Spatial contagion in mortgage defaults

Citation for published version:
Calabrese, R & Crook, J 2020, 'Spatial contagion in mortgage defaults: A spatial dynamic survival model with time and space varying coefficients', *European Journal of Operational Research*, vol. 287, no. 2, pp. 749-761. https://doi.org/10.1016/j.ejor.2020.04.031

Digital Object Identifier (DOI):
10.1016/j.ejor.2020.04.031

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
European Journal of Operational Research

Spatial contagion in mortgage defaults: a spatial dynamic survival model with time and space varying coefficients

Raffaella Calabrese
raffaella.calabrese@ed.ac.uk
and
Jonathan Crook
jonathan.crook@ed.ac.uk
University of Edinburgh Business School

Abstract

This paper proposes a spatial discrete survival model to estimate the time to default for UK mortgages. The model includes a flexible parametric link function given by the Generalised Extreme Value Distribution and a dynamic spatially varying baseline hazard function to capture neighbourhood effects over time. We incorporate time and space varying variables into the model. The gains of the proposed model are illustrated through the analysis of a dataset on around 74,000 mortgage loans issued in England and Wales from 2006 to 2015.

Keywords: conditional autoregressive model, survival model, spatial contagion, mortgage defaults.

1 Introduction

Although there is a vast literature on scoring models for mortgage loans (e.g. Kelly, 2011; Tong et al., 2012; Wagner, 2004), there is little work that addresses spillover effects in modelling mortgage risk. The last financial crisis showed the effects of contagion or so called 'spillover or contagion effects' - how the deterioration of a borrower's future ability to honour his/her mortgage debt obligation can affect the ability of other borrowers that usually live in the same neighbourhood. Different kinds of contagion effects for mortgage loans have been analysed so far in the literature. Goodstein et al. (2017) obtain evidence of spillover effects only between strategic defaulters (borrowers that can be influenced in their decision) but they are not significant for defaults that are the result of inability to pay (borrowers that had no choice). Guiso et al. (2013), Seiler et al. (2013), Towe and Lawley (2013) found that homeowners with negative equity are more likely to strategically default if they know or they are neighbours of others who have done so.

Gupta (2019) analyses contagion effects for foreclosures (taking possession of a mortgaged property when the borrower fails to keep up their mortgage payments). The author identifies a few potential mechanisms through which foreclosures can affect the propensity to default of their neighbours. First, foreclosures can reduce the market price of neighbouring homes, which represents an incentive to default due to the negative equity (Schuetz et al., 2008). Moreover, financial institutions may deny refinancing opportunities to borrowers from areas that have previously experienced foreclosure activity. Finally, foreclosures could lead to an increase in crime, vandalisation and other activities that could depreciate the property value of a specific area. Pence (2006) also analysed data on US foreclosures and obtained that in states where laws favoured borrowers, the supply of mortgage credit may decrease because lenders may face higher costs.

Not only mortgages, but also other kinds of debts can show spillover effects. For example, loans to firms could be characterised by contagion effects as the economic distress from one company could propagate to another one (Calabrese et al. 2019; Giesecke and Weber, 2006). Longstaff (2010) found strong evidence of contagion in different financial markets of collateralised debt obligations (CDOs) through liquidity and risk-premium channels.

To the best of our knowledge, this is the first paper that introduces spillover effects in survival analysis to predict the default probability of mortgage loans. To understand the importance of contagion, we highlight that there are several possible mechanisms through which a distressed property can affect the prices of nearby houses. The exterior appearance of distressed properties can deteriorate because such properties could experience neglect, abandonment or vandalism. Several papers (Harding et al. 2009; Lin et al., 2009; Immergluck and Smith, 2006) obtain a negative relationship between the number of nearby foreclosures and the prices of non-distressed properties. There is a vast literature on housing spillover effects that has been summarised by Schwartz et al. (2003). Some authors, i.e. Clauretie and Daneshvary (2009) and Agarwal et al. (2012), highlight the importance of controlling for neighbourhood and spatial effects to analyse the effects of foreclosed properties on nearly non-distressed properties.

One of the widely used approaches to capture spillover effects is to include fixed effects based on the property location. Agarwal et al. (2012) follow this approach using the concentration of foreclosures in the same zip code, finding that an increase in the local foreclosure rate raises the probability of borrower default. They find that subprime mortgages are highly concentrated in some zip codes, so showing significant neighbourhood effects. Harding et al. (2012) control for location fixed effects by including zip code dummy variables and they are significantly different from zero. The main disadvantage of this methodology is that it requires a high number of fixed effects to capture the local nature of spillover effects in real estate. To overcome this drawback, we analyse the underlying spatial process of the propensity to default, analogously to Zhu and Pace (2014). Using a cross-sectional analysis, Zhu and Pace (2014) show that a probit model with spatially dependent disturbances increases the predictive accuracy compared to the probit model with independent errors. We extend this approach in three main directions.

The first methodological innovation of this paper is to use a flexible asymmetric link function instead of the probit model as it is more suitable for binary unbalanced data, as few authors have already shown in a non-spatial context (Calabrese et al. 2015; Wang and Dey, 2010). As the number of defaulted properties in a portfolio is much lower than the frequency of non-distressed mortgage loans, the sample of good and bad loans is usually highly unbalanced. To extend this approach to a longitudinal framework, the second contribution of this paper is to propose a survival model with spatial dependence and the flexible skewed link function. Since the initial work of Narain (1992), the survival approach has been widely used in credit risk modeling (Andreeva et al., 2005, 2007; Banasik et al., 1999; Bellotti and Crook, 2009 and 2013; Crook and Bellotti, 2010; Djeundje and Crook, 2019; Leow and Crook, 2016). Divino and Rocha (2013) show that survival analysis improves the accuracy of a scoring model in comparison with that obtained by a cross-sectional logistic regression. Analogously to most of the models used in the literature (Dirick et al., 2016), we also include time-varying covariates in survival model.

The third methodological innovation of this paper is that the coefficients of the proposed model can vary over space and over time. Some authors have used models with time-varying coefficients to predict corporate defaults, for example Hwang (2012) used them to investigated the effects of macroeconomic variables on firm-specific characteristics. For retail banking, Leow and Crook (2016) built two survival models based on accounts opened before and after the financial crisis and show that the parameters of the two models are statistically significantly different. In a recent work, Djeundje and Crook (2019) show that time-varying coefficients in a survival model increase the goodness of fit and the predictive accuracy of a scoring model.

We call the model proposed in this paper the Spatial Generalised Extreme Value Survival (SGEVSUR) model. We apply our proposal to a large dataset of 74,081 mortgage loans issued in England and Wales from June 2006 to December 2015. The SGEVSUR model outperforms the probit model with temporal, spatial and spatial-temporal components over different time horizons (12, 24 and 36 months). In the empirical analysis we find that **the AUC of the spatial-temporal model with the GEV link function is always lower than the AUC with either spatial or temporal components for all the time horizons.** We cannot reject (with $\alpha = 0.1$) the null hypothesis that the **Area Under the Curve** (AUC) of the GEV model with only spatial component is statistically significantly different from the AUC of the GEV model with only the temporal component. However, we can reject this hypothesis (with $\alpha = 0.01$) if we compare the spatial-temporal and the temporal GEV or the spatial-temporal and the spatial model.

The rest of the article is structured as follows: Section 2 presents the SGEVSUR model and the estimation procedure. This is followed by data description, empirical results, model fit and performance in Section 3, while the last Section 4 concludes with discussion. Appendix 1 explains in details the sampling procedure for estimating the SGEVSUR model and Appendix 2 contains some tables and plots.

2 A spatial discrete survival model for rare events

Some authors (Banasik et al., 1999; Bellotti and Crook, 2009; Djeundje and Crook, 2019; Stepanova and Thomas, 2002) use survival analysis not only to predict the probability that a borrower will default but also to assess the dynamical behaviour of the probability of default over the future. Suppose we have a portfolio of n mortgages over a geographical region $S = \{s_1, s_2, ..., s_{n_s}\}$ with n_s areas. Let t_i be the observed number of months since the i-th mortgage account was open, known as duration time. As mortgage account records are discrete and usually monthly reported, we assume that the random variable T_i that represents the duration time has a discrete domain where $t_i \in \{1, 2, ..., n_t\}$. As we know the postcode area for each property, we consider a discrete domain for space S. Let $\mathbf{x}_{it} = [1, x_{1it_i}, x_{2it_i}, ..., x_{pit_i}]$ denote the vector of p time-dependent covariates for mortgage i at time t_i.

We define a binary random variable Y_{it} for the default event with $Y_{it} = 1$ if the borrower defaults at time t and $Y_{it} = 0$ otherwise. We assume that default is an absorbing state, this means that there are no cured cases. We consider the conditional default probability of a mortgage loan as

$$P\{Y_{it} = 1 | Y_{iq} = 0 \ \ \forall q < t; \mathbf{x}_{it}, s_i, t\} = p(\mathbf{x}_{it}, s_i, t).$$

In survival analysis, $p(\mathbf{x}_{it}, s_i, t)$ is known as a discrete-time hazard rate and it represents the probability of defaulting on the repayment of the mortgage for the property i at month t given that the property was not in a distressed state until the month $t - 1$.

Two widely used approaches to model the hazard rate $p(\mathbf{x}_{it}, s_i, t)$ for discrete time are the probit (Chang et al., 2013) and the logit model (Allison, 1982; Homes and Held, 2006). In a non-spatial cross-sectional framework, some papers (Calabrese et al. 2015; King and Zeng, 2001; Wang and Dey, 2010) show that the logit and the probit models are inaccurate if the binary classification is strongly unbalanced, such as in scoring models for the mortgage market. The characteristics of the minority class, represented by defaulters in our application, are more informative than those of the majority class (non-defaulters). The features of defaulters are given by the values of the response curve close to 1. If we use a symmetric link function that approaches the extreme values 0 and 1 at the same rate, the probability of default is underestimated for the actual defaulters, as shown by Calabrese and Osmetti (2013) on empirical data.

Several methods have been proposed to deal with this drawback. The widely used approach is to use sampling to obtain balanced classes (Sahare and Gupta, 2012). We cannot apply a sampling method with spatial data because it can change the spatial dependence structure in the data. In a non-spatial context, some authors suggest the use of the Generalised Extreme Value (GEV) cumulative distribution function (Calabrese et al. 2015; Wang and Dey, 2010) to increase the weight given to the event with lower frequency, represented by the distressed properties $Y_{it} = 1$ in this analysis. We choose this random variable because we focus the attention on the right tail of the response curve and the GEV distribution has been widely used in the literature to model the tail of a distribution (Kotz and Nadarajah, 2000). An important advantage of the GEV distribution is that it is very flexible with a parameter controlling the tail size and the shape (Dey and Yan, 2016). Calabrese and Elkink (2016) employ the GEV distribution for a spatial cross-sectional approach.

Li at al. (2016) use the GEV distribution to model the logarithm of time $\ln(T)$ in a spatial continuous time survival model. Instead, in this paper we consider a spatial discrete time survival framework and we suggest modelling the conditional probability of default $p(\mathbf{x}_{it}, s_i, t)$ using the GEV distribution as follows

$$p(\mathbf{x}_{it}, s_i, t) = F_{GEV}[\mathbf{x}'_{it}\boldsymbol{\beta}(s_i, t)] = \begin{cases} \exp\left\{-\left[1+\tau\left(\frac{\mathbf{x}'_{it}\boldsymbol{\beta}(s_i,t)-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\tau}}\right\} & \tau \neq 0 \\ \exp\left[-\exp\left(-\frac{\mathbf{x}'_{it}\boldsymbol{\beta}(s_i,t)-\mu}{\sigma}\right)\right] & \tau = 0 \end{cases} \tag{1}$$

where τ denotes the shape parameter, $\mu \in R$ the location parameter, $\sigma \in R^+$ the scale parameter and $x_+ = max(x, 0)$. It is important to highlight that the GEV distribution has a support that depends on the parameter τ.

- If $\tau > 0$, the Fréchet distribution is obtained with a finite lower end-point at $\mu - \sigma/\tau$.

- If $\tau < 0$, the Weibull distribution is obtained with a finite upper end-point at $\mu - \sigma/\tau$.

- If $\tau = 0$, the Gumbel distribution is obtained with infinite support.

Without lost of generality, analogously to Andreeva et al. (2016) and Calabrese et al. (2015), we consider $\mu = 0$ and $\sigma = 1$ as the coefficients $\boldsymbol{\beta}$ can be changed to take into account any choice of μ and σ. Let $\boldsymbol{\beta}(s_i, t) = [\beta_0(s_i, t), \beta_1(s_i, t), ..., \beta_p(s_i, t)]'$ be the vector of regression coefficients for the location s and time t.

In equation (1), for $(x_{1it} = 0, x_{2it} = 0, ..., x_{pit} = 0)$ we obtain the baseline risk $\beta_0(s, t)$ that varies with both space and time. We relate the baseline risk to a spatial spillover effect $\mu_0(s)$ and to a temporal effect $\gamma_0(t)$ by

$$\beta_0(s, t) = \eta_0 + \mu_0(s) + \gamma_0(t)$$

where η_0 is the overall average. Analogously, each covariate effect

$$\beta_j(s, t) = \eta_j + \mu_j(s) + \gamma_j(t) \tag{2}$$

for $j = 1, 2, ..., p$ is modelled similarly[1].

Two main models have been used to analyse the spillover effects $\mu_j(s)$ in equation (2) known as the conditionally and simultaneously autoregressive models, i.e. CAR and SAR models (Wall, 2004). We use a CAR model in this paper as it represents an attractive approach to handle complicated joint statistical relationships using a set of conditional dependencies and it is computationally very convenient (Barnerjee et al., 2015 p. 155).

[1] To limit the number of parameters, we do not include space-time interaction terms.

The CAR model was originally developed by Besag (1974) and it represents a larger class than SAR models (Barnerjee et al., 2015 p. 87). An additional difference between the CAR and the SAR specifications is that in the former the distribution for the dependent variable is specified and it induces a distribution for the disturbances. The latter specification reverses this designation providing a distribution for the disturbances which induces a distribution for the dependent variable (Barnerjee et al., 2015 p. 83).

Let W_s be an exogenous square matrix of order n_s known as a spatial adjacency matrix. The generic element $w_{s,s'}$ is equal to one when s and s' are neighbours, that is $s' \sim s$ and zero otherwise. A second matrix D_s is a diagonal matrix with elements on the main diagonal given by $\sum_{s' \neq s} w_{s,s'} = m_s$, where m_s is the number of neighbours[2] of region s. The joint distribution of the spatial effects $\boldsymbol{\mu}_j = [\mu_{j1}, \mu_{j2}, ..., \mu_{jn_s}]'$ is given by (Banerjee, 2004)

$$\boldsymbol{\mu}_j \sim MVN(\mathbf{0}, \sigma_{sj}^2 (D_s - \rho_{sj} W_s)^{-1})$$

with $s' \neq s$ and $j = 1, 2, ..., p$, where MVN stands for Multivariate Normal Distribution, $\rho_{sj} \in [0, 1]$ is the spatial autocorrelation parameter, $s' \sim s$ indicates that regions s and s' are neighbours, $m_s$is the number of neighbours of the region s and the parameter $\sigma_{sj}^2 > 0$ controls the amount of variation between the random effects. The conditional distribution of the spatial effect $\mu_j(\cdot)$ for the j-th covariate is assumed to be

$$\mu_j(s) | \{\mu_j(s')\}_{s' \sim s} \sim N\left(\rho_{sj} \frac{\sum_{s' \sim s} \mu_j(s')}{m_s}, \frac{\sigma_{sj}^2}{m_s}\right). \tag{3}$$

This is a standard definition of a CAR model widely used by scholars (e.g. Barnerjee et al., 2015; Wall, 2004; Zhu and Pace, 2014). Barnerjee et al. (2015, p. 82) explain that $\rho_{sj} \frac{\sum_{s' \sim s} \mu_j(s')}{m_s}$ can be viewed as a reaction function where ρ_{sj} is the expected proportional reaction of $\mu_j(s)$ to $\frac{\sum_{s' \sim s} \mu_j(s')}{m_s}$. Therefore, we are modelling the spatial effect $\mu_j(s)$ such that its mean is a proportion of the average of its neighbours' spatial effects. If $\rho_{sj} = 0$, the spatial effects $\mu_j(s)$ become independent. The conditional variance in equation (3) is inversely proportional to the number of neighbours, so that the more neighbours an area has, the greater the precision for the effect of that area.

[2] We assume that $m_s > 0$ for any $s \in S$.

We assume that the temporal effects $\boldsymbol{\gamma}_j = [\gamma_j(1), \gamma_j(2), ..., \gamma_j(T)]'$ of the j-th covariate in equation (2) **follow** a first-order autoregressive model. Let t and t' denote different periods of time and W_t be the temporal adjacency matrix where the element $w_{t,t'}$ is equal to one if $|t - t'| = 1$, otherwise zero. We define D_t as a diagonal matrix where the elements on the main diagonal are given by $\sum_{t' \neq t} w_{t,t'} = 2$.

The temporal effects $\boldsymbol{\gamma}$ are jointly distributed as a multivariate normal

$$\boldsymbol{\gamma}_j \sim MVN(\mathbf{0}, \sigma_{tj}^2 (D_t - \rho_{tj} W_t)^{-1}),$$

where ρ_{tj} is the temporal autocorrelation parameter and σ_{tj}^2 is the variance parameter with $j = 1, 2, ..., p$.

The conditional distribution of the temporal effect $\gamma_j(\cdot)$ for the j-th covariate is assumed to be

$$\gamma_j(t) | \{\gamma_j(t')\}_{t' \sim t} \sim N\left(\rho_{tj} \frac{\gamma_j(t-1) + \gamma(t+1)}{2}, \frac{\sigma_{tj}^2}{2}\right)$$

with $t' \neq t$, where $\rho_{tj} \in [0, 1]$ is the temporal autocorrelation parameter, $t' \sim t$ indicates that $|t - t'| = 1$ and the parameter $\sigma_{tj}^2 > 0$ controls the amount of variation between the random effects.

To avoid identifiability issues, we add sum-to-zero constraints for the random effects (Gelfand and Sahu, 1999). We call the proposed model the Spatial GEV survival (SGEVSUR) model.

2.1 Prior, likelihood and posterior distributions

As time to default can be subject to right censoring, let δ_i be the right-censoring indicator, where $\delta_i = 1$ if the borrower is observed to default on the mortgage and $\delta_i = 0$ if the time to default is right-censored (the mortgage debt is fully paid or the payments are still being made). Let T_i denote the corresponding failure or censoring time. We assume that the censoring mechanism is random and noninformative as defined by Kalbfleisch and Prentice (2002, pp 53 and 195).

Let the parameter vector be $\boldsymbol{\theta} = (\boldsymbol{\rho}_s', \boldsymbol{\rho}_t', \boldsymbol{\eta}', (\boldsymbol{\sigma}_s^2)', (\boldsymbol{\sigma}_t^2)')'$. Analogously to Andreeva et al. (2016) and Calabrese et al. (2015), we fix a value of the parameter τ and we estimate the parameter vector $\boldsymbol{\theta}$ using a Gibbs sampling. We explain in the following section that we use a deviance criterion to choose τ.

The survival function $S(\cdot)$ for the i-th right-censored observation is

$$P(\tilde{T}_i > t) = S(t, \boldsymbol{\theta}) = \prod_{t_i=1}^{t} [1 - p(t_i, \boldsymbol{\theta})], \tag{4}$$

where $\tilde{T}_i$ is the underlying uncensored time to default and the probability of default $p(t_i, \boldsymbol{\theta})$ is defined in equation (1). Let $D = (n, \boldsymbol{y}, X, \boldsymbol{\delta})$ denote the observed data. Therefore, the likelihood function is given by

$$\mathcal{L}(D|\boldsymbol{\theta}) = \prod_{i=1}^{n} \{[p(t_i, \boldsymbol{\theta})]^{\delta_i} [S(t_i, \boldsymbol{\theta})]^{1-\delta_i}\}$$

and the log-likelihood function is

$$\ell(D|\boldsymbol{\theta}) = log[\mathcal{L}(D|\boldsymbol{\theta})] = \sum_{i:defaulted} p(t_i, \boldsymbol{\theta}) + \sum_{i:nondefaulted} S(t_i, \boldsymbol{\theta}).$$

The posterior distribution of $\boldsymbol{\theta}$ is thus given by

$$\pi(\boldsymbol{\theta}|D) = \frac{\mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

where $\pi(\cdot)$ is the prior for the parameter vector $\boldsymbol{\theta}$. As $\int_{\boldsymbol{\theta}} \mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ does not have an analytic closed form, we use the Gibbs sampling algorithm (Ibrahim et al. 2001, p.19) to sample from the posterior distribution $\pi(\boldsymbol{\theta}|D)$.

Under the assumption that the prior distributions for parameters $\boldsymbol{\rho}_s, \boldsymbol{\rho}_t$ $\boldsymbol{\eta}, \boldsymbol{\sigma}_s^2, \boldsymbol{\sigma}_t^2$ are independent, we use proper and weak informative priors on all the parameters to assure parameter identifiability, in line with Chang et al. (2013), Li et al. (2016) and Wang et al. (2010). Particularly, we assign priors for the individual parameters of the model as follows:

- the mean η_j defined in equation (2) is distributed as $\eta_j \sim N(0, 100^2)$ (Chang et al., 2013);
- the **precision parameters** $1/\sigma_{sj}^2$ **and** $1/\sigma_{tj}^2$ are distributed as Gamma($a_1 = 0.5, b_1 = 0.005$)[3] (Chang et al., 2013);

[3] We also considered two additional priors for $1/\sigma_{sj}^2$ and $1/\sigma_{tj}^2$ given by Gamma($a_1 = 0.03, b_1 = 0.005$) and Gamma($a_1 = 0.1, b_1 = 0.0001$). We obtained nearly identical results to the original prior.

- we discretise the prior to 1,000 equally-spaced points over [0,1] for ρ_{sj} and ρ_{tj} to facilitate the MCMC sampling. We choose uninformative priors where $\rho_{sj}, \rho_{tj} \sim Beta(1,1)$ (LeSage and Pace, 2009, Chapter 5).

Wang and Dey (2010) show that the posterior distributions under the GEV link are proper for many non-informative priors.

In the following section we explain the procedure to estimate the $2(n_s + n_t) + p$ parameters $(\boldsymbol{\rho}_s, \boldsymbol{\rho}_t, \boldsymbol{\eta}, \boldsymbol{\sigma}_s^2, \boldsymbol{\sigma}_t^2)$ that represent the parameter vector $\boldsymbol{\theta}$.

2.2 The estimation procedure

The binary dependent variable Y_i in a Bayesian approach can be considered as an indicator of a continuous latent variable Y_{it}^* (LeSage and Pace, 2009, p.281) such that

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Considering Y^* as an additional set of parameters to be estimated, the posterior distribution for the parameters $\boldsymbol{\theta}$ conditioning on both Y and Y^* becomes a Bayesian regression model with a continuous dependent variable (Albert and Chib, 1993). This approach has been already used for different link functions, such as the Student-t (Albert and Chib, 1993), a generalised Student-t (Kim et al. 2008), the Gosset and the Pregibon functions (Koenker and Yoon, 2009) and a new skewed link model (Chen et al. 1999).

A GEV link function leads to a truncated GEV distribution (TGEV) for the latent variable y_{it}^* given by

$$\begin{aligned} Y_{it}^* &\sim GEV(\mathbf{x}_{it}'\boldsymbol{\beta}(s_i,t), \tau, \mu = 0, \sigma = 1) I(y_{it}^* > 0) \qquad \text{if } y_i = 1 \\ Y_{it}^* &\sim GEV(\mathbf{x}_{it}'\boldsymbol{\beta}(s_i,t), \tau, \mu = 0, \sigma = 1) I(y_{it}^* < 0) \qquad \text{if } y_i = 0. \end{aligned} \tag{6}$$

The first step of the estimation procedure for the SGEVSUR model is to choose a value for the parameter τ. The skewness and approaching rate to 1 and 0 of the link function depend on τ (Calabrese et al., 2015). If τ is negative, the conditional probability of default $p(\mathbf{x}_{it}, s_i, t)$ defined in equation (1) approaches 0 slowly and 1 more rapidly compared to the log-log curve. If τ is positive, we obtain the opposite. As the sample is highly unbalanced with a low percentage of 1, we need a curve which approaches 1 more sharply, given by negative values for the parameter τ. For this reason, we try out the values of τ in the set $(-1, -0.9, -0.6, -0.3)$.

For a fixed value of τ, we apply the Gibbs sampling (Casella and George, 1992) to obtain the posterior distributions of the parameter set $\boldsymbol{\theta}$. We consider 10,000 iterations and we ignore the first 2,000 as burn-in.

We start the MCMC algorithm choosing the following starting values for the parameters[4] analogously to Chang at al. (2013)
$\beta_j(s,t) = 0$ for all j, s, t, $\boldsymbol{\rho}_s^{(0)} = \mathbf{0}$, $\boldsymbol{\rho}_t^{(0)} = \mathbf{0}$, $\boldsymbol{\eta}^{(0)} = \mathbf{0}$, $\boldsymbol{\sigma}_s^{(0)} = \mathbf{1}$, $\boldsymbol{\sigma}_t^{(0)} = \mathbf{1}$.

At each iteration $r = 1, 2, ..., R$, each parameter $\boldsymbol{\theta}_i^{(r)}$ of the vector $\boldsymbol{\theta}$ is sampled from $f(\boldsymbol{\theta}_i^{(r)} | \boldsymbol{\theta}_{-i}^{(r-1)}, X)$ conditional on both the covariates X and the vector of the remaining parameters at their current values $\boldsymbol{\theta}_{-i}^{(r-1)} = [\boldsymbol{\theta}_1^{(r)}, ..., \boldsymbol{\theta}_{i-1}^{(r)}, \boldsymbol{\theta}_{i+1}^{(r-1)}, ..., \boldsymbol{\theta}_d^{(r-1)}]'$. At each iteration r, we compute the residual without the parameter $\boldsymbol{\theta}_i^{(r)}$

$$\epsilon_{it}^{(r)} = Y_{it}^* - \boldsymbol{x}_{it}' \boldsymbol{\beta}^*(s_i, t)$$

where $\boldsymbol{\beta}^*(s_i, t)$ denote the coefficient vector with the term under investigation set to zero. For example, if we are investigating η_j, the equation (2) becomes $\beta_j^*(s_i, t) = \mu_j(s) + \gamma_j(t)$

We provide the conditional density functions of the parameters in the SGEVSUR model and the sequence followed to sample them in Appendix 1.

After estimating the vector parameter $\boldsymbol{\theta}$ for a fixed τ, we apply the described estimation procedure for different τs and we choose the value of τ that minimises the Deviance Information Criterion (DIC) (Zhu and Carlin, 2000; Spiegelhalter et al., 2002) defined as

$$DIC = G + F$$

where G is the posterior expectation of the deviance DE ($G = E_{\boldsymbol{\theta}/D}[DE]$) and represents the fit of the model, while F is the difference between the expected deviance and the

[4]We estimate the models for different starting values: $\beta_j(s,t) = 1$ for all j, s, t, $\boldsymbol{\rho}_s^{(0)} = \mathbf{0.25}, \mathbf{0.5}$; $\boldsymbol{\rho}_t^{(0)} = \mathbf{0.25}, \mathbf{0.5}$; $\boldsymbol{\eta}^{(0)} = \mathbf{1}$; $\boldsymbol{\sigma}_s^{(0)} = \mathbf{2}$; $\boldsymbol{\sigma}_t^{(0)} = \mathbf{2}$ and we increase the burn-in number of iterations to 5,000. We obtain that the results are robust for different starting values and burn-in number of iterations. These results are available upon request to the authors.

deviance evaluated at the posterior expectations and captures the complexity of the model given by the effective number of parameters. The posterior distribution of the deviance statistic is

$$DE(\boldsymbol{\theta}) = -2\log[\mathcal{L}(D|\boldsymbol{\theta})] + 2\log[h(D)]$$

where $\mathcal{L}(D|\boldsymbol{\theta})$ is the likelihood function and $h(D)$ is a standardising function of the data.

3 Empirical analysis

3.1 Data description

We have a large dataset of 74,081 mortgage loans provided by a financial institution in England and Wales covering the period from June 2006 to December 2015. The data consist of monthly behavioural data. Consistent with Basel II (BCBS, 2005), a mortgage loan is defined in default when it fails to make the payments for at least three consecutive months. We apply this definition of default and we compute the number of defaulted loans over the total number of mortgages (74,081) and we obtain a percentage of default equal to 1.806%. The default is considered as an **absorbing state**, this means that there are not cured cases in the dataset. The dependent variable Y is coded as 1 if the borrower is in default on her mortgage loan, 0 otherwise.

The postcodes in the UK are alphanumeric codes. The first part of the postcode has between two and four characters and the second three characters. One or two letters of the first part of the postcode indicates a postcode area. It could represent a city (such as L for Liverpool) or a region (such as HS for Outer Hebrides) or a part of London (such as W for part of central and part of west London). We know the postcode area of the property. The dataset is spread over 106 postcode areas. We report the mortgage distribution in Table 8 and the map in Figure 1 in Appendix 2.

We estimate the model using the data observed from June 2006 to December 2012[5].

[5]Given the last financial crisis in 2008, we check if the results are robust for a different time period. We estimate the model on a training sample from January 2009 to December 2012 and we test it on a

Table 1: Description of the explanatory variables. The letter (T) in the second column denotes a time-varying variable. The data source of the macroeconomic variables is the Office of National Statistics, UK.

Variable	**Description**
LTV	Ratio of the loan amount over the valuation of the property at completion
Applicants	Number of mortgage applicants
Balance	Mortgage balance (T)
Repayment	Minimum contractual repayment (T)
Interest	Interest rate on the mortgage loan (T)
Property	Estimated value of the property (T)
Marital	Marital status of the first applicant at origination:
	married (1) or single, separated, divorced, widowed, cohabiting or other
Buy to let	Property is buy to let: yes (1) or no (0)
Type	Type of mortgage repayment: Repayment (1) or interest only and split (0)
Fixed	Type of mortgage product:
	Fixed (1) or flexible, tracker, variable, discount and further advance (0)
Flexible	Type of mortgage product:
	Flexible (1) or fixed, tracker, variable, discount and further advance (0)
Tracker	Type of mortgage product:
	Tracker (1) or fixed, flexible, variable, discount and further advance (0)
HPI	House price index (T)
Production	Index of all UK production, not seasonally adjusted (T)
UN	Unemployment rate for people aged 16 and over, seasonally adjusted (T)
IR	Interest rate, selected UK retail banks base rate (T)
Consumer	Consumer price index (% change) (T)

Table 2: Descriptive statistics for the training sample.

Variable	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**
LTV	59.864	21.689	61.290	0.273	96.888
Applicants	1.540	0.516	2.000	1.000	4.000
Balance	97,469	69,402	83,028.720	-10,8740.060	2,502,093
Repayment	503	3,357	438.910	0.000	5,705,002
Interest	0.043	0.011	0.047	0.000	0.218
Property	186,584	214,700	151,104.210	0.000	46,385,190
Marital	0.484	0.500	0.000	0.000	1.000
Buy to let	0.244	0.429	0.000	0.000	1.000
Type	0.620	0.485	1.000	0.000	1.000
Fixed	0.540	0.498	1.000	0.000	1.000
Flexible	0.102	0.303	0.000	0.000	1.000
Tracker	0.149	0.356	0.000	0.000	1.000
HPI	2.434	6.678	3.700	-15.600	10.800
Production	102.102	5.124	99.400	95.200	111.700
UN	6.781	1.166	7.300	5.100	8.500
IR	1.761	2.537	0.500	0.500	5.750
Consumer	2.537	1.262	2.500	-0.100	5.200

Table 3: Goodness of fit for the SGEVSUR model for τ=-0.30.

SGEVSUR model	*G*	*F*	*DIC*
Spatial-Temporal	39,755	681	40,436
Temporal	41,113	315	41,428
Spatial	41,709	398	42,106

To avoid a spurious indication of performance we assess the forecast accuracy using an out-of-time sample relating to January 2013 to December 2015. This provides up to 36 months of test data, which represents a long period for forecasts. To ensure that the test sample is also out-of-sample, mortgage loans originated from 2006 to 2012 and still active after 2012 are considered right-censored data in the training set. All the mortgage loans in the test sample are issued from January 2013. This procedure gives 46,386 and 27,695 loans in the training and test samples respectively. Table 1 provides the description of the explanatory variables and Table 2 presents some descriptive statistics on the training sample.

3.2 Estimation results

To choose the value of τ, we compute the DIC[6] for values of the parameter τ in the set $(-1, -0.9, -0.6, -0.3)$. We obtain that the model with the best goodness of fit is for $\tau = -0.30$. As Table 3 shows, the SGEVSUR model with both the spatial and temporal components shows the highest fit to the training sample.

Table 4 around here

We report in Table 4 the average value over time and/or region of the posterior mean of the parameter estimates for the SGEVSUR model with τ=0.-30. Some predictors show

sample from January to December 2013. The results, available upon request to the authors, are similar to those reported in the paper.

[6]We report the DIC results for the SGEVSUR model with spatial-temporal components for different values of τ in Table 10 in Appendix 2.

similar behaviours in the three models (temporal, spatial and spatial-temporal models), such as the *Buy to Let* dummy variable that is significant in the spatial and spatial-temporal models with an inverse relationship with the default risk. Also, the balance owing on a mortgage is an important risk factor in all the analysed models but the sign of this parameter estimate is coherent with the expectations only for the model with both spatial and temporal components.

The most interesting result is for the variable *Loan to valuation*: this variable becomes significant only if we consider the spatial component in the scoring model . Several studies in the US have shown the importance of this variable (e.g. Kau et al., 2011; Zhu and Pace, 2014) as home-owners may be most likely to decide to default on their mortgage in neighbourhoods where the majority of the borrowers have high loan to valuation ratios (Agarwal et al., 2011 and Harding et al., 2009).

When the coefficients of the scoring model are space-varying, the *property* value shows a negative significant relationship with the propensity to default on the mortgage loan. Otherwise, this relationship becomes positive and non-significant if we consider only time-varying coefficients. Also Lin et al. (2009), Immergluck and Smith (2006) documented a negative relationship between sales prices and the number of nearby foreclosures for US properties.

As the macroeconomic variables vary only over time and not over space in this scoring model, most of them (*HPI*, *UN*, *IR*) are significant risk factors only if we introduce the temporal component. The importance of these variables is highlighted by several studies, Ambrose and Diop (2014), Bellotti and Crook (2009) among others. The first plot in Figure 3 in Appendix 2 shows the posterior mean of the parameter β for West London over time and the second plot the posterior mean of the parameter β on June 2006 over space for the variable *Contractual Repayment*.

Table 5 around here

We report the estimates of the spatial ρ_s and temporal ρ_t autocorrelation parameters in Table 5. If the risk factor does not change over time or over space, we cannot compute the temporal or the spatial autocorrelation parameter. Table 4 shows several interesting results. The variables *N of applicants*, *Married*, *Buy To Let* and the dummies related to the type of mortgage repayment (*Repay*, *Fixed* , *Flexible*, *Tracker*) show all significant spatial autocorrelation parameters. We can expect that similar number of applicants and marriage status are concentrated in the same neighbourhoods, for example families prefer to buy a property in residential areas. Similar considerations are valid also for the dummies *Buy To Let*, *Repay*, *Fixed*, *Flexible* and *Tracker*. Consistent with expectations, most of the macroeconomic variables show a significant temporal first order autocorrelation parameter ρ_t.

We report in Table 9 in Appendix 2 the I-statistic for the parameter ρ_s and ρ_t[7]. This is a convergence diagnostic of the parameter estimates proposed by Raftery and Lewis (1992) that should be smaller than 5.

Figure 2 in Appendix 2 shows the posterior distributions of the spatial ρ_s and temporal ρ_t autocorrelation parameters for the variables *Contractual payment and Property value* in the SGEVSUR model with spatial-temporal components.

3.3 Model performance

In this section we compare the performance of the SGEVSUR with those of the probit models with spatial, temporal and spatial-temporal components. We estimate the probit models using the same procedure described in Section 2.2 where the equation (6) becomes

$$\begin{aligned} Y^*_{it} &\sim N(\mathbf{x}'_{it}\boldsymbol{\beta}(s_i,t), \sigma=1) I(y^*_{it} > 0) \qquad \text{if } y_i = 1 \\ Y^*_{it} &\sim N(\mathbf{x}'_{it}\boldsymbol{\beta}(s_i,t), \sigma=1) I(y^*_{it} < 0) \qquad \text{if } y_i = 0. \end{aligned} \tag{7}$$

The expression (7) represents a univariate normal distribution with mean $\mathbf{x}'_{it}\boldsymbol{\beta}(s_i,t)$ and variance 1 that is truncated to the left at 0 if $y_i = 1$ and to the right at 0 if $y_i = 0$ (LeSage and Pace, 2009 pp. 283).

As we explained in Section 3.1, to avoid sample dependency we estimate the models using observations from June 2006 to December 2012 and we evaluate the predictive accuracy on data from January 2013 to December 2015. Because the forecasting horizon

[7] The authors have also computed the I-statistic for $\boldsymbol{\eta}, \boldsymbol{\sigma}^2_s, \boldsymbol{\sigma}^2_t$ and they are always smaller than 5. The results are available upon request from the authors.

is shorter than the time horizon used to estimate the model, we know the values of the parameters $\beta_j(s,t)$ in equation (2) where t represents duration time.

For assessing the predictive accuracy, we compute a few standard measures used in the literature (Tong et al., 2012) and in the industry such as the AUC and the Kolmogorov-Smirnov (KS) statistic. We also consider the H measure proposed by Hand (2009, 2010) as it is not sensitive to the empirical score distributions of the default and non-default groups[8]. Analogously to Bellotti and Crook (2009), we choose 0.05 as severity ratio that represents the ratio between the misclassification cost of a non-default and that of a defaulter. Following Zhu and Pace (2014), we compute also the misclassification rates (M) for defaults where the true status is default but the model predict non-default. We compute the cut-off using the equation in Krzanowski and Hand (2009, p. 172) that accounts for imbalanced data.

Table 6 around here

Table 6 reports the AUC, KS, H and M measures for different forecasting time horizons (12, 24 and 36 months). Using a GEV link function instead of a Gaussian function drastically increases the predictive accuracy of a scoring model for a short time horizon as 12 months. Figure 4 shows the ROC curves for the SGEVSUR and the probit models with spatial-temporal components for a time horizon of 12 months.

Figure 4 around here

The difference between the performance measures of these two models decreases as the time horizon increases. Table 6 also shows that the predictive accuracy of the SGEVSUR model decreases faster than that for the probit model as the time horizon increases. **If we focus our attention on the GEV link function, Table 6 shows that the AUC for the models with spatial-temporal components is always lower than those with only spatial or temporal components. However, the spatial-temporal models show the lowest misclassification rates (M) for defaults. This means that GEV models with only spatial or only temporal components perform better than spatial-temporal models in classifying both**

[8]We use the R package *hmeasure* to compute the AUC, KS and H measure.

defaulters and non-defaulters for different thresholds. Conversely, the spatial temporal approach outperforms the other GEV models in classifying defaults for a specific threshold. We obtained similar in-sample results[9].

To understand if the AUCs of the spatial-temporal (ST), temporal (T) and spatial (S) GEV models are statistically different, we apply the DeLong-DeLong test[10] (DeLong et al., 1988). We perform a two-sided test for the difference in AUC where the null hypothesis is that the AUCs of the two models are equal. If we define

$$u_{D,ND} = \begin{cases} 1, & if s_D < s_{ND} \\ 0 & if s_D \geq s_{ND}, \end{cases}$$

then the test statistic $\hat{U}$ of Mann-Whitney is

$$\hat{U} = \frac{1}{N_D N_{ND}} \sum_{(D,ND)}^{n} u_{D,ND}$$

where the sum is over all pairs of defaulters (D) and non-defaulters (ND) in the sample. The DeLong-DeLong test statistic T is defined as

$$T = \frac{\hat{U}_1 - \hat{U}_2}{\sqrt{Var(\hat{U}_1) + Var(\hat{U}_2) - 2cov(\hat{U}_1, \hat{U}_2)}}$$

where $Var(\hat{U}_1)$, $Var(\hat{U}_2)$ and $cov(\hat{U}_1, \hat{U}_2)$ are computed in Engelmann et al. (2003). This test statistic is asymptotically distributed as a normal distribution.

Table 7 shows that if we compare the S and the T models, the p-value for each time horizon increases as the time horizon increases. Also the p-value of the difference in AUC between the ST and S models increases as the forecasting horizon increases. Finally, we obtain stronger evidence against the null hypothesis if we compare the ST and T models instead of the S and T models as the p-value in the first case is always lower than that in the latter comparison. The p-value of the comparison between ST and S models is lower than the p-value for comparing S and T models for all the time horizons.

Table 7 around here

[9]The in-sample results are available upon request to the authors.

[10]We use the function *roc.test* in the R package pROC.

If we perform the DeLong-DeLong test for comparing the AUC of the SGEVSUR and the AUC of the probit model, we obtain that we reject the null hypothesis at a significance level of 0.05 for the time horizon of 12 months. We can also reject the null hypothesis at the 0.1 significance level for 24 and 26 months.

4 Conclusion

In this paper we propose a scoring model for mortgage loans introducing spillover effects in survival analysis. As the sample of good and bad loans is usually highly unbalanced, the first innovation of this paper is to use a flexible asymmetric link function. The second innovation is to include spatial dependence in the longitudinal framework to present the strong empirical evidence of contagion effects in distressed properties (e.g. Agarwal et al., 2012; Harding et al. 2009; Lin et al., 2009; Immergluck and Smith, 2006). The third methodological innovation of this manuscript is to consider coefficients that can vary both over time and over space. We call the proposed model the Spatial Generalised Extreme Value Survival (SGEVSUR) model.

From an applied perspective, we analyse a large dataset of 74,081 mortgage loans issued in England and Wales from June 2006 to December 2015. The time horizon is very interesting since it includes the financial crisis of 2008. In order to capture the economic cycle, we include some macroeconomic variables in the scoring model (Bellotti and Crook, 2009). A crucial result of this analysis is that we improve the forecasting accuracy of classic alternatives, such as probit model with spatial and temporal components. Finally, we perform hypothesis tests to check if the differences between the AUCs of spatial-temporal, temporal and spatial GEV models are statistically significantly different from zero. We find that the spatial-temporal model predicts more accurately than either the temporal or the spatial models alone, but that there is no difference in accuracy between the purely spatial and the purely temporal models.

5 Bibliography

Agarwal, S., Ambrose, B.W., Chomsisengphet, S., Sanders, A.B. (2012). Thy neighbors mortgage: Does living in a subprime neighborhood affect ones probability of default? *Real Estate Economics*, 40(1), 1-22.

Andreeva G., Calabrese R., Osmetti S. A. (2016) A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models. *European Journal of Operational Research*, 249 (2), 506-516.

Albert, J. H., Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88 (422), 669-679.

Allison P. D.(1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology*, 13(1982), 61-98

Ambrose, B. W., Diop, M. (2014). Spillover effects of subprime mortgage originations: The effects of single-family mortgage credit expansion on the multifamily rental market. *Journal of Urban Economics*, 81, 114-135.

Banasik, J., Crook J., Thomas L. (1999). Not if but when will borrowers default. *Journal of Operational Research*, 50, 1185-1190.

Banerjee, S., Carlin, B. P., Gelfand, A. E. (2004)*Hierarchical modeling and analysis for spatial data*. New York: Chapman & Hall/CRC, 2004.

Bellotti, T., Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12), 1699-1707.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society: Series B*, 36, 192-236.

Bholat, D., Lastra, R., Markose, S., Miglionico, A., Sen, K. (2016) Non-performing loans: regulatory and accounting treatments of assets. Working Paper 594, Bank of England.

Calabrese, R., Osmetti, S. A. (2013) Modelling small and medium enterprise loan defaults as rare events: the generalized extreme value regression model. *Journal of Applied Statistics*, 40 (6), 1172-1188.

Calabrese, R., Marra, G., Osmetti, S. A. (2015) Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model. *Journal of the Operational Research Society*, 67 (4), 604-615.

Calabrese, R., Johan, A. E., (2016) Estimating Binary Spatial Autoregressive Models for Rare Events *Spatial Econometrics: Qualitative and Limited Dependent Variables*, 145-166.

Calabrese, R., Andreeva, G., Ansell, J., (2019) Birds of a Feather Fail Together: Exploring the Nature of Dependency in SME Defaults. *Risk Analysis*, 39 (1), 71-84.

Casella, G., George, E. I. (1992) Explaining the Gibbs sampler. *American Statistician*, 6, 167-174.

Chang, H. H., Reich, B. J. and Miranda, M. L. (2013), A spatial time-to-event approach for estimating associations between air pollution and preterm birth. *Journal of the Royal Statistical Society: Series C*, 62, 167-179.

Chen, M., Dey, D. K., Shao, Q. (1999) A new skewed link model for dichotomous quantal response data. *Journal of American Statistical Association*, 94 (448), 1172-1186.

Clauretie, T., Daneshvary, N., (2009). Estimating the house foreclosure discount corrected for spatial price interdependence and endogeneity of marketing time. *Real Estate Economics*, 37 (1), 43-67.

DeLong, E. R., DeLong, D. M., Clarke-Pearson, D. L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44, 837-845.

Dey, K. D., Yan, J. (2016) *Extreme Value Modeling and Risk Analysis.* Methods and Applications. Taylor-Francis/CRC press.

Djeundje, V.A.B., Crook, J. (2019) Survival models for credit risks with time-varying coefficients. *European Journal of Operational Research*, 275 (1), 319-333.

Engelmann, B., Hayden, E., Tasche, D. (2003). Testing rating accuracy. *Risk*, 16(1), 82-86.

Fitzpatrick, T., Mues, C. (2016) An empirical comparison of classification algorithms for mortgage default prediction: evidence from distressed mortgage market. *European Journal of Operational Research*, 249, 427-439.

Gelfand A. E. , Sahu S. K. (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of American Statistical Association*, 94(445), 247-253.

Giesecke K., Weber S. (2006) Credit contagion and aggregate losses. *Journal of Economic Dynamics and Control*, 30(5), 741-767.

Goodstein, R., Hanouna, P., Ramirez, P. M., Christof, W., Stahel, S. W. (2017) Contagion effects in strategic mortgage defaults. *Journal of Financial Intermediation*, 30, 50-60.

Guiso, L., Sapienza, P., Zingales, L. (2013) The determinants of attitudes toward strategic default on mortgages. *The Journal of Finance*, 68(4), 1473-1515.

Gupta, A. Foreclosure contagion and the neighborhood spillover effects of mortgage defaults. *Journal of Finance*, 74(5), 2249-2301.

Hand, D. J. (2009) Measuring classifier performance: a coherent alternative to the area under the ROC curve, *Machine Learning*, 77, 103-123.

Hand D.J. (2010) Evaluating diagnostic tests: the area under the ROC curve and the balance of errors. *Statistics in Medicine*, 29 , 1502-1510.

Harding, J.P., Rosenblatt, E., Yao, V.W. (2009) The contagion effect of foreclosed properties. *Journal of Urban Economics*, 66, 164-178.

Holmes, C. C., Held L. (2006) Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1, 145-168.

Hwang, R. C. (2012) A varying-coefficient default model. *International Journal of Forecasting*, 28, 675-688.

Ibrahim, J. G., Chen, M., Sinha, D. (2001) *Bayesian Survival Analysis.* Springer.

Immergluck, D., Smith, G. (2006). The external costs of foreclosure: the impact of single-family mortgage foreclosures on property values. *Housing Policy Debate*, 17 (1), 57-79.

Kau, J. B., Keenan, D. C., Li, X. (2011) An Analysis of Mortgage Termination Risks: A Shared Frailty Approach with MSA-Level Random Effects. *Journal of Real Estate and Financial Economics*, 42, 51-67.

Kelly, R. (2011) The Good, The Bad and The Impaired: A Credit Risk Model of the Irish Mortgage Market. *Conference Paper*, Central Bank of Ireland.

Kim, S., Chen, M., Dey, D. K. (2008) Flexible generalized t-link models for binary response data. *Biometrika*, 95 (1), 93-106.

King, G., Zeng, L. (2001) Logistic Regression in Rare Events Data. *Political Analysis*, 9, 321-354.

Koenker, R., Yoon, J. (2009) Parametric links for binary choice models: A Fisherian-Bayesian colloquy. *Journal of Econometrics*, 152 (2), 120-130.

Kotz, S., Nadarajah, S. (2000)*Extreme Value Distributions. Theory and Applications*, Imperial College Press, London.

Krzanowski W.J., Hand D. J. (2009) *ROC Curves for Continuous Data.* Taylor & Francis, Inc., Boca Raton.

Leow M., Crook J. (2016) The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis. *European Journal of Operational Research*, 249 (2), 457-64.

LeSage J P, Pace R K. (2009) *Introduction to spatial econometrics.* Taylor-Francis/CRC press.

Li, D., Wang, X., Dey D. K. (2016) A flexible cure rate model for spatially correlated survival data based on generalized extreme value distribution and Gaussian process priors. *Biometrical Journal*, 58 (5), 11781197.

Lin, Z., Rosenblatt, E., Yao V. W., 2009. Spillover effects of foreclosure on neighborhood property values. *Journal of Real Estate Finance and Economics*, 38(4).

Longstaff, F. A. (2010) The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics*, 97(3), 436-450.

Pence, K. M. (2006) Foreclosing on opportunity: State laws and mortgage credit. *Review of Economics and Statistics*, 88(1), 177-182.

Raftery, A.E. and Lewis, S.M. (1992). One Long Run with Diagnostics: Implementation Strategies for Markov chain Monte Carlo. *Statistical Science*, 7, 493-497.

Schuetz, J., Been, V., Ellen, I. G. (2008) Neighborhood effects of concentrated mortgage foreclosures. *Journal of Housing Economics*, 17(4), 306-319.

Schwartz, A., Ellen, I., Voicu, I., Schill, M., 2003. Estimated External Effects of Subsidized Housing Investment on Property Values. Lincoln Institute of Land Policy. Working Paper WP03AS1.

Seiler, M., Collins, A., Fefferman, N. (2013) Strategic mortgage default in the context of a social network: an epidemiological approach. *Journal of Real Estate Research*, 35(4), 445-475.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*, 64, 583-639.

Stepanova, M., Thomas, L. (2002) Survival analysis methods for personal loan data. *Operations Research*, 50 (2), 277-289.

Sun, X., Xu W. (2014) Fast Implementation of DeLongs Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. *IEEE Signal Processing Letters*, 21, 13891393.

Tong, E. N. C., Mues, C., Thomas, L. C. (2012). Mixture cure models in credit scoring: if and when borrowers default. *European Journal of Operational Research*, 218(1), 132-139.

Towe, C., Lawley, C. (2013) The contagion effect of neighboring foreclosures. *American Economic Journal: Economic Policy*, 5(2), 313-35.

Wagner, H. (2004) The use of credit scoring in the mortgage industry. *Journal of Financial Services Marketing*, 9 (2), 179-183.

Wall, M. (2004) A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121, 311-324.

Wang, X., Dey, D. K. (2010) Generalized Extreme Value regression for binary response data: An application to B2B electronic payments system adoption. *Annals of Applied Statistics*, 4(4), 2000-2023.

Xu, S., Weichao, X. (2014) Fast Implementation of DeLongs Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. *IEEE Signal*

Processing Letters, 21, 1389-1393.

Zhu, L., Carlin, B. C. (2000) Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion. *Statistics in Medicine*, 19, 2265-2278.

Zhu, S., Pace, K. (2014) Modeling Spatially Interdependent Mortgage Decisions. The *Journal of Real Estate Finance and Economics*, 49(4), 598-620.

Appendix 1

We follow the following sequence to sample each parameter of the SGEVSUR model conditioned on all others:

(1) we sample the latent variable $Y_{it}^{*(r)}$ from a truncated GEV distribution $\text{TGEV}(\mathbf{x}_{it}'\boldsymbol{\beta}^{(r-1)}(s_i,t),\tau,\mu=0,\sigma=1)$ following equation (6);

(2) we sample the average $\eta_j^{(r)}$ for each covariate j from

$$N\left(\frac{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}\epsilon_{it}^{(r)}}{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2+1/100^4},\frac{1}{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2+1/100^4}\right) \tag{8}$$

where $\epsilon_{it}^{(r)} = Y_{it}^{*(r)} - \mathbf{x}_{it}'\boldsymbol{\beta}^{*(r)}(s_i,t)$ and $\boldsymbol{\beta}^{*(r)}(s_i,t)$ is $\boldsymbol{\beta}^{(r)}(s_i,t)$ with the element under consideration set to zero.

(3) we sample the spatial effect $\mu_j^{(r)}$ for each covariate j from $N([P_j^{(r)}+Q_j^{(r)}]^{-1}R_j^{(r)},[P_j^{(r)}+Q_j^{(r)}]^{-1})$ where $Q_j^{(r)}=(D_s-\rho_{sj}^{(r)}W_s)/\sigma_{sj}^{2(r)}$, P_j is a diagonal matrix where the generic element s-th (with $s=1,2,...,n_s$) is given by $\sum_{i|s_i=s}\sum_{t=1}^{t_i} x_{jit}^2$ and $R^{(r)}$ is the diagonal matrix where the generic element $s-th$ is given by $\sum_{i|s_i=s}\sum_{t=1}^{t_i} x_{jit}^2\epsilon_{it}^{(r)}$;

(4) we sample the temporal effect $\gamma_j^{(r)}$ for each covariate j from $N([P_j^{(r)}+Q_j^{(r)}]^{-1}R_j,[P_j^{(r)}+Q_j^{(r)}]^{-1})$ where $Q_j^{(r)}=(D_t-\rho_{tj}^{(r)}W_t)/\sigma_{tj}^{2(r)}$, $P_j^{(r)}$ is a diagonal matrix where the generic element t-th $(t=1,2,...,T)$ is given by $\sum_{i|t_i\geq t} x_{jit}^2$ and $R_j^{(r)}$ is the diagonal matrix where the generic element s-th is given by $\sum_{i|t_i\geq t} x_{jit}^2\epsilon_{it}^{(r)}$;

(5) we sample $\sigma_{sj}^{2(r)}$ from $\text{InvGamma}(n_s/2+a_1,\boldsymbol{\mu}_j'(D_s-\rho_{sj}^{(r)}W_s)\boldsymbol{\mu}_j/2+b_1)$;

(6) we sample $\sigma_{tj}^{2(r)}$ from InvGamma$(n_t/2 + a_1, \boldsymbol{\gamma}_j'(D_t - \rho_{tj}^{(r)} W_t)\boldsymbol{\gamma}_j/2 + b_1)$.

(7) we sample $\rho_{sj}^{(r)}$ and $\rho_{tj}^{(r)}$ from two discrete conditional distributions proportional to the product between the discrete $Beta(1,1)$ distribution function and the CAR density.

We obtain the posterior distribution (8) **as follows**

$$
\begin{aligned}
\pi(\eta_j^{(r)}|D) &\propto \prod_{i=1}^{n} \{[p(t_i, \boldsymbol{\theta})]^{\delta_i} [S(t_i, \boldsymbol{\theta})]^{1-\delta_i}\} \pi(\eta_j^{(r)}) \\
&\propto \exp\left\{ -\frac{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2}{2} \sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}\epsilon^{(r)} \right\} \exp\left\{ -\frac{1}{2\cdot 100^4} \eta_j^{2(r)} \right\} \\
&\propto \exp\left\{ -\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2 + 1/100^4 \right) \eta_j^{2(r)} + \eta_j^{(r)} \sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}\epsilon_{it}^{(r)} \right\} \\
&= \exp\left\{ -\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2 + 1/100^4 \right) \left[\eta_j^{2(r)} - 2\eta_j^{(r)} \frac{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}\epsilon_{it}^{(r)}}{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2 + 1/100^4} \right] \right\} \\
&= \exp\left\{ -\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2 + 1/100^4 \right) \left[\eta_j^{(r)} - \frac{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}\epsilon_{it}^{(r)}}{\sum_{i=1}^{n}\sum_{t=1}^{t_i} x_{jit}^2 + 1/100^4} \right]^2 \right\}.
\end{aligned}
$$

We compute the posterior distributions for the spatial effect $\mu_j^{(r)}$**, the temporal effect** $\gamma_j^{(r)}$ **and the variances** $\sigma_{sj}^{2(r)}$ **and** $\sigma_{tj}^{2(r)}$ **following Chang et al (2013) and Banerjee et al (2004).**

Appendix 2

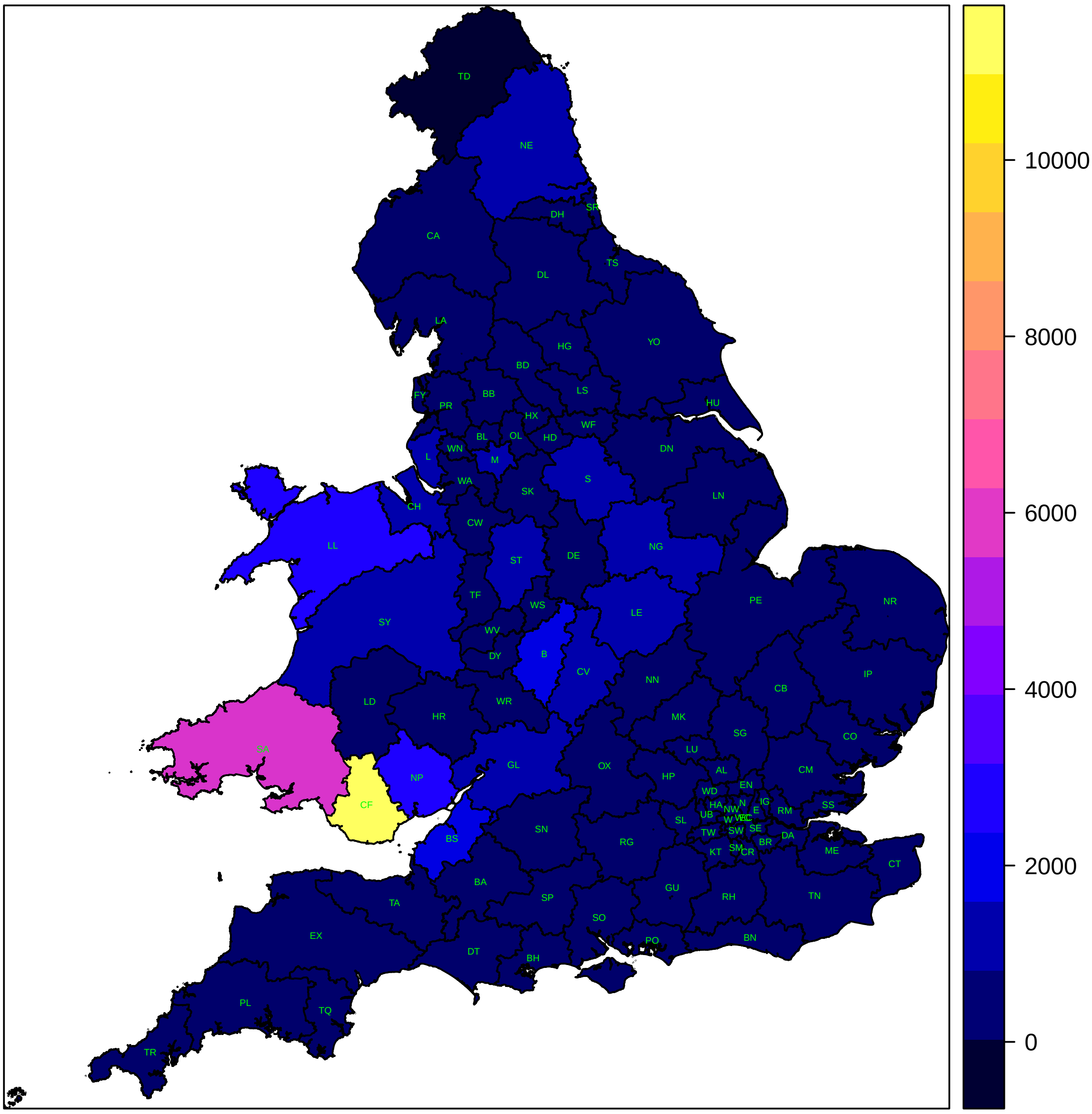


Figure 1: The mortgage distribution for postcode areas in England and Wales.

Table 4: Posterior mean of the parameter $\beta(s,t)$ for the SGEVSUR model (τ=-0.30) with temporal, spatial and spatial-temporal components. The significance levels are based on approximations under the normality assumption of the parameter estimates in conjunction with the mean and the standard deviation of each parameter chain. $^{*}p - value \leq 0.1$; $^{**}p - value \leq 0.05$

Variable	*Spatial-temporal model*	*Temporal model*	*Spatial model*
Intercept	-1.9515**	-1.9753**	-1.9495**
Loan to valuation	**0.0003****	0.0001	**0.0003****
N of applicants	-0.0026	0.0041	-0.0057
Balance	$\mathbf{3.4330{\cdot}10^{-6}}$ **	$\mathbf{-1.6861{\cdot}10^{-7}}$ **	$\mathbf{-5.4612{\cdot}10^{-8}}$ **
Contractual repayment	$-5.7456\cdot10^{-6}$	$-2.5280{\cdot}10^{-5}$	$\mathbf{3.8781{\cdot}10^{-5}}$ **
Interest rate	**0.3470****	**0.3244** **	**0.1439***
Property value	$\mathbf{-6.4591{\cdot}10^{-8}}$ **	$2.1821{\cdot}10^{-8}$	$\mathbf{-1.5957{\cdot}10^{-7}}$ **
Married	0.0002	0.0005	-0.0007
Buy To Let	**-0.0086***	-0.0071	**-0.0044** *
Repay	**-0.0095***	-0.0068	-0.0135
Fixed	-0.0118	-0.0064	-0.0025
Flexible	-0.0053	-0.0028	-0.0018
Tracker	-0.0070	-0.0018	$-3.2935{\cdot}10^{-8}$
HPI	**0.0013***	**0.0002***	0.0002
Production	-0.0003	0.0003	-0.0003
UN	0**.0015***	**0.0046** *	-0.0021
IR	**0.0011****	**0.0063***	-0.0030
Consumer	-0.0011	-0.0017	**-0.0030** *

Table 5: Posterior mean of the spatial $\boldsymbol{\rho}_s$ and temporal $\boldsymbol{\rho}_t$ autocorrelation parameters for the GEV model with temporal, spatial and spatial-temporal components. The significance levels are based on approximations under the normality assumption of the parameter estimates in conjunction with the mean and the standard deviation of each parameter chain. $^{*}p - value \leq 0.1$; $^{**}p - value \leq 0.05$

Variable	*Spatial-temporal model*		*Temporal model*	*Spatial model*
	Spatial aut par	*Temporal aut par*	*Temporal aut par*	*Spatial aut par*
Loan to valuation	0.2179			0.2336
N of applicants	**0.3797***			**0.3607***
Balance	0.2205	0.1668	0.1674	**0.2220***
Contractual repayment	0.2156	0.1687	0.1686	0.2142
Interest rate	**0.4494***	**0.4905****	**0.4018***	**0.4369****
Property value	**0.2092****	**0.1688***	**0.1737***	**0.2305***
Married	**0.3677***			**0.3530***
Buy To Let	**0.3749***			**0.3842***
Repay	**0.3812***			**0.4431***
Fixed	**0.3926***			**0.3890***
Flexible	**0.4228***			**0.4100***
Tracker	**0.3790****			**0.3647***
HPI		**0.2458***	0.2114	
Production		0.1758	0.1696	
UN		**0.2798***	**0.3197***	
IR		**0.3233***	**0.3417***	
Consumer		**0.2415***	0.2404	

Table 6: Forecasting accuracy measures for the SGEVSUR (τ=-0.30) and probit model with fixed parameters over time and over space (Independent) and with temporal, spatial and spatial-temporal components for different forecasting time horizons (12, 24 and 36 months).

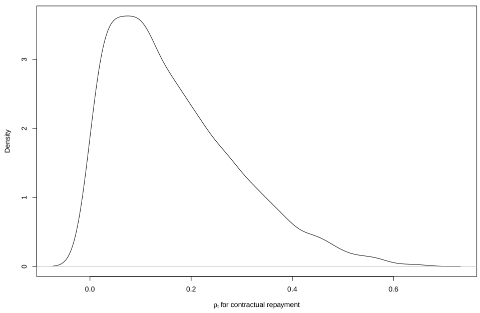
Model	*Time horizon*	*Measure*	*Spatial-Temporal*	*Temporal*	*Spatial*	*Independent*
SGEVSUR	12	AUC	0.8250	0.8328	0.8305	0.7853
		KS	0.5778	0.5765	0.5828	0.5521
		H	0.0984	0.1121	0.0983	0.0723
		M	0.3695	0.4049	0.4130	0.4239
probit	12	AUC	0.7267	0.7041	0.7162	0.6736
		KS	0.4044	0.3359	0.3907	0.3125
		H	0.0343	0.0365	0.0336	0.0215
		M	0.4130	0.4239	0.4293	0.4375
SGEVSUR	24	AUC	0.7332	0.7376	0.7362	0.7029
		KS	0.4736	0.4743	0.4784	0.4471
		H	0.0743	0.0845	0.0742	0.0624
		M	0.3109	0.3435	0.3658	0.3862
probit	24	AUC	0.6672	0.6651	0.6576	0.6263
		KS	0.3327	0.2876	0.3274	0.2710
		H	0.0263	0.0222	0.0273	0.0203
		M	0.3618	0.3658	0.3780	0.3841
SGEVSUR	36	AUC	0.7250	0.7275	0.7291	0.6945
		KS	0.4552	0.4469	0.4584	0.4268
		H	0.0697	0.0771	0.0695	0.0616
		M	0.3126	0.3427	0.3710	0.3823
probit	36	AUC	0.6610	0.6622	0.6638	0.6519
		KS	0.2865	0.2884	0.3158	0.2573
		H	0.0250	0.0261	0.0201	0.0192
		M	0.3653	0.3710	0.3804	0.3879

Table 7: The DeLong-DeLong test for comparing the AUC of the SGEVSUR (τ=-0.30) with temporal, spatial and spatial-temporal components for different forecasting time horizons (12, 24 and 36 months). The p-values are computed following the approach described in Sun and Xu (2014) and implemented in the function *roc.test* of the R package pROC.

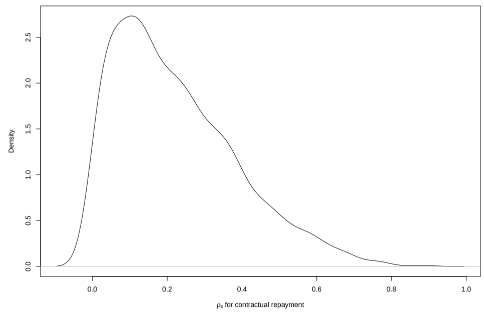
Time horizon	*Compared models*	p-value
12	Spatial vs Temporal	0.192
	Spatial-Temporal vs Temporal	0.061
	Spatial-Temporal vs Spatial	0.068
24	Spatial vs Temporal	0.226
	Spatial-Temporal vs Temporal	0.098
	Spatial-Temporal vs Spatial	0.077
36	Spatial vs Temporal	0.234
	Spatial-Temporal vs Temporal	0.072
	Spatial-Temporal vs Spatial	0.115

Table 8: The mortgage distribution for postcode areas in England and Wales.

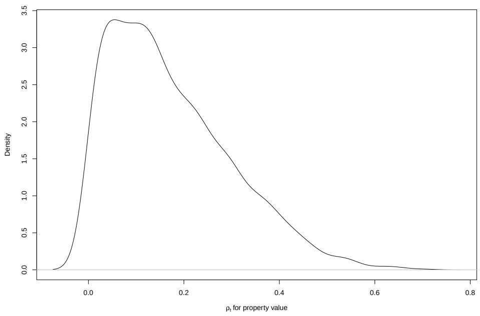
Postcode	*Frequency*	*Postcode*	*Frequency*	*Postcode*	*Frequency*	*Postcode*	*Frequency*
AL	162	E	582	MK	460	SP	150
B	2246	EC	22	N	511	SR	214
BA	626	EN	216	NE	1193	SS	330
BB	367	EX	736	NG	1166	ST	925
BD	492	FY	289	NN	492	SW	734
BH	459	GL	1304	NP	3122	SY	1124
BL	335	GU	478	NR	688	TA	346
BN	625	HA	267	NW	279	TD	13
BR	207	HD	259	OL	469	TF	479
BS	1678	HG	152	OX	581	TN	435
CA	304	HP	294	PE	718	TQ	412
CB	293	HR	471	PL	628	TR	526
CF	10985	HU	440	PO	608	TS	522
CH	1344	HX	181	PR	577	TW	299
CM	487	IG	182	RG	635	UB	211
CO	313	IP	514	RH	344	W	337
CR	224	KT	374	RM	310	WA	729
CT	327	L	830	S	1090	WC	12
CV	1007	LA	281	SA	5706	WD	215
CW	446	LD	255	SE	760	WF	381
DA	260	LE	949	SG	319	WN	357
DE	740	LL	2660	SH	1	WR	463
DH	270	LN	266	SK	751	WS	492
DL	386	LS	701	SL	238	WV	425
DN	752	LU	216	SM	149	YO	668
DT	204	M	1130	SN	504		
DY	550	ME	379	SO	466		



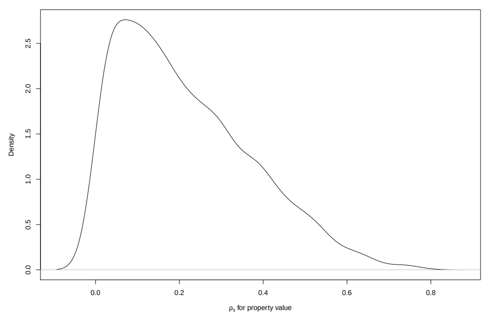
(a) ρ_t for contractual repayment



(b) ρ_s for contractual repayment



(c) ρ_t for property value



(d) ρ_s for property value

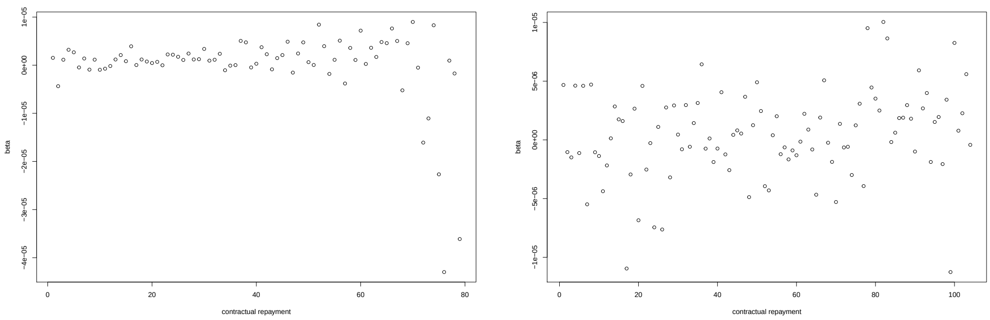
Figure 2: Posterior distributions for the temporal ρ_t and the spatial ρ_s autocorrelation parameters in the SGEVSUR model with spatial-temporal components.

Table 9: The I-statistic for the parameter ρ_s and ρ_t . This is a convergence diagnostic of the parameter estimates proposed by Raftery and Lewis (1992) that should be smaller than 5. We use the function *Raftery.Diagnostic* in the R package *coda*.

Variable	*Spatial-temporal model*		*Temporal model*	*Spatial model*
	Spatial aut par	*Temporal aut par*	*Temporal aut par*	*Spatial aut par*
Loan to valuation	1.2751			1.3046
N of applicants	1.3245			1.3632
Balance	1.1858	1.2427	1.3332	1.4874
Contractual repayment	1.2101	1.3648	1.4385	1.5516
Interest rate	1.3524	1.2638	1.2486	1.3823
Property value	1.1742	1.2651	1.3045	1.2807
Married	1.2513			1.3468
Buy To Let	1.4027			1.4471
Repay	1.3213			1.3458
Fixed	1.2592			1.2844
Flexible	1.3731			1.4266
Tracker	1.2513			1.2236
HPI		1.3249	1.3888	
Production		1.1271	1.1733	
UN		1.2715	1.2932	
IR		1.4838	1.3944	
Consumer		1.3711	1.2546	

Table 10: Goodness of fit for the SGEVSUR models with spatial-temporal components for τ=-0.30, -0.6,-0.9,-1 for the training sample.

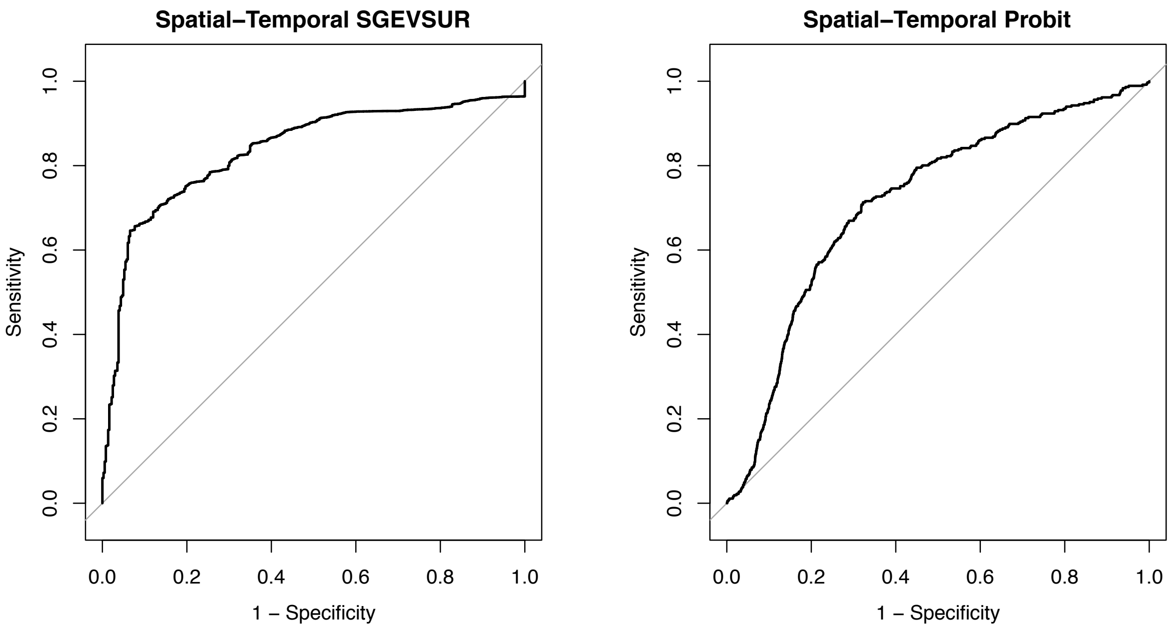
τ	DIC
-1	40,562
-0.9	40,505
-0.6	40,481
-0.3	40,436



(a) β over time for West London

(b) β over regions on June 2006

Figure 3: Posterior mean of the parameter β varying over time and over space for the variable *contractual repayment* in the SGEVSUR model.



(a) (b)

Figure 4: ROC curves for the SGEVSUR and probit models with spatial-temporal components for a time horizon of 12 months.