Statistical Modelling 2021; **21(1–2)**: 137–160

Spatial survival modelling of business re-opening after Katrina: Survival modelling compared to spatial probit modelling of re-opening within 3, 6 or 12 months

Roger S Bivand[1] and Virgilio Gómez-Rubio[2]

[1]Department of Economics, Norwegian School of Economics, Bergen, Norway
[2]Department of Mathematics, School of Industrial Engineering - Albacete, University of Castilla-La Mancha, Albacete, Spain

Abstract: Zhou and Hanson; Zhou and Hanson; Zhou and Hanson (2015, *Nonparametric Bayesian Inference in Biostatistics*, pages 215–46. Cham: Springer; 2018, *Journal of the American Statistical Association*, **113**, 571–81; 2020, spBayesSurv: Bayesian Modeling and Analysis of Spatially Correlated Survival Data. *R package version 1.1.4*) and Zhou et al. (2020, *Journal of Statistical Software, Articles*, **92**, 1–33) present methods for estimating spatial survival models using areal data. This article applies their methods to a dataset recording New Orleans business decisions to re-open after Hurricane Katrina; the data were included in LeSage et al. (2011b, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **174**, 1007—27). In two articles (LeSage etal., 2011a, *Significance*, **8**, 160—63; 2011b, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **174**, 1007—27), spatial probit models are used to model spatial dependence in this dataset, with decisions to re-open aggregated to the first 90, 180 and 360 days. We re-cast the problem as one of examining the time-to-event records in the data, right-censored as observations ceased before 175 businesses had re-opened; we omit businesses already re-opened when observations began on Day 41. We are interested in checking whether the conclusions about the covariates using aspatial and spatial probit models are modified when applying survival and spatial survival models estimated using MCMC and INLA. In general, we find that the same covariates are associated with re-opening decisions in both modelling approaches. We do however find that data collected from three streets differ substantially, and that the streets are probably better handled separately or that the street effect should be included explicitly.

Key words: business decisions, spatial survival models, spatial probit models, Bayesian inference, integrated nested Laplace approximation, Markov chain Monte Carlo

Received: October 2019; **revised:** June 2020; **accepted:** September 2020

1 Introduction

Often it is far from clear how one should represent spatio-temporal data in seeking to model outcomes, not least when the covariates do not vary in time. Rather than

For coloured figures, please use the online version of this article.

Address for correspondence: Roger S Bivand, Department of Economics, Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway.
E-mail: roger.bivand@nhh.no

10.1177/1471082X20967158

proposing methods innovations, we are here concerned with the choice of methods that may be applied to a specific, well-known case: the re-opening of businesses in New Orleans following closure caused by Hurricane Katrina in 2005. Mandatory evacuation was ordered on 28 August 2005, and although by mid-September residents were starting to return, a further evacuation was ordered on 19 September because of the threat of Hurricane Rita to weakened levees. Shortly after Rita's passage, residents began to return to the city, and businesses began to repair storm and flood damage and to re-open following the gradual restoration of power.

In two articles (LeSage et al., 2011a,b), spatial dependence in New Orleans business decisions to re-open following Hurricane Katrina in 2005 was analysed using spatial probit models. A third article (Lam et al., 2012) uses data from three successive telephone surveys of businesses and is not directly comparable. The data (`https://rss.onlinelibrary.wiley.com/hub/journal/1467985x/series-a-datasets/pre_2016`) used for the analyses were included in LeSage et al. (2011b), and as detailed in LeSage et al. (2011a), 673 businesses were observed on three streets initially at weekly intervals starting 41 days after the impact of the hurricane (60 businesses open). Observations ceased after 449 days, leaving 175 businesses still closed. In the published spatial probit analyses of this dataset, decisions to re-open were aggregated to the first 90, 180 and 360 days from 29 August 2005.

If we re-cast the problem to one of examining the time-to-event records in the data (Ibrahim et al., 2001), left-censored as businesses could not re-open until not flooded areas began to be repopulated (about 6 weeks; observations began shortly afterwards; we omit businesses already open when observations began from all analyses), and right-censored as observations ceased before 175 businesses re-opened. We are interested in checking whether the conclusions about the covariates using aspatial and spatial probit models are modified when applying survival and spatial survival models (Zhou and Hanson, 2015, 2018, 2020; Zhou et al., 2020). We are also interested in exploring the hypothesis, raised in LeSage et al. (2011b) in the context of effects estimates, that the three streets observed in data collection should not be pooled. We find that while the relative contributions of pooled spatial survival model covariates match those of pooled spatial probit models, it is clear that these pooled effects seem to cancel out the outcomes for the three different streets.

2 The Katrina business re-opening dataset

LeSage et al. (2011b) provided data and code for their analysis, and this data is included in Wilhelm and de Matos (2013, 2015). The dataset is included in two versions; we will use the 'raw' version. The data as provided has columns for re-opening date, geographical coordinates, elevation and flood depth in whole feet, median household income of surrounding area, ownership, employment size and socio-economic status for the census block group at business locations—the latter three as categorical variables each with three levels. Figure 1 shows the counts

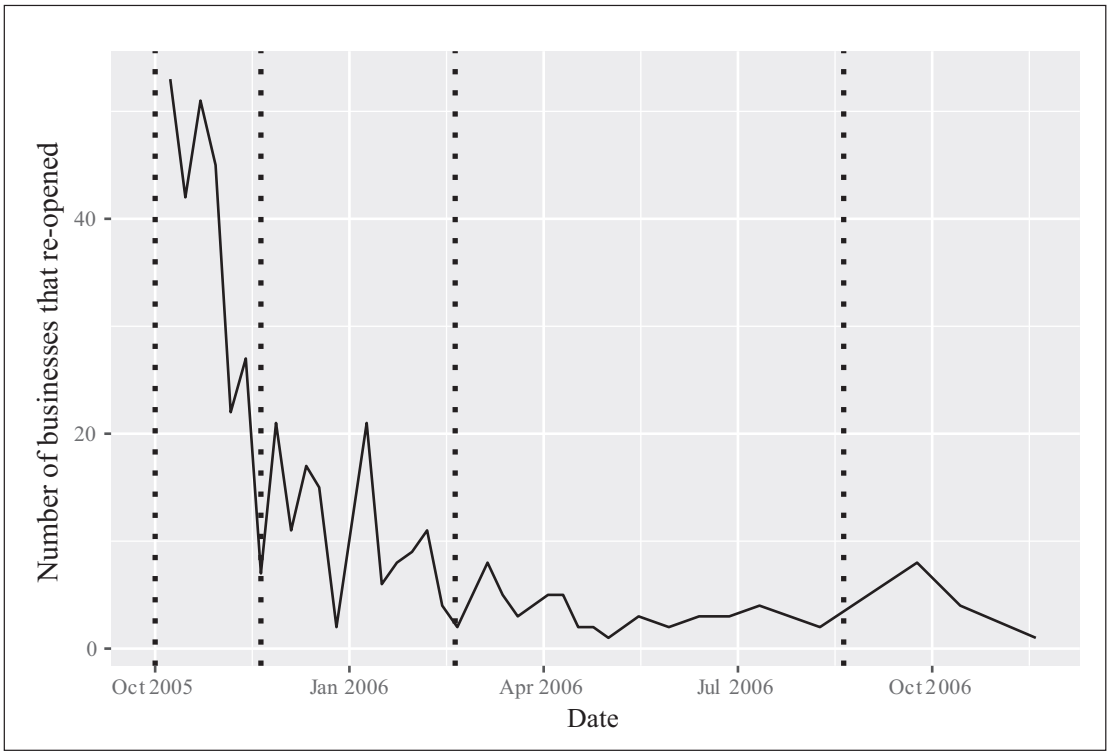

Figure 1 Counts of businesses that re-opened after Day 41 until the end of the data collection by week of observation. About 175 businesses were still closed when observation ceased; the 60 businesses that were already open on Day 41 are not included. The vertical dotted lines show Days: 41 when observations began, and 90, 180 and 360 used to divide the data in the original study

of businesses that re-opened at each observation as a time series. Observations were taken by cycling along the three streets and noting whether businesses had re-opened. In most cases until May 2006, observations were approximately weekly, and fortnightly or less frequently after that. This means that although time is reported in days in the dataset provided, it actually has a close-to-weekly reporting interval for most of the observation period, extending to observations every second or third week later.

LeSage et al. (2011b, p. 1016) discuss the reasons for casting the model in a cross-sectional form rather than a spatio-temporal form, based on information spillover prior to re-opening, and the fact that the covariates do not change over time. They chose to slice the observations at 90, 180 and 360 days, modelling decisions to re-open during the first three and six months, and the first year after Katrina. Figure 2 and Table 1 indicate that the three selected business streets differ in the numbers of businesses and in the proportion re-opening during the first three months. In particular, over half of the businesses were on Magazine Street, and of these, over half had re-opened during the first three-month period. Indeed,

Table 1 Counts of re-opened businesses by street and time intervals used in the original study: 0–90, 90–180, 180–360 and over 360 days within the observation period

	Magazine Street	Carrollton Avenue	St. Claude Avenue
Day 0–90	263	30	7
Day 90–180	83	26	23
Day 180–360	20	23	10
Day 360 +	4	5	4
Never	31	95	49
Sum	401	179	93

over 90% of the businesses on Magazine Street re-opened during the observation period, but less than 50% of the businesses on the other two streets re-opened in the same period.

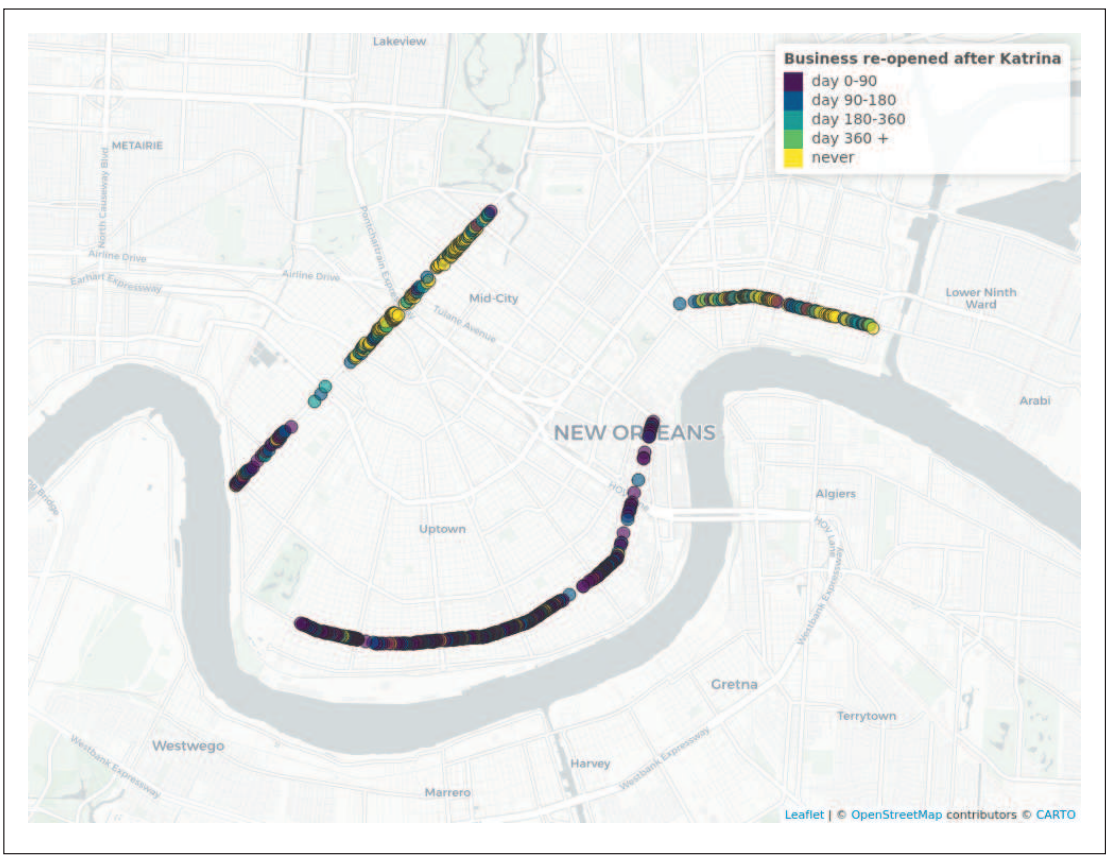

Figure 2 The location of the three streets in central New Orleans, with businesses marked using the time intervals used in the original study: 0–90, 90–180, 180–360 and over 360 days within the observation period

Table 2 'Still closed' status percentages by street after removing left-censored observations

	Magazine Street	Carrollton Avenue	St. Claude Avenue
Still closed	8.96	54.29	53.26
Re-opened	91.04	45.71	46.74

We already know that many businesses never re-opened within the observation period, so we must acknowledge that their status was surviving closed giving no information about whether or when they re-opened; in survival analysis these are right-censored. There is a more difficult problem regarding businesses that had re-opened before observations began (left-censored) and were recorded as open on Day 41. We choose to drop these observations, leaving a right-censored dataset; the percentages of still closed and re-opened businesses by street are shown in Table 2 for the dataset without left-censored observations, with this variable constituting survival status.

Of the 60 left-censored observations, 55 were on Magazine Street (16% of businesses on that street), 4 on Carrollton Street (2%), and 1 on St. Claude Avenue (1%). Almost 90% of the left-censored businesses had sole proprietors; almost half of the large businesses by employment size on Magazine Street were left-censored.

The first section of the Appendix shows some summary statistics about the covariates included in the models. As it can be seen, because of the lack of covariate variability by street presented in the Appendix, it became necessary to recode the three categorical variables to sole proprietor versus chain for ownership type, to small versus average for employment size, and lower versus average for socio-economic status. These are the recoded categorical covariates used in analysis below, with chain ownership type, average employment size and socio-economic status being base categories.

3 Survival modelling

3.1 Introduction

Turning to survival modelling (Therneau, 2019; Therneau and Grambsch, 2000; Moore, 2016), we may be able to disentangle the relationships contained in the business re-opening dataset but perhaps hidden by the temporal aggregation of decisions by time intervals. Event times are expressed in fractions of 7-day units starting from the date observations began, thus removing left-censoring and aligning the observational units with the time intervals between observations.

In survival analysis, if T denotes the survival time, the interest is in estimating the survival function at time t, $S(t) = \pi(T > t)$, the probability of the event occurring at least at time t. The survival function can be estimated in a number of ways, with the Kaplan–Meier estimator (Kaplan and Meier, 1958) a commonly used non-parametric method.

Parametric estimates can be made by using a convenient model for the data. We will use a Weibull likelihood as this is a common and convenient choice:

$$f(t \mid \alpha, \lambda) = \lambda\alpha t^{\alpha-1} \exp\{-\lambda t^{\alpha}\}. \tag{3.1}$$

Here, α is a shape parameter and λ a scale parameter. The hazard function $h(t \mid \alpha, \lambda)$ is the probability of observing the event right after time t and it holds that $h(t) = f(t)/S(t)$. For a Weibull likelihood, we have that $h(t \mid \alpha, \lambda) = \lambda\alpha t^{\alpha-1}$.

As covariates are available, we will rely on accelerated failure time (AFT) models (Ibrahim et al., 2001), which can be regarded as generalized linear models for survival data. AFT models consider $t^* = \log(T)$ and model it as $t^* = \mathbf{x}^\top\boldsymbol{\beta} + \sigma\varepsilon$. Here, σ is an error parameter that scales the error term ε, which can take different distributions. When an extreme value distribution is used, the resulting likelihood is Weibull with hazard distribution equal to

$$h(t \mid \alpha, \lambda) = \lambda\alpha t^{\alpha-1} \exp\{\mathbf{x}^\top\boldsymbol{\beta}\}.$$

Here, $\mathbf{x}$ is a column vector of covariates with associated coefficients $\boldsymbol{\beta}$. Note that the linear predictor on the covariates $\mathbf{x}^\top\boldsymbol{\beta}$ can also include other terms such as different types of random effects.

A key feature of survival models is that they account for censoring when the likelihood of the models is built. For each time t_i, $i = 1, \ldots, n$, δ_i represents the censoring status (with 1 for observed event, and 0 for censored), with n the number of businesses equal to 613 (as the left-censored businesses are not considered in the analysis). The likelihood of the model becomes

$$\mathcal{L}(\mathbf{t} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i \mid \boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta})^{1-\delta_i} = \prod_{i=1}^{n} h(t_i \mid \boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta}).$$

Here, $\mathbf{t} = (t_1, \ldots, t_n)$ and $\boldsymbol{\theta}$ is the vector of model parameters, which will include the scale and shape parameters and, possibly, covariates, latent effects and other hyperparameters.

Frailty models (Ibrahim et al., 2001) are survival models that include random effects with different structures. These effects are often used to model within-subject variation or cluster effects. In this case, the linear predictor in the covariates includes a random effect u_j, that is, it is now $\mathbf{x}_i^\top\boldsymbol{\beta} + u_j$. In its simplest form, the distribution of u_j is a Gaussian distribution with zero mean and precision τ_u^2. However, correlated random effects can be used, as explained below.

Note that random effects u_j are indexed over $j = 1, \ldots, J$, where the value of J depends on how the random effects are defined. When the random effects are set on the streets, then $J = 3$. If a different random effect is considered for each business, as when spatial frailties are included in the model (see below), then $J = n$.

In a Bayesian framework, the interest is in computing the posterior distribution of the model parameters $\pi(\boldsymbol{\theta} \mid \mathbf{y})$, that is, the distribution of $\boldsymbol{\theta}$ given the ensemble of data $\mathbf{y}$ (that includes observed times, censorship status, covariates and any other data).

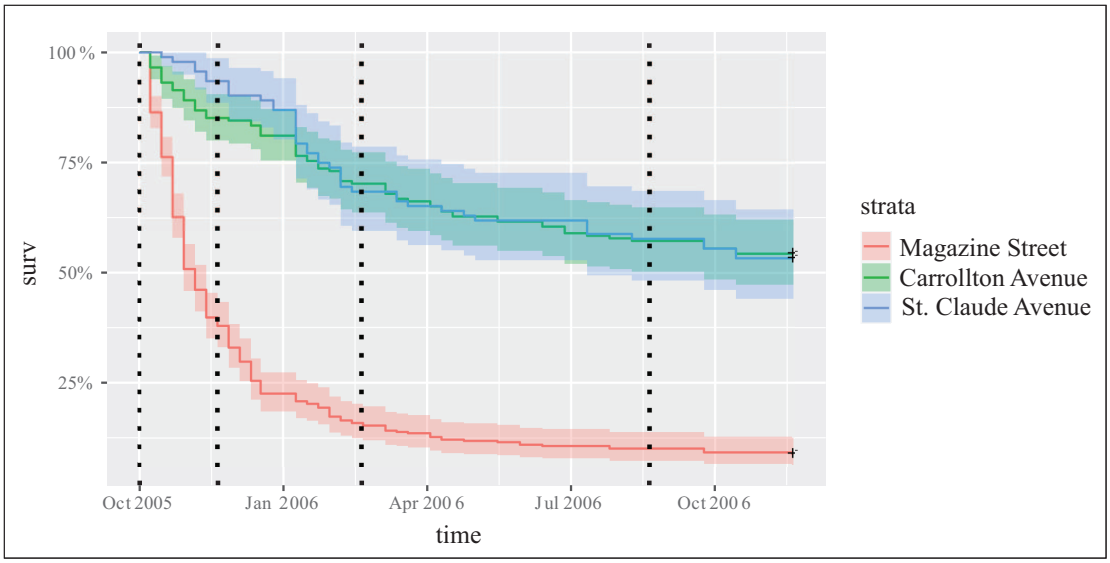

Figure 3 Kaplan–Meier survival curves for observations from Day 41 by street showing 95% confidence intervals. The dashed vertical lines show the left-censoring cut-off, and 90, 180 and 360 day slices after Katrina used in probit modelling

Furthermore, the posterior distribution of the survival function, $\pi(S(t) \mid \mathbf{y})$ is often computed from the joint posterior distribution of the model parameters.

The posterior distribution is obtained by using Bayes' rule, so that the posterior is $\pi(\boldsymbol{\theta} \mid \mathbf{y}) \propto \mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$. Here, $\pi(\boldsymbol{\theta})$ is the prior distribution of the ensemble of model parameters, and it can often be split into the product of lower-dimensional (often univariate) distributions. Vague prior distributions will be used for all the model parameters. In particular, coefficients of fixed effects are assigned a Gaussian prior with zero mean and small precisions, and precision parameters are assigned Gamma priors with a large variance.

Figure 3 shows the Kaplan–Meier survival curves with the three streets treated as a categorical covariate, where, before taking other covariates into account, the survival curve of Magazine Street differs greatly from the other two streets throughout the observation period. Businesses in Magazine Street have shorter survival times. Recall that survival means surviving still closed, so that re-opening changes the status of the observation from closed to open. Hence, businesses in Magazine Street tend to open faster than those in the other two streets. The survival curve confidence intervals of Carrollton Avenue and St. Claude Avenue overlap from February 2006, but before that more businesses re-open on Carrollton Avenue than St. Claude Avenue. The weekly observation sequence is clearly seen in the vertical steps in the curves shown in Figure 3, because business re-opening is recorded on the same day, although it may have occurred at any time since the previous observations were made.

3.2 Bayesian inference

We have chosen to fit a number of survival models for subsequent comparison with spatial survival, probit and spatial probit models. The survival models are all fitted using the Weibull distribution. They differ in using only the included covariates, and the included covariates and a street random effect. All these models have been fit using `inla()` from the INLA package (Rue et al., 2019) and with `survregbayes` from the spBayesSurv package (Zhou and Hanson, 2020). Time for models fit with INLA have been re-scaled (by dividing by 60) to avoid computational problems; however, this will not affect the estimates of the effects in the linear predictor but only those of the scale parameter λ. Furthermore, `survregbayes` estimates the baseline hazard function semi-parametrically, which differs from INLA, and the parameterization of the hyperparameters α and λ of the Weibull distribution is also differnt from INLA (see Zhou and Hanson, 2020, for details).

Priors used with `survregbayes` are vague, and they correspond to the default ones (see Zhou and Hanson, 2020, for details), that is, a Gaussian prior with zero mean and precision 10^{-10} for the fixed effects. Priors for INLA models are taken accordingly whenever possible (see below). We take 200 000 draws, discarding 50 000 in all runs; fitting the frailty models showed that many draws were needed, so long runs are used here too for consistency. Finally, we have also tried pooled models with by-street dummy variables, but these lead to poor converge of the MCMC simulations ad we have discarded them in the pooled models.

The included covariates are elevation in whole feet from the lowest point (-3.5 feet) as non-negative feet, the square of this elevation value, the two-category ownership type variable (sole owner, chain), the two-category census block group socio-economic status variable (lower, average), the logarithm of census block group median income and the two-category employment size variable (small, average). Note also that estimates obtained with `inla()` rely on Bayesian inference with the integrated nested Laplace approximation method (Rue et al., 2009; Gómez-Rubio, 2020), which provides approximations to the posterior marginal distributions of the model parameters. As discussed below, INLA method for estimation is also based on slightly different parameterizations of the survival models. As a robustness check, we have included maximum likelihood (ML) estimates of these models in the Appendix.

The covariate coefficients of the INLA estimates had reversed signs (due to a different parameterization of the model), which for easier comparison have here been multiplied by -1. This sign shift is required because the parameterization of the Weibull survival distribution in the INLA package is that shown in Equation 3.1. In addition, when covariates and other effects are introduced in the model as a linear predictor η, this is linked to the λ parameter as $\log(\lambda) = \eta$. Priors used to fit this model are Gaussian with zero mean and precision of 10^{-10} for the fixed effects and a Gamma with parameters 0.001 and 0.001 for the precision of the frailty effect, that is, Gaussian i.i.d. random effects. These priors match those used by `spBayesSurv` for the same parameters. The prior on the α parameter is set in the log-scale and is a penalized complexity prior (Simpson et al., 2017), as described in the documentation in the INLA package that can be accessed with `inla.doc("pc.alphaw")`.

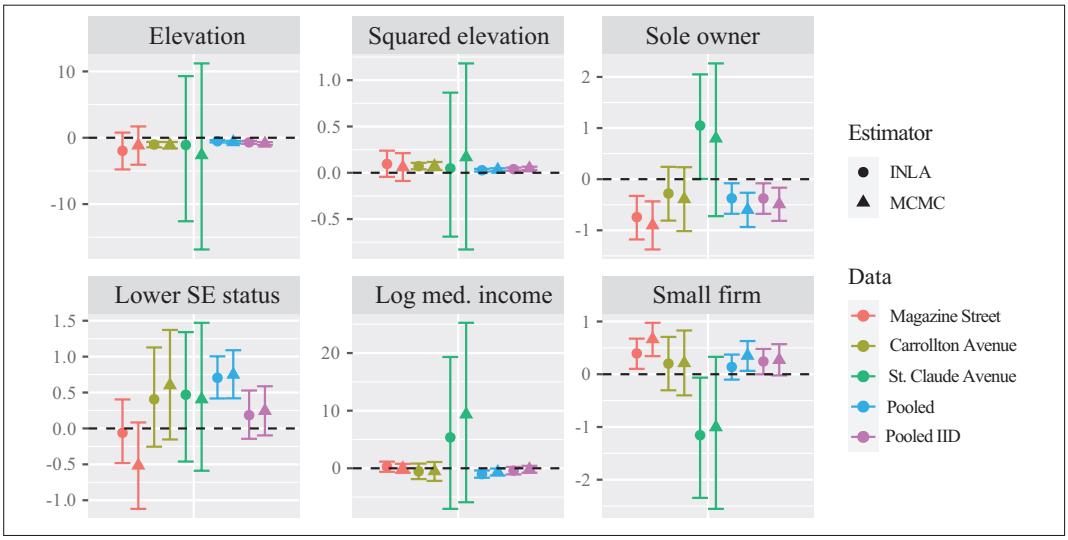

Figure 4 AFT Weibull survival model coefficients and 95% credible interval bars for models for each street, a pooled model and a pooled model with street IID random effects, estimated using INLA and MCMC

Figure 4 displays the estimates of the fixed effects for INLA and MCMC. Here and throughout the article, positive effects indicate a longer survival, that is, later re-opening. Figure 5 shows the posterior marginals of the by-street random effects. While the random effects estimates for Carrollton Avenue and Magazine Street do not seem to add much to the outcome, the random effect for St. Claude Avenue does matter. We have also reversed the signs here, and interpret the location of businesses on St. Claude Avenue as slowing re-opening.

Both Figures 4 and 5 show the differences between the streets, and between the by-street models and the pooled models. In many covariates, the credible intervals for St. Claude Avenue are much wider than for other models, with the differences

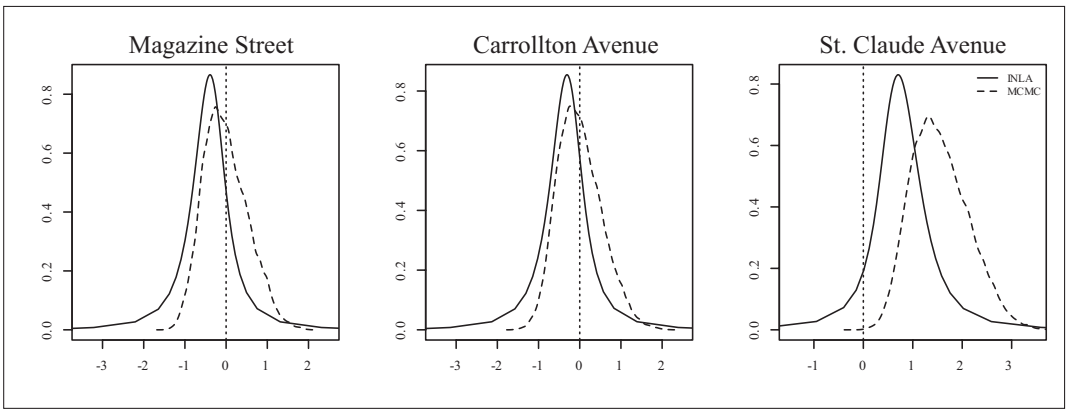

Figure 5 Three MCMC and INLA IID street random effects estimated using INLA and MCMC

far exceeding the effect of the differences in numbers of observations. In the case of ownership and firm employment size, the St. Claude Avenue coefficients also differ in sign from the other models. It is perhaps not surprising that the pooled models and the Magazine Street model are broadly similar, as there are many more businesses on Magazine Street than either of the other two streets. In the pooled models and the Magazine Street model, when the business is not part of a chain, re-opening seems to occur faster, but smaller firms by employment re-open slower. Businesses on St. Claude Avenue show reversed tendencies. We also feel that the predominance in the dataset of Magazine Street, where no businesses were flooded and where most of the businesses re-opening as soon as permitted were located, suggests that the use of the dataset to elicit clear relationships between covariates and the decision to re-open is in fact challenging. Had the data collection tried to match observations by covariate across flooding impact, more might have become clear. We find that the coefficient estimates on covariates vary substantially between streets, and show in the appended extended discussion below that the streets differ greatly by covariates.

In order to assess MCMC model fitting, Cox–Snell residuals (Cox and Snell, 1968) can be computed. These are defined as $r_i = -\log(S_i(t_i))$. If the model fits the data well, then the distribution of the Cox–Snell residuals follow an exponential distribution with rate equal to 1. This can be used to produce Cox–Snell diagnostic plots by plotting the estimates of r_i (from the model fit) against corresponding values for a exponential distribution of rate 1. If points are close to the identity line, then model fitting is good. Figure 6 shows Cox–Snell diagnostic plots for the MCMC models without frailty components (Zhou and Hanson, 2018). In general, model fitting looks better for the pooled models than for any of the three by-street models; naturally, model fitting is affected by the number of observations and the degree of right-censoring, so that the pooled model, the pooled model with a street IID random effect and the Magazine Street model .

3.3 Probit modelling

Next, we will consider a binary response variable z_i that indicates whether the business has re-opened ($z_i = 1$) or remains closed ($z_i = 0$) in a given time interval. This is modelled using a probit model, so that each z_i follows a Bernoulli distribution with probability p_i, and this is linked to the linear predictor as

$$\Phi^{-1}(p_i) = \mathbf{x}^\top \boldsymbol{\beta}, \ i = 1, \ldots, n.$$

Here, $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution of a standard Gaussian distribution (i.e., the probit function). For comparison, we also fitted probit models. The probit models were fitted using `inla` with a Binomial family and a probit link. The models consider the included covariates for each street and a pooled model for the 0–360 day time interval. As in previous analyses, left-censored observations are omitted for better comparison with the survival model outcomes. The response is coded 0 for re-opened and 1 for still closed at the end of the time interval.

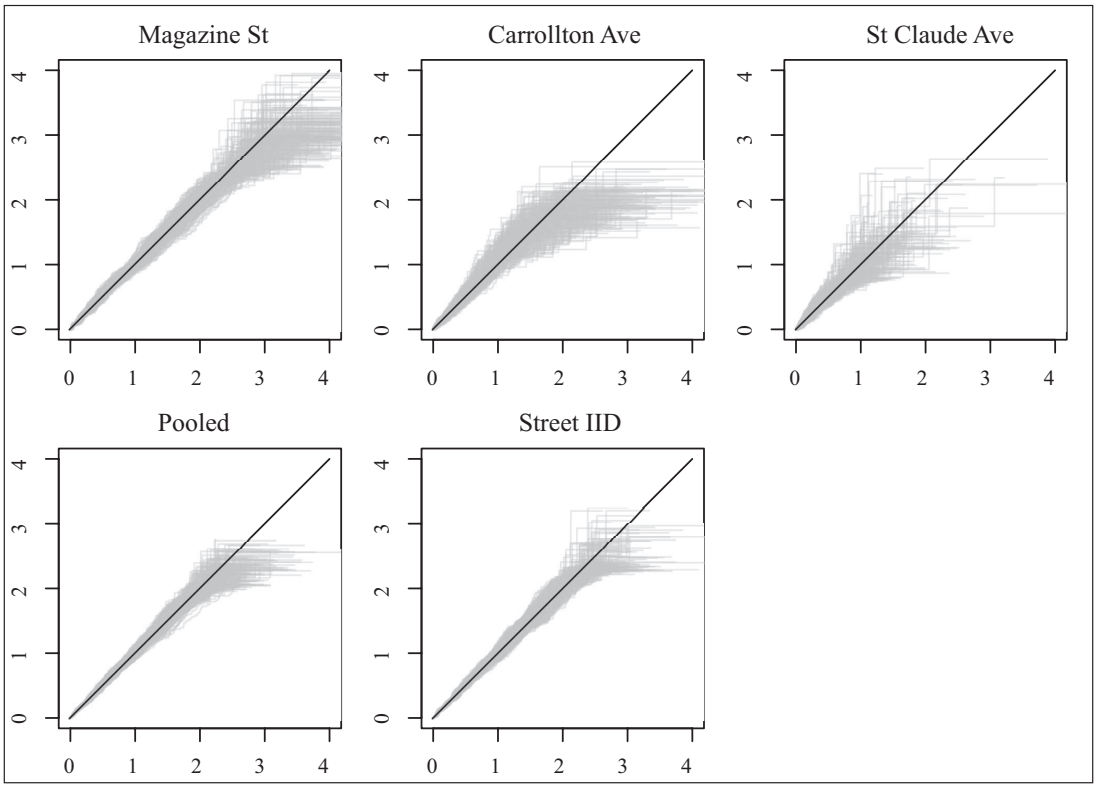

Figure 6 Cox–Snell diagnostic residual plots of MCMC AFT Weibull survival models for each street and pooled model and a pooled model with street random effects, and 250 sampled curves (Zhou and Hanson, 2018). Cox–Snell diagnostic plots are often used to compare different models fitted to the same data, to see which lies closer to the 45° line. Here, 250 lines are shown, sampled from the retained model draws, and represent the MCMC outcomes of a single model in summary. The *y*-axes of the plots are the estimated residuals, the *x*-axes are the expected residuals for an exponential distribution with rate equal to 1

Figure 7 summarizes the probit model outcomes, which are qualitatively similar to those of the Weibull AFT parametric survival models presented already in Figure 4. The similarities are both in coefficient sign, in the relative coefficient values and in the coefficient credible intervals.

4 Spatial survival modelling

Frailty models may include correlated terms to account for complex dependence structures. Before moving to spatial frailties, we will consider an AFT Weibull survival model with by-shop IID frailties, that is a frailty model with Gaussian distribution. Models have been fit using function `survregbayes`. In addition to the priors for

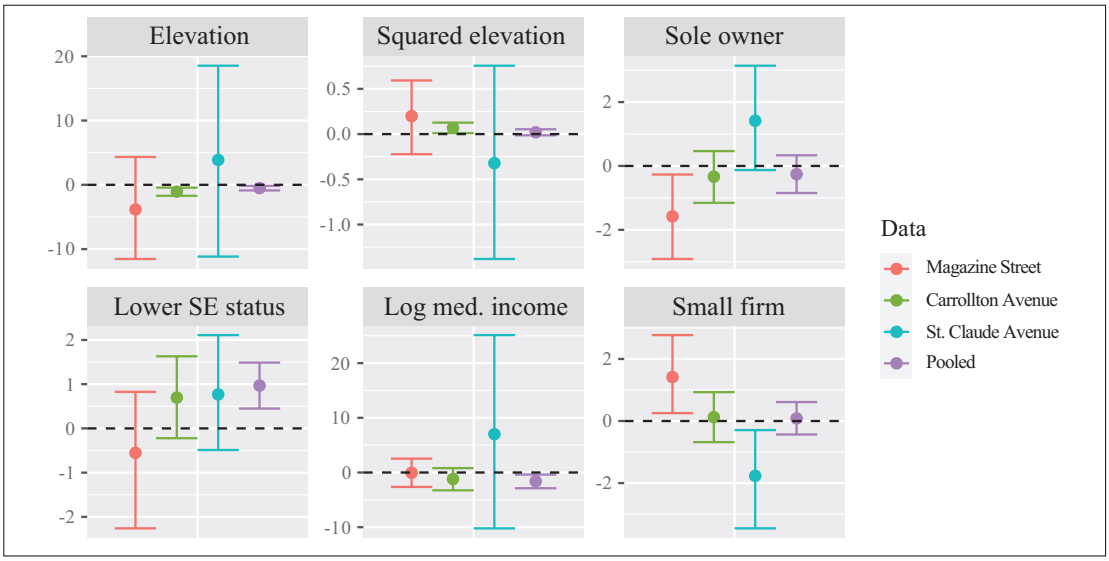

Figure 7 INLA probit model coefficients and 95% credible intervals for street models and a pooled model and for the 0–360 days interval (coefficients show the influence of the covariates on a business not re-opening for easier comparison with the survival model output)

the fixed effects and the α parameter stated above, the prior on the precision of the frailties is a Gamma prior with parameters 0.001 and 0.001.

As a robustness check, IID frailty models were fitted by street and to the pooled dataset. Figure 8 shows the estimated mean coefficient values and 95% credible intervals. In general, estimates of the fixed effects are very similar for all models fit, and INLA providing slightly higher estimates of the variances of the frailties for the model fit to Magazine Street and the pooled model. Figure 9 shows the Cox–Snell diagnostic plots for these models. Again, model fitting seems to be better for the pooled model, which also looks better than the results presented in Figure 6 for the same model.

Spatial frailties will be modelled using a multivariate Gaussian distribution with zero mean and precision matrix $\mathbf{\Sigma}$ that accounts for the spatial dependence of the frailties. This requires an adjacency or neighbourhood structure that defines dependence among observations and that can be computed in different way (see, for example, Bivand et al., 2013).

This adjacency is often expressed as an $n \times n$ adjacency matrix $\mathbf{W}$ with entries (i, j) higher than zero if and only if areas i and j are neighbours, and zero otherwise. Most often $\mathbf{W}$ is a binary matrix, with entries 0 or 1. In spatial econometrics, this binary matrix is often row-standardized in order to force the spatial autocorrelation parameter to be bounded in a convenient interval, usually $(-1, 1)$, so that it can be interpreted as an autocorrelation parameter.

There are different choices for the precision matrix $\mathbf{\Sigma}$ that will lead to particular spatial frailties (see, for example, Banerjee et al., 2014). For example, a conditionally

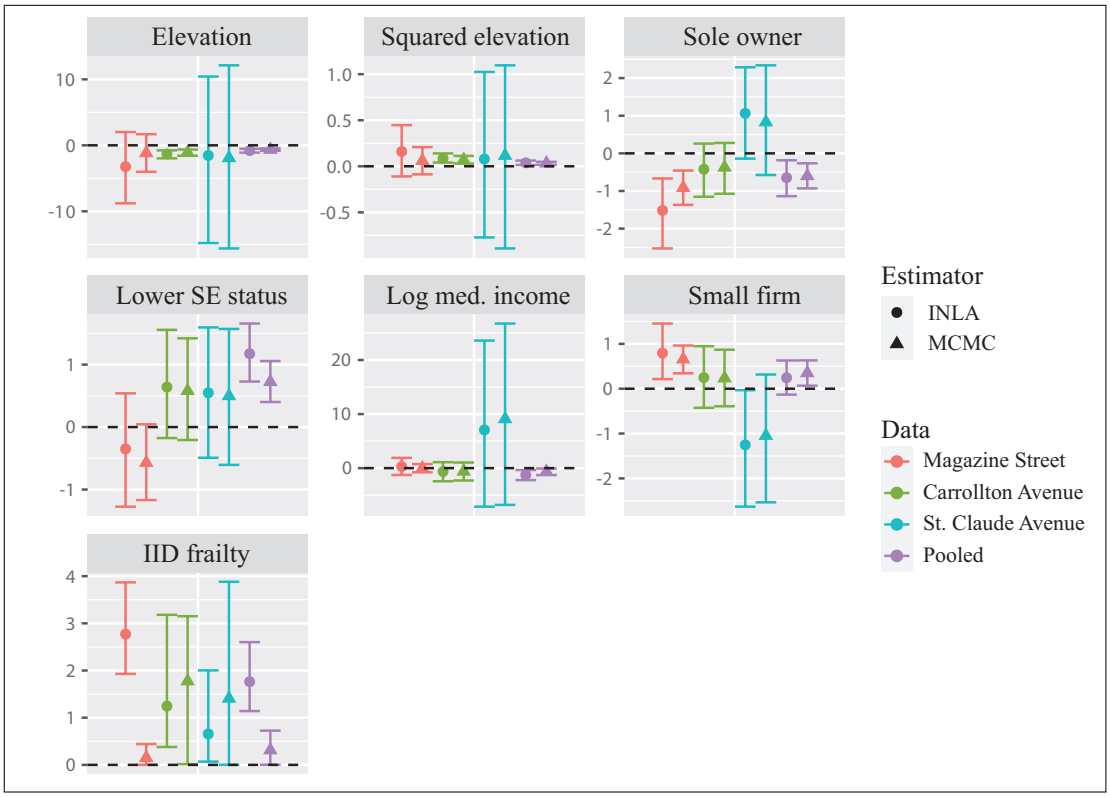

Figure 8 INLA and MCMC IID frailty AFT Weibull survival model coefficients and 95% credible interval bars for street models and a pooled model, and IID frailty variance by dataset and estimator

autoregressive specification (CAR) considers $\Sigma = D - \rho W$, with $\boldsymbol{D}$ a diagonal matrix with entries the number of neighbours, $\boldsymbol{W}$ the aforementioned adjacency matrix and ρ a spatial autocorrelation parameter. LeSage and Pace (2009) describe a spatial autoregresssive (SAR) specification by taking $\Sigma = (I - \rho W)^\top (I - \rho W)$, where now W is taken to be row-standardized. This forces ρ to be in the $(1/m, 1)$ for the model to be fit, with m the minimum eigenvalue of W.

For our particular models, we will consider an intrinsic conditional autoregressive (ICAR) specification (Banerjee et al., 2014) for the frailties. In this case, the marginal distribution of the spatial frailty u_j is Gaussian with mean $\sum_{k\sim j} u_k/n_j$ and precision $\tau_u n_j$, where $k \sim j$ denotes that area k is a neighbour of j, n_j is the number of neighbours of observation j and τ_u the precision parameter of the ICAR effect.

There appear to be two R packages for spatial survival modelling, spBayesSurv (Zhou and Hanson, 2015, 2018; Zhou et al., 2020; Zhou and Hanson, 2020) and spatsurv (Taylor and Rowlingson, 2017; Taylor et al., 2018). Both provide Gaussian Random Field (GRF) approaches, but only spBayesSurv provides an ICAR frailty prior; we feel that this matches at least partly the earlier business re-opening studies

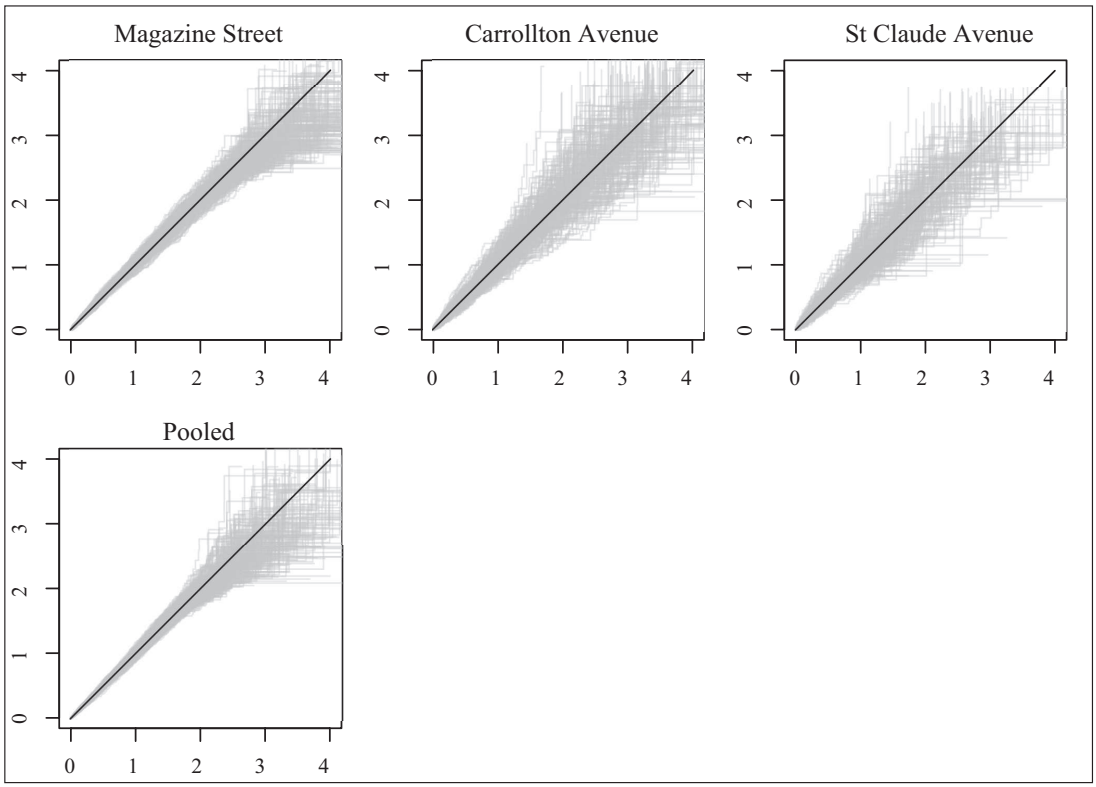

Figure 9 Cox–Snell diagnostic plots of MCMC AFT Weibull IID frailty survival models for each street and a pooled model, and 250 sampled curves (Zhou and Hanson, 2018); please refer to the caption of Figure 6 for a detailed explanation

better than GRF. Further, attempts to use GRF were hindered by the alignment of the businesses along streets, giving three linear objects, rather than an irregular pattern of observations on a surface; these fitting difficulties might also occur when data are transects. Because we are not concerned with developing model fitting methods in this study, and to save space, we refer to Zhou and Hanson (2015, 2018) and Zhou et al. (2020) for a detailed presentation of the underlying methods implemented in spBayesSurv. Models have been fit with function `survregbayes` again, and similar priors (as explained above) have been used in these models. Note that the spatial frailty based on the ICAR effect has only a precision parameter, which will be assigned a Gamma prior as the Gaussian frailty described before.

The analyses in this article were conducted by taking the 11 nearest neighbours as the proximate neighbours of an observation, the same choice as used in LeSage et al. (2011b); some business locations were identical, giving the same sets of 11 nearest neighbours. Distances between neighbours along a street were very short, but Great Circle distances were used anyway to ensure that the correct neighbours were found. This yielded an asymmetric (directed graph) object with three disconnected

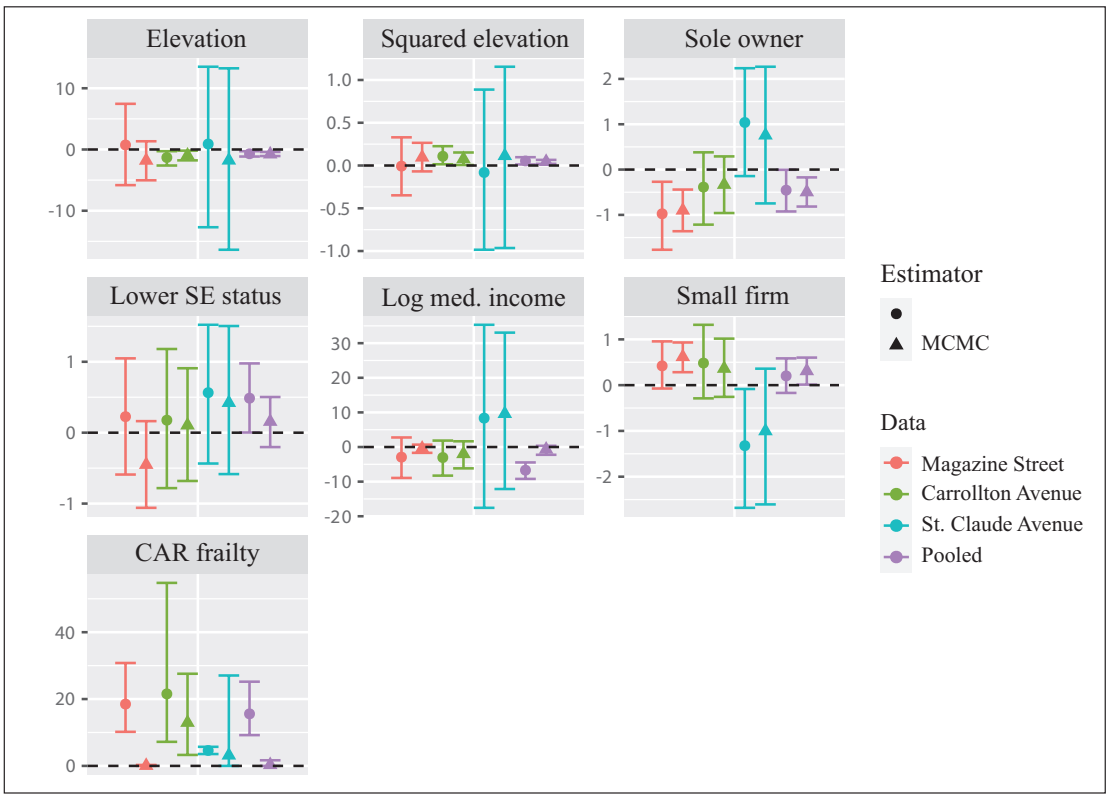

Figure 10 INLA and MCMC CAR frailty AFT Weibull survival model coefficients and 95% credible interval bars for street models and a pooled model, with INLA and MCMC CAR frailty variance

graph components (the streets) which we made symmetric by adding missing links. This step took us away from the original analyses in LeSage et al. (2011b), because CAR models require symmetric spatial weights, whereas simultaneous autoregressive models may utilize asymmetric spatial weights.

A further major conceptual difference was that the original analyses in LeSage et al. (2011b) positioned the spatial autoregressive process in the binary response, arguing that proximate neighbours of a business re-opened in a given cumulative time interval were more likely themselves to have re-opened. It was further argued that an interplay of expectations with regard to the decisions of proximate neighbours would mitigate the issue of posited causality working backwards in time (that a business re-opening on Day 90 might be construed as influencing a business already re-opened when observation began on Day 41). The logical equivalent might be an as-yet not elaborated survival model with time-dependent cross-correlation between survival status changes among proximate neighbours.

Our choice has then been that the spatial process takes the form of Markov Random Field random effects specified as an ICAR, added to the existing aspatial

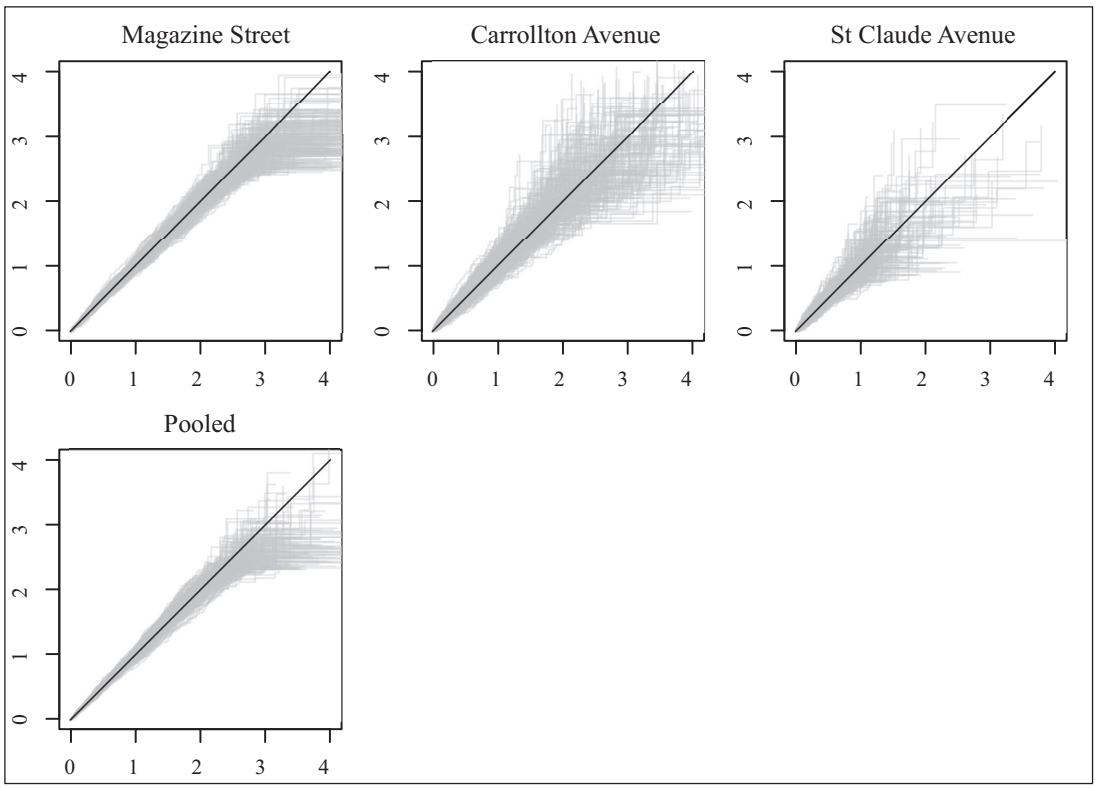

Figure 11 Cox–Snell diagnostic plots of MCMC AFT Weibull CAR frailty survival models for each street and a pooled model, and 250 sampled curves (Zhou and Hanson, 2018); please refer to the caption of Figure 6 for a detailed explanation

survival model. For this reason, the matching spatial probit model will also associate the spatial process with the error term rather than the response. Furthermore, given that there are three disconnected components in the adjacency graph. The spatial survival model implementation used here provides flexibility in choosing distribution and representation, we chose the Weibull distribution and AFT representation. Again, the covariate coefficient signs were reversed by multiplying by −1. The total number of Markov chain Monte Carlo (MCMC) draws was set to 200 000, with 50 000 discarded as burn-in.

Figure 10 is similar in form to Figure 4 and gives broadly very similar outcomes to the aspatial models. It reports coefficient and 95% credible intervals for the three street models and the pooled model, all using the $k = 11$ neighbour definition. In general, estimates of the fixed effects are very similar for all models fit, and INLA provides higher estimates of the variance for the model fit to Magazine Street and the pooled model.

Table 3 Deviance Information Criteria reported for MCMC and INLA Weibull AFT models. Note that time has been re-scaled for INLA models and, hence, INLA and MCMC DICs cannot be directly compared

	Magazine Street	Carrollton Avenue	St. Claude Avenue	Pooled
MCMC No frailty	2023.89	788.60	469.19	3350.72
MCMC IID frailty	1994.49	647.79	395.50	3271.63
MCMC CAR frailty	2026.86	703.42	459.94	3282.99
INLA No frailty	−386.04	144.51	121.96	−88.18
INLA IID frailty	−693.78	110.61	116.03	−364.17
INLA CAR frailty	−601.97	67.78	111.72	−404.10

Cox–Snell diagnostic plots have been produced for the CAR frailty survival model. Figure 11 shows these plots for the CAR frailty street and pooled models. Again, model fitting seems to be better for the pooled model than for the three by-street models presented.

In order to assess goodness of fit, the deviance information criterion (DIC, Spiegelhalter et al., 2002) has been computed for all survival models. The DIC can be defined as

$$DIC = D(\hat{\theta}) + 2p_D.$$

Here, $\boldsymbol{\theta}$ represents the vector of model parameters and effects and $\hat{\theta}$ their point estimates and $D(\hat{\theta})$ the deviance evaluated at these point estimates. p_D is the effective number of parameters, which can be computed as $\overline{D(\theta)} - D(\hat{\theta})$, with $\overline{D(\theta)}$ the posterior mean of the deviance. Lower values of the DIC indicate a better model fit. Note that INLA will use the posterior mode of the hyperparameters when computing $\hat{\theta}$ instead of their posterior means, whicy may lead to discrepancies in the values of the DIC reported by MCMC and INLA.

Values of the DIC were computed for the three by-street models (no frailty, IID frailty and CAR frailty), and for the three pooled models for MCMC and INLA (Table 3). We believe that the different ordering of the DIC values by dataset between INLA and MCMC is related to the use in INLA of the posterior modes of the hyperparameters (instead of the posterior means), while in this implementation, MCMC uses the posterior mean. The posteriors of the precisions (for example) usually have long right tails, that shift the posterior mean to the right (as compared to the posterior mode). The by-street and pooled MCMC models appear to favour the IID frailty, while the INLA models rather point to the CAR frailty with the exception of the Magazine street dataset.

4.1 Spatial probit modelling

The spatial probit model is similar to the probit model presented above, but including a spatial term based on the SAR specification. In particular the binary response y_i is modelled using a continuous latent variable z_i such as:

$$y_i = \begin{cases} 1 \text{ if } z_i \geq 0, \\ 0 \text{ if } z_i < 0. \end{cases}$$

Then, continuous latent variables $z = (z_1, \ldots, z_n)$ are modelled using a SAR specification as follows:

$$z = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon.$$

Here, $\boldsymbol{\varepsilon}$ is a vector of independent Gaussian errors with zero mean, spatial autocorrelation parameter ρ and precision τ_ε. In the absence of spatial effect this model reduces to an ordinary probit model.

The spatial error probit model takes the latent variable as follows:

$$z = X\beta + (I - \rho W)^{-1} \varepsilon$$

This is similar to considering a spatial error model (SEM, LeSage and Pace, 2009) on the latent variables z. Again, in the absence of spatial effect this model reduces to an ordinary probit model.

As with the spatial survival models, we will not present the development of the methods implemented in the software package used, referring the interested reader to Martinetti and Geniaux (2017) and LeSage et al. (2011b). In Martinetti and Geniaux (2017), the spatial error probit model is discussed more fully than in LeSage et al. (2011b), which concentrates on the impact of decisions to re-open

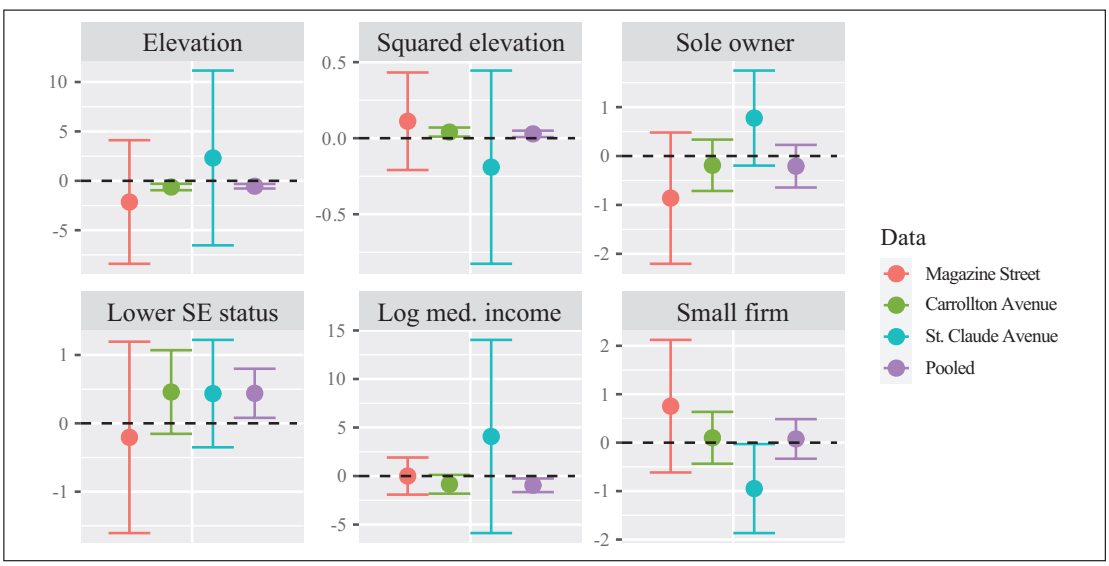

Figure 12 Fitted spatial error probit model coefficients and 95% error bars for by-street models and a pooled model for the 0–360 days interval

on decisions made at neighbouring businesses. This approach, using the decisions of neighbours to model own decision outcomes, also involves the calculation of impacts, as the effects of the covariates are filtered through the spatial process in the response, and are only equal to the coefficients themselves when the spatial coefficient is negligible.

Models have been fit with function `SpatialProbitFit()` from package ProbitSpatial (Martinetti and Geniaux, 2016), which relies on ML estimation. Figure 12 shows the outcomes of fitting spatial error probit model for time interval 0–360 days. There are minor shifts when compared to the equivalent aspatial model outcomes shown in Figure 7, but none of major importance; indeed, we would not in general expect the coefficients to shift, but we might expect broader coefficient standard errors in the spatial error probit model output, which we see to some extent.

5 Conclusions

The intention underlying this comparison has been to begin to investigate how far it is possible to approach the analysis of datasets which have a clear spatial time-to-event structure, using business re-opening after Hurricane Katrina as an example. As this also invites extension and deepening of the examination of the implementations used, a script to reproduce the analyses presented here is available at `https://github.com/rsbivand/survival_katrina`, and contributions through issues raised on the repository providing this script are welcome. A possible avenue might be to re-visit the intuition underlying the original spatial probit analysis, that decisions influence one-another directly through contagion or information spillovers to proximate neighbours rather than as here through additive random effects. We have not shown that the inferences drawn in the original study have been modified by the application of survival analysis in this case with these observations and covariates, but this perhaps suggests that in fact survival analysis might rather have been employed in the original study as it does respect the time-to-event nature of the data better than aggregated time intervals. Furthermore, models proposed in this article provide very similar estimates of the covariates. A possible way to further assess the different models is to explore their predictive accuracy, for which the dataset could be split into training data and test data. However, this is out of the scope of the current study.

Acknowledgement

We would like to thank the editor and referees for helpful and constructive suggestions.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Virgilio Gómez-Rubio has been supported by grants SBPLY/17/180501/000491 and PPIC-2014-001-P, funded by Consejería de Educación, Cultura y Deportes (Junta de Comunidades de Castilla-La Mancha, Spain) and FEDER and grant PID2019-106341GB-I00 from Ministerio de Ciencia e Innovación (Spain).

Appendix

In this appendix, we have included some summary statistics and a maximum likelihood (ML) analysis of the Katrina dataset.

Summary statistics

Tables A.1, A.2 and A.3 show the percentages of the three categories for each of the three categorical variables by streets. Carrollton Avenue appears to have been where

Table A.1 Business employment size category percentages by street.

	Magazine Street	Carrollton Avenue	St. Claude Avenue
Small	71.07	44.69	73.12
Average	25.69	46.93	23.66
Large	3.24	8.38	3.23

Table A.2 Business ownership category percentages by street

	Magazine Street	Carrollton Avenue	St. Claude Avenue
Sole	89.28	59.78	84.95
Local	8.73	35.20	11.83
National	2.00	5.03	3.23

Table A.3 Socio-economic status category percentages by street for census block group at business locations

	Magazine Street	Carrollton Avenue	St. Claude Avenue
Lower	7.23	27.37	77.42
Average	51.37	68.72	22.58
Upper	41.40	3.91	0.00

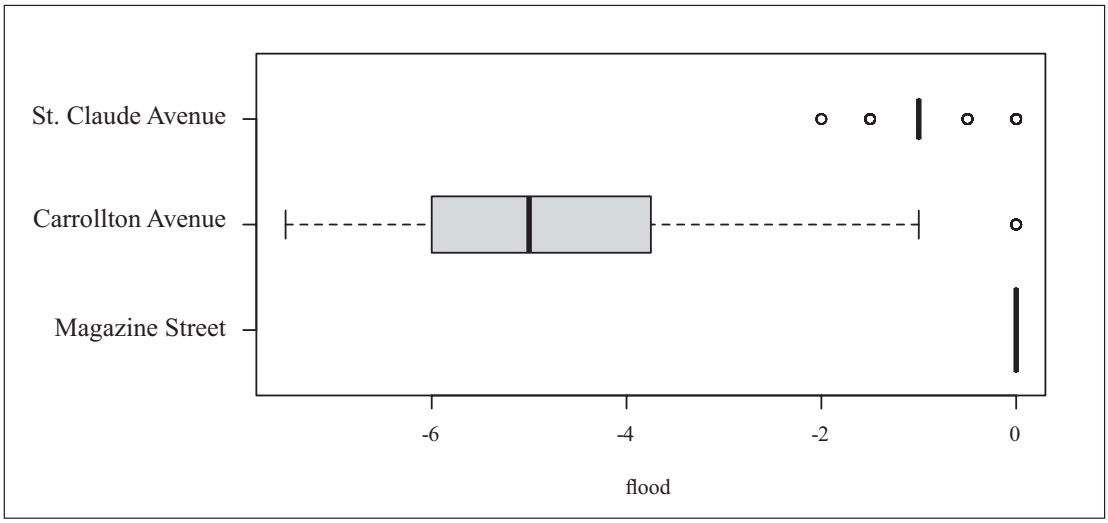

Figure A.1 Boxplots of reported flood depth measured in whole feet at business locations by street

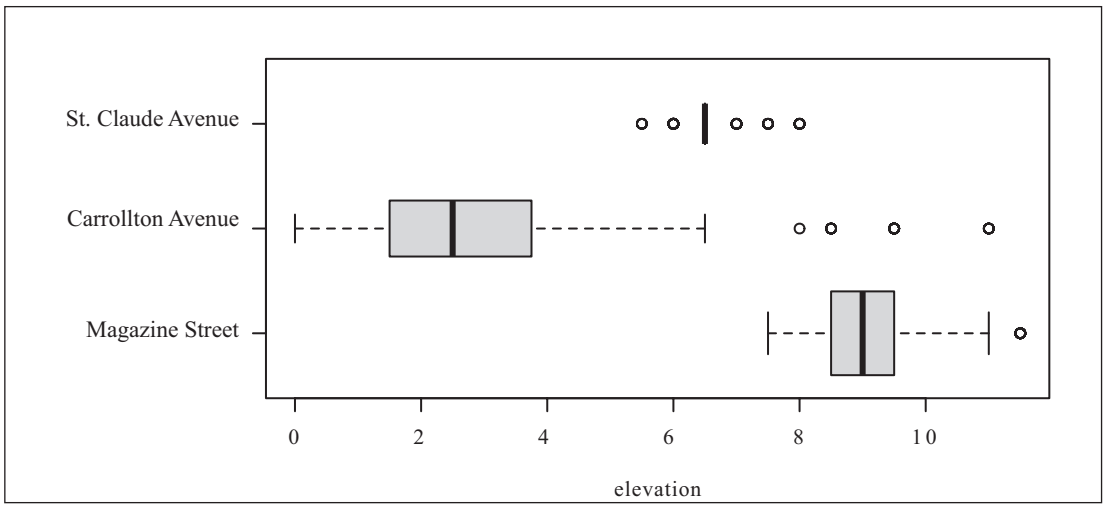

Figure A.2 Boxplots of elevation measured in whole feet at business locations by street; the minimum of 3.5 feet has been subtracted to permit use of squared elevation as a covariate

businesses were larger in terms of employment size category (>50% average and large compared to <30%), and where fewer (60%) were owned by sole proprietors than on Magazine Street and St. Claude Avenue (>80%). The census block groups read off at business locations also differed by street, with Carrollton Avenue >4% upper socio-economic status, St. Claude Avenue 0%, but Magazine Street >40% upper

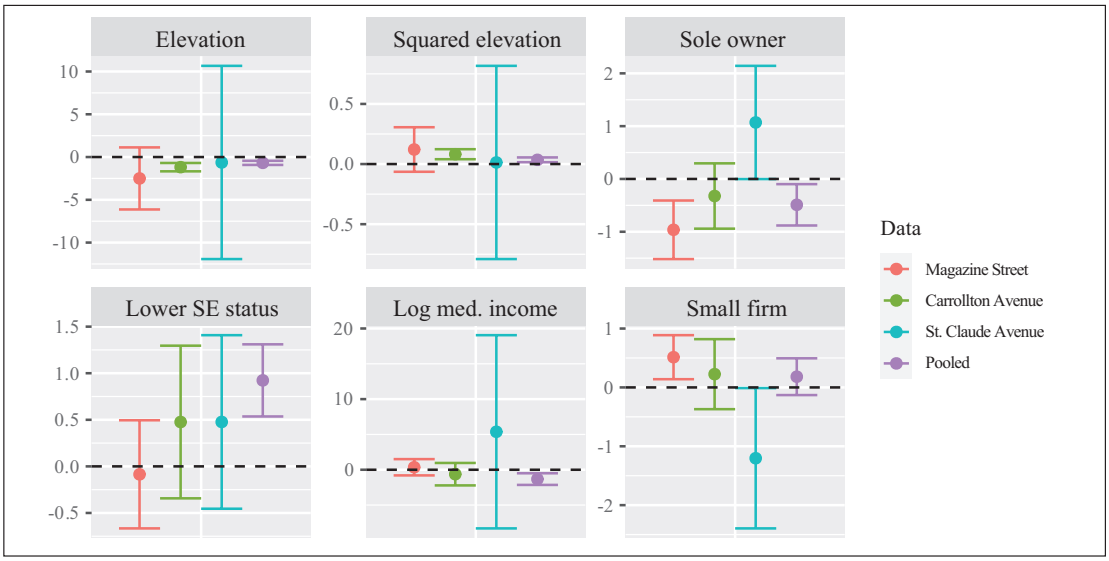

Figure A.3 Fitted Weibull survival regression model coefficients and 95% error bars for `survreg` models by street, for a pooled `survreg` model. Negative coefficient values mean that surviving closed is shortened, for example for higher elevation, positive values that surviving closed is extended

socio-economic status. St. Claude Avenue has >75% lower socio-economic status, and Carrollton Avenue almost 70% average socio-economic status. Median incomes, read from the same block groups, are similarly distributed.

In the original article, flood depth was used as a covariate, but as Figure A.1 shows, there is no variability in flood depth on Magazine Street, and little on St. Claude Avenue. Subtracting minimum elevation from the measure of elevation gives a local measure from the minimum. Figure A.2 shows more variability on Magazine Street and similar variability to flood depth on St. Claude Avenue. Setting the lowest elevation at zero also permits the use of squared elevation as an additional covariate to attempt to address possible non-linearity.

Maximum likelihood estimation

We present here the results obtained by fitting the models described in Section 3 using `survreg()` from the survival package (Therneau, 2019). Note that `survreg()` estimates are obtained by ML estimation, while those presented in other sections of the article are based on Bayesian inference via MCMC or INLA. The ML estimates can be regarded as a robustness check as well. Figure A.3 shows the coefficients from the three fitted aspatial Weibull survival regression models by street and a pooled model. Error bars shown correspond to approximate 95% confidence intervals.

References

Banerjee S, Carlin BP and Gelfand AE (2014) *Hierarchical Modeling and Analysis for Spatial Data, 2nd edition*. Boca Raton, FL: Chapman & Hall/CRC.

Bivand RS, Pebesma E and Gómez-Rubio V (2013) *Applied Spatial Data Analysis with R, 2nd edition*. New York, NY: Springer.

Cox D and Snell E (1968) A general definition of residuals (with discussion). *Journal of the Royal Statistical Society, Series B*, **30**, 248–75.

Gómez-Rubio V (2020) *Bayesian Inference with INLA*. Boca Racon, FL: Chapman and Hall/CRC.

Ibrahim JG, Chen M-H and Sinha D (2001). *Bayesian Survival Analysis*. New York, NY: Springer.

Kaplan EL and Meier P (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481. doi: 10.1080/01621459.1958.10501452. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452 (last accessed 2 November 2020).

Lam NSN, Arenas H, Pace K, LeSage J and Campanella R (2012) Predictors of business return in New Orleans after Hurricane Katrina. *PLOS ONE*, **7**, 1–8. doi: 10.1371/journal.pone.0047935. URL https://doi.org/10.1371/journal.pone.0047935 (last accessed 2 November 2020).

LeSage J and Pace RK (2009) *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman and Hall/CRC.

LeSage J, Pace RK, Campanella R, Lam N and Liu X (2011a) Do what the neighbours do. *Significance*, **8**, 160–163. doi: 10.1111/j.1740-9713.2011.00520.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1740-9713.2011.00520.x (lastb accessed 2 November 2020).

LeSage J, Pace RK, Lam N, Campanella R and Liu X (2011b) New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **174**, 1007–1027. doi: 10.1111/j.1467-985X.2011.00712.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2011.00712.x (last accessed 2 November 2020).

Martinetti D and Geniaux G (2016) ProbitSpatial: Probit with Spatial Dependence, SAR and SEM Models. *R package version 1.0*. URL https://CRAN.R-project.org/package=ProbitSpatial (last accessed 2 November 2020).

——— (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics*, **64**, 30–45. doi: https://doi.org/10.1016/j.regsciurbeco.2017.02.002. URL http://www.sciencedirect.com/science/article/pii/S0166046217300546 (last accessed 2 November 2020).

Moore D (2016) *Applied Survival Analysis Using R*. New York, NY: Springer.

Rue H, Lindgren F, Simpson D, Martino S, Teixeira Krainski E, Bakka H, Riebler A and Fuglstad G-A (2019) INLA: Full Bayesian Analysis of Latent Gaussian Models using Integrated Nested Laplace Approximations. *R package version 19.09.03*.

Rue H, Martino S and Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, **71**, 319–92.

Simpson DP, Rue H, Riebler A, Martins TG and Sørbye SH (2017) Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, **32**, 1–28.

Spiegelhalter DJ, Best NG, Carlin BP and Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–616.

Taylor BM and Rowlingson BS (2017) spatsurv: An R package for Bayesian inference

with spatial survival models. *Journal of Statistical Software*, **77**, 1–32. doi: 10.18637/jss.v077.i04.

Taylor BM, Rowlingson BS and Zheng Z (2018) spatsurv: Bayesian Spatial Survival Analysis with Parametric Proportional Hazards Models. *R package version 1.2*. URL https://CRAN.R-project.org/package=spatsurv (last accessed 2 November 2020).

Therneau T (2019) Survival: Survival Analysis. *R package version 2.441.1*. URL https://CRAN.R-project.org/package=survival (last accessed 2 November 2020).

Therneau T and Grambsch P (2000) *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer.

Wilhelm S and de Matos MG (2013) Estimating Spatial Probit Models in R. *The R Journal*, **5**, 130–43. URL https://journal.rproject.org/archive/2013/RJ-2013-013/index.html (last accessed 2 November 2020).

——— (2015) spatialprobit: Spatial Probit Models. *R package version 0.9-11*. URL https://CRAN.R-project.org/package=spatialprobit (last accessed 2 November 2020).

Zhou H and Hanson T (2015) Bayesian spatial survival models. In *Nonparametric Bayesian Inference in Biostatistics*, edited by R Mitra and P Müller, pages 215–46. Cham: Springer. URL https://doi-org.pva.uib.no/10.1007/978-3-319-19518-6_11 (last accessed 2 November 2020).

——— (2018). A unified framework for fitting Bayesian semiparametric models to arbitrarily censored survival data, including spatially referenced data. *Journal of the American Statistical Association*, **113**, 571–81. doi: 10.1080/01621459.2017.1356316.

——— (2020) spBayesSurv: Bayesian Modeling and Analysis of Spatially Correlated Survival Data. *R package version 1.1.4*. URL https://CRAN.R-project.org/package=spBayesSurv (last accessed 2 November 2020).

Zhou H, Hanson T and Zhang J (2020) spbayessurv: Fitting Bayesian spatial survival models using R. *Journal of Statistical Software, Articles*, **92**, 1–33. doi: 10.18637/jss.v092.i09. URL https://www.jstatsoft.org/v092/i09 (last accessed 2 November 2020).