be uncorrelated over time (the corresponding Ljung-Box test has p -value 0.661 when computed using \hat{e}_t and 0.373 when computed using \hat{e}_t^2).

We also evaluate the goodness of fit of the assumed Poisson conditional distribution of y_t by comparing the observed zero counts with the corresponding model-implied probabilities, $\hat{P}(y_t = 0 | \mathcal{F}_{t-1}) = e^{-\hat{\lambda}_t}$ ($t = 1, \dots, T$), *i.e.* the (conditional on the past) probability that a Poisson($\hat{\lambda}_t$) random variable equals zero under the selected model specification. Figure 4 shows the relation between the observed zeros and such model-implied probabilities. There is a clear correspondence between periods characterized by high number of zeros and the conditional probability of observing $y_t = 0$, given the specified model.

5.2 Structural Instabilities

We found in the previous subsection that the preferred model does a reasonable good job in terms of in-sample fit. To further examine the fit, we also perform a pseudo-out-of-sample forecasting exercise for the preferred model (the PARX(2, 1) with RV and $LI^{(-)}$ as included covariates) along the lines of, for example, Stock and Watson (1996): we split the sample in two with the first part of the sample of size $T_0 (= 120)$, $\{(y_t, x_{t-1}) : t = 1, \dots, T_0\}$ being used for initial estimation of the model, and the remaining observations $\{(y_t, x_{t-1}) : t = T_0 + 1, \dots, T\}$ being used for a forecasting exercise described below.

Let

$$\hat{\theta}_t = \arg \max_{\theta} L_t(\theta)$$

be the MLE using observations up to time $t \geq T_0$, where

$$L_t(\theta) = \sum_{s=1}^t l_s(\theta), \quad l_s(\theta) := y_s \log \lambda_s(\theta) - \lambda_s(\theta).$$

Given $\hat{\theta}_t$, we then compute the corresponding one-step-ahead forecast of λ_{t+1} using only information up to time t ,

$$\hat{\lambda}_{t+1|t} = \lambda_{t+1}(\hat{\theta}_t).$$

We then repeat the above exercise for $t = T_0 + 1, \dots, T$, thereby providing us with a time series of estimators, $\{\hat{\theta}_t : t = T_0, \dots, T\}$, and corresponding intensity forecasts, $\{\hat{\lambda}_{t+1|t} : t = T_0, \dots, T\}$. This procedure mimicks what a forecaster would obtain as (s)he starts forecasting

at time T_0 and updates his (her) estimates and forecasts as more data arrive. Given the forecast path $\hat{\lambda}_{t+1|t}$, we evaluate the performance of the preferred PARX specification by computing the following estimate of the mean-square forecasting error,

$$MSFE_t = \frac{1}{t - T_0} \sum_{s=T_0}^t (y_{s+1} - \hat{\lambda}_{s+1|s})^2, \quad t = T_0, \dots, T,$$

and the forecasting score (FS),

$$FS_t = \frac{1}{t - T_0} \sum_{s=T_0}^t (y_{s+1} \log \hat{\lambda}_{s+1|s} - \hat{\lambda}_{s+1|s}), \quad t = T_0, \dots, T.$$

In Figure 5, we plot $MSFE_t$ and FS_t as functions of time. The forecasting objectives vary a lot. In particular, there appears to be radical structural breaks around the outset of the financial crisis in the early 2000's and in 2008. This could seem to indicate that there are structural instabilities as caused by time-varying parameters during the sample period.

To formally test whether there indeed are structural breaks in the sample, we implement the Nyblom (1989) test (NT)

$$NT_{T,t} = S_t(\hat{\theta}_T)' H_t^{-1}(\hat{\theta}_T) S_t(\hat{\theta}_T) = (\hat{\theta}_t - \hat{\theta}_T)' H_t(\hat{\theta}_T) (\hat{\theta}_t - \hat{\theta}_T) + o_P(1/\sqrt{T}),$$

where $S_T(\theta)$ and $H_T(\theta)$ are defined in eqs. (A.2) and (A.3), respectively. We clearly reject parameter constancy using this test. This is also evident from the plots of $\hat{\theta}_t$ reported in Figure 4 where there appears to be structural breaks around the early 2000's and in 2007.

[Figure 5 and Table 4 about here]

Based on these findings, we split the full sample into three subsamples, 1982-1998, 1998-2007, and 2007-2011, and for each subsample re-do model selection and estimation. In Table 5, we report the preferred model with corresponding estimated parameters for each subsample. In the early period (1982-1998), all macro factors (incl. RV and $LI^{(-)}$) are irrelevant and there are strong contagion effects ($\hat{\alpha}_1 + \hat{\alpha}_2 = 0.65$). During the second subsample, RV is a very strong explanator of defaults while contagion effects are weak ($\hat{\alpha}_1 + \hat{\alpha}_2 = 0.09$). Finally, during the Great Recession (2007-2011), we find that RV and $LI^{(-)}$ are very strong explanatory variables and there are no contagion effects ($\hat{\alpha}_1 + \hat{\alpha}_2 = 0.00$). This last finding goes against much of the recent discussion of the financial crisis and how contagion effects and systemic risk played a big role in its evolution. One possible explanation is the relatively small sample used for the last set of estimates.

[Table 5 about there]

We also note that the estimated models for the three regimes match well with the parameter estimates we reported for the full sample, which are basically an average over the three different regimes.

6 Conclusions

In this paper, we have developed a class of Poisson autoregressive models with exogenous covariates (PARX) for time series of counts. Since PARX models allow for overdispersion arising from persistence, they are suitable to model count time series of corporate defaults, which are strongly correlated over time and exhibit high peaks, known as default clusters. Our empirical analysis, based on application of different PARX specifications (*i.e.* including different sets of covariates) to a thirty year-long time series of US default counts, reveals that our model is capable to capture the dynamic features of default counts very satisfactorily. Our PARX models also allow to test to what extent dependence over time in default counts can be explained by exogenous factors. We find that the lagged realized volatility of financial returns, together with macroeconomic variables, significantly explains the number of defaults. A full sample analysis shows that the estimated dependence over time is significant even when the exogenous covariates are included, hence indicating that the so-called "conditional independence" hypothesis on firm defaults is not supported by the data. However, a further econometric investigation reveals that such dependence is not constant over time. Specifically, while in the early period 1982–1998 all macro factors considered are not significant and, accordingly, default counts are strongly characterized by contagion effects over time, in the subsequent periods 1999–2006 and 2007–2011 we find that financial volatility and macroeconomic factors are strong explanators of defaults, while contagion effects (captured by the parameters linking current intensity to past default counts) become weak, or even absent during the Great Recession (2007–2011). The latter result, which contrasts with much of the recent literature on the role of contagion effects in the financial crisis, shows that while in general current defaults do indeed affect the probability of other firms defaulting in the future, in recent years economic and financial factors at the macro level are capable to explain a large portion of the correlation of US firms defaults over time.

Further issues are left to future research. First, our analysis is limited to defaults of U.S. industrial firms. It would be interesting to assess whether similar results characterize different sectors (e.g., financial) and/or countries. Second, the PARX specifications developed in this paper are univariate, in the sense that they can be used to model a single time series of default counts. The multivariate PARX case, which would permit to analyze the cross linkages

between different time series of defaults, represents an obvious extension of the econometric theory proposed in this paper and is currently under investigation by the authors.

A Appendix

A.1 Proof of Theorem 1

Define $\zeta := \max \left\{ \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i), \rho \right\} < 1$ and the norm $\|(x, \lambda)\|_w := w_x \|x\| + w_\lambda \|\lambda\|$, where $w_x, w_\lambda > 0$ are chosen below. Observe that the Markov chain $z_t = (x_t, \lambda_t)$ solves $z_t = F(z_{t-1}; \varepsilon_t, N_t)$, where, with $\alpha = (\alpha_1, \dots, \alpha_p)$ and $\beta = (\beta_1, \dots, \beta_q)$, and correspondingly, N of dimension p and λ of dimension q ,

$$F(x, \lambda; \varepsilon, N) := (g(x; \varepsilon), \omega + \alpha N + \beta \lambda + \gamma f(x))'.$$

We then wish to show that $F(x, \lambda; \varepsilon, N)$ is a stochastic contraction mapping w.r.t. (x, λ) . To this end, observe that, with $\bar{N}_t = (N_t, \dots, N_{t-p})'$,

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

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

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

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

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

Finally, $E \left[\left\| F(0, 0; \varepsilon_t, \bar{N}_t) \right\|_w \right] = w_x E \left[\|g(0; \varepsilon)\| \right] + w_\lambda (\gamma f(0) + \omega) < \infty$ by Assumption 2. This verifies the assumptions of Corollary 3.1 in Doukhan and Wintenberger (2008) with $\Phi(z) = \|z\|_w$, which yields that z_t is weakly dependent.

The next step is to show that $X_t = (y_t, z_t)$ is jointly stationary and weakly dependent using arguments similar to the ones for Meitz and Saikkonen (2008): First note that y_t is obviously stationary since λ_t is. Next, consider

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

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

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

Next, using the Markov chain property of z_t ,

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

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

To complete the proof, we verify existence of moments: Observe that $E[|y_t^*|^s] = \sum_{j=0}^s \binom{s}{j} E[(\lambda_t^*)^j]$ where, with $\bar{y}_t = (y_t, \dots, y_{t-p+1})'$ and $\bar{\lambda}_t = (\lambda_t, \dots, \lambda_{t-q+1})'$,

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

and

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

Hence,

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

with $r_{s-1}(y, \lambda, z)$ being an $(s-1)$ -order polynomial in $(\bar{y}, \bar{\lambda}, z)$ and so $E[r_{s-1}(\bar{y}_{t-1}^*, \bar{\lambda}_{t-1}^*, f(x_{t-1}^*))] < \infty$ by induction. Moreover, $E[(\omega + \gamma f(x_{t-1}^*))^s] < \infty$ by applying Doukhan and Wintenberger (2008, Theorem 3.2) on x_t together with Assumption 2. Thus, we are left with considering terms of the form,

$$\begin{aligned} E[(\alpha_i y_{t-1-i}^* + \beta_i \lambda_{t-1-i}^*)^s] &= \sum_{j=0}^s \binom{s}{j} \alpha_i^j \beta_i^{s-j} E[(y_{t-1-i}^*)^j (\lambda_{t-1-i}^*)^{s-j}] \\ &= \sum_{j=0}^s \binom{s}{j} \alpha_i^j \beta_i^{s-j} \sum_{k=0}^j \binom{j}{k} E[(\lambda_t^*)^{s+(k-j)}] \\ &= \sum_{j=0}^s \binom{s}{j} \alpha_i^j \beta_i^{s-j} E[(\lambda_t^*)^s] + C \\ &= (\alpha_i + \beta_i)^s E[(\lambda_t^*)^s] + C, \end{aligned}$$

where by induction $E [(\lambda_t^*)^k] < \infty$, for $k < s$. Collecting terms,

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

which, since $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$ (Assumption 3), has a well-defined solution.

A.2 Proof of Lemma 1

By the assumptions made, there exists a τ -weakly dependent solution to the dynamic system, c.f. Doukhan and Wintenberger (2008, Corollary 3.1). With X_t^* denoting the stationary solution, write

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

where, by the LLN for stationary and ergodic sequences, $\frac{1}{T} \sum_{t=1}^T h(X_t^*) \rightarrow^P E[h(X_t^*)]$. To show that the first term vanishes, first note that, by repeated use of iterated expectations and the contraction condition,

$$\begin{aligned} E [\|X_t - X_t^*\|^s] &\leq E [E [\|X_t - X_t^*\|^s | X_{t-1}, X_{t-1}^*]] \\ &\leq \rho E [\|X_{t-1} - X_{t-1}^*\|^s] \\ &\vdots \\ &= \rho^t E [\|X_0 - X_0^*\|^s], \end{aligned}$$

and

$$\begin{aligned} E [\|h(X_t)\|^{1+\delta}] &\leq E [\|X_t\|^s] = E [\|F(X_{t-1}, \xi_t)\|^s] \leq E [\|F(X_{t-1}, \xi_t) - F(0, \xi_t)\|^s] + E [\|F(0, \xi_t)\|^s] \\ &\leq \rho E [\|X_{t-1}\|^s] + E [\|F(0, \xi_t)\|^s] \\ &\vdots \\ &\leq \rho^t E [\|X_0\|^s] + E [\|F(0, \xi_t)\|^s] \sum_{i=0}^{t-1} \rho^i \\ &\leq \frac{E [\|X_0\|^s] + E [\|F(0, \xi_t)\|^s]}{1 - \rho} \\ &=: M. \end{aligned}$$

Now, with $c > 0$ given in the lemma, define $\mathbb{I}_{c,t} = \mathbb{I} \{\|X_t - X_t^*\| \leq c\}$ and write

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T [h(X_t) - h(X_t^*)] &= \frac{1}{T} \sum_{t=1}^T [h(X_t) - h(X_t^*)] \mathbb{I}_{c,t} + \frac{1}{T} \sum_{t=1}^T [h(X_t) - h(X_t^*)] [1 - \mathbb{I}_{c,t}] \\ &=: A_{T,1} + A_{T,2}. \end{aligned}$$

Using that, by assumption, $\|h(X_t) - h(X_t^*)\| I_{c,t} \leq L_c \|X_t - X_t^*\|$,

$$\begin{aligned} E[\|A_{T,1}\|] &\leq \frac{1}{T} \sum_{t=1}^T E[\|h(X_t) - h(X_t^*)\| \mathbb{I}_{c,t}] \leq \frac{L_c}{T} \sum_{t=1}^T \|X_t - X_t^*\| \leq \frac{L_c E[\|X_0 - X_0^*\|]}{T} \sum_{t=1}^T \rho^t \\ &\leq \frac{1}{T} \frac{L_c E[\|X_0 - X_0^*\|]}{1 - \rho}, \end{aligned}$$

while, using Holder's inequality,

$$\begin{aligned} E[\|A_{T,2}\|] &\leq \frac{1}{T} \sum_{t=1}^T E[\|h(X_t) - h(X_t^*)\| \mathbb{I}\{\|X_t - X_t^*\| > c\}] \\ &\leq \frac{2M}{T} \sum_{t=1}^T P(\|X_t - X_t^*\| > c)^{(1+\delta)/\delta} \\ &\leq \frac{2M}{c^{s(1+\delta)/\delta}} \frac{1}{T} \sum_{t=1}^T E[\|X_t - X_t^*\|^s]^{(1+\delta)/\delta} \\ &= \frac{2ME[\|X_0 - X_0^*\|^s]}{c^{s(1+\delta)/\delta}} \frac{1}{T} \sum_{t=1}^T \rho^t \\ &\leq \frac{2ME[\|X_0 - X_0^*\|^s]}{c^{s(1+\delta)/\delta} (1 - \rho)} \frac{1}{T}. \end{aligned}$$

Thus, $A_{T,k} \xrightarrow{P} 0$, $k = 1, 2$, which completes the proof.

A.3 Proof of Lemma 2

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

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

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

holds with $\delta_1(c) := \zeta_1(c) / \left(1 - \sum_{j=1}^p [\alpha_j + \beta_j]\right)$. Next,

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

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

$$\begin{aligned}
E[(\lambda_t^c - \lambda_t)(y_s^c - y_s)] &= E[E((\lambda_t^c - \lambda_t)(y_s^c - y_s) | \mathcal{F}_{s-1})] \\
&= E[(\lambda_t^c - \lambda_t) E(N_s[\lambda_s, \lambda_s^c])] = E(\lambda_t^c - \lambda_t)(\lambda_s^c - \lambda_s),
\end{aligned}$$

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

$$\begin{aligned}
E[(y_t^c - y_t)(y_s^c - y_s)] &= E[E((y_t^c - y_t)(y_s^c - y_s) | \mathcal{F}_{s-1})] \\
&= E[(y_t^c - y_t) E((y_s^c - y_s) | \mathcal{F}_{s-1})] \\
&= E(y_t^c - y_t)(\lambda_s^c - \lambda_s),
\end{aligned}$$

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

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

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

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

Collecting terms, one finds $E [(\lambda_t^c - \lambda_t)^2]$ is bounded by, $C \sum_{j=1}^t \psi_j E (e_{t-j}^c)^2$ 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,

$$E [(y_t^c - y_t)^2] \leq E [(\lambda_t^c - \lambda_t)^2] + |E [\lambda_t^c - \lambda_t]| \leq E [(\lambda_t^c - \lambda_t)^2] + \delta_1(c).$$

This completes the proof of Lemma 2.

A.4 Proof of Theorem 2

We consider here the case of $p = q = d_x = 1$ and write the model as

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

the following arguments are easily extended to the general case since this is alone complicated in terms of notation. We show consistency by verifying the general conditions provided in Kristensen and Rahbek (2005, Proposition 2). Given the LLN established in Lemma 1, these are easily verified apart from the condition, $E [\sup_{\theta \in \Theta} l_t^*(\theta)] < \infty$, and showing identification. Since $\lambda_t(\theta) \geq \omega_L$,

$$E \left[\sup_{\theta \in \Theta} l_t^*(\theta) \right] \leq \frac{1}{\omega_L} E \left[\lambda_t(\theta_0) \sup_{\theta \in \Theta} |\log \lambda_t^*(\theta)| \right].$$

Using Hölder's inequality, the right-hand side is finite if $E [\sup_{\theta \in \Theta} \lambda_t^*(\theta)] < \infty$, but this holds by Theorem 1. Regarding identification, we need to show that

$$L(\theta) := E [l_t^*(\theta)] = E [\lambda_t^*(\theta_0) \log \lambda_t^*(\theta) - \lambda_t^*(\theta)]$$

has a unique maximum at $\theta = \theta_0$. To this end, first note that

$$\begin{aligned} L(\theta) - L(\theta_0) &= E \left[\lambda_t^*(\theta_0) \log \left(\frac{\lambda_t^*(\theta)}{\lambda_t^*(\theta_0)} \right) + \lambda_t^*(\theta_0) - \lambda_t^*(\theta) \right] \\ &\leq E \left[\lambda_t^*(\theta_0) \left(\frac{\lambda_t^*(\theta)}{\lambda_t^*(\theta_0)} - 1 \right) + \lambda_t^*(\theta_0) - \lambda_t^*(\theta) \right] \\ &= 0, \end{aligned}$$

with equality if and only if

$$\lambda_t^*(\theta_0) = \lambda_t^*(\theta) \text{ almost surely.} \quad (\text{A.1})$$

The stationary solution can be represented as

$$\lambda_t^*(\theta) = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} a_i(\theta) y_{t-i}^* + \sum_{i=1}^{\infty} b_i(\theta) f(x_{t-i}^*),$$

where $a_i(\theta) = \alpha\beta^{i-1}$ and $b_i(\theta) = \gamma\beta^{i-1}$. Suppose now that there exists $\theta \in \Theta$ so that eq. (A.1) holds. We then claim that $\omega_0 = \omega$ and $c_i(\theta_0) = c_i(\theta)$ for all $i \geq 1$, where $c_i(\theta) = (a_i(\theta), b_i(\theta))$, which in turn implies $\theta = \theta_0$. We show this by contradiction: Let $m > 0$ be the smallest integer for which $c_i(\theta_0) \neq c_i(\theta)$ (if $c_i(\theta_0) = c_i(\theta)$ for all $i \geq 1$, then obviously $\omega_0 = \omega$). Eq. (A.1) can then be rewritten as

$$a_0 y_{t-m}^* + b_0 f(x_{t-m}^*) = \omega - \omega_0 + \sum_{i=1}^{\infty} a_i y_{t-m-i}^* + \sum_{i=1}^{\infty} b_i f(x_{t-m-i}^*),$$

where $a_i := \alpha_0 \beta_0^{i-1} - \alpha \beta^{i-1}$ and $b_i := \gamma_0 \beta_0^{i-1} - \gamma \beta^{i-1}$, $i = 1, 2, \dots$. The right hand side belongs to \mathcal{F}_{t-m-1} and so $a_0 y_{t-m}^* + b_0 f(x_{t-m}^*) | \mathcal{F}_{t-m-1}$ is constant. This is ruled out by Assumption 5.

To establish asymptotic normality we follow Kristensen and Rahbek (2005, proof of Theorem 2) and analyse the asymptotic behaviour of the score and information which is done below.

A.4.1 Score

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

$$S_T(\theta) = \sum_{t=1}^T s_t(\theta), \quad \text{where } s_t(\theta) = \left(\frac{y_t}{\lambda_t(\theta)} - 1 \right) \frac{\partial \lambda_t(\theta)}{\partial \theta}. \quad (\text{A.2})$$

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

$$\begin{aligned} \frac{\partial \lambda_t(\theta)}{\partial \eta} &= v_t + \beta \frac{\partial \lambda_{t-1}(\theta)}{\partial \eta} \\ \frac{\partial \lambda_t(\theta)}{\partial \beta} &= \lambda_{t-1}(\theta) + \beta \frac{\partial \lambda_{t-1}(\theta)}{\partial \gamma} \end{aligned}$$

In particular, with $\lambda_t = \lambda_t(\theta_0)$ and $\dot{\lambda}_t = \partial \lambda_t(\theta) / (\partial \theta)_{\theta=\theta_0}$, $s_t(\theta_0) = \dot{\lambda}_t (y_t / \lambda_t - 1)$. The score function is a Martingale difference w.r.t. the filtration \mathcal{F}_t satisfying $E[s_t(\theta_0) s_t(\theta_0)' | \mathcal{F}_{t-1}] = \dot{\lambda}_t \dot{\lambda}_t' / \lambda_t$. Note that $\dot{\lambda}_t = (v_t', \lambda_{t-1})' + \beta \dot{\lambda}_{t-1}$, with $\dot{\lambda}_0 = 0$. Thus, by the same arguments as in

the proof of Theorem 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 1, is weakly dependent with finite second moments. Furthermore, since $\lambda_t \geq \omega$, $\left\| \dot{\lambda}_t \dot{\lambda}'_t / \lambda_t \right\| \leq \left\| \dot{\lambda}_t \right\|^2$. It now follows by the remark following Lemma 1 that $\sqrt{T} S_T(\theta_0) \xrightarrow{d} N(0, \Omega(\theta_0))$ where, with $H(\theta)$ defined in the theorem,

$$\Omega(\theta) = E \left[s_t^*(\theta) s_t^{*\prime}(\theta) \right] = E \left[\dot{\lambda}_t^* (\dot{\lambda}_t^*)' / \lambda_t^* \right] = -H(\theta).$$

A.4.2 Information

The information is defined as

$$H_T(\theta) = -\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}, \quad (\text{A.3})$$

where

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

and

$$\begin{aligned} \frac{\partial^2 \lambda_t(\theta)}{\partial \eta \partial \beta} &= \frac{\partial \lambda_{t-1}(\theta)}{\partial \eta} + \beta \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial \eta \partial \beta}, \\ \frac{\partial^2 \lambda_t(\theta)}{\partial \beta^2} &= 2 \frac{\partial \lambda_{t-1}(\theta)}{\partial \beta} + \beta \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial \beta^2}, \\ \frac{\partial^2 \lambda_t(\theta)}{\partial \eta^2} &= \beta \frac{\partial^2 \lambda_t(\theta)}{\partial \eta^2} = \dots = 0. \end{aligned}$$

These recursions can be used to show that the augmented process $\tilde{X}_t(\theta) := \left(X'_t(\theta), \dot{\lambda}'_t(\theta), \text{vec}(\ddot{\lambda}_t(\theta)) \right)'$ is weakly dependent with second moments for $\theta \in \Theta$ in the same way that Theorem 1 was proved. In particular, for all $\theta \in \Theta$, we can apply Lemma 1 to obtain

$$H_T(\theta) = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \xrightarrow{P} E \left[\frac{\partial^2 l_t^*(\theta)}{\partial \theta \partial \theta'} \right].$$

Moreover, $\theta \mapsto \partial^2 l_t(\theta) / (\partial \theta \partial \theta')$ is continuous and, with $\bar{\theta} = (\omega_U, \alpha_U, \beta_U, \gamma_U)$ containing the maximum values of the individual parameters, we obtain

$$\begin{aligned} \frac{\partial \lambda_t(\theta)}{\partial \beta} &= \lambda_{t-1}(\theta) + \beta \frac{\partial \lambda_{t-1}(\theta)}{\partial \gamma} \leq \sum_{i=0}^{t-1} \beta_U^i \lambda_{t-1-i}(\bar{\theta}) = \frac{\partial \lambda_t(\bar{\theta})}{\partial \beta}, \\ \frac{\partial^2 \lambda_t(\theta)}{\partial \beta^2} &= 2 \frac{\partial \lambda_{t-1}(\theta)}{\partial \beta} + \beta \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial \beta^2} \leq 2 \sum_{i=0}^{t-1} \beta_U^i \dot{\lambda}_{t-1-i}(\bar{\theta}) = \frac{\partial^2 \lambda_t(\bar{\theta})}{\partial \beta^2}, \end{aligned}$$

and similar for the other second order derivatives. while, by the same arguments as in Han and Kristensen (2014), there exists a function $B(\tilde{x})$ so that $\lambda_t(\theta_0)/\lambda_t(\theta) \leq B(\tilde{X}_t)$ for all θ in a neighbourhood of θ_0 , where

$$E \left[B(\tilde{X}_t^*) \left\| \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta} \right\|^2 \right] < \infty, \quad E \left[B(\tilde{X}_t) \left\| \frac{\partial^2 \lambda_t(\bar{\theta})}{\partial \theta \partial \theta'} \right\| \right] < \infty.$$

In total,

$$\left\| \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq \bar{D}(\tilde{X}_t(\bar{\theta})), \quad \bar{D}(\tilde{X}_t(\bar{\theta})) := B(\tilde{X}_t) \left\{ \left\| \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta} \right\|^2 + \left\| \frac{\partial^2 \lambda_t(\bar{\theta})}{\partial \theta \partial \theta'} \right\| \right\},$$

where $E[\bar{D}(\tilde{X}_t^*(\theta))] < \infty$ with \tilde{X}_t^* denoting the stationary version of \tilde{X}_t . It now follows by Proposition 1 in Kristensen and Rahbek (2005) that $\sup_{\theta \in \Theta} \|H_T(\theta) - H(\theta)\| \xrightarrow{P} 0$ with $H(\theta)$ defined in the theorem.

Finally, we show that $H(\theta_0)$ is non-singular. To see this, we use the same arguments as in the proof of identification that we provided as part of showing consistency: First note that $H(\theta_0) = E[\dot{\lambda}_t^* (\dot{\lambda}_t^*)' / \lambda_t^*]$ is singular if and only if there exists $a \in \mathbb{R}^4 \setminus \{0\}$ and $t \geq 1$ such that $a' \dot{\lambda}_t^* = 0$ a.s. Since $\dot{\lambda}_t^*$ is stationary, this must hold for all t . Recall that $\dot{\lambda}_t^* \in \mathbb{R}^4$ can be written as $\dot{\lambda}_t^* = V_t^* + \beta \dot{\lambda}_{t-1}^*$, where $V_t^* = (1, y_{t-1}^*, f(x_{t-1}^*), \lambda_{t-1}^*)'$ is a vector of positive elements. So $a' \dot{\lambda}_t^* = 0$ a.s. holds if and only if $a' V_t^* = 0$ a.s. for all $t \geq 1$. However, this is ruled out by Assumption 5, c.f. proof of identification.

A.5 Proof of Theorem 3

The proof follows by noting that Lemmas 3.1-3.4 in Fokianos, Rahbek and Tjøstheim (2009) carry over to our setting with only minor modifications. The only difference is that the parameter vector θ here include γ as related to the link function $f(x_{t-1})$. However, as $E[f(x_{t-1}^*)] < \infty$, all arguments remain identical as is easily seen upon inspection of the proofs of the lemmas in Fokianos, Rahbek and Tjøstheim (2009).

A.6 Proof of Corollary 1

It suffices to verify the regularity conditions of Andrews (1999, Theorem 3). First, in the proof of Theorem 2 we establish consistency and classic convergence of the score and information. Second, the parameter set satisfies the geometric conditions needed by arguments identical to the ones in Francq and Zakoian (2009).

References

- [1] Andrews, D.W.K. (1999), "Estimation when a parameter is on a boundary", *Econometrica*, 67, 1341–1383.
- [2] Andrews, D.W.K. (2000), "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space", *Econometrica*, 68, 399–405.
- [3] Berkes, I., L. Horváth and P. Kokoszka (2003), "GARCH processes: Structure and estimation," *Bernoulli* 9, 201–227.
- [4] Brown, B. (1971), "Martingale central limit theorems", *Annals of Mathematical Statistics*, 42, 59–66.
- [5] Christou, V. and K. Fokianos (2013), "Quasi-likelihood inference for negative binomial time series models," *Journal of Time Series Analysis* 35, 55–78.
- [6] Das, S.R., D. Duffie, N. Kapadia, and L. Saita (2007), "Common failings: How corporate defaults are correlated," *Journal of Finance* 62, 93–117.
- [7] Demos, A. and E. Sentana (1998), "Testing for GARCH effects: A one-sided approach," *Journal of Econometrics* 86, 97–127.
- [8] Doukhan, P. and O. Wintenberger (2008), "Weakly dependent chains with infinite memory," *Stochastic Processes and their Applications* 118, 1997–2013.
- [9] Duffie, D., A. Eckner, G. Horel, and L. Saita (2009), "Frailty correlated default," *Journal of Finance* 64, 2089–2123.
- [10] Ferland, R., A. Latour and D. Oraichi (2006), "Integer-valued GARCH processes", *Journal of Time Series Analysis*, 27, 923–942.
- [11] Fokianos, K., A. Rahbek and D. Tjøstheim (2009), "Poisson autoregression", *Journal of the American Statistical Association*, 104, 1430–1439.
- [12] Francq, C. and J.M. Zakoïan (2009), "Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons", *Journal of the American Statistical Association*, 104, 313–324.
- [13] Gallo, G.M. and B. Pacini (2000), "The effects of trading activity on market volatility," *European Journal of Finance* 6, 163–175
- [14] Giesecke, K., F. A. Longstaff, S. Schaefer and I. Strebulaev (2011), "Corporate bond default risk: A 150-year perspective," *Journal of Financial Economics* 102, 233–250.

- [15] Gouriéroux, C., A. Monfort, E. Renault and A. Trognon (1987), "Generalised residuals," *Journal of Econometrics*, 34, 5–32.
- [16] Gouriéroux, C., A. Monfort and A. Trognon (1984), "Pseudo maximum likelihood methods: Applications to Poisson models", *Econometrica*, 52, 701–720.
- [17] Han, H., and Kristensen, D. (2014), "Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates," *Journal of Business and Economic Statistics* 32, 416-429.
- [18] Hansen, P.R., Z. Huang and H.W. Shek (2012) "Realized GARCH: A joint model for returns and realized measures of volatility," *Journal of Applied Econometrics* 27, 877–906.
- [19] Kedem, B., and K. Fokianos (2002), *Regression Models for Time Series Analysis*, Hoboken, NJ: Wiley.
- [20] Koopman, S.J., A. Lucas and B. Schwaab (2012), "Dynamic factor models with macro, frailty, and industry effects for U.S. default counts: the credit crisis of 2008," *Journal of Business and Economic Statistics* 30, 521-532.
- [21] Kristensen, D. and A. Rahbek (2005), "Asymptotics of the QMLE for a class of ARCH(q) models", *Econometric Theory*, 21, 946–961.
- [22] Lamoureux, C.G., and W.D. Lastrapes (1990), "Heteroskedasticity in stock return data: Volume versus GARCH effects", *Journal of Finance*, 45, 221–229.
- [23] Lando, D., and M. Nielsen (2010), "Correlation in corporate defaults: Contagion or conditional independence?", *Journal of Financial Intermediation*, 19, 355-372.
- [24] Lando, D., M. Medhat, M. Stenbo Nielsen and S.F. Nielsen (2013), "Additive intensity regression models in corporate default analysis," *Journal of Financial Econometrics* 11, 443-485.
- [25] Meitz, M., and P. Saikkonen (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.
- [26] Nyblom, J. (1989), "Testing for the constancy of parameters over time," *Journal of The American Statistical Association* 84, 223–30.

- [27] Rydberg, T. H., and N. Shephard (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. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young, Cambridge: Isaac Newton Institute and Cambridge University Press, pp. 217–246.
- [28] Shephard, N. and K. Sheppard (2010) Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* 25, 197-231.
- [29] Stock, J. and M. Watson (1997) "Evidence on structural instability in macroeconomic time series relations," *Journal of Business and Economic Statistics* 14, 11–30.
- [30] Streett, S. (2000), “Some observation driven models for time series of counts,” Ph.D. thesis, Colorado State University, Dept. of Statistics.
- [31] Tay, A.S and K.F. Wallis (2000) “Density forecasting: A survey”, *Journal of Forecasting*, 19, 235-254.
- [32] White, H. (1982) "Maximum likelihood estimation of misspecified models", *Econometrica*, 50, 1-25.

Table 1: Results of simulations for PARX(1,1) with DGP 1.

		Scenario 1 ($\beta = 0$)				Scenario 2 ($\beta = 0.2$)			Scenario 3 ($\beta = 0.7$)		
T		True	Mean	RMSE	KS	Mean	RMSE	KS	Mean	RMSE	KS
100	ω	0.10	0.09	0.16	0.36	0.10	0.18	0.01	0.15	0.30	0.07
	α	0.30	0.28	0.13	0.32	0.27	0.11	0.97	0.18	0.15	0.00
	β	0.00	0.02	0.15	0.31	0.22	0.14	0.34	0.77	0.15	0.00
	γ	0.50	0.51	0.07	0.85	0.51	0.07	0.32	0.51	0.11	0.84
250	ω	0.10	0.09	0.07	0.85	0.10	0.08	0.19	0.13	0.21	0.13
	α	0.30	0.30	0.07	0.87	0.29	0.07	0.99	0.23	0.06	0.72
	β	0.00	0.00	0.08	0.93	0.21	0.08	0.63	0.72	0.06	0.64
	γ	0.50	0.50	0.04	0.49	0.50	0.04	0.92	0.50	0.05	0.81
500	ω	0.10	0.10	0.05	0.66	0.10	0.05	0.35	0.11	0.13	0.21
	α	0.30	0.30	0.04	0.33	0.30	0.04	0.87	0.24	0.04	0.86
	β	0.00	0.00	0.05	0.17	0.20	0.05	0.16	0.71	0.04	0.96
	γ	0.50	0.50	0.02	0.34	0.50	0.02	0.75	0.50	0.02	0.95
1000	ω	0.10	0.10	0.03	0.38	0.10	0.04	0.42	0.10	0.10	0.24
	α	0.30	0.30	0.03	0.52	0.30	0.03	0.61	0.24	0.02	0.79
	β	0.00	0.00	0.03	0.98	0.20	0.03	0.71	0.71	0.02	0.81
	γ	0.50	0.50	0.02	0.74	0.50	0.02	0.32	0.50	0.02	0.99

Table 2: Results of simulations for PARX(1,1) with DGP 2.

		Scenario 1 ($\beta = 0$)				Scenario 2 ($\beta = 0.2$)			Scenario 3 ($\beta = 0.7$)		
T		True	Mean	RMSE	KS	Mean	RMSE	KS	Mean	RMSE	KS
100	ω	0.10	0.12	0.20	0.00	0.11	0.18	0.00	0.16	0.30	0.02
	α	0.30	0.29	0.13	0.47	0.27	0.13	0.43	0.17	0.16	0.00
	β	0.00	-0.01	0.23	0.16	0.21	0.19	0.31	0.78	0.16	0.00
	γ	0.50	0.51	0.13	0.50	0.51	0.12	0.32	0.51	0.14	0.81
250	ω	0.10	0.10	0.09	0.14	0.12	0.12	0.08	0.18	0.25	0.00
	α	0.30	0.30	0.07	0.70	0.29	0.07	0.57	0.23	0.05	0.58
	β	0.00	0.00	0.10	0.33	0.20	0.10	0.81	0.71	0.06	0.84
	γ	0.50	0.50	0.06	0.39	0.50	0.07	0.85	0.51	0.14	0.30
500	ω	0.10	0.10	0.07	0.22	0.10	0.07	0.54	0.13	0.14	0.00
	α	0.30	0.30	0.05	0.95	0.30	0.05	0.96	0.24	0.04	0.47
	β	0.00	0.00	0.07	1.00	0.20	0.07	0.90	0.71	0.04	0.29
	γ	0.50	0.50	0.04	0.59	0.50	0.05	0.97	0.51	0.07	0.46
1000	ω	0.10	0.10	0.05	0.73	0.10	0.05	0.14	0.12	0.11	0.02
	α	0.30	0.30	0.03	0.81	0.30	0.03	0.56	0.24	0.02	0.95
	β	0.00	0.00	0.05	0.82	0.20	0.05	0.80	0.70	0.03	0.97
	γ	0.50	0.50	0.03	0.74	0.50	0.03	0.43	0.51	0.05	0.77

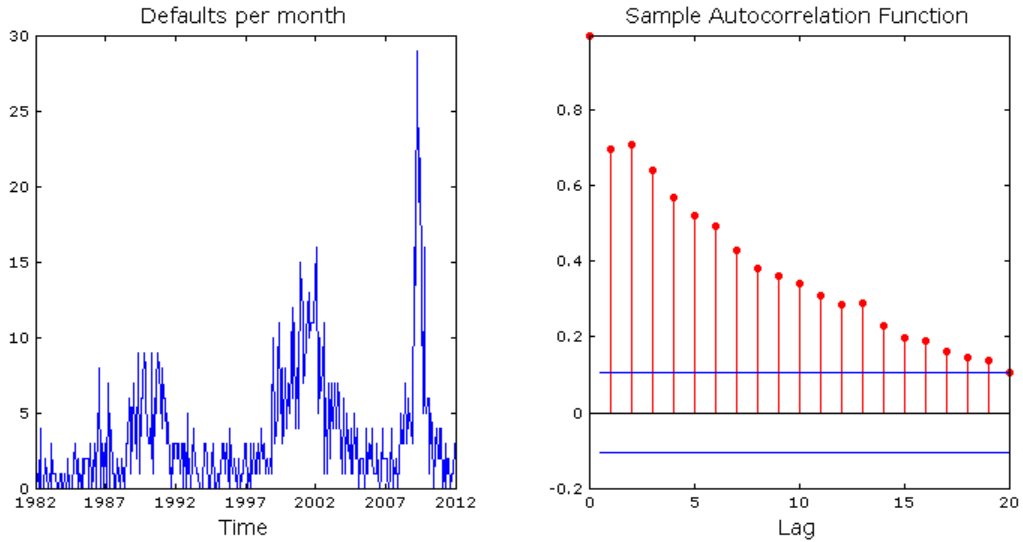


Figure 1: (a) Number of defaults per month among Moody's rated US industrial firms in the period 1982-2011. (b) Autocorrelation function of the default data.

Table 3: Estimation results of different PARX models.

	PAR	RV	SP	DG	NB	IP	LI	RV & LI ⁽⁻⁾	All
$\hat{\omega}$	0.301 (3.625)	0.169 (2.467)	0.116 (0.716)	0.206 (2.219)	0.289 (3.551)	0.202 (2.142)	0.295 (2.013)	0.232 (3.242)	0.208 (1.001)
$\hat{\alpha}_1$	0.241 (5.441)	0.197 (4.395)	0.227 (5.159)	0.221 (4.933)	0.228 (5.119)	0.213 (4.716)	0.193 (4.265)	0.185 (4.109)	0.180 (3.944)
$\hat{\alpha}_2$	0.215 (3.221)	0.179 (2.908)	0.2217 (3.348)	0.198 (3.026)	0.206 (3.138)	0.145 (2.262)	0.198 (3.117)	0.188 (3.039)	0.183 (2.898)
$\hat{\beta}$	0.459 (6.094)	0.526 (7.939)	0.4298 (5.430)	0.455 (6.063)	0.469 (6.296)	0.552 (8.173)	0.498 (6.881)	0.518 (7.547)	0.512 (7.087)
RV		63.99 (4.111)						28.09 (2.057)	24.31 (1.692)
SP			0.241 (2.802)						0.000 (0.000)
DG				0.017 (1.893)					0.006 (0.640)
NB					0.419 (2.229)				0.000 (0.000)
IP						0.695 (3.287)			0.000 (0.000)
LI							0.941 (4.194)	0.729 (3.733)	0.754 (1.561)
$\hat{\alpha}_1 + \hat{\alpha}_2$	0.465 (7.452)	0.376 (8.069)	0.449 (9.635)	0.419 (8.743)	0.434 (9.249)	0.358 (7.380)	0.391 (7.726)	0.373 (6.679)	0.363 (7.235)
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 statistic	29.81	11.02	19.99	26.96	24.89	17.32	2.79	0.33	

Notes: t statistics in parentheses. For any significance level $\eta < 1/2$ standard critical values for one sided t tests apply, see Remark 4.

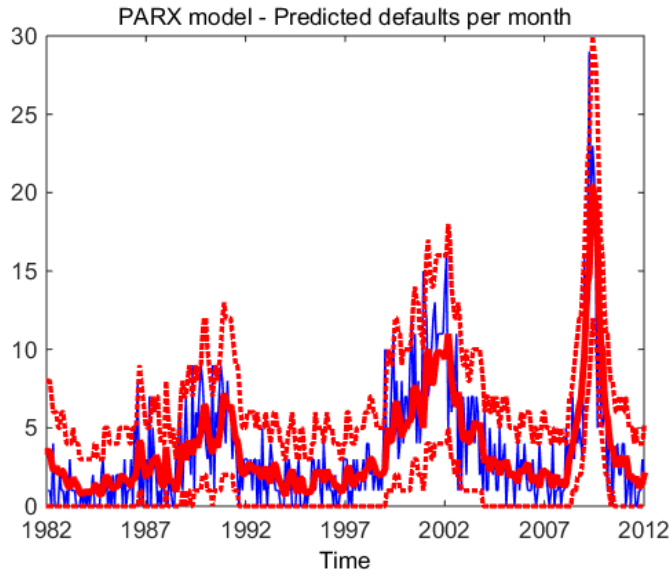


Figure 2: Actual number of defaults (blue) and estimated intensity (red).

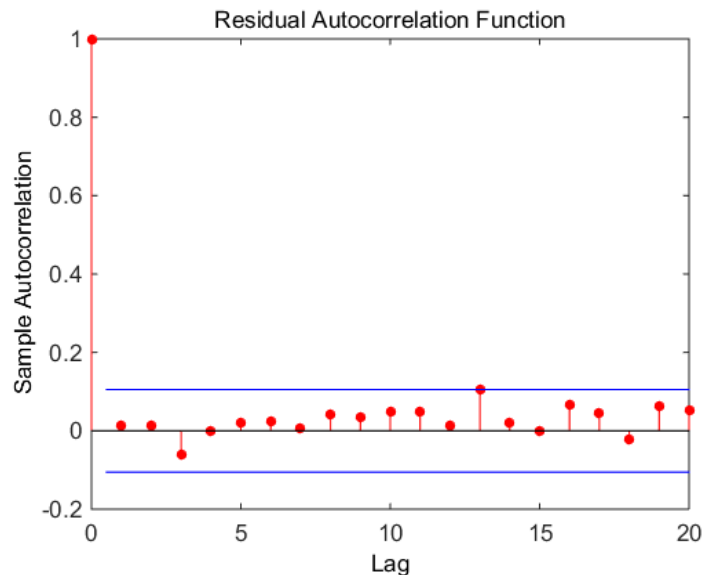


Figure 3: Sample autocorrelation function of Pearson residuals.

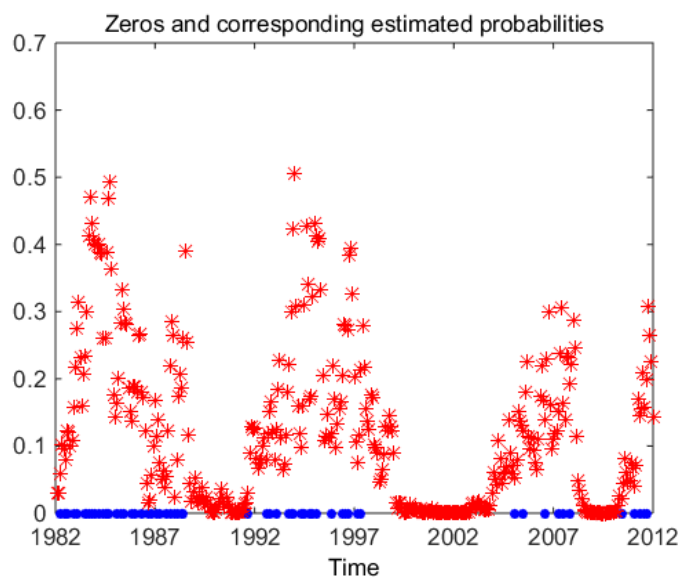


Figure 4: Empirical zero counts (asterisks) and probability of having a zero count under the estimated model (crosses).

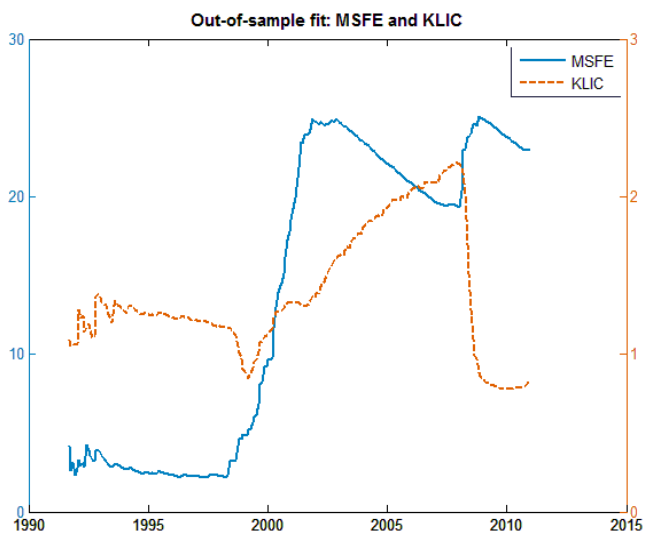


Figure 5: Out-of-sample forecasting performance of preferred PARX model.

Table 4: Time-variation of parameter estimates

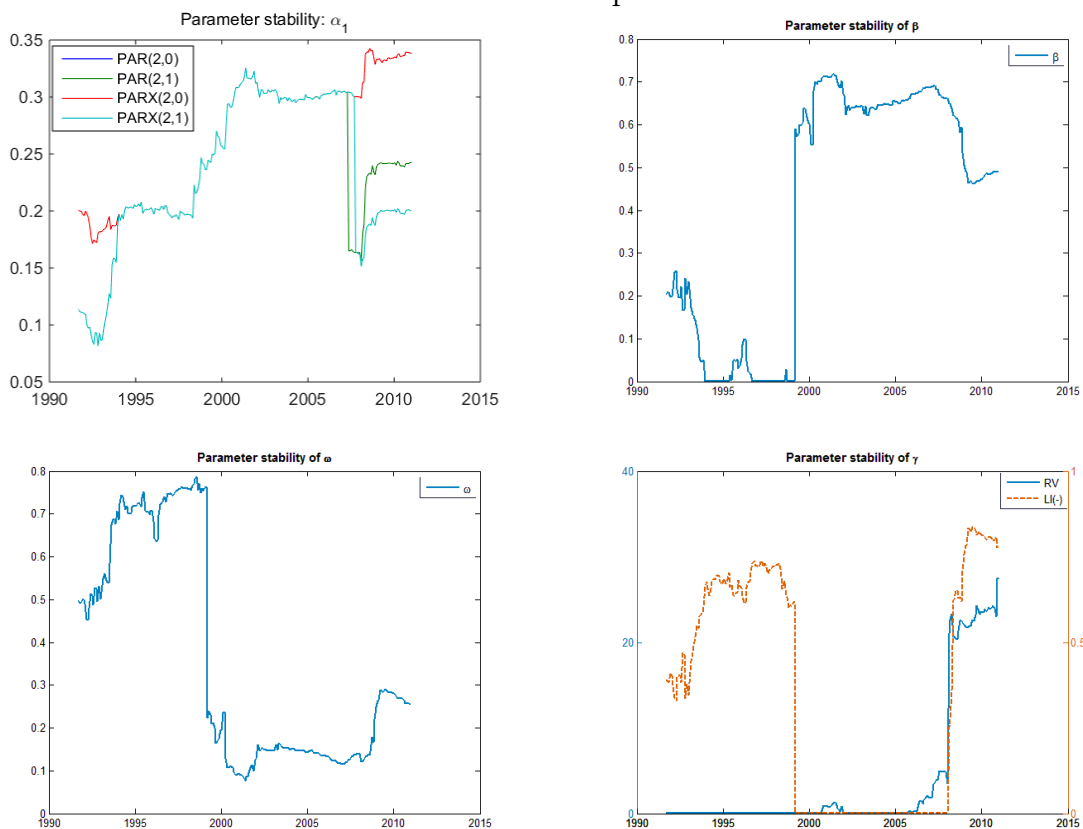


Table 5: Preferred models and their parameter estimates, 1982-1998, 1998-2007 and 2007-2011

	ω	α_1	α_2	β	RV	LI ⁽⁻⁾
1982-1998 - PAR(1,1)	0.80	0.22	0.43	-	-	-
<i>t</i> -stats	(7.04)	(5.29)	(8.32)	-	-	-
1998-2007 - PARX(1,1)	0.00	0.09	-	0.73	168.34	-
<i>t</i> -stats	0.10	(2.04)	-	(7.31)	(8.65)	-
2007-2011 - PARX(2,1)	0.00	-	-	0.82	99.23	0.70
<i>t</i> -stats	(0.00)	-	-	(6.81)	(2.17)	(2.38)

Research Papers 2013



- 2014-52: Tobias Fissler and Mark Podolskij: Testing the maximal rank of the volatility process for continuous diffusions observed with noise
- 2014-53: Cristina M. Scherrer: Cross listing: price discovery dynamics and exchange rate effects
- 2014-54: Claudio Heinrich and Mark Podolskij: On spectral distribution of high dimensional covariation matrices
- 2014-55: Gustavo Fruet Dias and Fotis Papailias: Forecasting Long Memory Series Subject to Structural Change: A Two-Stage Approach
- 2014-56: Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Risk Premia Embedded in Index Options
- 2014-57: Eduardo Rossi and Paolo Santucci de Magistris: Indirect inference with time series observed with error
- 2014-58: Anders Bredahl Kock and Haihan Tang: Inference in High-dimensional Dynamic Panel Data Models
- 2015-01: Tom Engsted, Simon J. Hviid and Thomas Q. Pedersen: Explosive bubbles in house prices? Evidence from the OECD countries
- 2015-02: Tim Bollerslev, Andrew J. Patton and Wenjing Wang: Daily House Price Indices: Construction, Modeling, and Longer-Run Predictions
- 2015-03: Christian M. Hafner, Sebastien Laurent and Francesco Violante: Weak diffusion limits of dynamic conditional correlation models
- 2015-04: Maria Eugenia Sanin, Maria Mansanet-Bataller and Francesco Violante: Understanding volatility dynamics in the EU-ETS market
- 2015-05: Peter Christoffersen and Xuhui (Nick) Pan: Equity Portfolio Management Using Option Price Information
- 2015-06: Peter Christoffersen and Xuhui (Nick) Pan: Oil Volatility Risk and Expected Stock Returns
- 2015-07: Peter Christoffersen, Bruno Feunou and Yoontae Jeon: Option Valuation with Observable Volatility and Jump Dynamics
- 2015-08: Alfonso Irarrazabal and Juan Carlos Parra-Alvarez: Time-varying disaster risk models: An empirical assessment of the Rietz-Barro hypothesis
- 2015-09: Daniela Osterrieder, Daniel Ventosa-Santaulària and Eduardo Vera-Valdés: Unbalanced Regressions and the Predictive Equation
- 2015-10: Laurent Callot, Mehmet Caner, Anders Bredahl Kock and Juan Andres Riquelme: Sharp Threshold Detection Based on Sup-norm Error rates in High-dimensional Models
- 2015-11: Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen and Anders Rahbek: Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)