Vogler, Jan; Liesenfeld, Roman; Richard, Jean-Francois

Conference Paper

Likelihood based inference and prediction in spatio-temporal panel count models for urban crimes

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Microeconometric Modelling, No. D22-V3

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Likelihood based inference and prediction in spatio-temporal panel count models for urban crimes

Roman Liesenfeld
Institute of Econometrics and Statistics, Universität Köln, Germany

Jean-François Richard
Department of Economics, University of Pittsburgh, USA

Jan Vogler
Institute of Econometrics and Statistics, Universität Köln, Germany

(June 15, 2015)

Abstract

We develop a panel data count model combined with a latent Gaussian spatio-temporal heterogenous state process to analyze monthly severe crimes at the census tract level in Pittsburgh, Pennsylvania. Our data set combines Uniform Crime Reporting data with socio-economic data from the 2000 census. The likelihood of the model is accurately estimated by adapting recently developed efficient importance sampling techniques applicable to high-dimensional spatial models with sparse precision matrices. Our estimation results confirm socio-economic explanations for crime and, foremost, the broken-windows hypothesis, whereby less severe crimes in a region is a leading indicator for severe crimes. In addition to ML parameter estimates, we compute several other statistics of interest for law enforcement such as elasticities (idiosyncratic, total, short-term as well as long-term) of severe crimes w.r.t. less severe crimes, one-month-ahead out-of-sample forecasts, predictive cumulative distribution functions and validation test statistics based on these cdf's.

JEL classification: C15; C23; C25; C51; C53; K42; R15.

Keywords: Broken-windows hypothesis; Efficient Importance sampling; Empirical crime model; Out-of-sample crime forecasts; Spatio-temporal econometrics.

1 Introduction

The spatio-temporal urban distribution of crimes is receiving growing attention not only from researchers (criminologists, sociologists, economists, geographers,...) but also from law enforcement agencies. See, e.g., Ratcliffe (2013), Bernasco and Elffers (2013), Tita and Radil (2013), Roth et al. (2013) or Li et al. (2014) for recent contributions, surveys and extensive lists of references. Additional references are discussed in Section 2 below. Two quotes from the literature are directly relevant to the present paper. In his overview of current crime analysis, Ratcliffe (2013, p. 14) states that: "At present, the most under-researched area of spatial criminology is that of spatio-temporal crime patterns". Among their discussion of unmet needs for spatio-temporal crime analysis, Roth et al. (2013, p. 238) highlight a need to "integrate geographic and temporal representation and analyses".

In the present paper, we aim at addressing such needs by proposing a spatio-temporal latent panel model for high-dimensional urban crime count data, which we then apply to a data set consisting of monthly (2008-2013) counts of severe crimes (classified as Part I crimes) for the 138 census tracts in Pittsburgh, Pennsylvania. By taking advantage of an Efficient Importance Sampling (EIS) approach recently developed by Liesenfeld et al. (2015) for likelihood evaluations in spatial latent models, we estimate a latent panel count data model that includes temporal as well as spatial dependence, socio-economic census tract characteristics and unobserved heterogeneity (random effects) across census tracts. Not only can the model be estimated by Maximum Likelihood (ML) but a wide range of important auxiliary statistics (for out-of-sample prediction, model validation, model selection) can also be routinely computed.

A further motivation for our paper originates from three earlier papers who analyzed a variety of forecasting models for severe crime in Pittsburgh (Gorr et al., 2003, and Cohen et al., 2007) and in Pittsburgh and Rochester, New York (Cohen and Gorr, 2005). In the absence of potential socio-economic covariates and spatially and temporally lagged dependent variables, they find that a linear multivariate leading indicator model is typically best for forecasting large decreases in

crime volume, while various forms of exponential smoothing provide the best average forecast point accuracy according to mean squared and mean absolute percentage forecast error. Relative to these contributions, we benefit from two key advantages. First, as mentioned above, we are able to generalize the numerical EIS procedure developed by Liesenfeld et al. (2015) for likelihood functions in latent spatial count models to account not only for spatial but also for temporal lags as well as unobserved and observed (socio-economic) heterogeneity. Second, we benefit from access to highly disaggregated and, foremost, internally consistent data at the census tract level combining Uniform Crime Reporting (UCR) data classified according to the handbook of the U.S. Department of Justice (2004) with socio-economic data from the 2000 census.

As a preview of our main results, we test alternative socio-economic explanations for severe crime intensities and find strong confirmation of the 'broken-windows' phenomenon, whereby the intensity of less severe crime in a census tract provides a leading indicator for more severe crimes. See, e.g., Wilson and Kelling (1982), Anselin et al. (2000, p. 225), Cohen and Gorr (2005) and Cohen et al. (2007). In particular, our model allows us to compute both short-term and long-term elasticities of severe crimes w.r.t. less severe crimes. Foremost, by exploiting the spatial component of our model, we can compute 'total' elasticities for the city of Pittsburgh in response to a reduction in less severe crimes in any given census tract. Such results highlight the critical importance of fully accounting for urban spatial dependence but could also provide a useful tool for efficient allocation of law enforcement resources.

We also run sequential one-month-ahead out-of-sample forecasts that demonstrate the superior predictive performance of our model relative to exponential smoothing (a widely used and typically hard to beat benchmark). Moreover, we can produce complete predictive distributions, not just point predictions, from which predictive forecasts intervals can be obtained. Last but not least, we also use these predictive distributions to produce statistical validation of our forecasting model.

The paper is organized as follows. In Section 2 we provide a review of the literature on the socio-economic determinants of variations in crime rates across geographic regions and their spatial and temporal dependence. In Section 3 we describe the data. Section 4 presents the spatio-temporal

panel count data model used to analyze severe crimes and in Section 5 we outline the numerical spatial EIS procedure which we use for likelihood inference. The empirical results are discussed in Section 6 and conclusions are drawn in Section 7.

2 Predictors and dependence in time and space of crime rates

2.1 Predictors

Empirical research in criminology commonly applies regression models to explain observed variations in crime rates across geographic regions with fixed boundaries such as counties (Baller et al., 2001), police precincts (Gorr et al., 2003), census tracts (Helbich and Arsanjani, 2014), or census block groups (Willits et al. 2013). The theoretical background consists of sociological theories of crime including social ecology theories and place-based theories (see, e.g., Anselin et al., 2000). Social ecology theories such as the *social disorganization theory* (Shaw and McKay, 1942) explain the geographical variation in crime levels in terms of varying social conditions of the population. Under place-based theories, including the *routine activities theory* (Cohen and Felson, 1979) and the *rational choice theory* (Cornish and Clarke, 1986), the geographical variation of crime levels is determined by the intersection in time and space of suitable targets, motivated offenders and the absence of crime suppressors. Those theories point to several indicators of structural (environmental, urbanistic, sociological and economic) conditions which may help predicting the geographical distribution of crime rates. Structural predictors used in empirical studies include measures for population structure (size and density), composition of the resident population (percentage of white and African-American population, age structure), family cohesion (percentage of female-headed households, divorce rate), socio-economic structure (income figures, unemployment rates), and condition of buildings and houses (rental and homeowner vacancy rates) – see, e.g., Land et al. (1990), Tolnay at al. (1996), Baller et al. (2001), Kubrin (2003), and Helbich and Arsanjani (2014).

2.2 Spatial dependence

A common stylized fact in empirical criminology is that levels of violence and crime activity are not randomly distributed across geographical regions. Instead similar levels cluster in space, which implies positive spatial autocorrelation, consistent with a fundamental property of spatial data whereby there is a tendency for observations that are in close geographic proximity to be more alike than those that are further apart (Tobler's, 1970, first law of geography). If criminal activities are solely determined by the structural factors included in a regression model there should be no spatial dependence beyond that generated by structural similarities of regions that are in close geographic proximity (Baller et al., 2001). However, spatial clustering typically cannot be completely explained by common measures of structural similarity between geographical regions (see, e.g., Morenoff and Sampson, 1997, Morenoff et al., 2001, Baller, et al., 2001, and Tita and Radil, 2013, p. 107).

Spatial correlation has attracted growing attention in empirical criminological research for two reasons (see, e.g., Baller et al., 2001). The first one is statistical. If spatial correlation in the data is ignored, then estimates for the effects of covariates and the corresponding standard errors may be biased and inconsistent (see Anselin, 1988). Thus, modelling spatial correlation is critical when assessing the marginal effects of structural covariates on crime rates. The second reason is that spatial dependence by itself is of substantive importance in crime analysis since positive spatial autocorrelation is interpreted as evidence of the spatially diffusive and contagious nature of certain types of crime (see, Tolnay et al., 1996 and Morenoff and Sampson, 1997). Such effects may reflect interacting criminals and gangs linked together by rivalry networks (Tita and Greenbaum, 2008) and/or 'subcultural' processes in which case violence spreads throughout the population and regional areas via direct social contacts (Loftin, 1986).

In order to explicitly incorporate spatial correlation, regression models used to explain the variation in crime rates are typically generalized to include either spatially correlated errors (spatial error models) or spatially lagged dependent variables (spatial lag models) – see, e.g. Baller et al. (2001), Morenhoff et al. (2001) and Kubrin (2003). Spatial error models are appropriate in cases where spatial dependence is treated as a nuisance resulting from omitted spatially correlated

structural factors rather than being of substantive interest of its own. Instead spatial lag models are more compatible with the notion of diffusion and contagion processes. However, it is important to recognize that these models are designed to find only indirect evidence of diffusion and contagion and that they are unable to discover the actual mechanisms through which criminal events in one geographical area influence events in other areas at later times (see Baller et al. 2001, p. 567).

2.3 Temporal dependence

When observations are available in the form of panel data for a cross-section of regional areas with fixed boundaries at different points in time, it becomes possible to model complex combinations of spatial heterogeneity and or spatial and temporal dependence in crime rates (see Anselin et al., 2000, p. 241). A dynamic panel-data setting also provides a natural basis for joint time-space forecasting of crime activity which can support tactical deployments of police resources (Gorr and Harris, 2003, Gorr et al., 2003, Cohen and Gorr, 2005, and Roth et al., 2013).

There are several reasons to expect not only spatial but also temporal dependencies as well as seasonal regularities in levels of criminal activity in areal units. For example, convergence of crime opportunities in space and time, as emphasized by place-based routine activity theories, could be facilitated by various physical and social features that provide a setting more or less conductive to crime, such as local population composition or urbanistic conditions (Anselin et al., 2000, p. 220). Temporal persistence and the slowly changing nature of such (observed or unobserved) features lead to systematic temporal dependence in time series of frequency counts of crime observed for geographic areas (Li et al., 2014, p. 181). If some of those physical and social features are stable over time but unobservable, the implied spatial heterogeneity cannot be controlled for in a panel regression analysis and gives rise to temporally correlated errors. In order to account for such unobserved spatial heterogeneity we shall use a random effect approach with an area-specific time-invariant error component[1].

[1] An alternative way to capture time invariant spatial differences would be to use a model with fixed-effect area-specific dummy variables. However, such a model requires the estimation of a larger number of additional parameters, leading to a significant loss of degrees of freedom.

The interpretation of spatial autocorrelation in terms of diffusion and contagion also implies inter-temporal linkages of crime rates within regional areas since these concepts inherently assert time-sequential processes (see, Bernasco and Elffers, 2013, p. 710). For example, spatial diffusion reflecting interacting gangs implies a chronology of criminal actions and counter-reactions – due to the retaliatory nature of gang violence – and leads to temporal persistence of crime rates. Such persistence can also be expected if the contagious nature of crime is the result of 'subcultural' processes whereby violence spreads across the population via social contacts so that an increase in assaults in a given area may set-off a chain reaction of criminal events extending to neighboring areas (Tita and Radil, 2013, p. 107). In a panel regression model such spatio-temporal persistence can be accounted for by including not only spatially but also temporally lagged dependent variables.

Furthermore, high levels of urban crime are typically concentrated in relatively few small areas. Empirical evidence suggests that such crime hot spots may arise first as a concentration of soft crimes (e.g., vandalism, gambling and public order disturbances) that later hardens into more serious crimes (e.g., assaults, robbery and homicides) – see, e.g., Anselin et al. (2000) and the references cited therein. Explanations for this temporal development of crime hot spots emphasize that public signs of disorder like vandalism, gambling and 'broken windows' foster increases in more serious crime, since they signal a loss in the ability to exercise social control, further attracting and perpetuating crimes (Wilson and Kelling, 1982). This 'broken-windows' phenomenon suggests that the intensity of soft crimes observed in a given regional area may serve as a leading indicator for the number of serious crimes in that area (Anselin et al., 2000, p. 225 and Cohen and Gorr, 2005, Cohen et al., 2007). In order to make use of the leading-indicator properties of soft crimes when specifying a predictive panel model for severe crimes we shall include them as lagged explanatory variables.

Finally, crime appears to be a seasonal phenomenon and empirical evidence suggests that, for example, frequencies of violent crimes such as assaults and homicides are higher in summer and lower in winter (see, Cohen, 1941). Under the 'temperature aggression hypothesis' this seasonality is attributed to weather which increases violent crimes via ambient temperature and anger arousal

(see, Gorr et al., 2003). Within the routine activity theory regular temporal patterns can be viewed as a result of seasonality which gives rise to changes in one or more of the conditions propitious to crime (such as existence of suitable targets and motivated offenders and absence of crime suppressors). In order to take potential seasonality into account we shall include monthly seasonal dummies in the panel regression model.

3 Data

3.1 Data sources

Our crime data set includes monthly (January 2008 to December 2013) counts of Part I and Part II offenses for each of the 138 2000-census tracts in Pittsburgh. Part I and Part II offenses are defined in the Uniform Crime Reporting (UCR) handbook of the U.S. Department of Justice (2004, p. 8). Part I offenses, also known as index crimes regroup serious felonies in the following eight categories: criminal homicide, forcible rape, robbery, aggravated assault, burglary, larceny-theft (except motor vehicle theft), motor vehicle theft and arson. The UCR program's founder selected these offenses because "they are serious crimes, they occur regularly in all areas of the country and they are likely to be reported to police". The count data for Part I offenses consist of the numbers of offenses in these categories that are known to law enforcement. Part II offenses includes 21 categories of non serious felonies and misdemeanors ranging from other assaults to runaways (persons under 18) for which only arrest data were collected.

Our dependent variable y_{it} is defined as the number of Part I offenses in census tract i ($i = 1, \ldots, 138$) in month t ($t = 1, \ldots, 72$) for a total of 9,936 individual observations. In Table 1 we provide a list of the Pittsburgh census tracts with their i-ordering as used throughout the paper. As discussed above and further below, the lagged number of Part II offenses will prove to be a key predictor variable for y_{it}.

In addition, in order to account for heterogeneity across census tracts, we collected data from the

Census 2000 (US Census Bureau and Social Explorer Tables) on the following 15 socio-economic variables ($i = 1, \ldots, 138$): Log of total population (*Ltp*), log of population density per square mile (*Lpd*), log of median income (*Lmi*), dropout rate age 16-19 (*Dra*), civilian unemployment rate (*Cur*), poverty rate (*Pvr*), percentage of total population under 18 (*U18*), group quarter proportion (*Gqp*), percentage of total population that is African-American (*Paa*), percentage of population (age 25 and over) with less than a high school degree (*Hdl*), percentage of population (age 25 and over) with a bachelor degree or higher (*Bdh*), rental housing units as percentage of occupied housing units (*Rhu*), percentage of households having been in the same house for more than one year (*Sh1*), percentage of female headed households (*Fhh*), and housing units vacancy rate (*Hvr*). Occasional missing data for census tracts 203, 708, 9800, 9801, and 9803-9812, which do not have a regular resident population, were replaced by corresponding census tract dummies[2].

3.2 Some descriptive statistics

As we are dealing with a total of 9,936 observations, we computed time and census tracts averages of Part I crimes $\bar{y}_i$ ($i = 1, ..., 138$) and $\bar{y}_t$ ($t = 1, \ldots, 72$). In Figure 1 we provide a color coded map of the time averages $\bar{y}_i$, highlighting significant spatial clustering and heterogeneity across census tracts (time averages and standard deviations of Part I crimes are reported in Table 1). In Panel (a) of Figure 2, we provide a time plot of $\bar{y}_t$ and of the corresponding averages for Part II crimes. These plots illustrate significant seasonal patterns in line with the findings of Gorr et al. (2003). Less apparent at this aggregate level but more visible from the individual census tract time series (not presented here), we observe a tendency for Part II crimes to lead Part I crimes. This is a significant observation and one that will be carefully tested and fully confirmed by our statistical analysis. It has important policy implications in line with the 'broken-windows' hypothesis (Wilson and Kelling, 1982, Anselin et al., 2000, and Cohen and Gorr, 2005).

In Panel (c) of Figure 2, we present the Moran's I statistics for spatial correlation period by

[2]These census tracts include business districts and hospitals (203, 9800), zoo and parks (708, 9801-9805), industrial parks (9806-9809), cemeteries (9810-9811), and stadiums (9812).

period (see Moran, 1948 or Cliff and Ord, 1972). They are computed under the spatial weight matrix used for our spatial models (see Section 4 below). These statistics are standardized in Panel (d) of Figure 2, with a critical value of 1.96 at the 5 percent level. We find that the null hypothesis of no spatial correlation is rejected in all but two months.

Next we computed the temporal autocorrelation function of y_{it} for the 138 census tracts. In Panel (b) of Figure 2, we reproduce the mean autocorrelations across census tracts. The results unambiguously point toward the inclusion of a lagged dependent variable in our count model.

In summary, our descriptive analysis highlights the need for a count model that allows for temporal as well as spatial correlation, together with significant heterogeneity across census tracts. Part of that heterogeneity will be captured by selected regressors (including lagged Part II crimes) and the remainder by random effects.

4 Model specification

Our specification for the number of severe crimes y_{it} reported for census tract i in month t consists of a panel data count model combined with a latent Gaussian spatio-temporal state process. It assumes that the y_{it}'s are mutually independent conditional on the latent state variables λ_{it}, with Poisson distributions whose non-negative mean is specified as $\theta_{it} = \exp(\lambda_{it})$. The corresponding conditional probability density function (pdf) of y_{it} given λ_{it} is given by

$$f(y_{it}|\lambda_{it}) = \frac{\exp\{y_{it}\lambda_{it} - \exp(\lambda_{it})\}}{y_{it}!}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where λ_{it} represents the log of the conditional mean of y_{it}. The latent process is assumed to be a linear Gaussian dynamic panel model in space and time as discussed, e.g., by Elhorst (2010, 2012) and Baltagi et al. (2014). It has the following form:

$$\lambda_t = \kappa\lambda_{t-1} + \rho W\lambda_t + X_t\gamma + \epsilon_t, \tag{2}$$

where $\lambda_t = (\lambda_{it})$ denotes the $N \times 1$ vector of the state variables in period t, $\epsilon_t = (\epsilon_{it})$ the $N \times 1$ vector of error terms, and X_t the $N \times K$ matrix whose ith row $x_{it} = (x_{ikt})$ consists of K covariates observed for census tract i in period t. Potential covariates include (i) observed (time-invariant) socio-economic variables which might affect the rate of severe delinquency, (ii) monthly dummy variables that capture potential seasonality in severe crime activity, and (iii) the log of less severe crime lagged by one month which, according to the broken-windows hypothesis, serves as a potential leading indicator for severe crimes. The $N \times N$ matrix $W = (w_{ij})$ represents the contiguity relations across the N census tracts, where we set $w_{ij} > 0$ only if the borders of tract i and j share at least one common point and $w_{ij} = 0$, otherwise ('queen contiguity'). Following the usual convention, the diagonal elements w_{ii} are set to zero. The spatially lagged state variable $W\lambda_t$ captures potential diffusion and contagion effects in severe crimes and the parameter ρ measures the intensity of the resulting global spatial correlation. The temporally lagged state variable λ_{t-1} with persistence parameter κ accounts for potential census-tract specific intertemporal linkages of crime rates. Moreover, the spatially and temporally lagged states imply that the covariates in x_{it} have not only a contemporaneous and census-tract specific impact but also time-persistent and spatially expanding effects on severe crime rates.

In order to account for potential unobserved time-invariant heterogeneity in the crime rates across census tracts we assume that the error terms in Equation (2) follow a Gaussian random-effect specification of the form

$$\epsilon_t = \tau + e_t, \qquad \text{with} \quad e_t | X_t \sim \mathrm{N}_N(0, \sigma_e^2 I_N), \quad \tau | X_t \sim \mathrm{N}_N(0, \sigma_\tau^2 I_N), \tag{3}$$

where the $N \times 1$ vector $\tau = (\tau_i)$ contains census tract-specific effects that are not accounted for by variables in X_t, and $e_t = (e_{it})$ is the $N \times 1$ vector of idiosyncratic disturbance terms. $\mathrm{N}_N(\cdot, \cdot)$ denotes an N-dimensional Gaussian distribution and I_N the N-dimensional identity matrix. Moreover it is assumed that τ and ϵ_t are independent of each other, conditionally on X_t.

Conditions for invertibility of $(I_N - \rho W)$ and temporal stability of the dynamic spatial panel

specification for the state variables as given by Equation (2) are discussed in Elhorst (2012) and Baltagi et al. (2014). For a W matrix with real eigenvalues, a sufficient condition for invertibility is that $\rho \in (1/\zeta_{\max}, 1/\zeta_{\min})$, where $\zeta_{\min}$ and $\zeta_{\max}$ are the extreme eigenvalues. For an invertible $(I_N - \rho W)$, the model for the state variables in Equations (2) and (3) implies the following Gaussian distribution for λ_t given $(\lambda_{t-1}, \tau, X_t)$:

$$\lambda_t | \lambda_{t-1}, \tau, X_t \sim \mathrm{N}_N(K^* \lambda_{t-1} + m_t^* + \tau^*, H^{-1}), \tag{4}$$

with

$$K^* = \kappa (I_N - \rho W)^{-1}, \qquad m_t^* = (I_N - \rho W)^{-1} X_t \gamma, \qquad \tau^* = (I_N - \rho W)^{-1} \tau, \tag{5}$$

$$H = (I_N - \rho W)'(I_N - \rho W)/\sigma_e^2,$$

and the temporal stability (conditional on the covariates) is guaranteed if the largest absolute eigenvalue of the persistence matrix K^* is smaller than one. As discussed further below in Section 5.1, likelihood based inference requires an appropriate, yet operational treatment of the initial latent state λ_1 in Equations (2) and (4). Finally, note that since census tracts share a common boundary only with a small number of other tracts, the spatial weight matrix W and the resulting precision matrix H as given in Equation (5) are sparse with a very large proportion of zero entries, as typical for spatial applications.

In order to account for missing observations for socio-economic covariates in the 14 census tracts without a regular resident population (parks, cemeteries, etc.) we specify the regression function $X_t \gamma = (x_{it}' \gamma)$ in Equation (2) using a dummy-variable approach. Specifically, lets partition x_{it} and γ into

$$x_{it}' = ([1 - \iota_i] x_i^{(1)'} \;\; \iota_i \;\; x_{it}^{(2)'}), \qquad \gamma' = (\gamma_1^{(1)'} \;\; \gamma_2^{(1)} \;\; \gamma^{(2)'}), \tag{6}$$

where $x_i^{(1)}$ denotes the vector of the (time invariant) socio-economic covariates, $x_{it}^{(2)}$ the remaining covariates and ι_i a dummy variable ($\iota_i = 1$ if tract i has missing observations for $x_i^{(1)}$ and $\iota_i = 0$, otherwise). The associated slope vectors are $\gamma_1^{(1)}$, $\gamma_2^{(1)}$ and $\gamma^{(2)}$. Under this partitioning the

regression function for census tract i is specified as

$$x'_{it}\gamma = (1-\iota_i)x_i^{(1)'}\gamma_1^{(1)} + \iota_i\gamma_2^{(1)} + x_{it}^{(2)'}\gamma^{(2)}. \tag{7}$$

5 EIS based likelihood inference

Likelihood evaluation for the spatio-temporal Poisson panel model discussed in Section 4 requires high-dimensional integration in order to marginalize the joint distribution of the y_{it}'s w.r.t. the latent state and random-effect variables, for which we use the spatial EIS approach introduced in Liesenfeld et al. (2015). Spatial EIS combines the original EIS principle developed by Richard and Zhang (2007) for high-dimensional MC integration with sparse matrix algebra, which allows for fast computation on the large sparse precision matrices typically found in high-dimensional spatial applications. In fact, sparse matrix operations significantly reduce operation counts and memory requirements relative to the corresponding operations on dense matrices (see, e.g., Gilbert et al., 1992, LeSage and Pace, 2009, Pace and LeSage, 2011). This combination of EIS with sparse matrix algebra ensures that accurate MC likelihood evaluations remain computationally feasible even in high-dimensional latent spatial Gaussian models. EIS also enables us to compute MC estimates for the conditional mean of functions of the latent state variables λ_{it} and random effects τ_i given the observed data (smoothing) as well as moments and probabilities of (out-of-sample) predictive distributions for the dependent variable. As we shall see below, those MC estimates are instrumental in the selection of the covariates to be included in the model and the assessment of the predictive performance of the model. Spatial EIS is outlined in the next sections: In Section 5.1 we discuss our assumption for the initial condition λ_1 and present the resulting form of the likelihood integral; Section 5.2 briefly outlines the EIS principle (details of its spatial implementation for the models are regrouped in Appendix 2); In Section 5.3 we present smoothing and prediction based on EIS.

5.1 Likelihood

Likelihood inference based on the unconditional likelihood function would require the stationary distribution of the first-period state variables $f(\lambda_1|X_1,\tau)$ as implied by the conditional distribution (4). However, that stationary distribution is not available due to the presence of time-varying covariates in X_t [3]. In order to cope with this 'initial-condition problem' we treat $\lambda_1 = (\lambda_{i1})$ as a non-stochastic vector and set λ_{i1} equal to $\ln(y_{i1})$, which represents our best guess for the latent log conditional mean $\lambda_{i1} = \ln \mathrm{E}(y_{it}|\lambda_{it})$. This treatment of the initial-condition problem allows for an easy implementation of (conditional) ML based on spatial EIS and appears to be justified in our application for two reasons. First, with $T = 72$ the likelihood contribution of the first period is a relatively small part of the total likelihood; second, as we shall see in the application discussed below, the estimated largest eigenvalue of the persistence matrix K^* is of a moderate size, implying that the impact of λ_1 on future λ_t-values dies out quickly.

An alternative treatment of the initial condition consists of approximating $f(\lambda_1|X_1,\tau)$ by taking the conditional distribution of $\lambda_1|X_1, X_0, X_{-1}, ..., \tau$ and capturing the spatial heterogeneity generated by the unobservable pre-sample values $X_1, X_0, X_{-1}, ...$ by a latent random-effect variable (see Appendix 1). At the cost of additional modifications of the spatial EIS algorithm presented in Section 5.2 and Appendix 2 below, we also implemented ML based on this approximation to the unconditional likelihood. However, since the results we obtained for an initial baseline model indicate that the conditional and approximated unconditional ML lead to essentially the same conclusions, we decided to continue with the simpler conditional ML approach.

Evaluation of the conditional likelihood function requires integrating the joint density of the counts, states and random-effects w.r.t. the latent $N(T-1)$ state and the N random effect variables. Let $\lambda_{(t)} = (\lambda_2', \ldots, \lambda_t')'$, $X_{(t)} = (X_2, \ldots, X_t)$, and $y_{(t)} = (y_2', \ldots, y_t')'$, where $y_t = (y_{it})$ denotes the $N \times 1$ vector of the count response variables in period t. Let ψ denote the parameters to be

[3] For a purely linear spatio-temporal panel model with a small time dimension (T of the order of five) Elhorst (2012) implements an approximation to the unconditional ML-estimator as proposed by Bhargava and Sargan (1983). However, the application of this procedure to our non-linear spatio-temporal Poisson panel model with a large time dimension would amount to estimating an unfeasible large number of additional (auxiliary) parameters.

estimated. The likelihood integral to be evaluated is of the form

$$L(\psi) = \int_{\mathbb{R}^{NT}} \varphi(\lambda_{(T)}, \tau)\, d\lambda_{(T)} d\tau, \quad \text{with} \quad \varphi(\lambda_{(T)}, \tau) = f(y_{(T)}|\lambda_{(T)}) \cdot f(\lambda_{(T)}|X_{(T)}, \tau, \lambda_1) \cdot f(\tau), \quad (8)$$

and

$$f(y_{(T)}|\lambda_{(T)}) = \prod_{t=2}^{T} \prod_{i=1}^{N} f(y_{it}|\lambda_{it}), \qquad f(\lambda_{(T)}|X_{(T)}, \tau, \lambda_1) = \prod_{t=2}^{T} f(\lambda_t|\lambda_{t-1}, X_t, \tau) \quad (9)$$

$$f(\tau) = \prod_{i=1}^{N} f(\tau_i), \quad (10)$$

where $f(y_{it}|\lambda_{it})$ is the Poisson density given in Equation (1) while $f(\tau_i)$ and $f(\lambda_t|\lambda_{t-1}, X_t, \tau)$ are the Gaussian densities defined by Equations (3) and (4), respectively. Note that the likelihood integral (8) can not be evaluated analytically since the response variables are non-Gaussian with a mean that is non-linear in the temporally and spatially correlated latent states.

5.2 Spatial EIS

Following Richard and Zhang (2007), the EIS procedure for MC-evaluation of the likelihood function in Equation (8) is based upon an auxiliary importance sampling (IS) density for $\lambda_{(T)}$ and τ of the form

$$m(\lambda_{(T)}, \tau; a) = \frac{k(\lambda_{(T)}, \tau; a)}{\chi(a)}, \quad \text{with} \quad \chi(a) = \int_{\mathbb{R}^{NT}} k(\lambda_{(T)}, \tau; a) d\lambda_{(T)} d\tau, \quad (11)$$

where $\{k(\lambda_{(T)}, \tau; a), a \in \mathcal{A}\}$ denotes a preselected class of parametric density kernels indexed by a vector of (auxiliary) parameters a and with known integrating factors denoted by $\chi(a)$. For any given a, the corresponding IS MC estimate of the likelihood integral (8) obtains as

$$\hat{L}_S(\psi; a) = \chi(a) \cdot \frac{1}{S} \sum_{s=1}^{S} \omega(\lambda_{(T)}^{[s]}, \tau^{[s]}; a), \quad \text{with} \quad \omega(\lambda_{(T)}, \tau; a) = \frac{\varphi\left(\lambda_{(T)}, \tau\right)}{k\left(\lambda_{(T)}, \tau; a\right)}, \quad (12)$$

where $\{(\lambda_{(T)}^{[s]}, \tau^{[s]})\}_{s=1}^{S}$ denotes S i.i.d. (independently and identically distributed) draws from the IS density $m(\lambda_{(T)}, \tau; a)$. For any given ψ, EIS aims at selecting a value of a that minimizes the MC

sampling variance of the IS ratio ω under these draws.

A (near) optimal value $\hat{a}$ obtains by selecting an initial IS density $m(\lambda_{(T)}, \tau; \hat{a}^{[0]})$ and solving iteratively the following sequence of least squares (LS) approximation problems

$$\hat{a}^{[\ell+1]} = \arg\min_{a \in \mathcal{A}} \sum_{s=1}^{S} \left\{ \ln \varphi \left(\lambda_{(T)}^{[s,\ell]}, \tau^{[s,\ell]} \right) - \ln k \left(\lambda_{(T)}^{[s,\ell]}, \tau^{[s,\ell]}; a \right) \right\}^2, \qquad \ell = 0, 1, 2, ..., L, \tag{13}$$

where $\{(\lambda_{(T)}^{[s,\ell]}, \tau^{[s,\ell]})\}_{s=1}^{S}$ are S i.i.d. draws from $m(\lambda_{(T)}, \tau; \hat{a}^{[\ell]})$. The optimal $\hat{a}$ is the fixed-point solution to the resulting sequence $\{\hat{a}^{[\ell]}\}_{\ell=0}^{L}$. In order to ensure convergence to a (near) fixed-point solution[4] it is critical that all $(\lambda_{(T)}, \tau)$-draws generated for the sequence $\{\hat{a}^{[\ell]}\}$ be produced by using transformations of a single set of canonical random numbers $\{z^{[s]}\}_{s=1}^{S}$ (common random numbers, CRNs). Moreover, since $\hat{a}$ is an implicit function of the model parameters ψ, the EIS LS approximation in Equation (13) has to be rerun for each new value of ψ and the use of CRNs across reruns ensures continuity of the EIS likelihood estimation $\hat{L}_S(\psi; \hat{a})$ with respect to ψ.

In order to select an EIS kernel k, which approximates the target $\varphi(\lambda_{(T)}, \tau)$ defined in Equations (8)-(10), we note that the product $f(\lambda_{(T)}|X_{(T)}, \tau, \lambda_1) \cdot f(\tau)$ in φ defines a Gaussian kernel in $(\lambda_{(T)}, \tau)$. Hence, it is natural to construct k as a Gaussian kernel in $(\lambda_{(T)}, \tau)$ which includes that product together with a Gaussian kernel approximation to the product of independent Poisson densities $f(y_{(T)}|\lambda_{(T)}) = \prod_{t=2}^{T} \prod_{i=1}^{N} f(y_{it}|\lambda_{it})$, which is the sole non-Gaussian term in φ. It follows that k is constructed as

$$k(\lambda_{(T)}, \tau; a) = k^*(\lambda_{(T)}; a^*) \cdot f(\lambda_{(T)}|X_{(T)}, \tau, \lambda_1) \cdot f(\tau), \tag{14}$$

with

$$k^*(\lambda_{(T)}; a^*) = \prod_{t=2}^{T} \prod_{i=1}^{N} k_{it}^*(\lambda_{it}; a_{it}^*), \qquad k_{it}^*(\lambda_{it}; a_{it}^*) = \exp\left\{ -\frac{1}{2}(\alpha_{it}\lambda_{it}^2 - 2\beta_{it}\lambda_{it} + \kappa_{it}) \right\}, \tag{15}$$

where $a^* = (a_{it}^*)$ and $a_{it}^* = (\alpha_{it}, \beta_{it}, \kappa_{it})$. The product of the three Gaussian factors in Equation (14)

[4] In the present application, the number of iterations L is preset at $L = 20$, which proves sufficient for convergence to a close fixed-point approximation $\hat{a} = \hat{a}^{[21]}$. This relatively large value of L is necessitated by the fact that, as explained below, the LS problem in (13) is transformed further into a total of $NT = 9,936$ independent low-dimensional auxiliary LS problems.

produces an EIS kernel k for a Gaussian EIS density m, whose parameter a obtains analytically from a^* and ψ. Since factors common to φ and k cancel out in the EIS regression (13), it follows that the latter simplifies into $N(T-1)$ independent low-dimensional linear LS regressions of

$$\left\{\ln f(y_{it}|\lambda_{it}^{[s,\ell]})\right\}_{s=1}^{S} \quad \text{on}: \quad \left\{\left(\lambda_{it}^{[s,\ell]}\right)^2,\ \lambda_{it}^{[s,\ell]},\ \text{intercept}\right\}_{s=1}^{S}, \quad \text{for } i=1,...,N,\ t=2,...,T. \quad (16)$$

The initial EIS kernels k_{it}^* in Equation (15) are defined as second-order Taylor-series approximations of $\ln f(y_{it}|\lambda_{it})$ in λ_{it}. This produces an initial value for the a^*-vector which can be combined with ψ into $\hat{a}^{[0]}$ [5].

Note that the resulting optimal joint EIS density $m(\lambda_{(T)}, \tau; \hat{a})$ to be used for the MC likelihood estimation according to Equation (12) is an NT-dimensional Normal, where $\hat{a}$ consists of its $NT \times 1$ mean vector and the $NT(NT+1)/2$ parameters in its covariance/precision matrix (in our application NT equals $9,936$). Their construction requires operations on a recursive sequence of auxiliary parameter matrices with terminal size $O(N_*^2)$, where $N_* = N(T-1)$ denotes the dimension of $\lambda_{(T)}$. Hence, a 'brute force' implementation of EIS becomes rapidly computationally prohibitive as NT increases. However, the precision matrix H in our application is a sparse matrix with a large portion of zero entries. Thus, the key to a computationally feasible high-dimensional EIS implementation for our model lies in a recursive construction of the joint EIS density that operates on the sparse precision matrix H (instead of the dense covariance matrix H^{-1}) and, foremost, preserve its sparsity throughout the entire recursion. A description of our specific spatial EIS implementation is provided in Appendix 2.

[5] As discussed in Liesenfeld et al. (2015), in more general latent Gaussian spatial models allowing for truncated, censored or degenerated (Probit) response densities $f(y_{it}|\lambda_{it})$, the EIS parameters $\hat{a}$ and $\hat{a}^*$ would need to be obtained by a recursive sequence of interdependent LS-EIS regressions. Such a sequential EIS implementation becomes necessary in order to account for the integrating factors associated with response densities that depend on latent state variables and, therefore, enter the targets to be approximated by the EIS kernels. For the Poisson model under consideration with a response density which is neither degenerated nor censored or truncated we can implement EIS based on independent EIS regressions (see Equation 16), so that we can construct the joint EIS density $m(\lambda_{(T)}, \tau; a)$ directly rather than from a sequence of conditional EIS densities. As reported in Liesenfeld et al. (2015), this reduces significantly the computing time relative to a sequential EIS implementation.

5.3 Smoothing and Prediction

Once the parameters ψ have been estimated by ML-EIS, the EIS procedure can be used to compute additional statistics of interest such as the smoothed (posterior) estimates of the random effects τ (as in Section 6.2) and functions of the state variables λ as well as moments and probabilities of the out-of sample one-step-ahead predictive distribution of the dependent variable y (as in Section 6.4). All such computations rely upon a straightforward extension (propagation) of the EIS procedure implemented for likelihood evaluations in Section 5.2.

Let $g(\lambda_{T+1}, \lambda_{(T)}, \tau)$ denote a function of interest (dependence on X_{T+1}, $X_{(T)}$ and the model parameters ψ is ignored in the notation). Its conditional mean given $y_{(T)}$ is given by

$$\begin{aligned}\Gamma(y_{(T)}) &= \mathrm{E}\left[g(\lambda_{T+1}, \lambda_{(T)}, \tau)|y_{(T)}\right] \\ &= \frac{\int_{\mathbb{R}^{N(T+1)}} g(\lambda_{T+1}, \lambda_{(T)}, \tau) \cdot f(\lambda_{T+1}|\lambda_T, \tau) \cdot \varphi(\lambda_{(T)}, \tau)\, d\lambda_{T+1} d\lambda_{(T)} d\tau}{\int_{\mathbb{R}^{NT}} \varphi(\lambda_{(T)}, \tau)\, d\lambda_{(T)} d\tau},\end{aligned} \tag{17}$$

where $\varphi(\cdot)$ denotes the EIS target given in Equation (8) and $f(\lambda_{T+1}|\lambda_T, \tau)$ is the Gaussian transition density as defined by Equation (4). The conditional mean (17) is to be evaluated at the ML-EIS estimates of ψ. The denominator is the likelihood function in Equation (8), estimated using the EIS kernel $k(\lambda_{(T)}, \tau; \hat{a})$ and IS ratios $\omega(\lambda_{(T)}, \tau; a)$ as defined in Equation (12). Since the transition density $f(\lambda_{T+1}|\lambda_T, \tau)$ is already a Gaussian density it can be used to draw directly $\lambda_{T+1}|\lambda_T, \tau$ and propagate the EIS trajectories $\{\lambda_{(T)}^{[s]}, \tau^{[s]}\}_{s=1}^S$ into $\{\lambda_{T+1}^{[s]}, \lambda_{(T)}^{[s]}, \tau^{[s]}\}_{s=1}^S$. Thus, the ratio of integrals in Equation (17) is estimated by

$$\hat{\Gamma}_S = \sum_{s=1}^S g\left(\lambda_{T+1}^{[s]}, \lambda_{(T)}^{[s]}, \tau^{[s]}\right) \cdot \bar{\omega}\left(\lambda_{(T)}^{[s]}, \tau^{[s]}; \hat{a}\right), \tag{18}$$

where $\bar{\omega}$ denotes the normalized EIS ratio

$$\bar{\omega}\left(\lambda_{(T)}^{[s]}, \tau^{[s]}; \hat{a}\right) = \frac{\omega\left(\lambda_{(T)}^{[s]}, \tau^{[s]}; \hat{a}\right)}{\sum_{s=1}^S \omega\left(\lambda_{(T)}^{[s]}, \tau^{[s]}; \hat{a}\right)}. \tag{19}$$

Note that this EIS ratio does not depend on λ_{T+1} since $f(\lambda_{T+1}|\lambda_T, \tau)$ being already Gaussian need not be EIS approximated. Also the fact that we use the same EIS draws to estimate both the numerator and denominator in Equation (17) induces positive correlation between their respective EIS estimates, thereby reducing further the MC variance of $\hat{\Gamma}_S$. Next we discuss how this MC-EIS estimate of the conditional expectation of a function $g(\lambda_{T+1}, \lambda_{(T)}, \tau)$ can be used to compute several statistics of interest.

(i) The conditional (posterior) mean of the random effects $\hat{\tau} = \mathrm{E}(\tau|y_{(T)})$ and the dependent variables $\mathrm{E}(y_{it}|y_{(T)})$ $(i = 1, \ldots, N; t = 2, \ldots, T)$ given the sample information obtain from

$$g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv \tau, \quad \text{and} \quad g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv \mathrm{E}(y_{it}|\lambda_{(T)}, \tau) = \exp(\lambda_{it}), \tag{20}$$

respectively.

(ii) The one-step-ahead predictive mean $\mathrm{E}(y_{iT+1}|y_{(T)})$ $(i = 1, \ldots, N)$ obtains from

$$g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv \mathrm{E}(y_{iT+1}|\lambda_{T+1}, \tau) = \exp(\lambda_{iT+1}). \tag{21}$$

(iii) The one-step-ahead predictive cumulative distribution function (cdf) $P(y_{iT+1} \leq y|y_{(T)})$ for any integer value $y \geq 0$ obtain from

$$g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv P(y_{iT+1} \leq y|\lambda_{T+1}, \tau) = \sum_{\ell=0}^{y} \exp\{\ell\lambda_{iT+1} - \exp(\lambda_{iT+1})\}/(\ell!). \tag{22}$$

(iv) We can also use the predictive cdf in Equation (22) to compute randomized Probability Integral Transformation (PIT) residuals that can be used to test the validity of our predictive model (see Jung et al., 2006 and Czado et al., 2009). Specifically, let y^o_{iT+1} denote the (ex-post) observed value of y_{iT+1} and

$$P_a(y^o_{iT+1}) = P(y_{iT+1} \leq y^o_{iT+1} - 1|y_{(T)}), \qquad P_b(y^o_{iT+1}) = P(y_{iT+1} \leq y^o_{iT+1}|y_{(T)}), \tag{23}$$

with $P_a(0) = 0$. If the predictive model is valid, then the randomized predictive PIT residuals, defined as

$$\xi_{iT+1} = P_a(y^o_{iT+1}) + \upsilon_i \left[P_b(y^o_{iT+1}) - P_a(y^o_{iT+1})\right], \quad \upsilon_i \sim \text{i.i.d.} U_{[0,1]}, \tag{24}$$

are uniformly distributed on $[0, 1]$ ($U_{[0,1]}$). Using the inverse of a standardized Gaussian cdf, denoted by Φ, the variable ξ_{iT+1} can be transformed into a $\mathrm{N}(0, 1)$ variable

$$\xi^*_{iT+1} = \Phi^{-1}(\xi_{iT+1}). \tag{25}$$

Thus, if the predictive model is valid, the normalized predictive PIT residuals ξ^*_{iT+1} should follow a standard normal distribution.

6 Empirical results

Our descriptive analysis in Section 3.2 suggests that, in addition to spatio-temporal dependence and random effects for heterogeneity, our model needs to include seasonal dummies as well as lagged Part II crimes as covariates. Furthermore, we propose to explore whether the socio-economic census tract covariates we collected might explain some of the heterogeneity across census tracts, thereby reducing the variance of the random effects. However, with a (potential) total of 11 seasonal dummies and 15 covariates, a full joint specification search would prove computationally expensive. Moreover, it would fail to take full advantage of the orthogonality between the seasonal dummies (constant across census tracts) and the socio-economic variables (constant over time). Thus, we follow instead a sequential specification search consisting of the four following steps:

Step 1: We estimate a 'baseline model' in which the matrix of covariates X_t consists solely of an intercept, lagged log Part II crimes and eleven months dummies, where we select February (typically the lowest crime month) as the base month. *Step 2:* We rely upon Equations (18)-(20) to compute MC-EIS estimates for the posterior means of the random effects $\hat{\tau}_i$ ($i = 1, \ldots, N$) under the baseline model. Next, we regress those estimates on our 15 socio-economic covariates together

with the time averages of log part II crimes, whose inclusion follows from Mundlak (1978) and Wooldridge (2002, p. 487), with the objective of selecting a parsimonious subset of random effect covariates. *Step 3:* We re-estimate a 'full model' by adding the selected covariates to the baseline model. *Step 4:* We produce a 'predictive model' by eliminating insignificant seasonal dummies and covariates from the full model.

We now discuss in turn the results obtained from these four steps.

6.1 Baseline model

The ML-EIS estimates for the baseline model based on an EIS simulation sample size of $S = 500$ and $L = 20$ EIS iterations are reported in Table 2. We note immediately that the key parameters of interest (coefficients for the temporal lag κ, the spatial lag ρ, random effect standard deviation σ_τ and slope coefficient for lagged Part II crime) are all highly significant. Also, the maximum eigenvalue of the persistence matrix K^* is less than 1, which is a sufficient condition for the spatio-temporal stability of the model. These results indicate substantial positive spatial and temporal dependence in Part I crime rates, after controlling for lagged Part II crime and seasonal effects and, also substantial unobserved census-tract specific heterogeneity. The significant positive coefficient for lagged Part II crime is fully in line with the broken-windows hypothesis (Wilson and Kelling, 1982 and Cohen and Gorr, 2005). As for the seasonal dummies, which capture the difference in the seasonality between Part I and Part II crime (see Panel (a) of Figure 2), it is apparent that we can reduce their number since some of their slope estimates are fairly close to each other. However, as mentioned above, such a reduction is essentially irrelevant for steps 2 and 3 and will be postponed until step 4. Another important reason for such postponement is that log Part II crime being itself seasonal, we do not want to restrict the seasonal pattern of Part I crime before we produce our final (predictive) estimate of the coefficient of lagged log Part II crime in step 4.

6.2 Random effects covariates

In Table 3 (panel 1 to 4), we report the results of four auxiliary regressions for the estimated random effects $\hat{\tau}_i$ as obtained from the baseline model by Equations (18)-(20). We use the abbreviation *Alp-k* to denote average log Part k crime (k=*I,II*). The first regression uses *Alp-II* as sole regressor and produces a R^2 of 0.48. The second regression uses all 15 socio-economic covariates but excludes *Alp-II* for an R^2 of 0.42. It indicates that population size (*Ltp*), percentage of African-Americans (*Paa*) and housing units vacancy rate (*Hvr*) have a significantly positive effect on the estimated random effect, while the effect of group quarters proportion (*Gqp*) is significantly negative. These results are in line with those reported in the literature suggesting that concentration of violence typically occur in disadvantaged communities and regions with a large population size (see, e.g., Baller et al., 2001, Helbich and Arsanjani, 2014, Kubrin, 2003, and Tita and Radil, 2013, p. 106).

We also note that *Alp-II* alone provides a better fit than our 15 socio-economic covariates, with an adjusted R^2 of 0.48 versus 0.34, suggesting that *Alp-II* might provide a good proxy for the latter. This is fully confirmed by the results of a regression of *Alp-II* on the 15 socio-economic variables with an R^2 of 0.70, as reported in panel 5 of Table 3. Moreover, it reveals that the significant socio-economic factors for the estimated random effect in panel 2 are also key predictors for less severe crimes. Additional significant determinants for less severe crime are population density (*Lpd*), percentage of population under 18 (*U18*), percentage with a bachelor degree of higher (*Bdh*), all with a negative impact, and rental housing units (*Rhu*) and female headed households (*Fhh*) with a positive effect.

In panel 3 of Table 3, we report a benchmark regression of the $\hat{\tau}_i$'s on *Alp-II* as well as the 15 socio-economic covariates. In line with the previous results, we find a fairly high R^2 of 0.68. For interpretation of the slope coefficients it is important to note that with the inclusion of *Alp-II* in the auxiliary regression the coefficients of the socio-economic variables in panel 3 represent differential effects between the intensity of Part I and Part II crime. Specifically, panels 3, 2 and 5 correspond

respectively to the following auxiliary regressions:

$$\hat{\tau} = \pi_1 \textit{Alp-II} + \pi_2' X^{(\text{se})}, \qquad \hat{\tau} = \pi_3' X^{(\text{se})}, \qquad \hat{\textit{Alp-II}} = \pi_4' X^{(\text{se})}, \tag{26}$$

where $X^{(\text{se})}$ denotes the matrix of the 15 socio-economic covariates. Equation (26) implies that the coefficients of the socio-economic variables in panel 3 are given by $\pi_2 = \pi_3 - \pi_1 \pi_4$. The majority of those differential effects are statistically insignificant; the most significant difference we observe for *Bdh*, suggesting that less severe crime is substantially more concentrated in regions with a low education level than severe crime.

Finally, we proceed to a 'general-to-specific' simplification search, whereby we eliminate from the regression in panel 3 the least significant covariates one at a time. As reported in panel 4 of Table 3, this produces a parsimonious regression with only 5 covariates (*Alp-II*, *Lmi*, *U18*, *Bdh*, *Fhh*) for a minimal loss of fit (R^2 of 0.676 versus 0.683, but adjusted R^2 of 0.662 versus 0.636).

6.3 Full model

Next, we extend the baseline model by adding *Alp-II*, *Lmi*, *U18*, *Bdh* and *Fhh* to the list of covariates and a dummy for census tracts with missing observations for socio-economic covariates according to Equation (7). The results are reported in the right panel of Table 2. The estimated coefficients of the additional covariates are consistent with those obtained in the auxiliary $\hat{\tau}_i$-regression in panel 4 of Table 3 but are less significant since they are now unconditional on $\{\hat{\tau}_i\}_{i=1}^N$, an issue to be addressed in the next section. Compared to the baseline model we find that $\hat{\sigma}_\tau$ has been greatly reduced from 0.486 to 0.256, corresponding to a substantial 77 percent heterogeneity variance reduction. In parallel with this variance reduction the estimated spatial lag coefficient ρ has also been reduced from 0.442 to 0.374 but remains statistically highly significant indicating that socio-economic similarities between census tract help to explain part, but not all, of the spatial clustering in severe crime rates. Also, it confirms the spatially diffusive and contagious nature of severe crime (Tolnay et al., 1996, Morenoff and Sampson, 1997 and Loftin, 1996).

In contrast, the estimated temporal lag coefficient κ and the idiosyncratic standard deviation σ_e are essentially unchanged, as expected from the fact that the added covariates are time invariant. Also the coefficient of lagged Part II crime is nearly unchanged and remains statistically highly significant. As apparent from the results and confirmed further in the next section, the substantial log-likelihood gain of 76.6 (full versus baseline model) is essentially due to the inclusion of *Alp-II* and *Bdh* whose coefficients are both highly significant. We already discussed the interpretation of the *Bdh* coefficient.

All in all, the full model provides a sound statistical basis to analyze the spatio-temporal distribution of Part I crimes and, pending the elimination of unnecessary seasonal dummies and insignificant covariates, a solid basis for Part I crime predictions and policy implementation of the 'broken-windows' hypothesis.

In order to illustrate the flexibility of our model in this respect, we computed short-term and long-term elasticities of Part I crimes w.r.t. changes in lagged Part II crimes. The analytical forms of these elasticities are derived in Appendix 3. Short-term elasticities consider only a 1-period change in lagged Part II crimes while long-term elasticities consider a permanent change in Part II crimes, thereby covering both lagged and average Part II crimes. We computed these elasticities for all 138 census tracts w.r.t. Part II crime changes in each census tract. We distinguish between idiosyncratic elasticities (impact on the tract where the change of Part II takes place) and total elasticities (impact on the entire city). The individual short-term elasticities are provided in Panels (a) and (b) of Figure 3 and in Panels (c) and (d) the corresponding long-term elasticities. As expected from the data, we find that downtown Pittsburgh ($i = 2$ with census tract number 201) has the largest short-term total elasticity. The largest long-run total elasticities are found for downtown and the census tracts $i = 45, 48, 49, 65, 124$ (the corresponding census tract numbers are 1301, 1304, 1306, 1702 and 5632). The first three correspond to a relatively high Part I crime area at the east end of Pittsburgh (East Hills). Tract 1702 corresponds to the south-side (Carson street) and 5632 to the North Side (adjacent to the stadiums), see Figure 1. In Figure 4 we display a color map of the short-term and long-term elasticities for all the Pittsburgh census tracts w.r.t. a

reduction in Part II crime in downtown Pittsburgh. From a crime reduction policy viewpoint these elasticities could play an important role for efficient allocation of scarce police resources.

6.4 Predictive Model

For prediction purposes, it is important that we simplify further the model by eliminating unnecessary seasonal dummies and insignificant covariates, as long as it minimally impacts the goodness of fit. The slope estimates for the seasonal dummies under the full model (see Table 2) suggests that we can safely capture seasonality with only three dummies (March, April to August and September to January). Moreover, we can also eliminate the insignificant covariates *Lmi*, *U18*, and *Fhh*. This leaves us with a fairly parsimonious total of 8 variables in X_t (constant, three seasonal dummies, three covariates and one missing data dummy). The ML-EIS results for the simplified model reported in Table 4, indicate that this elimination of 11 variables produces a log-likelihood ratio test statistics of only 7.6, which is insignificant at the 0.25 level. Moreover, the impact of that elimination on the remaining coefficients is minimal. Last but not least, the remaining coefficients are now all highly significant even at the 0.1 percent significance level except for the missing data dummy which is essentially irrelevant for policy analysis. Such high significance is important from a classical perspective since it guarantees that parameter uncertainty will be essentially negligible relative to model uncertainty for prediction purposes.

One important by-product of our model are one-step-ahead predictions. In order to assess its predictive performance and validity we conducted an extensive out-of-sample predictive exercise for the last year of our sample 2013. Specifically, for each month $t = T' + 1$ in 2013, we re-estimated the model using data up to period T' and produced for each census tract a one-month-ahead point prediction for month $T' + 1$ using the MC-EIS estimate of the predictive mean $\mathrm{E}(y_{iT'+1}|y_{(T')})$ in Equations (18), (19) and (21). This provides us with a total of $138 \times 12 = 1,656$ point predictions which we then compare with two sets of benchmark values obtained by univariate exponential smoothing with values of 0.7 and 0.8 for the smoothing parameter (as recommended by Gourieroux and Monfort, 1997, Section 4.1.3). For all three methods we then compute Mean Squared Forecast

Errors (MSFE) and Mean Absolute Forecast Error (MAFE) for each month in 2013 as well as for the entire year 2013. The results are reported in the right panel of Table 5. We find that monthly as well as yearly our model outperforms exponential smoothing, except for May and June, where the latter performs slightly better. This is a very positive outcome since univariate exponential smoothing is considered to be a competitive benchmark (see, Gorr et al., 2003, or Cohen and Gorr, 2005).

Next, we test the predictive validity of our model over the same forecast period by computing the one-step-ahead predictive PIT residuals according to Equations (22) to (25). A quantile-quantile (QQ) plot of the normalized predictive PIT residuals $\xi^*_{iT'+1}$ in Equation (25) for the full year is shown in Figure 5 and indicates no significant deviation from standard normality. This is confirmed further by the following auxiliary statistics for $\xi^*_{iT'+1}$ we report in the left panel of Table 5: mean, standard deviation, Jarque-Bera test statistic for normality (see, e.g., Lütkepohl, 2007) and corresponding P-values which indicate that the Null hypothesis of normality is never rejected at the 10 percent level (5 percent for July 2013).

As shown by Equation (22) and illustrated by the computation of the predictive PIT residuals, we could also trivially produce complete one-step-ahead predictive cdfs for all 138 census tracts, as well as predictive probabilities for relevant count intervals.

7 Conclusions

Our findings are important at three different levels: Computations, criminology and law enforcement. From a computational viewpoint, we confirm the feasibility of (numerically) accurate likelihood evaluation for a high-dimensional spatio-temporal heterogeneous state-space count model. It also allows for evaluation of a wide range of additional statistics of empirical relevance such as elasticities w.r.t. to covariates, out-of-sample predictive distributions and model validation test statistics.

From a criminological perspective, our results relative to the impact of local socio-economic

covariates on severe and less severe crimes largely support prevailing conjectures in the literature. Moreover, they strongly confirm the 'broken-windows' hypothesis and enables us to use less severe crimes as a key leading indicator of more severe crimes. This implies that the coefficients of the retained covariates in our predictive model represent differential impacts between the intensity of severe and less severe crimes.

Last but not least, the computation of idiosyncratic as well as total elasticities enables us to quantify the impact of a reduction of less severe crimes on severe crime, both locally and globally through spatial diffusion. In combination with the immediate availability of one-month ahead forecast statistics (means, cumulative distributions, predictive intervals) we believe that our model could play a useful role in the efficient allocation of scarce law enforcement resources, in line with but more detailed than in the pioneering results of Cohen and Gorr (2005) and Cohen et al. (2005).

Acknowledgement

The authors are particularly grateful to Chris Briem (UCSUR, University of Pittsburgh) for numerous discussions, clarifications and suggestions as well as for having provided us with the socio-economic census tract data from the Census 2000. R. Liesenfeld and J. Vogler acknowledge support by the Deutsche Forschungsgemeinschaft (grant LI 901/3-1). We thank seminar and conference participants at 2015 Pittsburgh Economics Medley Conference (University of Pittsburgh) for the helpful comments and suggestions.

Appendix 1. Approximation to the first-period density $f(\lambda_1|X_1, \tau)$

According to Equations (4) and (5), λ_t can be written as

$$\lambda_t = K^*\lambda_{t-1} + m_t^* + \tau^* + e_t^*, \qquad e_t^* \sim \mathrm{N}_N(0, H^{-1}). \tag{A-1}$$

Repeated infinite back-substitution starting at $t = 1$ yields

$$\lambda_1 = \ (I - K^*)^{-1}\tau^* + m_1^* + \sum_{\ell=1}^{\infty}(K^*)^{\ell} m_{1-\ell}^* + \sum_{\ell=0}^{\infty}(K^*)^{\ell} e_{1-\ell}^*, \tag{A-2}$$

so that

$$\lambda_1|(X_1, X_0, ..., \tau) \sim \mathrm{N}_N\Big[\ (I - K^*)^{-1}\tau^* + m_1^* + \sum_{\ell=1}^{\infty}(K^*)^{\ell} m_{1-\ell}^*\ ,\ \Omega\Big], \tag{A-3}$$

where $\Omega = \mathrm{Var}\left[\sum_{\ell=0}^{\infty}(K^*)^{\ell} e_{1-\ell}^*\right]$. A closed-form expression for the covariance Ω is given by $\mathrm{vec}(\Omega) = [I_{N^2} - K^* \otimes K^*]^{-1}\mathrm{vec}(H^{-1})$, where $\otimes$ denotes the Kronecker product and $\mathrm{vec}(\cdot)$ the operator that stacks the columns of a matrix into a vector (see, e.g. Hamilton, 1994, p. 265).

The heterogeneity in the mean of the normal distribution (A-3) which is generated by the unobserved pre-sample values of the covariates via the term

$$\sum_{\ell=1}^{\infty}(K^*)^{\ell} m_{1-\ell}^* = \sum_{\ell=1}^{\infty}(K^*)^{\ell}(I_N - \rho W)^{-1} X_{1-\ell}\gamma, \tag{A-4}$$

can be approximated by an $N \times 1$ vector of latent independent Gaussian random variables, say $\xi = (\xi_i)$ with $\xi_i \sim \mathrm{N}(\mu_\xi, \sigma_\xi^2)$. This produces the following random-effect specification for λ_1

$$\lambda_1|(X_1, \tau, \xi) \sim \mathrm{N}_N\left[\ (I_N - K^*)^{-1}\tau^* + m_1^* + \xi\ ,\ \Omega\right], \quad \text{with} \quad \xi \sim \mathrm{N}_N(\iota\mu_\xi, \sigma_\xi^2 I_N), \tag{A-5}$$

where $\iota = (1, \ldots, 1)'$. Integrating the resulting joint Gaussian density for $(\lambda_1, \xi)|(X_1, \tau)$ with respect to ξ leads to the following operational density for $\lambda_1|(\tau, X_1)$ which can be used to approximate the full unconditional likelihood:

$$\lambda_1|(X_1, \tau) \sim \mathrm{N}_N\left[(I_N - K^*)^{-1}\tau^* + m_1^* + \iota\mu_\xi\ ,\ \Omega + \sigma_\xi^2 I_N\right]. \tag{A-6}$$

Appendix 2. Implementation of spatial EIS

This appendix first details the recursive construction of the joint (E)IS kernel $k(\lambda_{(T)}, \tau; a)$ in Equation (14) and of the corresponding density $m(\lambda_{(T)}, \tau; a)$ in Equation (11). Next, we discuss how to exploit the sparsity of the precision matrix H in Equation (5).

Derivation of the EIS kernel in Equation (14). We use the generic notation $f_L(z|\mu, \Sigma)$ to denote a multivariate Normal density for a random vector $z \in \mathbb{R}^L$ with mean vector μ and covariance matrix Σ. Our first step in the construction of the EIS kernel $k(\lambda_{(T)}, \tau; a)$ consists in the recursive derivation of the joint conditional density $f(\lambda_{(T)}|\tau, \lambda_1)$ in Equation (14), where we omit $X_{(T)}$ for the ease of notation. Also we use

$$D = (I_N - \rho W)^{-1} \tag{A-7}$$

as a short-hand notation for the spatial multiplier.

Lemma 1. *The joint density for $\lambda_{(t)}$ given (τ, λ_1) implied by Equation (4) is*

$$f(\lambda_{(t)}|\tau, \lambda_1) = f_{N(t-1)}\left(\lambda_{(t)} \mid c_{(t)} + C_{(t)}\tau \ , \ H_{(t)}^{-1}\right), \qquad t = 2, \ldots, T, \tag{A-8}$$

where, for $t = 2$

$$c_{(2)} = K^*\lambda_1 + m_2^*, \qquad C_{(2)} = D, \qquad H_{(2)} = H = D^{-1'}D^{-1}/\sigma_e^2, \tag{A-9}$$

and, for $t > 2$

$$c_{(t)} = \begin{pmatrix} c_{(t-1)} \\ m_t^* + G_t c_{(t-1)} \end{pmatrix}, \qquad C_{(t)} = \begin{pmatrix} C_{(t-1)} \\ D + G_t C_{(t-1)} \end{pmatrix}, \tag{A-10}$$

$$H_{(t)}^{-1} = \begin{pmatrix} H_{(t-1)}^{-1} & H_{(t-1)}^{-1}G_t' \\ G_t H_{(t-1)}^{-1} & H^{-1} + G_t H_{(t-1)}^{-1} G_t' \end{pmatrix}, \tag{A-11}$$

with

$$G_t = (0 \cdots 0 \; K^*) \in \mathbb{R}^{N \times N(t-2)}. \tag{A-12}$$

Proof: The initial parameter values in Equation (A-9) follow from Equation (4) for $t = 2$. The recursions for the subsequent parameter values in Equations (A-10) and (A-11) obtain by application of standard Gaussian algebra to the product of $f(\lambda_t|\lambda_{t-1}, \tau) \equiv f(\lambda_t|\lambda_{(t-1)}, \tau, \lambda_1)$ in Equation (4) and $f(\lambda_{(t-1)}|\tau, \lambda_1)$ in Equation (A-8). □

Note that the partitioned covariance matrix in Equation (A-11) implies the partitioned precision matrix:

$$H_{(t)} = \begin{pmatrix} H_{(t-1)} + G_t' H G_t & -G_t' H \\ -HG_t & H \end{pmatrix}, \tag{A-13}$$

with determinant $|H_{(t)}| = |H_{(t-1)}| \cdot |H|$. Hence, $|H_{(t)}| = |H|^{t-1}$, $t = 2, \ldots, T$.

In order to combine $f(\lambda_{(T)}|\tau, \lambda_1)$ as obtained from Lemma 1 with $f(\tau) \equiv f(\tau|\lambda_1)$ in Equation (10) and the Gaussian kernel $k^*(\lambda_{(T)}; a^*)$ defined in Equation (15), we write the Gaussian density in Equation (A-8) as

$$f(\lambda_{(T)}|\tau, \lambda_1) = \exp\left\{-\frac{1}{2}(\eta' P^* \eta - 2\eta' q^* + s^*)\right\}, \tag{A-14}$$

with $\eta' = (\lambda_{(T)}', \tau')$, and

$$P^* = \begin{pmatrix} H_{(T)} & B_{(T)} \\ B_{(T)}' & F_{(T)} \end{pmatrix}, \qquad B_{(T)} = -H_{(T)} C_{(T)}, \qquad F_{(T)} = C_{(T)}' H_{(T)} C_{(T)}, \tag{A-15}$$

$$q^* = \begin{pmatrix} b_{(T)} \\ a_{(T)} \end{pmatrix}, \qquad b_{(T)} = H_{(T)} c_{(T)}, \qquad a_{(T)} = -C_{(T)}' H_{(T)} c_{(T)}, \tag{A-16}$$

$$s^* = c_{(T)}' H_{(T)} c_{(T)} + N(T-1)\ln(2\pi) - (T-1)\ln|H|. \tag{A-17}$$

Next, the kernel $k^*(\lambda_{(T)}; a^*)$ in Equation (15) is written as a joint Gaussian kernel in $\lambda_{(T)}$

$$k^*(\lambda_{(T)}; a^*) = \exp\left\{-\frac{1}{2}(\lambda_{(T)}' A \lambda_{(T)} - 2\lambda_{(T)}'\beta + \kappa)\right\}, \tag{A-18}$$

where $A = \text{diag}(\alpha_{it})$ for $i = 1, \ldots, N$ and $t = 2, \ldots, T$, $\beta' = (\beta_2', \ldots, \beta_T')$ with $\beta_t' = (\beta_{1t}, \ldots, \beta_{Nt})$ and $\kappa = \sum_{t=2}^{T} \sum_{i=1}^{N} \kappa_{it}$. Also (see Equation 3)

$$f(\tau) = (2\pi\sigma_\tau^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_\tau^2}\tau'\tau\right\}. \tag{A-19}$$

Finally, the Gaussian kernel $k(\lambda_{(T)}, \tau; a)$ in Equation (14) obtains as the product of the three densities/kernel in Equations (A-14), (A-18), and (A-19) and can be written as

$$k(\eta; a) = \exp\left\{-\frac{1}{2}(\eta' P \eta - 2\eta' q + s)\right\}, \tag{A-20}$$

with

$$P = \begin{pmatrix} H_{(T)} + A & B_{(T)} \\ B_{(T)}' & F_{(T)} + I_N/\sigma_\tau^2 \end{pmatrix}, \quad q = \begin{pmatrix} b_{(T)} + \beta \\ a_{(T)} \end{pmatrix}, \quad s = s^* + \kappa + N[\ln(2\pi) + \ln(\sigma_\tau^2)]. \tag{A-21}$$

Its integrating factor in Equation (11) is given by

$$\chi(a) = \int_{\mathbb{R}^{NT}} k(\eta; a) d\eta = (2\pi)^{\frac{NT}{2}} |P|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(s - q' P^{-1} q)\right\}. \tag{A-22}$$

Exploiting sparsity of H using sparse matrix algebra. Since H and G_t are sparse, it follows from recursion (A-13) that $H_{(T)}$ and subsequently P in Equation (A-21) are themselves sparse. Hence, we can significantly accelerate the computations required to obtain the EIS density by running the recursion using sparse matrix functions. Moreover, we can also exploit sparsity to accelerate sampling from the large NT-dimensional EIS density $m(\eta; a)$ associated with the kernel $k(\eta; a)$ in Equation (A-20). Specifically, let $\eta = \mu + u$, with $\mu = P^{-1}q$ and $u \sim \text{N}_{NT}(0, P^{-1})$, where μ obtains directly by solving the sparse system $P\mu = q$, thereby avoiding computationally costly inversions of P. Next, we reorder the elements in u according to a Symmetric Approximate Minimum Degree (SAMD) permutation of P (see, Amestoy et al. 1996). The resulting SAMD-permuted u-vector denoted by $\tilde{u}$ with $\tilde{u} \sim \text{N}_{NT}(0, \tilde{P}^{-1})$ has a precision matrix $\tilde{P}$ with a sparse

Cholesky factorization $\tilde{P} = L'L$, where L is upper triangular. Draws from $\tilde{u}$ obtain as follows: We first draw $z^{[s]}$ from a $\mathrm{N}_{NT}(0, I_{NT})$-distribution, then solve the sparse linear system $L\tilde{u}^{[s]} = z^{[s]}$ and, finally, invert the SAMD permutation to retrieve $u^{[s]}$ from $\tilde{u}^{[s]}$. The corresponding draw from the EIS density $m(\eta; a)$ is given by $\eta^{[s]} = \mu + u^{[s]}$. It is important to note that neither the EIS parameter a^* nor the model parameters ψ impact the SAMD permutation which is, therefore, only computed once. Moreover, we note that $|P| = |\tilde{P}| = |L|^2$, so that $\ln|P|$ in the integrating factor in Equation (A-22) can be computed fast from $\ln|P| = 2\sum_{j=1}^{NT} \ln L_{jj}$, where L_{jj} denotes the j-th diagonal element of L.

All in all, reliance upon sparse matrix operations as outlined above, results in computing time for EIS likelihood evaluations that are $O(N_*^\delta)$ with $N_* = N(T-1)$ and δ close to one instead of $O(N_*^3)$ under a 'brute-force' EIS implementation.

Appendix 3. Elasticities w.r.t. lagged Part II crimes

Short-term elasticities. From Equations (2)-(3) it follows that

$$\lambda_t = \kappa D\lambda_{t-1} + DX_t\gamma + D\epsilon_t, \qquad \epsilon_t \sim \mathrm{N}_N(0, \sigma_\epsilon^2 I_N), \qquad \sigma_\epsilon^2 = \sigma_\tau^2 + \sigma_e^2, \tag{A-23}$$

where $D = (d_{ij})$ denotes the $N \times N$ matrix of spatial multipliers as defined in Equation (A-7). Therefore,

$$\mu_{it} = \mathrm{E}(y_{it}|X_t, \lambda_{t-1}) = \exp\{n_{it}\}, \quad \text{with} \quad n_{it} = \kappa d_i'\lambda_{t-1} + d_i'X_t\gamma + \frac{1}{2}\sigma_\epsilon^2 d_i'd_i, \tag{A-24}$$

where d_i' denotes the ith row of D. Let z_{jt} denote lagged Part II crime for census tract j and, for convenience let its logs be stacked in the first column of X_t with slope coefficient γ_1, so that

$$x_{j1t} = \ln(z_{jt}), \quad j = 1, \ldots, N. \tag{A-25}$$

It immediately follows that

$$\frac{\partial \mu_{it}}{\partial z_{jt}} = \mu_{it} \frac{\partial n_{it}}{\partial z_{jt}} = \frac{\mu_{it}}{z_{jt}} d_{ij} \gamma_1. \tag{A-26}$$

Therefore, the $N \times N$ matrix of the (time invariant) short-term elasticities of the μ_{it}'s w.r.t. the z_{jt}'s $(i, j = 1, ..., N)$ is given by

$$\Phi^{\text{short}} = \left(\frac{\partial \mu_{it}}{\partial z_{jt}} \frac{z_{jt}}{\mu_{it}} \right) = D\gamma_1. \tag{A-27}$$

Note that the diagonal elements of Φ^{short} represent the idiosyncratic short-term elasticities of Part I crime in the N census tracts w.r.t. their respective lagged Part II crime. We can also compute the corresponding 'total' elasticities for Pittsburgh. Let $\bar{\mu}_t = \sum_{i=1}^{N} \mu_{it}$. It follows that

$$\frac{\partial \bar{\mu}_t}{\partial z_{jt}} \frac{z_{jt}}{\bar{\mu}_t} = \sum_{i=1}^{N} \left(\frac{\partial \mu_{it}}{\partial z_{jt}} \frac{z_{jt}}{\mu_{it}} \right) \frac{\mu_{it}}{\bar{\mu}_t}, \tag{A-28}$$

where μ_{it} as given in Equation (A-24) is evaluated by setting X_t to its time average and λ_{t-1} to its approximate time average $\ln[(1/T) \sum_{t=1}^{T} y_t]$.

Long-term elasticities. To find long-term elasticities, we proceed by (infinite) forward substitution in the λ_t-Equation (A-23) under a fixed scenario $X_t = X \ \forall t$ which yields

$$\lambda^{\infty} = \lim_{L \to \infty} \lambda_{t+L} = \lim_{L \to \infty} \sum_{\ell=0}^{L} \kappa^{L-\ell} D^{L-\ell+1} X\gamma + \upsilon_t, \qquad \upsilon_t = \lim_{L \to \infty} \sum_{\ell=0}^{L} \kappa^{L-\ell} D^{L-\ell+1} \epsilon_{t+\ell}. \tag{A-29}$$

It follows that the limiting distribution of the vector $\lambda^{\infty} = (\lambda_i^{\infty})$ is a Normal with (conditional) mean $\Delta X\gamma$, where

$$\Delta = \lim_{L \to \infty} \sum_{\ell=0}^{L} \kappa^{L-\ell} D^{L-\ell+1} = D(I_N - \kappa D)^{-1} = [(1-\kappa) I_N - \rho W]^{-1}, \tag{A-30}$$

and its (conditional) stationary covariance matrix Ψ is that of the infinite sequence υ_t. Its closed

form is given by (see, e.g. Hamilton, 1994, p. 265)

$$\text{vec}(\Psi) = \sigma_\epsilon^2[I_{N^2} - \kappa^2(D \otimes D)]^{-1}\text{vec}(DD'). \tag{A-31}$$

Therefore,

$$\lambda_i^\infty | X \sim \text{N}_1(\delta_i' X\gamma, \psi_{ii}), \tag{A-32}$$

where δ_i' denotes the i'th row of Δ and ψ_{ii} the i'th diagonal element of Ψ. It follows that

$$\mu_i^\infty = \lim_{L\to\infty} \text{E}(y_{it+L}|X) = \exp\{n_i^\infty\}, \quad \text{with} \quad n_i^\infty = \delta_i' X\gamma + \frac{1}{2}\psi_{ii}. \tag{A-33}$$

When computing the long-term impact of a permanent change in z_{jt} (lagged Part II crime in census tract j) we need to account for the fact that such a change also affects its time average. Assuming that these two variables are stacked in columns 1 and 2 of X, we have $x_{j1} = x_{j2} = \ln(z_j)$, where z_j denotes the (hypothetical) permanent value assigned to z_{jt}. It immediately follows that the long-term elasticities are given by Equation (A-27), where D is replaced by Δ and γ_1 by $(\gamma_1 + \gamma_2)$, i.e.

$$\Phi^{\text{long}} = \Delta(\gamma_1 + \gamma_2). \tag{A-34}$$

Long-term elasticities for total Pittsburgh are then given by

$$\frac{\partial \bar{\mu}^\infty}{\partial z_j}\frac{z_j}{\bar{\mu}^\infty} = \sum_{i=1}^{N}\left(\frac{\partial \mu_i^\infty}{\partial z_j}\frac{z_j}{\mu_i^\infty}\right)\frac{\mu_i^\infty}{\bar{\mu}^\infty}, \tag{A-35}$$

with $\bar{\mu}^\infty = \sum_{i=1}^N \mu_i^\infty$. For the computation of those long-term elasticities we evaluate μ_i^∞ as given in Equation (A-33) by setting z_j and the remaining variables in X equal to their corresponding time averages.

References

Amestoy, P.R., Davis, T.A., Duff, I.S., 1996. An approximate minimum degree ordering algorithm. SIAM Journal on Matrix Analysis and Applications 17, 886-905.

Anselin, L., 1988. Spatial Econometrics. Kluwer Academic, Boston Massachusetts.

Anselin, L., Cohen, J., Cook, D., Gorr, W., Tita, G., 2000. Spatial analysis of crime. In: Duffee, D. (ed) Measurement and Analysis of Crime and Justice. National Institute of Justice/NCJRS, Rockville, 213–166.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., Hawkins, D.F., 2001. Structural covariates of U.S. county homicide rates: incorporating spatia effects. Criminology 39, 561–590.

Baltagi, B.H., Fingleton, B., Pirotte, A., 2014. Estimating and forecasting with a dynamic spatial panel data model. Oxford Bulletin of Economics and Statistics 76, 112–138.

Bernasco, W., Elffers, H., 2013. Statistical analysis of spatial crime data. In: Piquero, A.R., Weisburd, D. (eds) Handbook of Quantitative Criminology. Springer, 699–724.

Bhargava, A., Sargan, J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51, 1635–1659.

Cliff, A., Ord, J.K., 1972. Testing for spatial autocorrelation among regression residuals. Geographical Analysis 4, 267–284.

Cohen, J., 1941. The geography of crime. Annals of the American Academy of Political and Social Science 217, Crime in the United States, 29–37

Cohen, J., Gorr, W.L., 2005. Development of crime forecasting and mapping systems of use by police. Report, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh.

Cohen, J., Gorr, W.L., Olligschlaeger,. A.M., 2007. Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. Geographical Analysis 39, 105–127.

Cohen, L.E., Felson, M., 1979. Social change and crime rate trends: A routine activity approach. American Sociological Review 44, 588–605.

Cornish, L.E., Clark, R.V., 1986. The Reasoning Criminal: Rational Choice Perspectives on Offending. Springer-Verlag.

Czado, C., Gneiting, T., Held, L., 2009. Predictive model assessment for count data. Biometrics 65, 1254–1261.

Elhorst, J.P., 2010. Dynamic panels with endogenous interaction effects when T is small. Regional Science and Urban Economics 40, 272–282.

Elhorst, J.P., 2012. Dynamic spatial panels: models, methods, and inferences. Journal of Geographical Systems 14, 5–28.

Gilbert, J.R., Moler, C., Schreiber, R., 1992. Sparse matrices in MATLAB: Design and implementation. SIAM Journal on Matrix Analysis and Applications 13, 333–356.

Gorr, W., Harries, R., 2003. Introduction to crime forecasting. International Journal of Forecasting 19, 551–555.

Gorr, W., Olligschlaeger, A., Thompson, Y., 2003. Short-term forecasting of crime. International Journal of Forecasting 19, 579–594.

Gourieroux, C., Monfort, A., 1997. Time Series and Dynamic Models. Cambridge University Press, Cambridge.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, New Jersey.

Helbich M., Arsanjani, J.J., 2014. Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. Cartography and Geographic Information Science 42, 1–15.

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

Kubrin, C.E, 2003. Structural covariates of homicide rates: Does type of homicide matter? Journal of Research in Crime and Delinquency 40, 139–170.

Land, K.C., McCall, P.L., Cohen, L.E., 1990. Structural covariates of homicide rates: Are there any invariances across time and social space? American Sociological Review 95, 992–963.

LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Taylor and Francis Group.

Li, G., Haining, R., Richardson, S., Best, N., 2014. Space-time variability in burglary risk: A Bayesian spatio-temporal modelling approach. Spatial Statistics 9, 180–191.

Liesenfeld R., Richard, J.-F., Vogler, J., 2015. Likelihood evaluation of high-dimensional spatial latent Gaussian models with non-Gaussian response variables. Working paper, University of Cologne (available at SSRN: http://ssrn.com/abstract=2196041).

Loftin, C, 1986. Assaultive violence as a contagious process. Bulletin of New York Academy of Medicine 62, 550–555.

Lütkepohl, H., 2007. New Introduction to Multiple Time Series Analysis. Springer, Berlin.

Moran, P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society B 10, 243–251.

Morenoff, J., Sampson, R.J., 1997. Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970-1990. Social Forces 76, 31–64.

Morenoff, J., Sampson, R.J., Raudenbush, S., 2001. Neighborhood inequality, collective efficacy, and the spatial dynamics of urban crime violence. Criminology 39, 517–560.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.

Pace, R.K., LeSage, J.P., 2011. Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. Working paper, Louisiana State University, Baton Rouge.

Ratcliffe, J.H., 2013. Crime mapping: Spatial and temporal challenges. In: Piquero, A.R., Weisburd, D. (eds) Handbook of Quantitative Criminology. Springer, 5–24.

Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. Journal of Econometrics 141, 1385–1411.

Roth, R.E., Ross, K.S., Finch, B.G., Luo, W., MacEachren, A.M., 2013. Spatiotemporal crime analysis in U.S. law enforcement agencies: Current practices and unmet needs. Government Information Quarterly 30, 226–240.

Shaw, C.R., McKay, H.D., 1942. Juvenile Delinquency and Urban Areas: A study of Rates of Delinquents in Relation to Differential Characteristics of Local Communities in American Cities. University of Chicago Press.

Tita, G.E., Greenbaum, R., 2008. Crime neighborhood and units of analysis: Putting space in its place. In: Weisburd, D., Bernasco, W., Bruinsma, G.J.N., (eds) Putting Crime in its Place: Units of Analysis in Spatial Crime Research. Springer, 145–170.

Tita, G.E., Radil, S.M., 2013. Spatial regression models in criminology: Modeling Social processes in the spatial weights matrix. In: Piquero, A.R., Weisburd, D. (eds) Handbook of Quantitative Criminology. Springer, 101–121.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46, 234–240.

Tolnay, S.E., Deane, G., Beck, E.M., 1996. Vicarious violence: Spatial effects on southern lynchings, 1890-1919. American Journal of Sociology 102, 788–815.

U.S. Department of Justice, Federal Bureau of Investigation, 2004. Uniform Crime Reporting Handbook.

Willits, D., Broidy, L., Dennman, K., 2013. Schools, neighborhood risk factors, and crime. Crime & Delinquency 59, 292–315.

Wilson, J.Q., Kelling, G.L., 1982. Broken windows: The police and neighborhood safety. Atlantic Monthly 249, 29–38.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge.

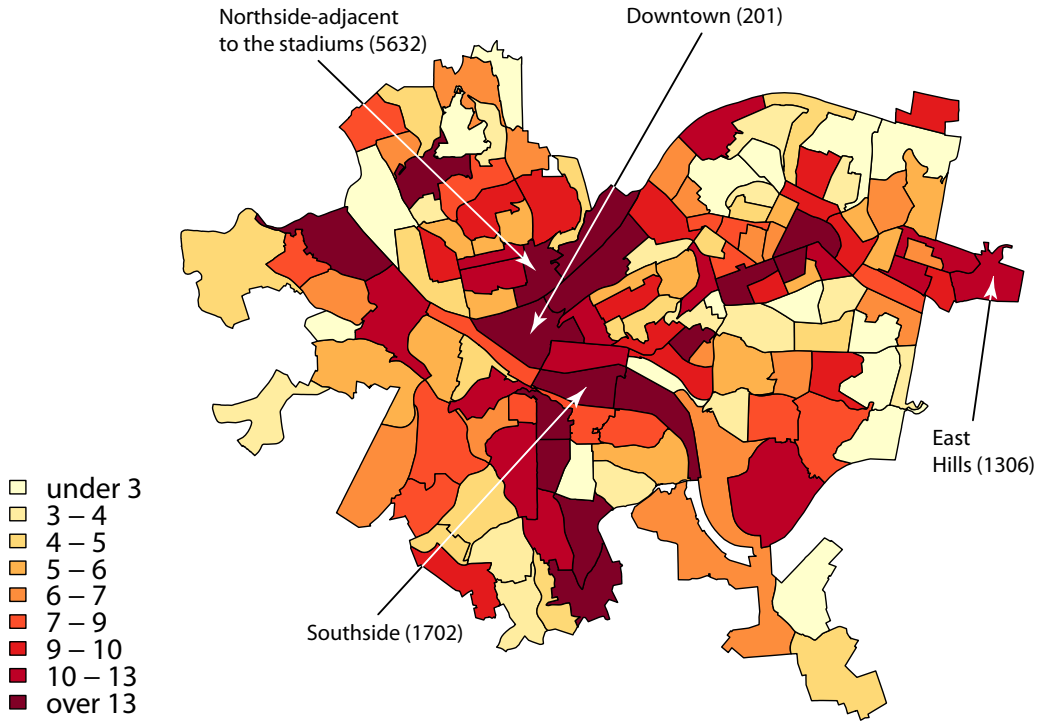

Figure 1. Time averages of part I crimes $\bar{y}_i$ in the $N = 138$ Pittsburgh census tracts.

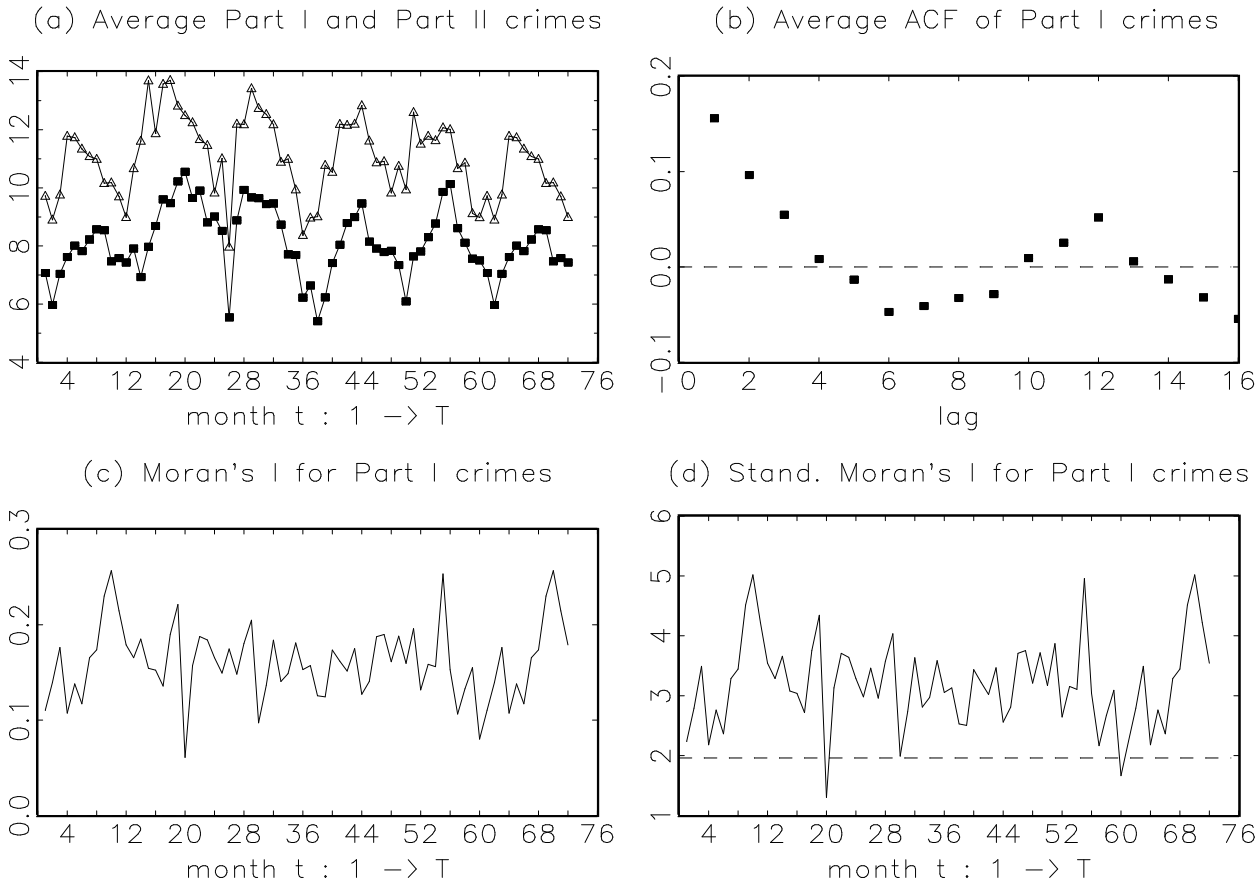

Figure 2. Panel (a): Time plot of Part I crimes $\bar{y}_t$ (solid squares) and Part II crimes (triangles) averaged across census tracts; Panel (b): Temporal autocorrelations averaged across census tracts; Panel (c): Period-by-period time plot of Moran's I for Part I crimes; Panel (d): Period-by-period time plot of the standardized Moran's I for Part I crimes (dotted line: critical value at the 5 percent level);

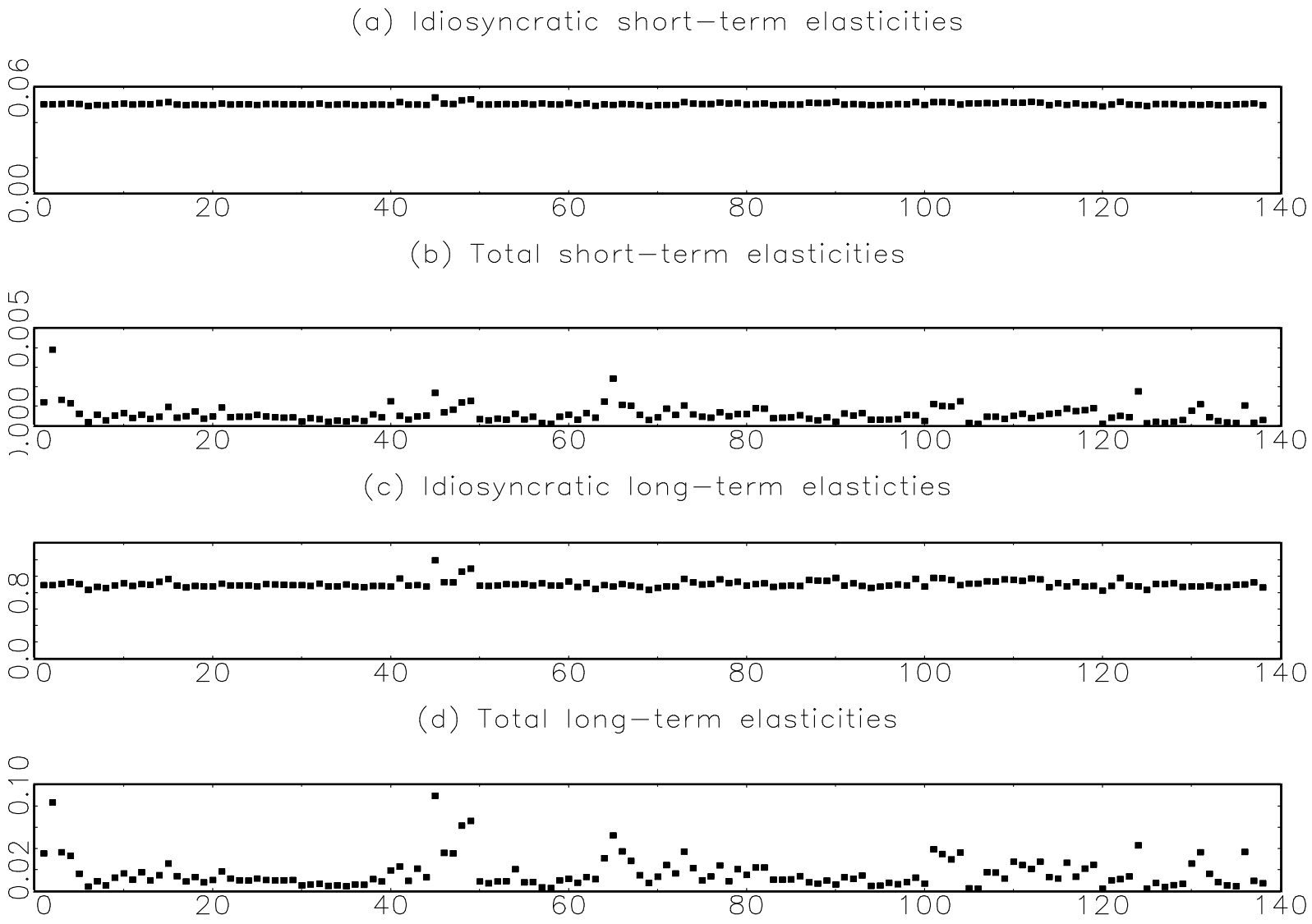

Figure 3. Elasticity of Part I crime w.r.t. a change in lagged Part II crime in tract i ($i = 1, \ldots, 138$); Panel (a): Idiosyncratic short-term elasticity; Panel (b): Total short-term elasticity; Panel (c): Idiosyncratic long-term elasticity; Panel (d): Total long-term elasticity.

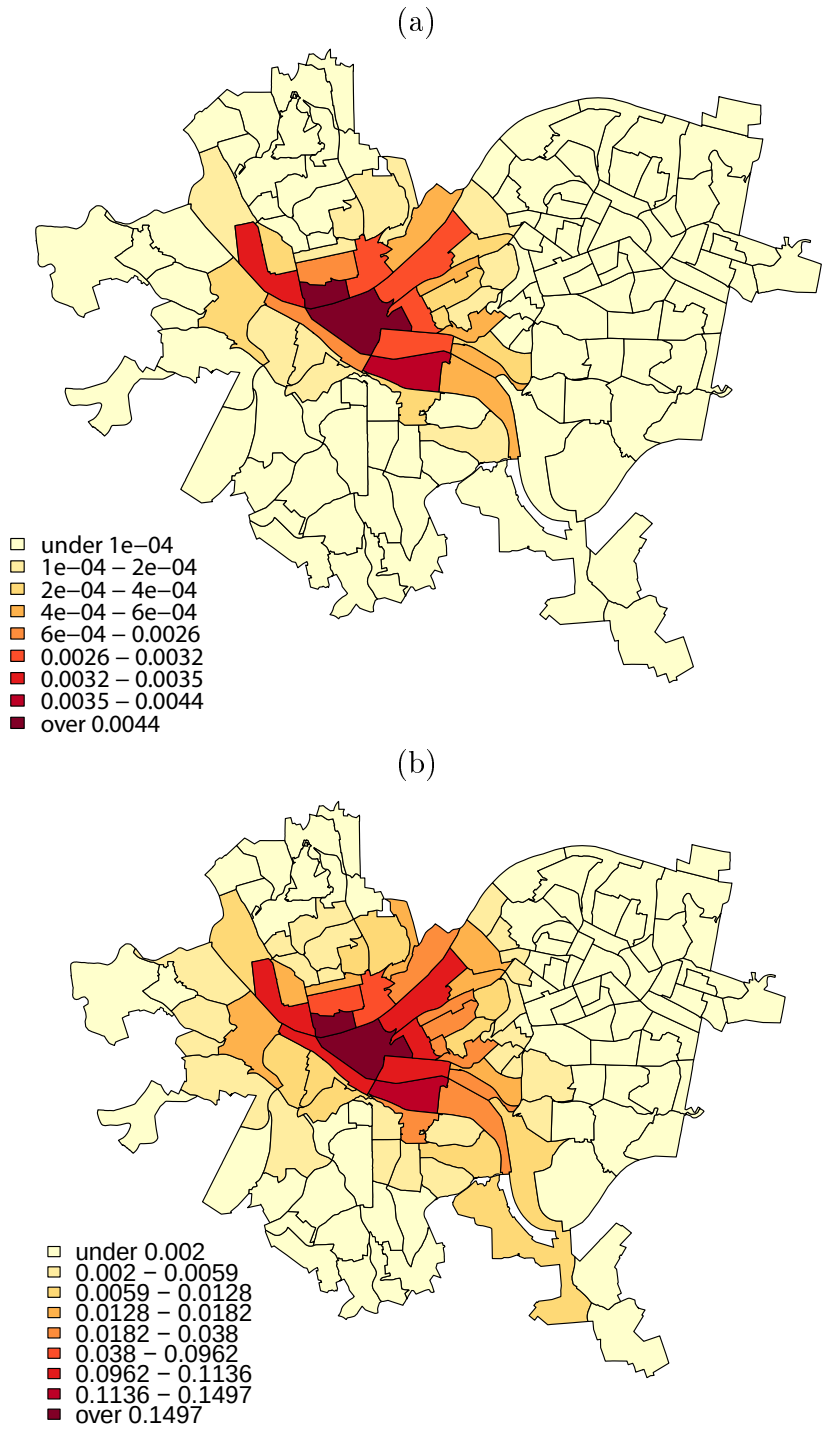

Figure 4. Short-term (panel a) and long-term (panel b) Part I crime elasticities for the Pittsburgh census tracts w.r.t. a change in lagged Part II crime in Downtown ($i = 2$, census tract number 201).

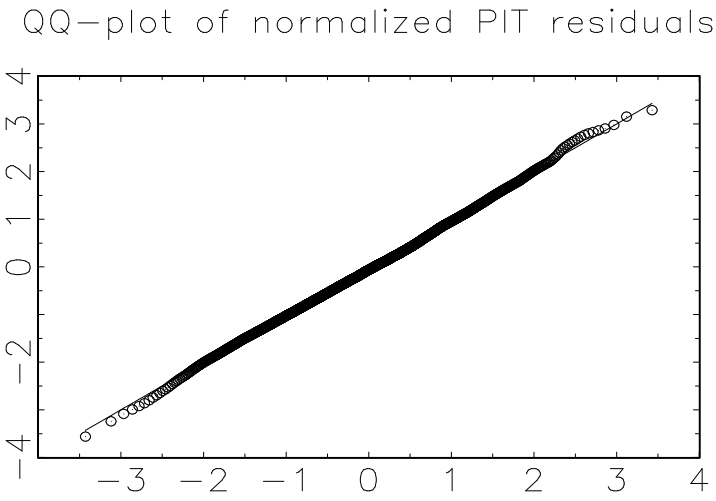

Figure 5. Quantile plots of the normalized one-step-ahead predictive PIT residuals $\xi^*_{iT'+1}$ *computed for* $i = 1, ..., N$ *and* $T' + 1 = 01/13, ..., 12/13$*. The solid line plots the quantiles of a standard normal distribution against the quantiles of standard normal distribution and the dotted line plots the sorted* $\xi^*_{iT'+1}$*'s against the quantiles of standard normal distribution.*

Table 1. List of the Pittsburgh census tracts

i	Pgh census tract	Aver. Part I crime	Std. Part I crime	i	Pgh census tract	Aver. Part I crime	Std. Part I crime
1	103	10.3	3.3	45	1301	11.1	4.4
2	201	74.6	12.7	46	1302	7.1	3.8
3	203	15.4	5.4	47	1303	11.5	4.4
4	305	10.6	4.7	48	1304	9.3	3.4
5	402	8.9	4.6	49	1306	12.4	4.5
6	404	2.9	2.3	50	1401	3.2	1.9
7	405	14.3	4.7	51	1402	2.5	2.2
8	406	6.2	3.1	52	1403	5.3	2.4
9	409	8.6	3.9	53	1404	3.4	2.5
10	501	9.6	5.1	54	1405	7.1	3.1
11	506	5.7	2.4	55	1406	6.3	3.2
12	509	5.4	2.7	56	1408	9.7	3.9
13	510	4.9	3.1	57	1410	3.1	2.2
14	511	4.4	3.6	58	1411	1.7	1.9
15	603	9.5	3.8	59	1413	6.8	3.4
16	605	3.3	2.2	60	1414	7.5	3.9
17	703	10.0	4.5	61	1516	3.9	2.7
18	705	12.5	5.4	62	1517	8.3	3.2
19	706	5.9	4.1	63	1608	8.0	3.5
20	708	11.3	3.6	64	1609	19.3	7.8
21	709	13.5	5.8	65	1702	41.3	10.7
22	802	4.6	2.8	66	1706	8.6	4.8
23	804	7.6	2.7	67	1803	15.5	5.6
24	806	6.9	3.7	68	1807	8.2	3.8
25	807	6.8	3.2	69	1903	5.4	3.1
26	809	8.5	3.6	70	1911	5.6	2.6
27	901	7.0	3.2	71	1914	12.3	4.3
28	902	6.6	2.9	72	1915	7.0	3.1
29	903	8.1	3.1	73	1916	8.3	3.1
30	1005	2.5	1.7	74	1917	5.0	2.3
31	1011	10.6	4.1	75	1918	9.8	4.1
32	1014	5.4	2.8	76	1919	5.0	3.1
33	1016	2.9	2.2	77	1920	7.4	4.0
34	1017	5.2	2.8	78	2022	8.5	4.3
35	1018	3.1	1.9	79	2023	6.7	3.4
36	1102	8.8	4.4	80	2107	8.7	3.2
37	1106	3.7	2.6	81	2206	11.1	3.6
38	1113	9.7	3.3	82	2406	12.8	4.8
39	1114	6.1	2.9	83	2412	4.6	2.5
40	1115	25.7	8.1	84	2503	5.4	2.7
41	1203	7.0	3.0	85	2507	5.4	2.7
42	1204	5.6	3.2	86	2509	6.1	3.5
43	1207	5.7	2.7	87	2602	6.5	3.0
44	1208	8.9	3.9	88	2607	3.9	2.3

continued overleaf

Table 1 (continued). List of the Pittsburgh census tracts

i	Pgh census tract	Aver. Part I crime	Std. Part I crime	i	Pgh census tract	Aver. Part I crime	Std. Part I crime
89	2609	6.5	3.6	114	5620	11.5	5.2
90	2612	1.2	1.4	115	5623	12.5	4.3
91	2614	9.8	3.9	116	5624	10.5	4.3
92	2615	7.2	3.3	117	5625	16.1	6.2
93	2620	9.0	3.3	118	5626	11.0	4.0
94	2701	7.2	3.7	119	5627	10.7	4.5
95	2703	6.5	2.5	120	5628	3.1	2.0
96	2704	3.9	2.4	121	5629	6.2	3.1
97	2708	5.3	3.3	122	5630	5.2	3.1
98	2715	16.4	5.9	123	5631	6.1	3.4
99	2814	7.1	3.4	124	5632	23.7	7.2
100	2815	2.5	1.6	125	9800	9.5	3.7
101	2901	11.5	4.0	126	9801	1.9	1.8
102	2902	13.5	4.4	127	9803	0.6	0.7
103	2904	14.4	5.5	128	9804	0.7	0.9
104	3001	19.0	5.8	129	9805	6.0	5.0
105	3102	4.6	2.5	130	9806	5.1	2.6
106	3103	0.7	1.0	131	9807	8.0	3.5
107	3204	4.7	2.8	132	9808	0.7	0.9
108	3206	4.1	1.9	133	9809	0.8	1.0
109	3207	4.3	2.9	134	9810	0.3	0.6
110	4810	0.8	0.9	135	9811	0.2	0.5
111	5616	5.9	3.4	136	9812	6.0	4.1
112	5617	2.9	2.5	137	9818	0.5	0.8
113	5619	5.8	2.5	138	9822	4.4	2.4

Table 2. ML-EIS results for the spatio-temporal Poisson panel model

	Baseline model		Full model	
Variable	Estimate	Asy. s.e.	Estimate	Asy. s.e.
Temporal lag (κ)	.372***	.027	.387***	.029
Spatial lag (ρ)	.442***	.030	.374***	.037
σ_τ	.486***	.038	.256***	.025
σ_e	.202***	.006	.210***	.006
Constant	.082	.053	−.627	1.028
Jan	.121***	.020	.135***	.022
Mar	.212***	.023	.230***	.025
Apr	.169***	.019	.192***	.021
May	.164***	.019	.190***	.021
Jun	.145***	.019	.171***	.021
Jul	.170***	.019	.199***	.022
Aug	.167***	.019	.197***	.022
Sep	.108***	.018	.132***	.020
Oct	.101***	.018	.120***	.020
Nov	.115***	.018	.133***	.020
Dec	.109***	.018	.124***	.019
Lagged log Part II	.058***	.009	.049***	.009
Averg. log Part II			.449***	.037
Lmi			−.039	.101
U18			.829	.539
Bdh			.825***	.164
Fhh			−.765	.568
Missing data dummy			−.118	.105
Max. eigenvalue of K^*	.666		.618	
Log-likelihood	−24,259.96		−24,183.34	
LR-stat. $H_0 : \kappa = \rho = 0$			429.13	
LR-stat. $H_0 : \kappa = 0$			233.16	
LR-stat. $H_0 : \rho = 0$			110.71	

NOTE: The ML-EIS estimates are based on a MC sample size of $S = 500$ and $L = 20$ EIS fixed-point iterations; The asymptotic standard errors (Asy. s.e.) are obtained from a numerical approximation of the Hessian; Values are statistically significant at the 10% (*), 5% (**), and 1% (***) significance level.

Table 3. Auxiliary regressions for the estimated random effects $\hat{\tau}_i$ obtained from the baseline model and for average log Part II crimes

	Dependent variable:				
	$\hat{\tau}_i$	$\hat{\tau}_i$	$\hat{\tau}_i$	$\hat{\tau}_i$	Averg. log Part II crime
Averg. log Part II	.332***		.446***	.423***	
	(.044)		(.067)	(.034)	
Ltp		.474***	−.051		1.177***
		(.089)	(.107)		(.104)
Lpd		−.070	.014		−.188*
		(.081)	(.053)		(.110)
Lmi		−.075	−.128	−.084*	.119
		(.145)	(.096)	(.046)	(.264)
Dra		−.074	−.226		.342
		(.302)	(.207)		(.361)
Cur		−.109	−.085		−.053
		(.470)	(.386)		(.831)
Pvr		−.244	−.047		−.441
		(.321)	(.253)		(.554)
U18		−.552	1.194**	1.076***	−3.916***
		(.966)	(.475)	(.409)	(1.263)
Gqp		−.592**	.104		−1.560**
		(.281)	(.279)		(.630)
Paa		.284*	−.001		.640**
		(.153)	(.115)		(.302)
Hdl		−.316	−.049		−.599
		(.988)	(.657)		(1.353)
Bdh		.129	.927***	.855***	−1.789***
		(.336)	(.302)	(.124)	(.475)
Rhu		.325	−.151		1.069**
		(.275)	(.200)		(.487)
Sh1		.008	−.077		.191
		(.569)	(.386)		(.777)
Fhh		−.073	−1.016**	−1.019**	2.116**
		(.695)	(.404)	(.404)	(.920)
Hvr		1.164***	−.031		2.679***
		(.368)	(.313)		(.656)
R^2	.481	.419	.683	.676	.695
R^2-adjusted	.477	.339	.636	.662	.652

NOTE: Sample size after excluding the census tracts with missing observations for the socio-economic variables is $\tilde{N} = 124$; Heteroscedasticity-robust standard errors are given in parentheses; Values are statistically significant at the 10% (*), 5% (**), and 1% (***) significance level.

Table 4. ML-EIS results for the predictive spatio-temporal Poisson panel model

Variable	Estimate	Asy. s.e.
Temporal lag (κ)	.391***	.028
Spatial lag (ρ)	.368***	.035
σ_τ	.256***	.024
σ_e	.210***	.006
Constant	−.923***	.096
Mar	.232***	.024
Apr – Aug	.190***	.018
Sep – Jan	.129***	.016
Lagged log Part II	.049***	.009
Averg. log Part II	.437***	.035
Bdh	.778***	.118
Missing data dummy	.196**	.095
Max. eigenvalue of K^*	.618	
Log-likelihood	$-24,187.14$	

NOTE: The ML-EIS estimates are based on a MC sample size of $S = 500$ and $L = 20$ EIS fixed-point iterations; The asymptotic standard errors (Asy. s.e.) are obtained from a numerical approximation of the Hessian; Values are statistically significant at the 10% (*), 5% (**), and 1% (***) significance level.

Table 5. Out-of-sample one-step-ahead predictions

	Normalized PIT residuals				Point predictions					
					Predictive model		Exponential smoothing (0.7)		Exponential smoothing (0.8)	
Period	Mean	Std. dev.	JB Stat.	JB *P*-value	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE
01/13	−.02	.99	4.05	.13	8.70^{+}	2.23^{+}	12.45	2.72	12.25	2.70
02/13	.04	.98	1.75	.42	6.95^{+}	2.02^{+}	14.24	2.85	14.88	2.87
03/13	−.12	1.07	1.29	.52	11.73^{+}	2.48^{+}	14.02	2.71	14.07	2.70
04/13	−.08	1.07	.43	.81	13.75^{+}	2.69	14.02	2.67	13.95	2.67^{+}
05/13	−.09	1.04	.96	.62	14.47	2.77	14.00	2.73	13.65^{+}	2.71^{+}
06/13	−.18	1.10	1.88	.39	14.72	2.85	14.44	2.89	13.59^{+}	2.81^{+}
07/13	−.07	.91	5.00	.08	8.39^{+}	2.30^{+}	10.82	2.44	10.43	2.41
08/13	−.11	.99	4.48	.11	12.71^{+}	2.51^{+}	17.12	2.66	17.55	2.68
09/13	.11	.92	2.98	.23	13.87^{+}	2.42^{+}	17.38	2.65	16.62	2.60
10/13	−.11	.99	.20	.90	10.23^{+}	2.38^{+}	11.93	2.55	11.19	2.49
11/13	.07	.92	1.72	.42	10.32^{+}	2.36^{+}	12.06	2.52	11.75	2.51
12/13	.01	1.02	2.94	.23	13.25^{+}	2.59^{+}	15.72	2.85	15.06	2.78
01/13 - 12/13	−.05	1.00	3.81	.15	132.87^{+}	29.58^{+}	158.39	32.26	157.45	31.93

Note: $^{+}$ indicates the smallest value of the MSFE or MAFE.