Received: 2 February 2021 | Accepted: 8 February 2022

DOI: 10.1111/rssc.12551

ORIGINAL ARTICLE

Modelling clusters of corporate defaults: Regime-switching models significantly reduce the contagion source

Geir D. Berentsen[1] | **Jan Bulla**[2,3] | **Antonello Maruotti**[2,4] | **Bård Støve**[2]

[1]Department of Business and Management Science, Norwegian School of Economics, Bergen, Norway

[2]Department of Mathematics, University of Bergen, Bergen, Norway

[3]Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

[4]Dipartimento di Giurisprudenza, Economia, Politica e Lingue Moderne (GEPLI), Libera Università Maria Ss Assunta, Rome, Italy

Correspondence
Geir D. Berentsen, Department of Business and Management Science, Norwegian School of Economics, Bergen, Norway.
Email: Geir.Berentsen@nhh.no

Funding information
This work was supported by the Finance Market Fund, Norway, project number 309218; 'Statistical modelling and inference for (high-dimensional) financial data' and the Norwegian Research Council through the Centre for Research-Based Innovation 'Climate Futures', project number 309562

Abstract
In this paper, we report robust evidence that the process of corporate defaults is time-dependent and can be modelled by extending an autoregressive count time series model class via the introduction of regime-switching. That is, some of the parameters of the model depend on the regime of an unobserved Markov chain, capturing the model changes during clusters observed for count time series in corporate defaults. Thus, the process of corporate defaults is more dynamic than previously believed. Moreover, the contagion effect—that current defaults affect the probability of other firms defaulting in the future—is reduced compared to models without regime-switching, and is only present in one regime. A two-regime model drives the counts of monthly corporate defaults in the United States. To estimate the model, we introduce a novel quasi-maximum likelihood estimator by adapting the extended Hamilton–Gray algorithm for the Poisson autoregressive model.

KEYWORDS
extended Hamilton-Gray algorithm, integer valued time series, Poisson autoregression, RS-INGARCHX

1 INTRODUCTION

Major causes for modelling and forecasting corporate defaults have been both the aspiration and the need to explain the observed clustering of defaults. In essence, two justifications have been proposed for this stylized fact. First, each firm can be considered exposed to a 'systematic risk', which is represented by shared economic or financial variables causing changes to a firm's conditional default probabilities; and second, the default of one firm may increase the likelihood of other firms defaulting, resulting in contagion.

While several existing models accommodate the possible impact of these two sources, fitting such models during different periods can produce significantly different parameter estimates (see, e.g. Agosto et al., 2016). Consequently, there may exist varying ways in which systematic risk and contagion effects impact the number of defaults during periods of clustering and periods with a low number of defaults. This paper explores such possibilities by using a regime-switching (RS) Poisson autoregressive model that allows parameter values to change over time.

Several studies have examined default clustering in the past and have done so using two main categories. For the first one, firm-level data are available in addition to macroeconomic variables and dates of default. These default times are then usually modelled by Poisson processes with both types of covariates driving the default intensities. Studies belonging to this category are, for example Das et al. (2007), who provide evidence that the 'systemic risk' on its own cannot explain the degree of clustering observed in American industrial defaults. Moreover, Lando and Nielsen (2010) put a different type of test procedure to practice and did not report a contagion effect. Duffie et al. (2007) provide evidence that default probabilities of American companies increase with decreasing short-term interest rates, whereas Duffie et al. (2009) present strong evidence for the presence of common latent factors.

Studies in the second category use aggregate data, consisting of the number of defaults and macroeconomic variables in given periods. Several papers have used this approach, for example Koopman et al. (2012), Agosto et al. (2016), Sant'Anna (2017), and Azizpour et al. (2018). In particular, Azizpour et al. (2018) find strong evidence that the contagion effect represents a main source for the observed clustering behaviour, even when including a so-called frailty factor (see also Bai et al., 2015). The papers most related to our approach are those by Agosto et al. (2016) and Sant'Anna (2017). Agosto et al. (2016) introduce a class of Poisson autoregressive models with covariates to model default counts. They find evidence of a contagion effect, which has diminished in recent years. Sant'Anna (2017) introduces a new test procedure that permits model checks for dynamic count models and finds evidence of a contagion effect as well. None of the existing studies of corporate defaults considers a regime-switching modelling approach to the best of our knowledge.

In this paper, we model the default counts using a specific class of regime-switching integer-valued generalized autoregressive conditional heteroscedasticity models with exogenous covariates (RS-INGARCHX). The model can be seen as an RS extension of the INGARCH models studied in Rydberg and Shephard (2000), Ferland et al. (2006), and Fokianos et al. (2009), and in particular the INGARCHX model used for studying defaults in Agosto et al. (2016). For an overview of count time series models and regime-switching models, see Appendix A. In detail, we consider a general m-regime RS-INGARCHX model in a frequentist framework, and we approximate the likelihood using an adaptation of the extended Hamilton–Gray (EHG) algorithm. Originally, an analogous algorithm was used for estimating a regime-switching autoregressive moving-average (RS-ARMA) model in Chen and Tsay (2011), which builds on the original algorithms of Hamilton (1989) and Gray (1996). Our version of the estimation algorithm also contains a backward procedure, which provides smoothing probabilities that we use to infer the clustering structure underlying the data.

Our model allows for the investigation of several features regarding the characteristics of default counts. In the economic and financial regime-switching literature, certain regimes often correspond to certain crisis periods (see, e.g. Ang & Timmermann, 2012; Bernardi et al., 2017; Gray, 1996). We establish the presence of one regime linked to low counts of defaults and a second one linked to clusters with high default counts. A posteriori, the second regime also coincides with periods of economic crisis. In comparison to existing models, our approach improves the model fit and forecasting performance. Furthermore, we investigate the extent to which contagion represents a significant source of default clustering. Our approach improves the ability to identify this effect and the impact of macroeconomic variables, thus enabling a deeper understanding of the dynamics of default counts.

The paper is organized as follows. Section 2 describes the American default counts data that we analyse. Section 3 contains the formulation of the RS-INGARCHX model and provides an interpretation of the model's inherent contagion and macroeconomic risk components as well as the implications of regime-switching. Furthermore, we present algorithms permitting the estimation of model parameters, state inference and prediction. Section 4 contains results and starts with a presentation of the considered models in Section 4.1. Subsequently, Sections 4.2 and 4.3 examine the link of regime-switching an economic crisis and the sources of clustering under it, respectively. We shed light on regime-specific dynamics in Section 4.4. After that, Section 4.5 presents various aspects related to the model fit and in-sample performance of multiple models under investigation. We offer details on the out-of-sample performance, including forecast accuracy and robustness checks in Section 4.6. Finally, Section 5 presents our conclusions. Also, the appendices (found in the supplementary files) presents further aspects and details. These include a more in-depth literature review, algorithmic information, a simulation study and various robustness checks.

2 | DATA

The American default counts data consist of the monthly number of bankruptcies filed in the US bankruptcy courts and is available from the UCLA-LopPucki Bankruptcy Research database (see http://lopucki.law.ucla.edu). We study data covering the period from January 1985 to September 2017, a total of 393 monthly observations. The data bases on the counts of defaults of all companies that had declared assets of more than US$ 100 million measured in 1980 dollars in the year before the firm filed the bankruptcy case, and had reported to the Securities and Exchange Commission (SEC) in the three years before the bankruptcy. The counts of monthly bankruptcies are considered in terms of the month during which the bankruptcy was filed. A total of 1065 defaults is counted over the entire sample period. Figure 1 (left panel) displays the time series together with recession periods. NBER-based recession indicators for the United States (USREC) available from the St. Louis Fed online database FRED determine these recession periods. The figure clearly shows the default clusters related to the savings and loan crisis of the late 1980s and early 1990s, the burst of the dot-com bubble in 2000–2002, and the financial crisis of 2007–2008. However, the European sovereign debt crisis of 2011 did not impact the numbers of US defaults considerably. The peak of defaults in 2016 can be partly explained by the fact that the economic growth in the US slowed to a tepid 1.6% annual rate, which was a 5-year low and a sharp drop

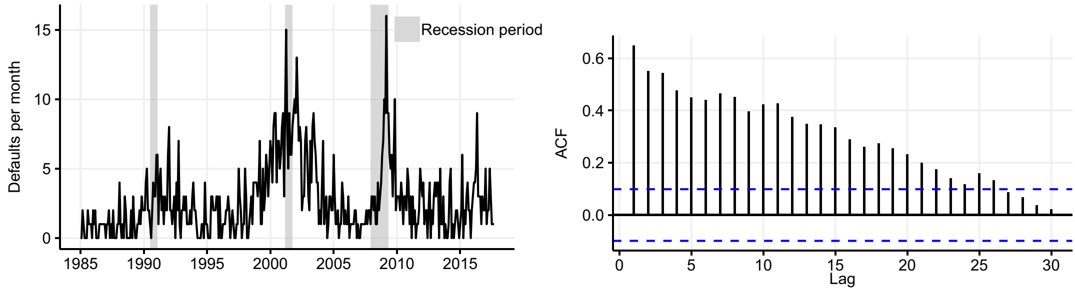

FIGURE 1 Monthly default data
Notes: The left figure shows the monthly number of defaults of American firms from January 1985 to September 2017. The grey rectangles represent US recession periods according to NBER. The right figure displays the autocorrelation function of these data.

TABLE 1 Overview and summary statistics of variables considered, including standard deviations (SD), skewness (Skew.) and Kurtosis (Kurt.). The last two columns shows the *p*-values for the augmented Dickey–Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt–Shin (KPSS) test for (level) stationarity. Monthly sampling frequency

Abbrev.	Description	Mean	SD	Skew.	Kurt.	ADF	KPSS
Y_t	Defaults per month	2.71	2.61	1.59	6.21	0.18	0.01
indpro	Industrial production index	1.89	7.32	−1.64	12.54	<0.01	0.08
ppifgsup	Produce price index:finished good	1.96	6.99	−0.84	6.94	<0.01	0.10
ppieng	Produce price index:fuels and related energy	1.71	48.43	−0.60	5.88	<0.01	0.10
unrate	Civilian unemployment rate	−0.09	0.93	2.05	8.21	<0.01	0.10
baa	Moody's seasoned baa corporate bond yield	−0.30	0.90	−0.19	3.79	<0.01	0.10
SP500ret	S&P500 yearly returns	3.53	7.06	−1.34	5.75	<0.01	0.09
GDP	Gross domestic product	2.66	2.34	−1.08	6.43	<0.01	0.01
SP500vol	S&P500 return volatility	0.94	0.59	3.76	25.74	<0.01	0.10

from the 2.9% annual growth of 2015. Figure 1 (right panel) shows the autocorrelation function of the observations, which exhibits signs of strong temporal dependence. Table 1 displays summary statistics related to the variables considered. These data have been analysed in the past (see e.g. Sant'Anna, 2017), while other studies have been using data from Moody's Default Risk Service (such as Agosto et al., 2016; Azizpour et al., 2018). However, both data sets exhibit comparable patterns, and no apparent reason is visible for preferring either.

We will study the impact of several macroeconomic and financial variables representing the common systematic risk corporations face on the monthly rate of defaults. Table 1 provides an overview of the considered covariates, along with summary statistics. These variables are the following: industrial production index (indpro), producer price index by commodity for final demand: finished goods (ppifgsup), producer price index: fuels and related energy (ppieng), civilian unemployment rate (unrate), Moody's seasoned baa corporate bond yield (baa), S&P500 yearly returns (SP500ret), gross domestic product (GDP) and S&P500 realized

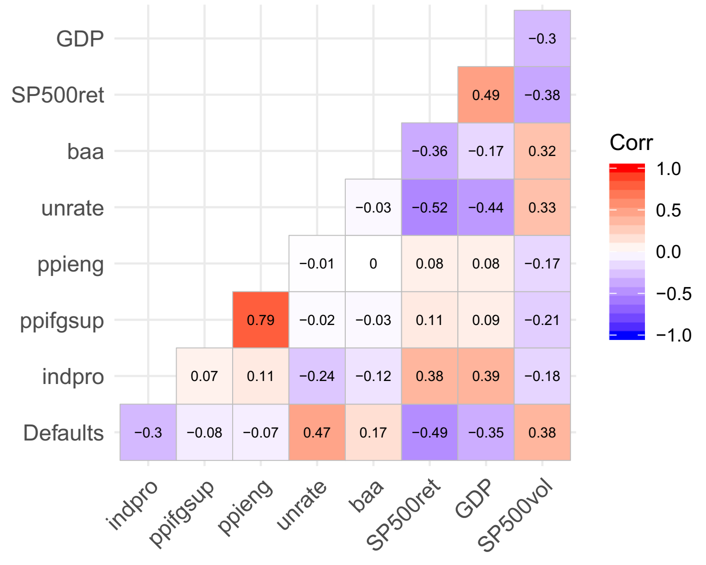

FIGURE 2 Cross-correlation matrix

Notes: This figure displays the cross-correlation (lower triangle) of the default counts and covariates considered. The covariates are lagged by one relative to the default counts.

volatility (SP500vol). Except for the variables SP500ret and SP500vol, which are collected from DataStream, the variables originate from the St. Louis Fed online database FRED. The variables indpro, ppifgsup and ppieng are expressed as yearly growth rates, whereas the variables unrate, baa and SP500ret are expressed as yearly differences. The SP500vol is monthly realized volatility defined as SP500vol $= \left(1/n_t \sum_{i=1}^{n_t} r_{i,t}^2\right)^{1/2}$, with $r_{i,t}$ denoting the i-th daily return of the S&P 500 index in month t and n_t being the number of trading days in month t. The GDP is the annualized quarterly growth rate of the United States. Thus, it is only observed on a quarterly basis, which we extrapolated backwards to obtain monthly observations. We will refer to all these variables as covariates during model development. They were selected because most of these covariates have been found significant and used in related studies, see for example Das et al. (2007), Duffie et al. (2009), Giesecke et al. (2011), Agosto et al. (2016) and Azizpour et al. (2018). Figure 2 shows the cross-correlation matrix of the default counts and covariates lagged by one month relative to the default counts. The last row in Figure 2 shows the cross-correlation between the default counts and the lagged covariates, which indicates the predictive power of several of the covariates. Note that several of the covariates also display high cross-correlation between each other.

Figure 3 (top) displays a three-year rolling window estimate of lag-one autocorrelation of the default counts. The rolling window estimate is highly variable throughout the sample period and is typically lower than the full sample lag-one autocorrelation, except for the maximum around the financial crisis 2007–2008. Similarly, the bottom panel of Figure 3 displays a three-year rolling window estimate of the cross-correlation between defaults and realized volatility of the S&P 500 (lagged by one), which is highly variable throughout the sample period as well, again with a maximum around the financial crisis. These variations suggest that past defaults and covariates impact the default counts in a time-varying manner.

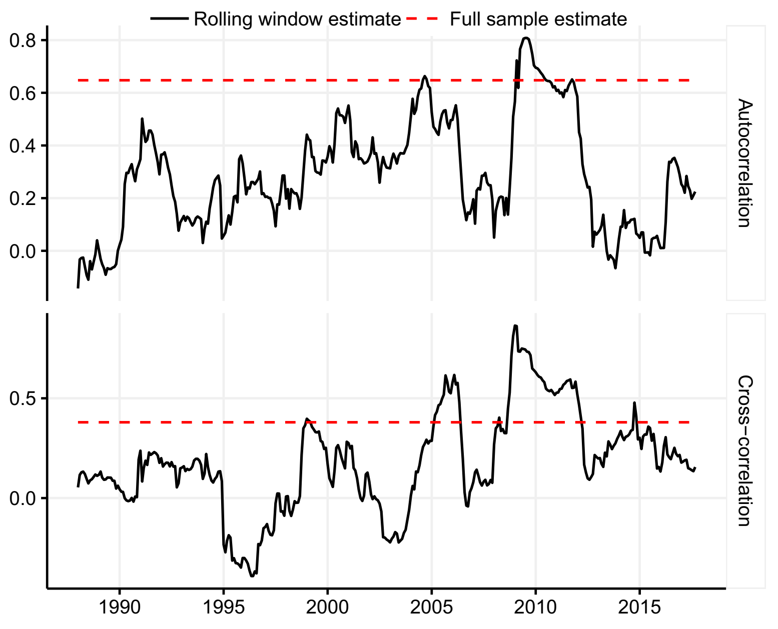

FIGURE 3 Rolling window estimates of correlation
Notes: The top panel displays the lag-one autocorrelation of the monthly number of default, and the bottom panel shows the cross-correlation between these counts and realized volatility on S&P 500. Both estimates are based on windows of size 36 (three years). The red dashed line represents the corresponding full sample (1985–2017) estimate.

3 | THE REGIME-SWITCHING INGARCHX MODEL

3.1 | Model formulation

Let Y_t be the monthly number of defaults and let $X_t = (X_{t,1}, \dots, X_{t,r})^t$ denote an r-dimensional time-varying covariate vector consisting of the macroeconomic and financial covariates described in Section 2. We model Y_t as a conditional Poisson distribution where the intensity λ_t at time t depends on both past default counts and past covariates. To capture possible regime changes in this dependence structure, we introduce an unobserved first-order Markov process S_t taking discrete values $1, \dots, m$. Let $\Gamma = \{\gamma_{ij}\}$ denote the $m \times m$ transition probability matrix of $\{S_t\}$, where the terms $\{\gamma_{ij}\}$ represent the probability of moving from the i^{th} regime at time $t-1$ to regime j at time t, where $i, j = 1, \dots, m$. We assume that the Markov chain is independent of previous counts and covariates. Suppose that $\mathcal{F}_{t-1}$ is the information set $\{Y_{t-1}, \lambda_{t-1}, X_{t-1}, S_t\}$. The model we consider is then given by

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = a_{S_t}\lambda_{t-1} + b_{S_t}Y_{t-1} + \exp\left(\alpha_{S_t} + \beta^t_{S_t}X_{t-1}\right), \quad t \geq 1. \tag{1}$$

The above specification implies m regime-specific sets of INGARCHX parameters $\{a_k, b_k, \alpha_k, \beta_k\}$, i.e. $k = 1, \dots, m$. For regime k, the coefficients a_k and b_k represent the effect of the intensity λ_{t-1} in the previous month and the number of defaults in the previous month Y_{t-1}, respectively. The last term captures the contribution of the covariates, which enter the intensity through a log-link function. In this function, the parameter α_k and the parameter vector $\beta_k = (\beta_{k,1}, \dots, \beta_{k,r})^t$

correspond to a scaling factor and covariate effects, respectively. This form ensures a positive contribution to the intensity when the covariates take negative values. For the same reason, we assume that all parameters a_k and b_k are greater or equal to zero. Moreover, in line with Agosto et al. (2016), see in particular Lemma 2 p. 646 allowing for non-Lipschitz functions of covariates, we suppose that the INGARCHX process is stationary within each regime such that

$$a_k + b_k < 1, \quad E\left(\exp\left(\alpha_k + \beta_k^t X_t\right)\right) < \infty, \quad k = 1, \dots, m. \tag{2}$$

Tests for stationarity (see Table 1) suggest that stationarity can be assumed for all covariates considered, perhaps with the exception of GDP. The same statement holds for the exponentially transformed covariates. Note that while the model defined by Equation (1) certainly belongs to the INGARCHX family for $m = 1$, the parametrization differs slightly from the one employed in Agosto et al. (2016). They avoid using a log-link function for implementing covariates, albeit entertaining the idea in a simulation study. Instead, they split covariates taking negative values into a positive and (absolute value of) a negative part. As most of the covariates we consider exhibit both negative and positive values, such an approach would almost double the number of considered covariates. Moreover, such a procedure could mask possible regimes driven by the covariates. An alternative approach would be to examine a RS extension of the log-linear INGARCH model of Fokianos and Tjøstheim (2011). However, the linear model has significant advantages in terms of interpretation since it allows for an additive decomposition of the intensity into an overall contagion and a macroeconomic risk component, as we will see below. Lastly, note that there is no additive intercept in the linear model part scale. The reason for this is that we experienced intrinsic aliasing for regimes with no or small covariate effects; multiple parameter sets of such an intercept and the parameter α_{S_t} implied models that are indistinguishable from each other. Even when fitting a simple INGARCHX model with additional additive intercept, the corresponding parameter estimate of such intercept equals practically zero. Moreover, we also attempted to replace the parameter α_{S_t} in the RS-INGARCHX model with an additive intercept in Equation (1). This did not lead to satisfactory results either, because the additive intercepts still converged towards zero, and the resulting models performed unsatisfactorily.

3.2 | Contagion and macroeconomic risk components

The joint impact of all past defaults and covariates propagates to future values of λ_t due to the feedback mechanism of the term $a_{S_t}\lambda_{t-1}$ in Equation (1). To illustrate this, assume that Y_0, X_0, and λ_0 are known quantities for simplicity. Furthermore, to ease notation, let $Y^{(t)}$ and $S^{(t)}$ represent the vector of observations $(Y_0, \dots, Y_t)'$ and the vector of hidden regimes $(S_1, \dots, S_t)'$, respectively. Similarly, let $X^{(t)}$ denote the $t+1$ times r matrix with row vectors $X_0, \dots, X_t$. By repeated substitution of Equation (1) we can rewrite the intensity in the form

$$\lambda_t = \lambda_0 \prod_{k=1}^{t} a_{S_k} + C\left(Y^{(t-1)}, S^{(t)}\right) + R\left(X^{(t-1)}, S^{(t)}\right), \tag{3}$$

where the first term of Equation (3) will approach zero with increasing t (since $0 \le a_k < 1$, $k = 1, \dots, m$). The second term of Equation (3) is given by

$$C\left(Y^{(t-1)}, S^{(t)}\right) = \sum_{i=0}^{t-1} b_{S_{t-i}} Y_{t-i-1} \prod_{k=0}^{i-1} a_{S_{t-k}}, \tag{4}$$

and represents the influence of all past defaults $Y_0, \dots, Y_{t-1}$ on the intensity at time t. Hence, this term can be interpreted as the overall contribution of the contagion effect as is of the form of a weighted sum $\sum_{i=0}^{t-1} w_i b_{S_{t-i}} Y_{t-i-1}$ with $w_i = \prod_{k=0}^{i-1} a_{S_{t-k}}$. The contribution of a past observation Y_{t-i-1} to the intensity at time t is associated with an effect $b_{S_{t-i}}$ that fades away with time (since $0 \leq w_{t-1} \leq w_{t-2} < \cdots < w_1 < w_0 = 1$). It is noteworthy that the rate at which this effect fades away is determined by both the current and past values of a_{S_t}. The third term of Equation (3) is given by

$$R\left(X^{(t-1)}, S^{(t)}\right) = \sum_{i=0}^{t-1} \exp\left(\alpha_{S_{t-i}} + \beta_{S_{t-i}}^{t} X_{t-i-1}\right) \prod_{k=0}^{i-1} a_{S_{t-k}} \tag{5}$$

and models the impact of all past covariates $X_0, \dots, X_{t-1}$ on the intensity at time t. This term represents the overall contribution of macroeconomic or financial risk and takes the form of a weighted sum $\sum_{i=0}^{t-1} w_i \exp\left(\alpha_{S_{t-i}} + X_{t-i-1}\beta_{S_{t-i}}\right)$ with the same weights introduced for the second term. Similar to Y_{t-i-1}, the contribution of X_{t-i-1} (via the log-link function) fades away with time at a rate determined by current and past values of a_{S_t}.

3.3 | Implications of regime-switching

Contagion can broadly be defined as a situation in which past defaults affect current defaults in some way, even when correcting for covariates. However, as pointed out by Agosto et al. (2016), this definition is specific to each model. In our case, contagion is explicitly represented by the term $C(Y^{(t-1)}, S^{(t)})$, which is not only driven by past defaults but also by the underlying Markov chain S_t. This implies that the effect of past defaults on the intensity at time t is time-varying. To illustrate this, consider as an example that $b_k = 0$ and $a_k \neq 0$ for some $k \in (1, \dots, m)$. Moreover, we assume that $S_t = k$ over some period $t \in (t_a, t]$. Regime k then represents a situation without contagion in the sense that defaults Y_t occurring during this period are independent of past defaults $Y_{t-1}, Y_{t-2}, \dots, Y_{t_a}$, conditional on the covariates $X_{t-1}, X_{t-2}, \dots, X_{t_a}$ and the intensity before entering the regime period λ_{t_a}. For further illustration, suppose $b_k > b_h$ in conjunction with $a_k \geq a_h$ for some other regime h. The overall contribution of contagion in regime k is then stronger than in regime h. Note that in the extreme case, that is when $b_1 = b_2 = \cdots = b_m = 0$, current and past defaults are independent conditional on all past covariates.

The impact of past covariates on current defaults is addressed by the term $R(X^{(t-1)}, S^{(t)})$, which is also driven by the underlying Markov chain S_t. Thus, we may observe covariates that have a larger impact on the intensity in one regime, while these effects remain small or negligible in another regime. For example, assume that $\beta_k = 0$ for some $k \in \{1, \dots, m\}$. If the process then sojourns in regime k for a long duration, the contribution of the macroeconomic variables to $R(X^{(t-1)}, S^{(t)})$ becomes negligible. Consequently, the terms involving α_k dominate $R(X^{(t-1)}, S^{(t)})$, picking up effects that are not captured by exogenous covariates. Nevertheless, those covariates observed before switching to regime k still possess an impact on $R(X^{(t-1)}, S^{(t)})$ at the beginning of such a sojourn. Empirical evidence for time-varying covariate effects are observed in Agosto et al. (2016), who reported rolling-window estimates. These effects changed substantially over the 20-year forecasting period, with no effects present at all during some interludes within this span.

In a case where the estimated parameters are different across regimes and a specific regime is persistent during periods of default clustering, one may conclude that the corresponding INGARCHX model under this regime provides a good description of the default clustering channels. Moreover, bias may be introduced from omitting relevant covariates (as pointed out in

Section 2.2. of Agosto et al., 2016), or from averaging parameter estimates by fitting a model with fewer states than would be appropriate. In a broader sense, regime-switching can also be interpreted as a mixed effect model, with time-varying random effects driven by the latent state process (Bartolucci & Farcomeni, 2009). Hence, the RS-INGARCHX model can account for certain types of unobserved heterogeneity. In our case, this concerns heterogeneity resulting from different parameter configurations during varying observation periods, but to a limited extent also omitted covariates through the time-varying term α_{S_t}.

3.4 | Estimation

Let θ denote the vector of parameters in the model given by

$$\theta = \left(\{a_k, b_k, \alpha_k, \beta_{1,k}, \ \dots \ , \beta_{r,k}\}_{k=1}^{m}, \{\gamma_{ij}\}_{i,j=1}^{m} \right),$$

where the terms $\{\gamma_{ij}\}$ represent the probability of moving from the i^{th} regime at time $t-1$ to regime j at time t. We estimate θ by a quasi-maximum likelihood procedure. Appendix B contains a detailed description, and we present the general concepts. As shown by Equation (3), λ_t depends on the entire regime path of S_t. For a sample of size T, evaluation of the likelihood involves summation over all m^T possible regime paths, which quickly becomes infeasible as the sample size increases. To avoid this problem, we use an adaptation of the EHG algorithm described in Chen and Tsay (2011), which has a computational cost similar to the recursive scheme proposed by Hamilton (1989). The EHG algorithm is implemented using the free and open-source R (R Development Core Team, 2020) package Template Model Builder (TMB, Kristensen et al., 2016), which provides a high-speed evaluation of the resulting quasi log-likelihood along with its gradient and Hessian matrix. In combination with TMB, the EHG algorithm renders feasible parameter estimation even for models with many regimes. It also provides approximate standard deviations of the parameters via the delta method. A simulation study suggests that the EHG algorithm generates consistent and approximately normal estimates of the parameters (see Appendix C). Following the approach of Kim (1994), we also provide a backward version of the EHG algorithm for computing the posterior marginal distribution of S_t given a comprehensive information set in Section 3.5. These smoothing probabilities provide useful information about if and when regime switches occur.

3.5 | Inference about the states

Several quantities involved in the EHG algorithm base on the information set Ω_τ, $\tau = \{1, \ \dots \ , T\}$, which is a reduced version of $\mathcal{F}_\tau$ involving conditional expectations of λ_τ and reduced state information (see Equation (B.3) in Appendix B). Given this information set, inference about the state (regime) S_t at any time t may be carried out via probabilities of the form $P(S_t = j \,|\, \Omega_\tau), j = 1, \ \dots \ , m$. Another core component of the EHG algorithm is the process S_t^*, which corresponds to a Markov chain tracing the m^2 possible two successive state visits of S_t (see Equation (B.2) in Appendix B). Using analogous probabilities for state inference for the process S_t^*, the quantity $P(S_t = j \,|\, \Omega_\tau)$ can be computed by

$$P(S_t = j \,|\, \Omega_\tau) = \sum_{i=1}^{m^2} P(S_t = j, S_t^* = i \,|\, \Omega_\tau) = \sum_{i=1}^{m^2} P(S_t^* = i \,|\, \Omega_\tau) \mathbb{1} \left[S_t(S_t^* = i) = j \right], \tag{6}$$

where $\mathbb{1}[\cdot]$ corresponds to the indicator function, and $S_t(S_t^* = i)$ equals the value of S_t given that S_t^* is in regime i. Thus, the filter probabilities $P(S_t = j \mid \Omega_t)$ and one-step-ahead probabilities $P(S_{t+1} = j \mid \Omega_t), j = 1, \ldots, m$, can be computed directly by Equation (6) once the corresponding probabilities for the process S_t^* have been obtained. This part is as well described in Appendix B, see in particular Equation (B.8) and (B.9), respectively.

Furthermore, the smoothing probabilities $P(S_t \mid \Omega_T), j = 1, \ldots, m$, can be derived. These represent the inference about S_t given the information set Ω_T, and are of particular interest when analysing data in in-sample settings. Similar to the filter and one-step-ahead probabilities, the smoothing probabilities can be computed by Equation (6), provided that the corresponding smoothing probabilities for S_t^* are available. For this purpose, we follow the approach of Kim (1994). First, the Markov property of S_t^* yields the identity

$$P(S_t^* = i \mid S_{t+1}^* = j, \Omega_T) = P(S_t^* = i \mid S_{t+1}^* = j, \Omega_t) = \frac{\gamma_{ij}^* P(S_t^* = i \mid \Omega_t)}{P(S_{t+1}^* = j \mid \Omega_t)},$$

where γ_{ij}^* corresponds to the transition probabilities of S_t^* analogously to γ_{ij}. Second, the smoothing probabilities for S_t^* can be represented as

$$\begin{aligned} P(S_t^* = i \mid \Omega_T) &= \sum_{j=1}^{m^2} P(S_{t+1}^* = j \mid \Omega_T) P(S_t^* = i \mid S_{t+1}^* = j, \Omega_T) \\ &= P(S_t^* = i \mid \Omega_t) \sum_{j=1}^{m^2} \frac{\gamma_{ij}^* P(S_{t+1}^* = j \mid \Omega_T)}{P(S_{t+1}^* = j \mid \Omega_t)} \end{aligned} \tag{7}$$

for $i, j = 1, \ldots m^2$. Third, using the already obtained filter and smoothing probabilities, we can iterate backward through Equation (7). For this part, the filter (smoothing) probabilities $P(S_T^* = j \mid \Omega_T), j = 1, \ldots, m^2$ serve as initial values. Hence, this recursive procedure permits to calculate the smoothing probabilities for S_t, $t = T - 1, \ldots, 1$ via Equation (6). Estimates of the filter, one-step-ahead, and smoothing probabilities result directly from replacing θ with the quasi-maximum likelihood estimate (QMLE) $\hat{\theta}$ in Ω_τ.

3.6 | Prediction

The one-step-ahead prediction $\hat{\lambda}_{T+1}$ for Y_{T+1} for the single-regime model follows directly from iterating through Equation (1) given the parameter estimates obtained via ordinary QMLE (Agosto et al., 2016). For the regime-switching models, a natural one-step-ahead prediction $\hat{\lambda}_{T+1}$ for Y_{T+1} is given by the expectation of Y_{T+1} conditional on the information set Ω_T. Consistent with the Poisson assumption follows

$$\begin{aligned} \hat{\lambda}_{T+1} = E(y_{T+1} \mid \Omega_T) &= \sum_{i=1}^{m^2} E\left(y_{T+1} \mid S_{T+1}^* = i, \Omega_T\right) P(S_{T+1}^* = i \mid \Omega_T) \\ &= \sum_{i=1}^{m^2} \hat{\lambda}_{T+1 \mid S_{T+1}^* = i, \Omega_T} P(S_{T+1}^* = i \mid \Omega_T), \end{aligned} \tag{8}$$

where the quantities $P(S^*_{T+1} = i \mid \Omega_T)$ and $\hat{\lambda}_{T+1 \mid S^*_{T+1}=i,\Omega_T}$, $i = 1, \dots, m^2$, are available from the T^{th} and $T+1^{th}$ recursion of the EHG algorithm, respectively. Note that these quantities also characterize the predictive distribution of Y_{T+1} via Equation (B.7) in Appendix B:

$$P(Y_{T+1} = y \mid \Omega_T) = \sum_{i=1}^{m^2} \frac{\left(\hat{\lambda}_{T+1 \mid S^*_{T+1}=i,\Omega_T}\right)^y \exp\left(-\hat{\lambda}_{T+1 \mid S^*_{T+1}=i,\Omega_T}\right)}{y!} P(S^*_{T+1} = i \mid \Omega_T). \tag{9}$$

Provided that we possess (or predict) covariate information up to time $T+k-1$, k-step-ahead predictions $\hat{\lambda}_{T+k}$ for Y_{T+k} can also be obtained. For achieving this, the EHG algorithm needs to be executed up to time $T+k$ while iteratively replacing the unobserved observations $Y_{T+1}, \dots, Y_{T+k-1}$ in $\Omega_{T+1}, \dots, \Omega_{T+k-1}$ by their respective one-step-ahead predictions $\hat{\lambda}_{T+1}, \dots, \hat{\lambda}_{T+k-1}$. In practice, θ is replaced by the QMLE $\hat{\theta}$ based on $y_1, \dots, y_T$ in $\Omega_{T+1}, \dots, \Omega_{T+k-1}$. Similarly, if covariates are not observed beyond T, these also need to be replaced by predicted values.

In an in-sample situation where $y_1, \dots, y_T$ have been observed, the in-sample predictions of $y_1, \dots, y_T$ provide valuable information about the model fit. Similar as in Equation (8), the in-sample predicted values are given by

$$\hat{\lambda}_t = \sum_{i=1}^{m^2} \hat{\lambda}_{t \mid S^*_t=i,\Omega_{t-1}} P(S^*_t = i \mid \Omega_T), \tag{10}$$

where $\hat{\lambda}_{t \mid S^*_t=i,\Omega_{t-1}}$, $i = 1, \dots, m^2$ is available from the t^{th} recursion of the EHG algorithm and $P(S^*_t = i \mid \Omega_T)$ corresponds to the smoothing probabilities given by Equation (7). Note that compared to Equation (8) we are using the smoothing probabilities rather than the one-step-ahead probabilities since these are directly available in a post-processing situation. Moreover, in analogy to Equation (9), the in-sample predictive distribution relying on smoothing probabilities is given by

$$P(Y_t = y \mid \Omega_T) = \sum_{i=1}^{m^2} \frac{\left(\hat{\lambda}_{t \mid S^*_t=i,\Omega_{t-1}}\right)^y \exp\left(-\hat{\lambda}_{t \mid S^*_t=i,\Omega_{t-1}}\right)}{y!} P(S^*_t = i \mid \Omega_T). \tag{11}$$

3.7 | Estimating the contribution of contagion and macroeconomic risk

To investigate the contribution of contagion and macroeconomic risk, we decompose the fitted values into quantities corresponding to the macroeconomic risk and contagion components in Equation (3). For the single-regime model (i.e. $m = 1$), the fitted values can directly be computed by Equation (3) for given parameter estimates. However, for $m \geq 2$ the necessary computations depend on the entire regime-path of S_t. The EHG algorithm is based on recursively deriving the evolution of the conditional means (distributions), and there is no simple way for decomposing the fitted values Equation (10). Following the idea of Gray (1996), we instead rely on iteratively replacing the intensity with its (unconditional) expectation. For $m = 2$, this involves iterating the following set of equations

$$\lambda_{t \mid 1} := a_1 \lambda^*_{t-1} + b_1 Y_{t-1} + \exp\left(\alpha_1 + \beta_1^t X_{t-1}\right)$$

$$\lambda_{t\,|\,2} := a_2\lambda^*_{t-1} + b_2 Y_{t-1} + \exp\left(\alpha_2 + \beta_2^t X_{t-1}\right)$$
$$\lambda^*_t = \lambda_{t\,|\,1} P(S_t = 1) + \lambda_{t\,|\,2} P(S_t = 2) \tag{12}$$

Once parameter estimates $(\hat{a}_i, \hat{b}_i, \hat{\alpha}_i, \hat{\beta}_i)$ and the smoothing-probabilities $P(S_t = i\,|\,\Omega_T)$ are obtained from the EHG algorithm and Equation (7), respectively, the above forwarding scheme provides the fitted values

$$\hat{\lambda}^*_t = a(t)\hat{\lambda}^*_{t-1} + b(t)Y_{t-1} + g(t). \tag{13}$$

Here, $a(t)$, $b(t)$, and $g(t)$ are the estimated temporal trajectories of a_{S_t}, b_{S_t}, and the link function component, respectively, given by

$$a(t) = \sum_{i=1}^{m} \hat{a}_i P(S_t = i\,|\,\Omega_T),$$
$$b(t) = \sum_{i=1}^{m} \hat{b}_i P(S_t = i\,|\,\Omega_T),$$
$$g(t) = \sum_{i=1}^{m} \exp\left(\hat{\alpha}_i + \hat{\beta}_i^t X_{t-1}\right) P(S_t = i\,|\,\Omega_T).$$

Similar to the decomposition carried out for λ_t in Equation (3), we can rewrite $\hat{\lambda}^*_t$ as

$$\hat{\lambda}^*_t = \hat{\lambda}^*_0 \prod_{k=1}^{t} a(k) + \hat{C}(t; Y^{(t-1)}) + \hat{R}(t; X^{(t-1)}), \tag{14}$$

by repeated substitution of Equation (13). Note that the estimators $\hat{C}(t; Y^{(t-1)})$ and $\hat{R}(t; X^{(t-1)})$ only depend on the previous observations and covariates, respectively:

$$\hat{C}\left(t; Y^{(t-1)}\right) = \sum_{i=0}^{t-1} b(t-i) Y_{t-i-1} \prod_{k=0}^{i-1} a(t-k) \tag{15}$$
$$\hat{R}\left(t; X^{(t-1)}\right) = \sum_{i=0}^{t-1} g(t-i) \prod_{k=0}^{i-1} a(t-k) \tag{16}$$

Our experience is that $\hat{\lambda}^*_t$ is close but inferior to $\hat{\lambda}_t$ given by Equation (10) in terms of prediction performance measured by the MSE. However, it permits to assess the contribution of the contagion component $\hat{C}(t; X^{(t-1)})$ and macroeconomic component $\hat{R}(t; X^{(t-1)})$, respectively, to each prediction of λ_t.

4 | ANALYSIS

4.1 | Considered models

As basis for our empirical analysis and for forming a benchmark for the later comparison to models with several regimes, we start by considering the single-regime INGARCHX model having the default intensity

TABLE 2 Estimation results of different single-regime models, estimated standard errors in parentheses. Estimated parameters significantly different from zero, on the 95% level, are marked with an asterisk

Parameter	INGARCH	indpro	ppifgsup	ppieng	unrate	baa	SP500ret	GDP	SP500vol
a	0.652*	0.751*	0.745*	0.750*	0.546*	0.778*	0.577*	0.641*	0.683*
	(0.080)	(0.045)	(0.045)	(0.041)	(0.131)	(0.037)	(0.100)	(0.087)	(0.065)
b	0.284*	0.194*	0.208*	0.200*	0.330*	0.198*	0.298*	0.262*	0.244*
	(0.059)	(0.036)	(0.037)	(0.034)	(0.076)	(0.033)	(0.061)	(0.058)	(0.048)
α	−1.743*	−1.944*	−2.157*	−2.374*	−1.097*	−3.007*	−0.968*	−0.927*	−2.140*
	(0.421)	(0.326)	(0.386)	(0.379)	(0.497)	(0.985)	(0.376)	(0.346)	(0.400)
β	–	−0.065*	−0.096*	−0.016*	0.263*	1.332*	−0.056*	−0.202*	0.494*
		(0.008)	(0.015)	(0.002)	(0.095)	(0.584)	(0.011)	(0.036)	(0.092)
log-likelihood	−733.54	−725.24	−726.78	−724.96	−732.06	−729.88	−728.89	−726.69	−727.41

$$\lambda_t = a\lambda_{t-1} + bY_{t-1} + \exp\left(\alpha + \beta' X_{t-1}\right). \tag{17}$$

This corresponds to the model specified by Equation (1) when $m = 1$, and is comparable to the model introduced in Agosto et al. (2016), as described earlier. Table 2 reports the estimation results for the model without covariates (denoted INGARCH) and all univariate INGARCHX model specifications. The univariate model specification with SP500vol as covariate (rightmost column) serves as a benchmark model for further comparisons due to various performance aspects, which are illustrated later.

Extending the single-regime model to a two-regime INGARCHX model leads to the default intensity given by

$$\lambda_t = \begin{cases} a_1\lambda_{t-1} + b_1 Y_{t-1} + \exp\left(\alpha_1 + \beta_1^t X_{t-1}\right) & \text{if } S_t = 1 \\ a_2\lambda_{t-1} + b_2 Y_{t-1} + \exp\left(\alpha_2 + \beta_2^t X_{t-1}\right) & \text{if } S_t = 2 \end{cases}. \tag{18}$$

We report the estimation results for this model with no covariates (denoted RS-INGARCH), and for all univariate RS-INGARCHX specifications in Table 3. The best performing model (rightmost column) in terms of various aspects that will be presented in the following sections contains SP500vol as the single covariate.

The selection of the correct number of regimes (m) is in general a challenging task. Several formal tests for different model specifications have already been introduced over the past decades, see e.g. Hansen (1992), Bartolucci (2006) and Holzmann and Schwaiger (2016) On the other hand, Zucchini et al. (2016) suggest relying on the Akaike information criteria (AIC) and Bayesian information criteria (BIC) as a feasible approach, see also Bacci et al. (2014). We follow an approach similar to the approach of Pohle et al. (2017). That is, we do not only rely on AIC and BIC but also consider other practical aspects of the model such as mean square error, forecasting performance and interpretability. As we will see in the following, a model with two regimes brings along various advantages over a one-regime model and that a further extension to a three-regime model is difficult to justify.

TABLE 3 Estimation results of different two-regime models. Estimated parameters significantly different from zero, on the 95% level, are marked with an asterisk. Note that we also report the stationary distribution of S_t, δ_1 and δ_2 (see Appendix B.2 for computational details). The significance of the stationary distribution and transition probabilities are relative to 0.5

Parameter	RS-INGARCH	indpro	ppifgsup	ppieng	unrate	baa	SP500ret	GDP	SP500vol
a_1	0.958*	0.913*	0.903*	0.894*	0.909*	0.919*	0.898*	0.883*	0.907*
	(0.008)	(0.025)	(0.039)	(0.027)	(0.030)	(0.020)	(0.029)	(0.034)	(0.027)
a_2	0.411*	0.955*	0.946*	0.902*	0.907*	0.982*	0.812*	0.923*	0.941*
	(0.154)	(0.087)	(0.117)	(0.117)	(0.098)	(0.064)	(0.164)	(0.099)	(0.102)
b_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
b_2	0.461*	0.076	0.083	0.112	0.110	0.035	0.175	0.102	0.073
	(0.103)	(0.072)	(0.102)	(0.091)	(0.084)	(0.055)	(0.115)	(0.084)	(0.083)
α_1	−2.992*	−2.538*	−2.823*	−2.913*	−2.563*	−2.571*	−2.698*	−1.466*	−2.878*
	(0.283)	(0.387)	(0.453)	(0.592)	(0.636)	(0.316)	(0.608)	(0.319)	(0.548)
α_2	−0.371	−2.546*	−2.235*	−1.730*	−1.861*	−2.878*	−1.042	−1.312*	−2.920*
	(0.568)	(1.195)	(1.073)	(0.614)	(0.912)	(0.763)	(0.758)	(0.656)	(0.941)
γ_{11}	0.978*	0.894*	0.891*	0.861*	0.900*	0.918*	0.865*	0.889*	0.879*
	(0.007)	(0.052)	(0.061)	(0.047)	(0.037)	(0.027)	(0.057)	(0.059)	(0.055)
γ_{21}	0.044*	0.169*	0.186*	0.239*	0.164*	0.139*	0.196*	0.196*	0.173*
	(0.023)	(0.068)	(0.064)	(0.071)	(0.061)	(0.045)	(0.073)	(0.067)	(0.064)
γ_{12}	0.022*	0.106*	0.109*	0.139*	0.100*	0.082*	0.135*	0.111 ∗	0.121 ∗
	(0.007)	(0.052)	(0.061)	(0.047)	(0.037)	(0.027)	(0.057)	(0.059)	(0.055)
γ_{22}	0.956*	0.831 ∗	0.814 ∗	0.761 ∗	0.836 ∗	0.861 ∗	0.804 ∗	0.804 ∗	0.827 ∗
	(0.023)	(0.068)	(0.064)	(0.071)	(0.061)	(0.045)	(0.073)	(0.067)	(0.064)
δ_1	0.666	0.614	0.630	0.633	0.620	0.628	0.593	0.639	0.589
	(0.126)	(0.078)	(0.092)	(0.073)	(0.077)	(0.076)	(0.083)	(0.091)	(0.087)
δ_2	0.334	0.386	0.370	0.367	0.380	0.372	0.407	0.361	0.411
	(0.126)	(0.078)	(0.092)	(0.073)	(0.077)	(0.076)	(0.083)	(0.091)	(0.087)
β_1	–	−0.037	0.091	0.023 ∗	−0.130	0.009	0.033	−0.246 ∗	0.265
		(0.040)	(0.072)	(0.006)	(0.648)	(0.189)	(0.057)	(0.041)	(0.261)
β_2	–	−0.071 ∗	−0.095 ∗	−0.013 ∗	0.690 ∗	1.437 ∗	−0.065 ∗	−0.235 ∗	0.809 ∗
		(0.026)	(0.033)	(0.003)	(0.347)	(0.335)	(0.027)	(0.088)	(0.204)
log-likelihood	−719.78	−713.66	−712.58	−708.55	−716.61	−711.35	−715.11	−713.55	−711.60

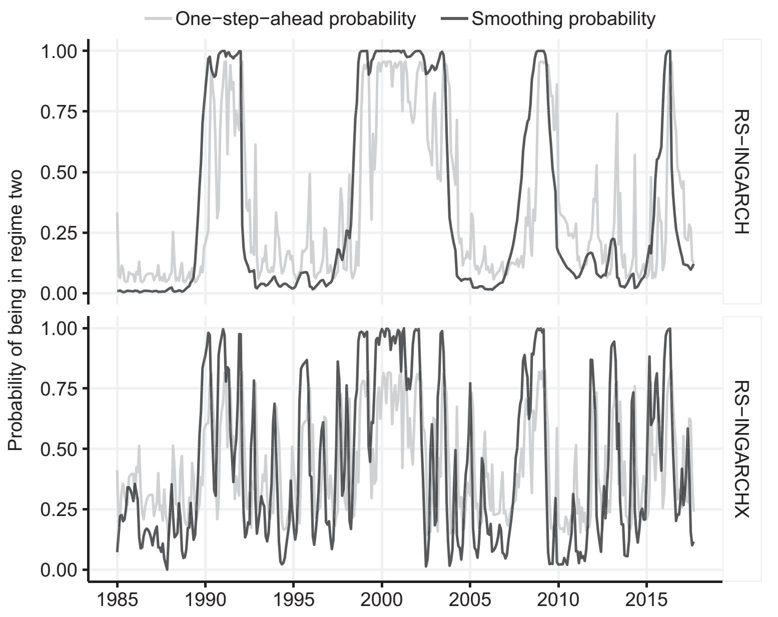

FIGURE 4 One-step-ahead and smoothing probabilities
Notes: The two panels show one-step-ahead and smoothing probabilities for the RS-INGARCH and RS-INGARCHX (SP500vol) model on the top and bottom, respectively. Recall that the one-step-ahead probabilities are based on the information available at time $t-1$ while the smoothing probabilities are based on the entire sample.

4.2 | Regimes and default clustering

As a first step for illustrating the implications of our model, we examine the occurrences of regime-switching. Figure 4 shows the probabilities of being in a regime using one-step-ahead and smoothing probabilities. The top panel is dedicated to the RS-INGARCH model, which exhibits highly persistent regimes. For this model, the second regime mainly corresponds to the recession periods, cfr. Figure 1, during which the clustering of defaults is most pronounced. Parts of 2016 also fall in the second regime, where a prominent peak of defaults is observed. This, however, is not the case for the period of the sovereign debt crisis of 2011, albeit the smoothing probabilities are estimated somewhat higher during this period compared to the, for example calm periods of 1993–1999 and 2004–2006. Analogously, the lower panel shows the corresponding probabilities for the RS-INGARCHX model with SP500vol as the covariate. This covariate inclusion results in a model with different dynamics inherent to the regimes. Most importantly, we infer a higher number of switches between the two regimes overall. More precisely, we observe occasional switches to the first regime during recession periods, although the second regime still mainly corresponds to these periods. Furthermore, some visits of the second regime also lie outside of periods of default clustering. This aspect needs to be kept in mind when discussing regimes across different models.

Summarizing, Figure 4 suggests that the parameters estimated for the second regime for both the RS-INGARCH and the RS-INGARCHX model in principle represent the dynamics during periods of default clustering. In contrast, the first regime of these models represents the dynamics in the periods without (notably) default clustering. Consequently, models ignoring the presence

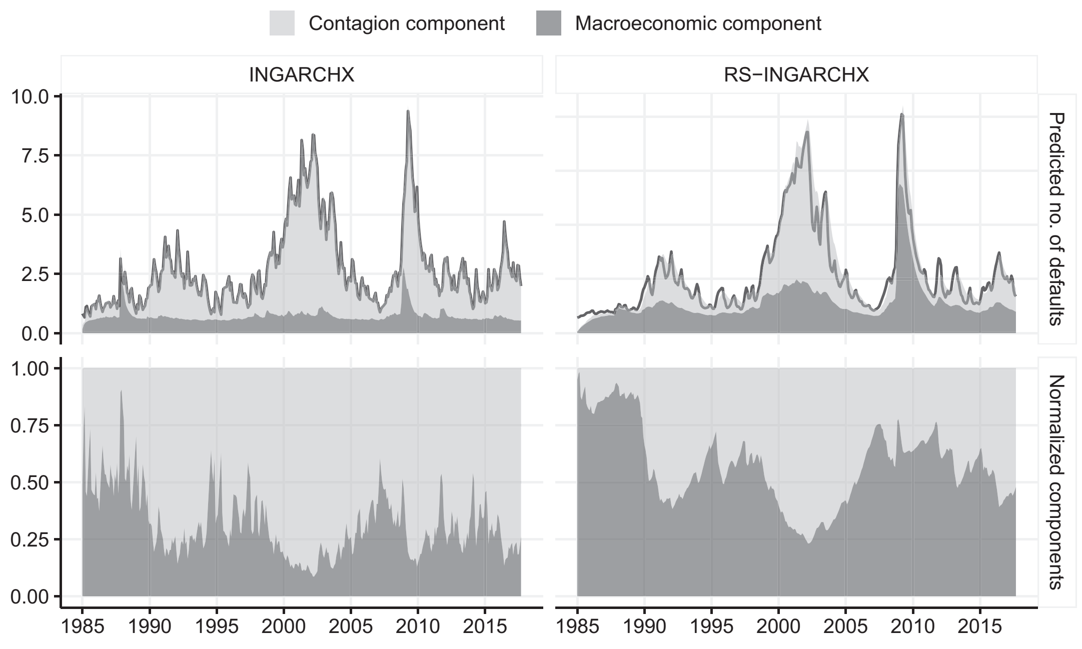

FIGURE 5 Comparison of sources of default clustering
Notes: The upper panel displays the estimated decomposition of predicted defaults into the contagion source and the macroeconomic source for the INGARCHX and RS-INGARCHX models with SP500vol as covariate. The lower panel shows a normalized decomposition. For the INGARCHX model, the estimates are plug-in estimators of Equations (4) and (5), while Equations (15) and (16) are used for estimating the RS-INGARCHX decomposition.

of any regime-switching may not be able to capture essential data features on various levels. Moreover, it should be kept in mind that the inclusion of covariates may affect regime dynamics. The subsequent sections illustrate these aspects in further detail. Last, it is noteworthy that the more dynamic regime-switching and correspondence of one regime to default clustering are present for most alternative RS-INGARCHX models with different covariates as well.

4.3 | Sources of clustering under regime-switching

The consideration of regime-switching provides a new view on the contribution of contagion and macroeconomic sources to default clustering. Figure 5 illustrates the impact of regime-switching on the contribution of these two sources. For the preferred INGARCHX model (left plots), the contagion component is the main driver, accounting for 75–85% of the intensity in the periods of default clustering. This is in line with the findings of Azizpour et al. (2018) (cfr. their Figure 9).

On the other hand, the decomposition of the two sources for the preferred RS-INGARCHX model paints a rather different picture. The impact of the same covariate is first higher, and second, more dynamic over the entire observation period. Furthermore, in the calm periods with few defaults (corresponding to regime one), the contagion source is small, and the default intensity is mainly driven by the macroeconomic source. In periods of default clustering (corresponding to the second regime), a further distinction is necessary. The contagion source dominated during the savings-and-loan crisis and the burst of the dot-com bubble. In contrast, the impact of the macroeconomic source remains on par with calm periods during the global financial crisis. During this latter period, the macroeconomic source accounts for 70%–75% compared to only 25%–40% in the course of the former two crises. The particular behaviour during the global financial crisis supports the presence of structural instabilities in the single-regime model. This has been commented by Agosto et al. (2016), who finds that macroeconomic and financial fundamentals mostly drove the financial crisis.

4.4 | Regime-specific dynamics

The findings presented in the previous two sections naturally lead to the question to which extent the two regimes—or more specifically, their dynamics—differ from each other and the one-regime model.

Investigation of the parameter estimates for the one-regime models reported in Table 2 shows that the effect of the covariates are in line with the results of previous studies, and all of them are significant in all univariate model specifications. In particular, the one-regime model confirms that the default intensity increases when the industrial production growth rate (indpro) falls, as shown, for example in Lando and Nielsen (2010). Furthermore, the default intensity rises in periods of GDP contraction, as described by Giesecke et al. (2011) and Azizpour et al. (2018). For the preferred single-regime model, the default intensity also increases with increasing realized volatility, a finding also reported by Agosto et al. (2016). Furthermore, compared to the INGARCH model, the b parameter decreases for all specifications where covariates are included (except for SP500ret). However, this parameter remains significantly positive, implying that contagion is still present (cfr. Section 3.3).

Several aspects of the estimated parameters of the regime-switching models (see Table 3) are interesting. First, we address regime one. Here, the estimated b_1 parameter lies on the zero boundary for all models, suggesting that this regime represents periods without contagion in the sense described in Section 3.3. The estimated covariate effect β_1 is not significant for the majority of models; GDP and pping are exceptions. Similarly, the feedback parameters a_1 and the scaling parameters α_1 are all significant. This causes the process to exhibit high dependence on past intensities, and observed contagion effects in this regime originate from previous visits of the second regime. Consequently, the macroeconomic sources drive default dynamics in this regime, despite their mostly non-significant coefficients.

Second, in regime two, the overall effect of both covariates and past defaults increases compared to regime one. More specifically, we observe non-zero estimates of the b_2 parameter in all cases, although the significance of the parameter is only evident for the RS-INGARCH model. Moreover, the covariate effects β_2 are significant in all model specifications. Both aspects result in a regime that is more dynamic compared to the first regime. The scaling parameter α_2, as well as the feedback parameter a_2, take values close to those in the first regime.

Note that the estimates of b_2 may seem slightly surprising at first, as they indicate little contagion effects for the second regime. From a statistical perspective, one could even argue that no significant contagion effects are present at all. However, one has to keep in mind that large values of a_2 in conjunction with non-zero values of b_2 lead to a considerable overall effect of past defaults. This explains the magnitude of contagion effects illustrated in the previous section.

Third, the sign and size of the covariate effects in regime two are mostly comparable to the single-regime models. However, in some cases, there are crucial differences. For instance, the estimated coefficient for the industrial production growth rate (indpro) in regime two is −0.071 and significantly different from zero. This corresponds to the single-regime estimate in Table 2, which equals −0.065, while the estimated coefficient in regime one is equal to −0.037 and not significantly different from zero. This observation also holds for the covariate SP500vol. Here we see a substantial increase of the estimated coefficient from 0.494 in the single-regime model to 0.809 in the second regime for the two-regime model. Notable exceptions exist, such as the models including ppieng and GDP.

Last, the diagonal elements of the transition probability matrix (γ_{11} and γ_{22}) are closest to one for the RS-INGARCH model. Although slightly lower, the corresponding probabilities for the RS-INGARCHX models indicate that all regimes are persistent. This is in line with the more dynamic regime-switching observed for the RS-INGARCHX model in Figure 4.

4.5 | In-sample performance

In the following, we report the in-sample performance for the different models fitted in Section 4.1. Table 4 shows the mean squared error of the Pearsons residuals $e_t = (y_t - \hat{\lambda}_t)^2/\hat{\lambda}_t$ given by $\mathrm{MSE} = \sum_{t=1}^{T} e_t^2/(T-p)$. Here p corresponds to the number of model parameters (see Kedem & Fokianos, 2005, Section 1.7). Moreover, we present the AIC and BIC.

Focusing on the MSE, the table clearly shows that all two-regime models possess a much lower MSE than the single-regime models. Furthermore, the MSE decreases from the models without covariates to the models with covariates. Note also that the two-regime model without covariates actually provides a better in-sample performance than the best single-regime model with covariates in terms of MSE. The overall best results are achieved by the three RS-INGARCHX models with ppifgsup, ppieng and SP500vol, respectively, as covariate. Looking at the AIC, the regime-switching models are all uniformly better than their corresponding single-regime counterparts. This finding inverses in almost all cases when considering BIC instead of AIC.

Summarizing, the three criteria MSE, AIC and BIC lead to varying conclusions, however, mostly supporting the two-regime models. One should keep in mind that the model

TABLE 4 In-sample performance, in terms of a standardized mean squared error (MSE) and the Akaike and Bayesian information criterion (AIC and BIC), of the different models

	MSE		AIC		BIC	
Covariate (s)	**INGARCHX**	**RS-INGARCHX**	**INGARCHX**	**RS-INGARCHX**	**INGARCHX**	**RS-INGARCHX**
–	1.2487	1.0465	1473.07	1455.56	1484.99	1487.35
indpro	1.2189	0.7789	1458.48	1447.33	1474.37	1487.07
ppifgsup	1.2239	0.7612	1461.55	1445.16	1477.45	1484.90
ppieng	1.2106	0.7648	1457.93	1437.10	1473.82	1476.84
unrate	1.2431	0.8052	1472.13	1453.22	1488.02	1492.96
baa	1.2516	0.7670	1467.76	1442.70	1483.66	1482.44
SP500ret	1.2160	0.8088	1465.78	1450.23	1481.68	1489.97
GDP	1.2299	0.7729	1461.38	1447.11	1477.28	1486.85
SP500vol	1.2045	0.7646	1462.82	1443.21	1478.72	1482.95

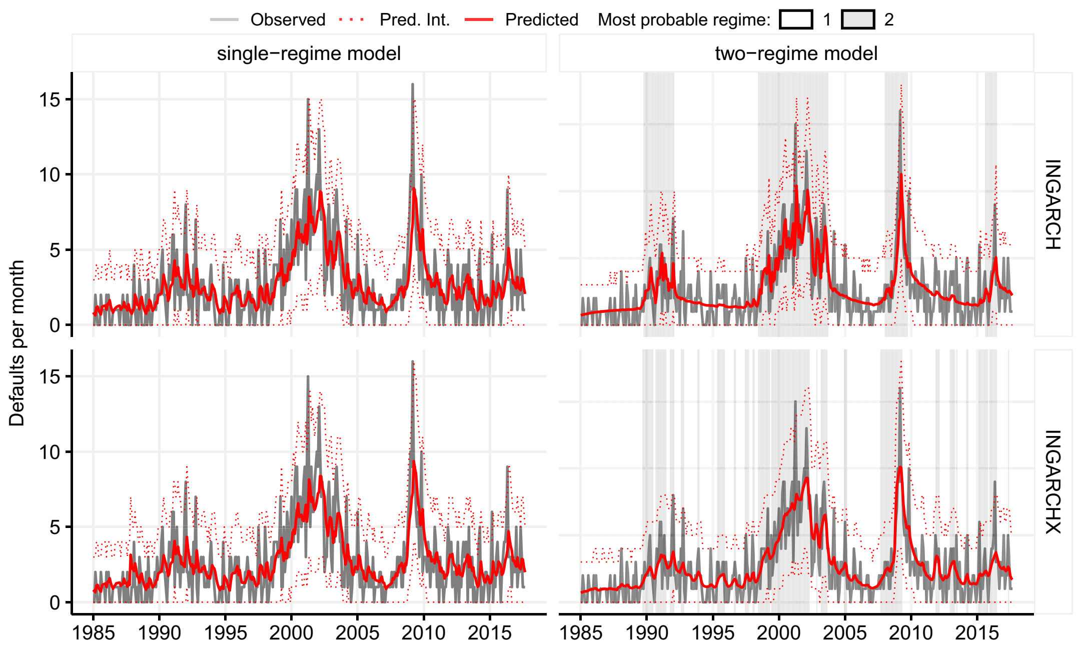

FIGURE 6 In-sample predictions

Notes: Observed number of defaults (grey line) plotted together with in-sample predictions (solid red line) and corresponding 95% prediction intervals (dotted red line). The figure displays models with no covariates (upper panel) and best models with covariates (lower panel) for the single-regime models (left side) and two-regime models (right side). Section 3.6 outlines details on the predicted values ($\hat{\lambda}_t$) as well as the in-sample prediction distribution which serves for constructing prediction intervals.

estimation was carried out via QMLE, which reduces the reliability of both AIC and BIC. Hence, we prefer to rely on the MSE as the main criterion. Taking into account the out-of-sample performance investigated in the following section as well, we prefer the two-regime model with SP500vol as covariate. The univariate counterpart is preferred among the single-regime models.

We visualize the in-sample fit for single- and two-regime models with and without the SP500vol covariate in Figure 6. All models capture the dynamics in the process reasonably well, but the regime-switching models (right panels) reproduce the variations in the process best. This becomes evident in particular during periods of economic crisis.

Last, the defaults contain 63 (16.0%) zero counts. In comparison, the expected number of zero counts predicted by the preferred INGARCHX and RS-INGARCHX model during the sample period is 54.26 (13.8%) and 67.39 (17.1%), respectively. For more details, see Appendix E.

4.6 | Out-of-sample performance and robustness

To assess the out-of-sample forecast accuracy of the different models, we start by estimating initial models with the data ranging up to T_0 = January 1, 1995, that is (Y_t, X_t), $t = 1, \dots, T_0$. Subsequently, these models forecast Y_{T_0+1} using the one-step-ahead prediction $\hat{\lambda}_{T_0+1}$ described in

Section 3.6. This procedure is then repeated for $t = T_0 + 1, \ldots, T - 1$ to obtain the predictions $\hat{\lambda}_{t+1}$. These predictions result from models fitted to updated data that are augmented successively by (Y_t, X_t). This imitates the situation of a forecaster that starts the predictions at time T_0 and updates the model and forecasts as more observations become available (Agosto et al., 2016; Stock & Watson, 1996).

Given these forecast paths for the different models, we evaluate the performance by means of two common forecasting measures. The first one is a standardized average mean squared forecasting error given by

$$\mathrm{MSFE}_t = \frac{1}{t - T_0} \sum_{s=T_0}^{t-1} \left(\frac{y_{s+1} - \hat{\lambda}_{s+1}}{\hat{\lambda}_{s+1}} \right)^2, \quad t = T_0 + 1, \ldots, T. \tag{19}$$

As the second measure serves the average logarithmic forecasting score (Amisano & Giacomini, 2007), defined by

$$\mathrm{FS}_t = \frac{1}{t - T_0} \sum_{s=T_0}^{t-1} \log \hat{P}(Y_{s+1} = y_{s+1}), \quad t = T_0 + 1, \ldots, T. \tag{20}$$

For the regime-switching models, $\hat{P}(Y_{s+1} = y_{s+1})$ corresponds to the one-step-ahead predictive probabilities (see Equation 9). For single-regime models, similar probabilities are obtained by evaluation of the Poisson density at y_{s+1} given $\hat{\lambda}_{s+1}$. Figure 7 display both these quantities, where the left part compares all single-regime models and their regime-switching counterparts. It is evident that regime-switching leads to a clear improvement in terms of both measures. The right panels focus on four selected models, including the RS-INGARCHX with SP500vol as the covariate. Again, regime-switching leads to an overall improvement, and our preferred model exhibits the best predictive performance. In particular, it shows a substantially better performance than the three other models during the two last periods of default clustering. For completeness, Table 5 displays MSFE_T and FS_T for all models considered. Again, the RS-INGARCHX model with SP500vol as covariate stands out as the preferred model.

In addition to the evaluation of forecast performance, we considered a series of alternative model specifications as supplementary robustness checks. First, we fitted models with three regimes with and without covariates. These models suggest the presence of a 'middle' regime, which can be interpreted as an intermediate step between a calm first regime and a dynamic second regime. While a transition to the additional regime may potentially serve as a warning before a crisis could occur, the main conclusions do not change. Moreover, the three-regime model was outperformed by the two-regime model in terms of AIC, BIC and forecasting, while only slightly improving the MSE. Another alternative model we fitted is motivated by the estimate of β_1 not being significantly different from zero (see Table 3). Hence it seemed natural to consider a model where β_1 is fixed to zero. Such a model is supported by AIC/BIC, and is slightly inferior to our preferred RS-INGARCHX model in terms of MSE. In short, interpretations from the reduced model are similar to those of the unrestricted alternative. Appendix F contains a detailed presentation of our findings. Second, we have evaluated models with combinations of several covariates. For these models, the main conclusions hold again, and we mostly identify two regimes with comparably similar economic interpretations. Nevertheless, multiple correlations between several covariates (see Figure 2) cause unreliable estimation results for several variable combinations, particularly when including more than three variables. This problem was addressed by first carrying out a

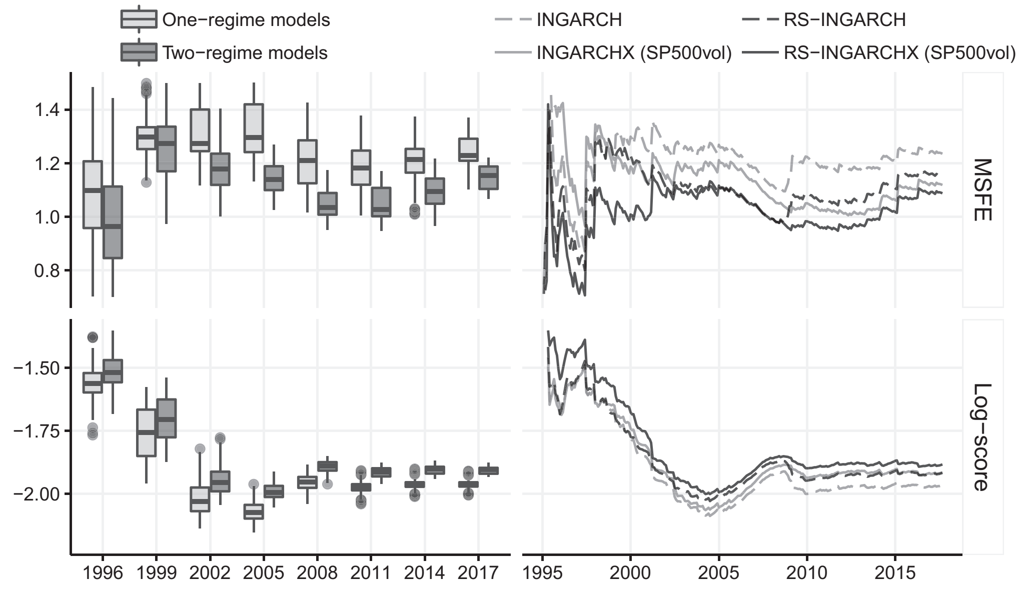

FIGURE 7 Out-of-sample forecast accuracy

Notes: The forecasting error and the forecasting score as a function of time is displayed in the top and bottom panels, respectively. The boxplots in the left panels depict these quantities in groups of single- and two-regime models over three-year windows. The right panels show the monthly values of the quantities for the one- and two-regime benchmark and the INGARCH and RS-INGARCH models.

TABLE 5 Out-of-sample performance, in terms of mean squared forecasting error and logarithmic score for monthly predictions for the complete forecasting period (1 January 1995 to 1 September 2017) for the fitted models

	MSFE		Log-score	
Covariate	**INGARCHX**	**RS-INGARCHX**	**INGARCHX**	**RS-INGARCHX**
-	1.235	1.151	−1.966	−1.917
indpro	1.287	1.178	−1.960	−1.911
ppifgsup	1.210	1.194	−1.952	−1.916
ppieng	1.342	1.194	−1.993	−1.921
unrate	1.286	1.155	−1.969	−1.897
baa	1.211	1.103	−1.955	−1.897
SP500ret	1.183	1.091	−1.948	−1.893
GDP	1.216	1.110	−1.952	−1.892
SP500vol	1.118	1.087	−1.914	−1.883

principal component analysis (PCA) on the covariates and subsequently using the two leading components as covariates in the model. In terms of model performance, the resulting model is slightly worse than our favoured model and with the same overall conclusions. A description is given in Appendix D. Third, we investigated the inclusion of higher lags of covariates in the autoregressive part. While on par with our preferred models in terms of MSE, the resulting alternative models were inferior in terms of AIC and BIC. Moreover, simultaneous inclusion of covariates of lag one and higher lags generally led to unstable estimation results. The outcomes from analogous extensions in the autoregressive part were similar. Last, we investigated the negative binomial distribution as an alternative for the conditional densities. The resulting models were, however, converging towards the limiting Poisson distribution in almost all cases.

5 | DISCUSSION

In this paper, we introduce a new class of regime-switching models for count time series, present a suitable estimation algorithm, and apply the model to default data from the United States. The fitted model suggests the presence of (at least) two regimes, corresponding to calm periods as well as periods of default clustering, respectively. Notably, all major financial crises belong to periods with defaults driven by the second regime.

More importantly, the regimes exhibit different dynamics. This sheds light on the role of contagion and macroeconomic variables and leads to an overall more dynamic description of the sources of default clustering. Compared to models without regime-switching, our approach indicates a reduced impact of contagion on default clustering and strengthens the impact of macroeconomic variables. In particular, our preferred model shows a dominant impact of the SP500vol covariate on default clustering during the financial crisis of 2007–08, which is not the case during previous crisis periods. Analogously to Sant'Anna (2017), knowledge about time-varying contagion and macroeconomic effects may also be beneficial for the correct specification of credit risk models, for example when managing portfolio credit risk at financial institutions.

The results of Azizpour et al. (2018) exhibit an interesting overlap with the results of this paper. Their data set extended to before 1985, which permitted an additional frailty component in their model. This component roughly corresponds to a regime-switching additive intercept in our model. In their analysis, the frailty component was estimated to be close to zero after around 1985 (less than 1 % contribution to the intensity), which could very well be why we were unable to include an additional additive intercept in our setup.

Several exciting aspects fall outside the scope of this paper. First, the data set we analyse only includes large firms (with more than 100 million USD in assets). We expect large companies to be interconnected with other firms and thus strongly affected by contagion risk. On the other hand, small, specialised firms may depend heavily on a few firms that they supply their products to and consequently be more exposed to contagion risk compared to larger firms with more customers. Likewise, macroeconomic conditions are possible of less relevance as long as the products of the small company are still in demand by other firms in its industry, or if some industries or geographic areas are less affected by economic and financial turmoil. However, we note that such effects could be captured using proper industry/firm-specific covariates. The findings of Azizpour et al. (2018), Hertzel et al. (2008), and Jorion and Zhang (2009) indicate that contagion can occur across industries, but a similar study taking into account the time dependence of the contagion and macroeconomic component does, to our knowledge, not exist in the literature.

Second, there might be an interaction between macroeconomic conditions and the risk of contagion. For example, there could be a feedback effect of past defaults into one or several of the covariates. This would require a more structural approach to capture, for instance, specifying a separate model for the covariates that allows dependence on past defaults. In a related direction, models where (combinations of) different covariates affect the conditional mean in different regimes may allow to capture more complex covariate effects. This would, however, require a significantly higher number of observations.

Third, recent results of Aknouche and Francq (2021) in a related setting without covariates suggest that the conditions formulated in Equation (2) may be relaxed. A detailed investigation of stationarity conditions for our model constitutes a possibility for future research.

Finally, we have assumed that the Markov chain is independent of past counts and past covariates. In theory, both past counts and covariates could be included additionally as covariates in the transition probability matrix Γ, see for example Banachewicz et al. (2008). This, however, would alter the interpretation of the model, see Bartolucci et al. (2012, p. 117).

ACKNOWLEDGEMENTS

The authors are grateful to the editor and two anonymous referees for their valuable comments that greatly improved the presentation of the paper.

DATA AVAILABILITY STATEMENT

The corporate default data that support the findings of this study are available from the UCLA-LoPucki Bankruptcy Research database (see https://urldefense.com/v3/__http://lopucki.law.ucla.edu__;!!N11eV2iwtfs!p6MxSMWLU7KfyXSvgFVBlWWvPNYg5Nvx1B3TRj-hT_hAqEJHC2O9jJBHsJl7aOpI7mosAlaJ0U4A2xrdCJA$). The financial returns data set (S&P 500) are available from DataStream. Restrictions apply to the availability of these data, which were used under license for this study. The remaining data that support the findings of this study are openly available from the St. Louis Fed online database FRED (https://urldefense.com/v3/__https://fred.stlouisfed.org/__;!!N11eV2iwtfs!p6MxSMWLU7KfyXSvgFVBlWWvPNYg5Nvx1B3TRj-hT_hAqEJHC2O9jJBHsJl7aOpI7mosAlaJ0U4A_J4o9VA$).

REFERENCES

Agosto, A., Cavaliere, G., Kristensen, D. & Rahbek, A. (2016) Modeling corporate defaults: poisson autoregressions with exogenous covariates(parx). *Journal of Empirical Finance*, 38, 640–663.

Aknouche, A. & Francq, C. (2021) Stationarity and ergodicity of Markov switching positive conditional mean models. *Journal of Time Series Analysis* (Accepted author manuscript).

Amisano, G. & Giacomini, R. (2007) Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25(2), 177–190.

Ang, A. & Timmermann, A. (2012) Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1), 313–337.

Azizpour, S., Giesecke, K. & Schwenkler, G. (2018) Exploring the sources of default clustering. *Journal of Financial Economics*, 129, 154–183.

Bacci, S., Pandolfi, S. & Pennoni, F. (2014) A comparison of some criteria for states selection in the latent MARKOV model for longitudinal data. *Advances in Data Analysis and Classification*, 8(2), 125–145.

Bai, J., Collin-Dufresne, P., Goldstein, R.S. & Helwege, J. (2015) On bounding creditevent risk premia. *The Review of Financial Studies*, 28(9), 2608–2642.

Banachewicz, K., Lucas, A. & Van Der Vaart, A. (2008) Modelling portfolio defaults using hidden Markov models with covariates. *The Econometrics Journal*, 11(1), 155–171.

Bartolucci, F. (2006) Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, 68(2), 155–178.

Bartolucci, F. & Farcomeni, A. (2009) A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104(486), 816–831.

Bartolucci, F., Farcomeni, A. & Pennoni, F. (2012) *Latent Markov models for longitudinal data*. Statistics in the Social and Behavioral Sciences. London: Chapman and Hall/CRC.

Bernardi, M., Maruotti, A. & Petrella, L. (2017) Multiple risk measures for multivariate dynamic heavy-tailed models. *Journal of Empirical Finance*, 43, 1–32.

Chen, C.-C. & Tsay, W.-J. (2011) A Markov regime-switching ARMA approach for hedging stock indices. *Journal of Futures Markets*, 31(2), 165–191.

Das, S.R., Duffie, D., Kapadia, N. & Saita, L. (2007) Common failings: how corporate defaults are correlated. *The Journal of Finance*, 62(1), 93–117.

Duffie, D., Saita, L. & Wang, K. (2007) Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635–665.

Duffie, D., Eckner, A., Horel, G. & Saita, L. (2009) Frailty correlated default. *The Journal of Finance*, 64(5), 2089–2123.

Ferland, R., Latour, A. & Oraichi, D. (2006) Integer-valued Garch process. *Journal of Time Series Analysis*, 27(6), 923–942.

Fokianos, K. & Tjøstheim, D. (2011) Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3), 563–578.

Fokianos, K., Rahbek, A. & Tjøstheim, D. (2009) Poisson autoregression. *Journal of the American Statistical Association*, 104(488), 1430–1439.

Giesecke, K., Longstaff, F.A., Schaefer, S. & Strebulaev, I. (2011) Corporate bond default risk: a 150-year perspective. *Journal of Financial Economics*, 102(2), 233–250.

Gray, S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1), 27–62.

Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.

Hansen, B.E. (1992) The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7(S1), S61–S82.

Hertzel, M.G., Li, F., Officer, M.S. & Rodgers, K.J. (2008) Inter-firm linkages and the wealth effects of financial distress along the supply chain. *Journal of Financial Economics*, 87(2), 374–387.

Holzmann, H. & Schwaiger, F. (2016) Testing for the number of states in hidden Markov models. *Computational Statistics & Data Analysis*, 100, 318–330.

Jorion, P. & Zhang, G. (2009) Credit contagion from counterparty risk. *The Journal of Finance*, 64(5), 2053–2087.

Kedem, B. & Fokianos, K. (2005) *Regression models for time series analysis*, volume 488. Hoboken: John Wiley & Sons.

Kim, C.-J. (1994) Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1), 1–22.

Koopman, S.J., Lucas, A. & Schwaab, B. (2012) Dynamic factor models with macro, frailty, and industry effects for us default counts: the credit crisis of 2008. *Journal of Business & Economic Statistics*, 30(4), 521–532.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H. & Bell, B. (2016) TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software, Articles*, 70(5), 1–21.

Lando, D. & Nielsen, M.S. (2010) Correlation in corporate defaults: contagion or conditional independence? *Journal of Financial Intermediation*, 19(3), 355–372.

Pohle, J., Langrock, R., van Beest, F.M. & Schmidt, N.M. (2017) Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3), 270–293.

R Development Core Team (2020). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rydberg, T.H. & Shephard, N. (2000) A modelling framework for the prices and times of trades made on the New York stock exchange. *Nonlinear and Nonstationary Signal Processing*, 217–246.

Sant'Anna, P.H.C. (2017) Testing for uncorrelated residuals in dynamic count models with an application to corporate bankruptcy. *Journal of Business & Economic Statistics*, 35(3), 349–358.

Stock, J.H. & Watson, M.W. (1996) Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1), 11–30.

Zucchini, W., MacDonald, I.L. & Langrock, R. (2016) *Hidden Markov for time series: an introduction using R* (2nd ed.). CRC Monographs on Statistics and Applied Probability. London: Chapman & Hall.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Berentsen, G.D., Bulla, J., Maruotti, A. & Støve, B. (2022) Modelling clusters of corporate defaults: Regime-switching models significantly reduce the contagion source. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1–25. Available from: https://doi.org/10.1111/rssc.12551