JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **32**: 600–620 (2017)
Published online 27 June 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2534

LIKELIHOOD-BASED INFERENCE AND PREDICTION IN SPATIO-TEMPORAL PANEL COUNT MODELS FOR URBAN CRIMES

ROMAN LIESENFELD,[a]* JEAN-FRANÇOIS RICHARD[b] AND JAN VOGLER[a]

[a] *Institute of Econometrics and Statistics, Universität Köln, Germany*
[b] *Department of Economics, University of Pittsburgh, PA, USA*

SUMMARY

We develop a panel count model with a latent spatio-temporal heterogeneous state process for monthly severe crimes at the census-tract level in Pittsburgh, Pennsylvania. Our dataset combines Uniform Crime Reporting data with socio-economic data. The likelihood is estimated by efficient importance sampling techniques for high-dimensional spatial models. Estimation results confirm the broken-windows hypothesis whereby less severe crimes are leading indicators for severe crimes. In addition to ML parameter estimates, we compute several other statistics of interest for law enforcement such as spatio-temporal elasticities of severe crimes with respect to less severe crimes, out-of-sample forecasts, predictive distributions and validation test statistics.

Received 19 August 2015; Revised 28 April 2016

Supporting information may be found in the online version of this article.

1. INTRODUCTION

The spatio-temporal urban distribution of crimes is receiving growing attention not only from researchers (criminologists, sociologists, economists, geographers, etc.) but also from law enforcement agencies; see, for example,, 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. 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.' 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 dataset consisting of monthly (2008–2013) counts of severe crimes for the 138 census tracts in Pittsburgh, Pennsylvania. Generalizing a spatial efficient importance sampling (EIS) approach recently developed by Liesenfeld *et al.* (2016), 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 the contributions of Gorr *et al.* (2003) and Cohen *et al.* (2007), which analyze a variety of forecasting models for severe urban crime. In the absence of potential socio-economic covariates and spatially and temporally lagged dependent variables, they find

---

* Correspondence to: Roman Liesenfeld, Institute of Econometrics and Statistics, Universität Köln, Universitätstr. 22a, D-50939,Köln, Germany. E-mail: liesenfeld@statistik.uni-koeln.de
Correction added on 25 July 2016: The 2008 crime data from UCSUR was previously erroneously coded and all empirical results have been re-computed in this current version.

that 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.* (2016) for latent spatial count models to account 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 US 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 (Wilson and Kelling, 1982; Anselin *et al.*, 2000; Cohen and Gorr, 2005; Cohen *et al.*, 2007). Foremost, we can compute spatio-temporal elasticities of severe crimes 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 1-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 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 spatial EIS procedure which we use for likelihood inference. Its implementation is detailed in the online Appendix (supporting information). 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 and place-based theories (Anselin *et al.*, 2000). Social ecology theories such as that of *social disorganization* (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* (Cohen and Felson, 1979) and the *rational choice* approach (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 conditions that may help in 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 Land *et al.* (1990), Tolnay *et al.* (1996), Baller *et al.* (2001), Kubrin (2003) and Helbich and Arsanjani (2014).

DOI: 10.1002/jae

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 they cluster in space. This implies positive spatial correlation, as observations that are in close geographic proximity tend to be more alike than those that are further apart (Tobler's, 1970, first law of geography). If criminal activities were solely determined by the structural factors included in a regression model, then 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 (Morenoff and Sampson, 1997; Morenoff *et al.*, 2001; Baller *et al.*, 2001; Tita and Radil, 2013, p. 107).

Spatial correlation has attracted growing attention in empirical criminological research for two reasons (Baller *et al.*, 2001). The first one is statistical. If spatial correlation in the data is ignored, then estimates of the effects of covariates and their standard errors may be inconsistent (Anselin, 1988). Thus modeling 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 correlation is interpreted as evidence of spatial diffusion of certain types of crime (Tolnay *et al.*, 1996; Morenoff and Sampson, 1997). Such effects may reflect criminals and gangs linked together by rivalry networks (Tita and Greenbaum, 2009) and/or 'subcultural' processes, in which case violence spreads throughout population and regional areas via direct social contacts (Loftin, 1986).

In order to incorporate spatial correlation, regression models for crime rates are typically generalized to include either spatially correlated errors (spatial error models) or spatially lagged dependent variables (spatial lag models); see Baller *et al.* (2001); Morenoff *et al.* (2001); 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 not designed to discover the actual mechanisms through which criminal events in one geographical area influence events in other areas at later times (Baller *et al.*, 2001, p. 567).

2.3. Temporal Dependence

When observations are available in the form of panel data, it becomes possible to model complex combinations of spatial heterogeneity and/or spatial and temporal dependence in crime rates (Anselin *et al.*, 2000, p. 241). A dynamic panel-data setting also provides a natural basis for joint time–space forecasting of crime activity that can support tactical deployments of police resources (Gorr and Harries, 2003; Gorr *et al.*, 2003; Cohen and Gorr, 2005; 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 geographic units. For example, convergence of crime opportunities in space and time, as emphasized by place-based routine activity theories, could be facilitated by physical and social features that provide a setting more or less conducive to crime, such as local population composition or urbanistic conditions (Anselin *et al.*, 2000, p. 220). The slowly changing nature of such (observed or unobserved) features leads to systematic temporal dependence in time series of geographic crime frequency counts (Li *et al.*, 2014, p. 181). If some of those physical and social features are stable over time but unobservable, the implied spatial heterogeneity 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.

The interpretation of spatial correlation in terms of diffusion and contagion also implies intertemporal linkages of crime rates within regional areas (Bernasco and Elffers, 2013, p. 710). For example,

DOI: 10.1002/jae

spatial diffusion reflecting interacting gangs implies a chronology of criminal actions and retaliations that induces 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, in which case 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 spatially and 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 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 subsequent 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 regional area may serve as a leading indicator for the number of serious crimes in that area (Anselin *et al.*, 2000, p. 225; Cohen and Gorr, 2005; Cohen *et al.*, 2007). In order to take advantage of the leading-indicator properties of soft crimes we shall include them as lagged explanatory variables in our predictive panel model for severe crimes.

Finally, crime appears to be a seasonal phenomenon and empirical evidence suggests that, for example, frequencies of violent crimes are higher in summer and lower in winter (Cohen, 1941). Under the 'temperature aggression hypothesis' this seasonality is attributed to weather that increases violent crimes via ambient temperature and anger arousal (Gorr *et al.*, 2003). Within the routine activity theory regular temporal patterns can be viewed as a result of seasonality that impacts some of the conditions propitious to crime (such as existence of suitable targets and motivated offenders or absence of crime suppressors). Thus we shall include monthly seasonal dummies in our panel regression model.

3. DATA

Our crime dataset includes 72 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 US 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 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 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 9936 individual observations. As discussed above, 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 with less than a high school degree (*Hdl*), percentage of population 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 1 year (*Sh1*), percentage of female-headed households (*Fhh*) and housing units vacancy

DOI: 10.1002/jae

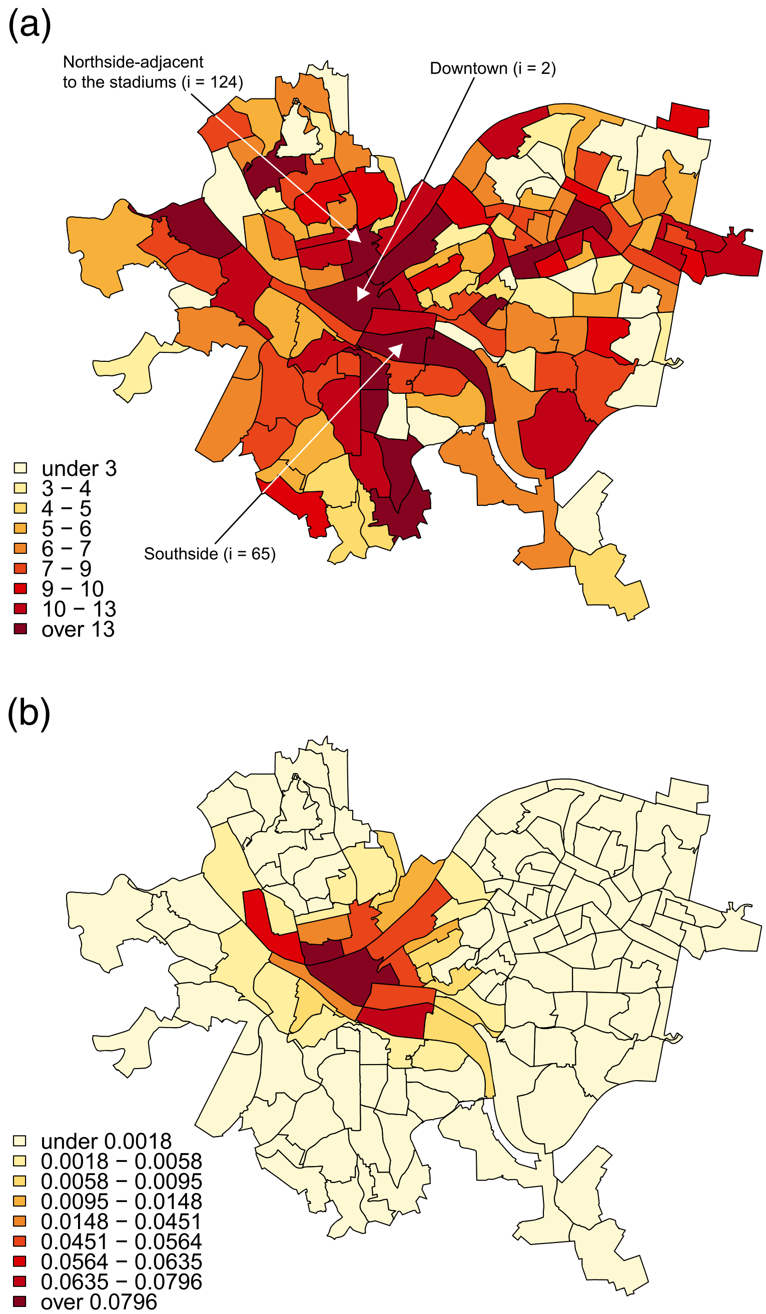

Figure 1. (a) Time averages of Part I crimes $\bar{y}_i$ in the $N = 138$ Pittsburgh census tracts. (b) Long-term Part I crime elasticities for the Pittsburgh census tracts w.r.t. a change in lagged Part II crime in Downtown ($i = 2$). [Colour figure can be viewed at wileyonlinelibrary.com]

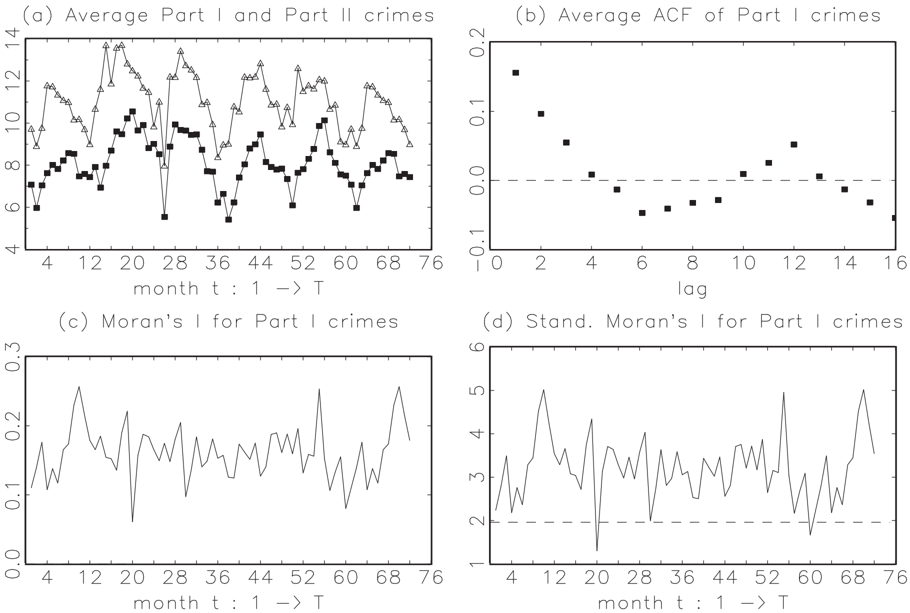

Figure 2. (a) Time plot of Part I crimes $\bar{y}_t$ (solid squares) and Part II crimes (triangles) averaged across census tracts. (b) Temporal autocorrelations averaged across census tracts. (c) Period-by-period time plot of Moran's I for Part I crimes. (d) Period-by-period time plot of the standardized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

rate (*Hvr*). Occasional missing data for 14 census tracts that do not have a regular resident population were replaced by census tract dummies.[1]

We computed time and census tracts averages of Part I crimes $\bar{y}_i$ ($i = 1, \ldots, 138$) and $\bar{y}_t$ ($t = 1, \ldots, 72$). In panel (a) of Figure 1 we provide a color-coded map of the time averages $\bar{y}_i$, highlighting significant spatial clustering and heterogeneity across census tracts. In Figure 2(a) 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; Cohen and Gorr, 2005).

In Figure 2(c) we present Moran's I statistics for spatial correlation period by period (Moran, 1948; Cliff and Ord, 1972). These are computed under the spatial weight matrix used for our spatial models (see Section 4 below). These statistics are standardized in Figure 2(d), with a critical value of 1.96 at the 5% 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 Figure 2(b) 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.

[1] These census tracts include business districts and hospitals, zoo and parks, industrial parks, cemeteries and stadiums.

4. SPATIO-TEMPORAL PANEL COUNT MODEL FOR CRIMES

4.1. 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, conditionally on the latent state variables λ_{it}, the y_{it}'s are mutually independent with Poisson distributions whose non-negative mean is specified as $\mu_{it} = \exp(\lambda_{it})$. The corresponding conditional probability density function (pdf) of y_{it} given λ_{it} is given by

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

where λ_{it} represents the log of the conditional mean of y_{it}. The latent state process is assumed to be a linear Gaussian dynamic panel model in space and time, as discussed, for example, by Elhorst (2012), Parent and LeSage (2012) and Debarsy *et al.* (2012). It has the following form:

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

where $\lambda_t = (\lambda_{it})$ denotes the $N \times 1$ vector of 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 1 month as leading indicator. 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 contemporaneous spatial correlation in severe crimes across the census tracts and the parameter ρ measures its intensity. The temporally lagged state variable λ_{t-1} with persistence parameter κ accounts for potential census-tract specific intertemporal linkages of crime rates and the term $W\lambda_{t-1}$ represents a component mixing spatial and intertemporal dependence with intensity measured by θ.

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

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

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. It is assumed that τ and ϵ_t are independent of each other, conditionally on X_t.

In order to account for missing values for our socio-economic covariates in 14 census tracts (parks, cemeteries, etc.) we use a dummy-variable approach for $X_t\gamma = (x'_{it}\gamma)$ in equation (2). Specifically, we use the following regression function for census tract i:

$$x'_{it}\gamma = x_{it}^{(1)'}\gamma^{(1)} + (1 - \iota_i)x_i^{(2)'}\gamma^{(2)} + \iota_i\gamma^{(3)}, \quad \text{with } \gamma' = \left(\gamma^{(1)'}, \gamma^{(2)'}, \gamma^{(3)}\right) \tag{4}$$

where $x_i^{(2)}$ denotes the vector of time-invariant socio-economic covariates, $x_{it}^{(1)}$ the remaining covariates and ι_i a dummy variable ($\iota_i = 1$ if tract i has missing observations for $x_i^{(2)}$ and $\iota_i = 0$, otherwise).

DOI: 10.1002/jae

4.2. Stationarity and Time–Space Separability

If the matrix $(I_N - \rho W)$ is invertible, then the state-space model in equations (2) and (3) implies the following Gaussian distribution for λ_t given $(\lambda_{t-1}, X_t, \tau)$:

$$\lambda_t|\lambda_{t-1}, X_t, \tau \sim \mathrm{N}_N\left(K\lambda_{t-1} + m_t^* + \tau^*, \Sigma\right) \tag{5}$$

with

$$K = D(\kappa I_N + \theta W), \quad m_t^* = DX_t\gamma, \quad \tau^* = D\tau, \quad \Sigma = \sigma_e^2 DD', \quad D = (I_N - \rho W)^{-1} \tag{6}$$

If the eigenvalues of the W matrix are real, then a sufficient condition for the invertibility of $(I_N - \rho W)$ is that $\rho \in (1/\zeta_{\max}, 1/\zeta_{\min})$, where $\zeta_{\min}$ and $\zeta_{\max}$ are the extreme eigenvalues of W (LeSage and Pace, 2009). Stationarity of the λ_t process (conditional on the covariates) is guaranteed if the characteristic roots of the persistence matrix K lie within the unit circle (Elhorst, 2012; Parent and LeSage, 2012).

In order to simplify the interpretation of the model in terms of the spatio-temporal effects of changes in the covariates, Debarsy *et al.* (2012) and Parent and LeSage (2012) consider the parameter restriction $\theta = -\rho\kappa$. Under this restriction the temporal persistence matrix in equations (5) and (6) simplifies to $K = \kappa I_N$, which implies that the impact of changes in the covariates on future values of the crime intensity λ_t can be decomposed into a contemporaneous spatial effect measured by the spatial multiplier $D = (I_N - \rho W)^{-1}$ and subsequent region-specific time effects captured by κI_N.

4.3. Spatio-temporal Effects of the Covariates

Using our model we can quantify the spatio-temporal responses of severe crimes to changes in the covariates in terms of marginal effects and elasticities. We follow the logic of Debarsy *et al.* (2012) for linear spatio-temporal models, but have to account for the nonlinearity of expected severe crimes under our Poisson model. (The derivations of the marginal effects are provided in the online Appendix.)

Let $x_{\cdot kt} = (x_{ikt})$ denote the kth column of the regressor matrix X_t and γ_k the corresponding slope coefficient. The $N \times N$ matrix with the partial derivatives of the L-step-ahead conditional expectations of severe crimes $\mu_{t+L} = \mathrm{E}(y_{t+L}|\lambda_{t-1}, X_t, \ldots, X_{t+L})$ w.r.t. a transitory change in the vector of the kth covariate $x_{\cdot kt}$ is

$$\Delta_{k,L} = \frac{\partial \mu_{t+L}}{\partial x'_{\cdot kt}} = (K^L D\gamma_k) \odot (\mu_{t+L}, \ldots, \mu_{t+L}) \tag{7}$$

where $\odot$ denotes the Hadamard (element-by-element) product. The diagonal elements of $\Delta_{k,L}$ measure the dynamic *direct* effects (impact on the tract where the change of the kth covariate takes place) and the non-diagonal elements the dynamic *indirect* spillover effects (impact on the other tracts). The sum of the ith column provides the *total* effect for the entire city resulting from a change of the kth covariate in tract i. The sequence $\{\Delta_{k,L}, L = 0, 1, \ldots\}$ provides the impulse response functions for a transitory one-period change that enables us to examine the responses' profile of decay over space and time. Under the time–space separability restriction $\theta = -\rho\kappa$, the spatio-temporal multiplier in $\Delta_{k,L}$ simplifies to $K^L D = \kappa^L D$, so that spatial dynamic multipliers decay geometrically over time within each census tract by the factor κ.

To find the long-term marginal effects of a permanent change in covariate k, we consider the long-run expectation $\mu = \lim_{L\to\infty} \mathrm{E}(y_{t+L}|\lambda_{t-1}, X_t, \ldots, X_{t+L})$ under a fixed scenario $X_s = X\ \forall s$. The $N \times N$ matrix regrouping the long-term effects of a permanent change in covariate k is given by

DOI: 10.1002/jae

$$\Delta_k = \frac{\partial \mu}{\partial x'_{\cdot k}} = \left[(I_N - K)^{-1} D\gamma_k\right] \odot (\mu, \ldots, \mu) \tag{8}$$

where $x_{\cdot k}$ denotes the kth column of X.

5. EIS-BASED LIKELIHOOD INFERENCE

5.1. Likelihood

For inference in our spatio-temporal Poisson panel model we rely on ML generalizing spatial EIS (Liesenfeld *et al.*, 2016) for likelihood evaluation. In order to obtain the unconditional likelihood function the stationary distribution of the first-period state variables $f(\lambda_1|X_1, \tau)$ would be required. However, that stationary distribution is not available due to the presence of time-varying covariates in X_t.[2] As an alternative, 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 its latent log conditional mean. This treatment of the initial-condition problem 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, the estimated largest eigenvalue of the persistence matrix K is approximately equal to 0.5, which implies 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 $\lambda_1|X_1, X_0, X_{-1}, \ldots, \tau$ and capturing the spatial heterogeneity generated by the unobservable pre-sample values $X_0, X_{-1}, \ldots$ by an additional latent random effect variable. We also implemented ML based on this approximation to the unconditional likelihood. 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 estimated. The likelihood integral to be evaluated is of the form

$$L(\psi) = \int_{\mathbb{R}^{NT}} \varphi(\lambda_{(T)}, \tau)\, \mathrm{d}\lambda_{(T)}\, \mathrm{d}\tau, \quad \text{with } \varphi(\lambda_{(T)}, \tau) = f(y_{(T)}|\lambda_{(T)}) \cdot f(\lambda_{(T)}|X_{(T)}, \tau, \lambda_1) \cdot f(\tau) \tag{9}$$

and

$$f(y_{(T)}|\lambda_{(T)}) = \prod_{t=2}^{T}\prod_{i=1}^{N} f(y_{it}|\lambda_{it}), \quad f(\lambda_{(T)}|X_{(T)}, \tau, \lambda_1) = \prod_{t=2}^{T} f(\lambda_t|\lambda_{t-1} X_t, \tau) \tag{10}$$

$$f(\tau) = \prod_{i=1}^{N} f(\tau_i) \tag{11}$$

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 (5), respectively. Note that evaluation of the likelihood

[2] 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 nonlinear spatio-temporal Poisson panel model with a large time dimension would amount to estimating an unfeasible large number of additional (auxiliary) parameters.

DOI: 10.1002/jae

integral (9) requires numerical integration. Moreover, the N-dimensional spatial state-transition density $f(\lambda_t|\lambda_{t-1}, X_t, \tau)$ has a non-Markovian structure, so that likelihood factorizations of the form given in equations (9)–(11) depend on large (auxiliary) parameter matrices. This poses substantial computational burdens for the likelihood evaluation (Wang, 2014; Pace, 2014). However, the spatial precision matrix Σ^{-1} of $f(\lambda_t|\lambda_{t-1}, X_t, \tau)$ obtained from equation (6) is a sparse parameter matrix with a large proportion of zero entries. Hence we can generalize the spatial EIS of Liesenfeld *et al.* (2016), which is specifically designed to exploit the sparsity of large spatial precision matrices. Spatial EIS combines the original EIS principle of Richard and Zhang (2007) for high-dimensional MC integration with sparse matrix algebra. Sparse matrix operations significantly reduce operation counts and memory requirements relative to the corresponding operations on dense matrices (Gilbert *et al.*, 1992; LeSage and Pace, 2009; Pace and LeSage, 2016). Thus the combination of EIS with sparse matrix algebra ensures that accurate MC likelihood evaluations remain computationally feasible even in high-dimensional latent spatial Gaussian models. The particular spatial EIS implementation used for the spatio-temporal likelihood (9) is provided in the online Appendix.

5.2. Smoothing and Prediction

Once the parameters ψ have been estimated by ML-EIS, the EIS procedure can also be used to compute MC estimates for the conditional mean of functions of the latent states 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 will be instrumental in the selection of the covariates to be included in the model (Section 6.2) and assessment of the predictive performance of the model (Section 6.4).

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

$$\mathrm{E}\left[g(\lambda_{T+1}, \lambda_{(T)}, \tau)|y_{(T)}\right] = \frac{\int_{\mathbb{R}^{N(T+1)}} g(\lambda_{T+1}, \lambda_{(T)}, \tau) \cdot f(\lambda_{T+1}|\lambda_T, \tau) \cdot \varphi(\lambda_{(T)}, \tau)\, \mathrm{d}\lambda_{T+1}\, \mathrm{d}\lambda_{(T)}\, \mathrm{d}\tau}{\int_{\mathbb{R}^{NT}} \varphi(\lambda_{(T)}, \tau)\, \mathrm{d}\lambda_{(T)}\, \mathrm{d}\tau} \tag{12}$$

where $\varphi(\cdot)$ is defined in equation (9) and $f(\lambda_{T+1}|\lambda_T, \tau)$ is the state-transition density defined in equation (5). The conditional mean (12) is to be evaluated at the ML-EIS estimates of ψ. For its MC estimation we rely upon a straightforward extension of the EIS procedure implemented for likelihood evaluations (see the online Appendix). MC-EIS estimation of the conditional mean of a function $g(\lambda_{T+1}, \lambda_{(T)}, \tau)$ can be used to compute several statistics of interest, such as:

(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 with

$$g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv \tau, \quad \text{and} \quad g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv \mathrm{E}(y_{it}|\lambda_{(T)}, \tau) = \exp(\lambda_{it}) \tag{13}$$

respectively;

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

$$g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv \mathrm{E}(y_{iT+1}|\lambda_{T+1}, \tau) = \exp(\lambda_{iT+1}) \tag{14}$$

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

DOI: 10.1002/jae

$$g(\lambda_{T+1}, \lambda_{(T)}, \tau) \equiv P(y_{iT+1} \leq y|\lambda_{T+1}, \tau) = \sum_{\ell=0}^{y} \exp\{\ell\lambda_{iT+1} - \exp(\lambda_{iT+1})\}/(\ell!) \qquad (15)$$

We can also use the predictive cdf in equation (15) to compute randomized probability integral transformation (PIT) residuals that can be used to test the validity of our predictive model (Jung *et al.*, 2006; Czado *et al.*, 2009). Specifically, let y^o_{iT+1} denote the (ex post) observed value of y_{iT+1} and define $P_a(y^o_{iT+1}) = P(y_{iT+1} \leq y^o_{iT+1} - 1|y_{(T)})$ and $P_b(y^o_{iT+1}) = P(y_{iT+1} \leq y^o_{iT+1}|y_{(T)})$ with $P_a(0) = 0$. If the predictive model is valid, then the randomized predictive PIT residuals, defined as

$$\xi_{iT+1} = P_a\left(y^o_{iT+1}\right) + \upsilon_i\left[P_b\left(y^o_{iT+1}\right) - P_a\left(y^o_{iT+1}\right)\right], \quad \text{with } \upsilon_i \sim \text{i.i.d. } U_{[0,1]} \qquad (16)$$

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

$$\xi^*_{iT+1} = \Phi^{-1}(\xi_{iT+1}) \qquad (17)$$

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 suggests that our model needs to include seasonal dummies as well as lagged Part II crimes as covariates. Furthermore, we propose exploring whether the socio-economic census tract covariates we collected might explain part 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 11 months dummies, where we select February (typically the lowest crime month) as the base month.

Step 2. We compute MC-EIS estimates for the posterior means of the random effects $\hat{\tau}_i$ ($i = 1, \ldots, N$) as given by equations (12) and (13) 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 $S = 500$ are reported in the left-hand panel of Table I. MC numerical standard deviations of the ML-EIS parameter

DOI: 10.1002/jae

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

Variable	Baseline model Estimate	Asy. SE	MC SD	Full model Estimate	Asy. SE	MC SD
Temporal lag (κ)	0.487***	0.036	0.0009	0.564***	0.044	0.0008
Temporal-spatial lag (θ)	−0.244***	0.059	0.0015	−0.387***	0.075	0.0013
Spatial lag (ρ)	0.563***	0.037	0.0009	0.556***	0.045	0.0009
σ_τ	0.388***	0.039	0.0007	0.151***	0.027	0.0003
σ_e	0.183***	0.008	0.0003	0.183***	0.010	0.0002
Constant	0.125***	0.046	0.0001	0.044	0.595	0.0003
Jan	0.097***	0.017	0.0002	0.099***	0.018	0.0002
Mar	0.157***	0.022	0.0003	0.143***	0.023	0.0003
Apr	0.132***	0.018	0.0003	0.137***	0.019	0.0003
May	0.132***	0.017	0.0003	0.144***	0.019	0.0003
Jun	0.119***	0.016	0.0002	0.135***	0.018	0.0002
Jul	0.139***	0.017	0.0003	0.156***	0.019	0.0003
Aug	0.140***	0.017	0.0002	0.160***	0.019	0.0003
Sep	0.094***	0.015	0.0002	0.116***	0.016	0.0002
Oct	0.085***	0.015	0.0002	0.099***	0.016	0.0002
Nov	0.093***	0.015	0.0002	0.102***	0.016	0.0002
Dec	0.088***	0.015	0.0002	0.093***	0.016	0.0002
Lagged log Part II	0.056***	0.008	0.0001	0.044***	0.008	0.00005
Ave. log Part II				0.324***	0.035	0.0005
Lmi				−0.058	0.064	0.0001
U18				0.306	0.370	0.0008
Bdh				0.585***	0.126	0.0009
Fhh				−0.416	0.385	0.0005
Missing data dummy				−0.424	0.636	0.0008
Max. eigenvalue of K	0.556			0.599		
Log-likelihood		−24,252.82			−24,166.07	
LR-stat. $H_0 : \kappa = \rho = \theta = 0$					463.84	[<0.001]
LR-stat. $H_0 : \kappa = 0$					180.99	[<0.001]
LR-stat. $H_0 : \theta = 0$					34.67	[<0.001]
LR-stat. $H_0 : \rho = 0$					112.35	[<0.001]
LR-stat. $H_0 : \theta = -\kappa\rho$					5.17	[0.023]

Note:The ML-EIS estimates are based on an MC sample size of $S = 500$. The asymptotic standard errors (Asy. SE) are obtained from a numerical approximation of the Hessian and numerical MC standard deviations (MC SD) from 50 ML-EIS estimations conducted under 50 different random number seeds. Values are statistically significant at the *10%, **5% and ***1% significance levels. Probability values for the LR tests are given in brackets.

estimates used as measures of their numerical precision are obtained from 50 repeated ML-EIS parameter estimations for the crime data conducted under different random number seeds. They indicate that the parameter estimates are numerically very accurate.

We note immediately that the key parameters of interest (coefficients for the temporal lag κ, spatial lag ρ, spatial and temporal lag θ, random effect standard deviation σ_τ and slope coefficient for lagged log 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 significant spatial and temporal dependence in Part I crime rates as well as substantial unobserved census-tract specific heterogeneity. The significant positive coefficient for lagged log Part II crime is fully in line with the broken-windows hypothesis (Wilson and Kelling, 1982; Cohen and Gorr, 2005). As for the seasonal dummies, which capture the difference in seasonality between Part I and Part II crime (see Figure 2(a)), it appears that we can reduce their number since several slope estimates are close to one another. However, such reduction is essentially irrelevant for steps 2 and 3 and will be postponed until step 4. Moreover, since log Part II crime is itself seasonal, we do not want to restrict the seasonal pattern of Part I crime before producing our final (predictive) estimate of the coefficient of lagged log Part II crime in step 4. As for the time–space separability restriction $\theta = -\kappa\rho$ (Debarsy *et al.*, 2012; Parent and LeSage, 2012), the results for the baseline model in Table I reveal that the ML estimates of θ and $-\kappa\rho$ are close to one another (-0.244 versus -0.272), suggesting an additional potential simplification of the model. Since this simplification is also not critical for the selection of the random effect covariates, its formal test is postponed as well and will be conducted in the context of the full model in step 3.

6.2. Random Effects Covariates

In Table II (columns 1–4), we report the results of four auxiliary regressions for the estimated random effects $\hat{\tau}_i$ obtained from the baseline model. 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 an R^2 of 0.50. The second regression uses all 15 socio-economic covariates but excludes *Alp-II* for an R^2 of 0.44. 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 concentrations of violence typically occur in disadvantaged communities and regions with a large population size (Baller *et al.*, 2001; Helbich and Arsanjani, 2014; Kubrin, 2003; 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 0.50 versus 0.36, 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 column 5 of Table II. Moreover, it reveals that the significant socio-economic factors for the estimated random effect in column 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 or higher (*Bdh*), all with a negative impact, and rental housing units (*Rhu*) and female-headed households (*Fhh*) with a positive effect.

In column 3 of Table II, 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 0.71. It is important to note that with the inclusion of *Alp-II* in the auxiliary regression the coefficients of the socio-economic variables in column 3 represent differential effects between the intensity of Part I and Part II crime. Specifically, columns 3, 2 and 5 correspond respectively to the following auxiliary regressions:

$$\hat{\tau} = \pi_1 Alp\text{-}II + \pi_2' X^{(\mathrm{se})}, \quad \hat{\tau} = \pi_3' X^{(\mathrm{se})}, \quad \widehat{Alp}\text{-}II = \pi_4' X^{(\mathrm{se})} \tag{18}$$

where $X^{(\mathrm{se})}$ denotes the matrix of the 15 socio-economic covariates. Equation system (18) implies that the coefficients of the socio-economic variables in column 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 obtains for *Bdh*, which suggests 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 column 3 the least significant covariates one at a time. As reported in column 4 of Table II, this produces a parsimonious regression with only five covariates (*Alp-II*, *Lmi*, *U18*, *Bdh*, *Fhh*) for a minimal loss of fit (R^2 virtually unchanged and adjusted R^2 of 0.698 instead of 0.667).

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

DOI: 10.1002/jae

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

	Dependent variable				
	$\hat{\tau}_i$	$\hat{\tau}_i$	$\hat{\tau}_i$	$\hat{\tau}_i$	Ave. log Part II crime
Ave. log Part II	0.273***		0.362***	0.345***	
	(0.035)		(0.049)	(0.026)	
Ltp		0.384***	−0.042		1.177***
		(0.065)	(0.077)		(0.104)
Lpd		−0.057	0.011		−0.188*
		(0.060)	(0.038)		(0.110)
Lmi		−0.063	−0.106	−0.068*	0.119
		(0.108)	(0.069)	(0.035)	(0.264)
Dra		−0.057	−0.181		0.342
		(0.225)	(0.152)		(0.361)
Cur		−0.105	−0.086		−0.053
		(0.348)	(0.278)		(0.831)
Pvr		−0.216	−0.057		−0.441
		(0.236)	(0.180)		(0.554)
U18		−0.579	0.840**	0.736**	−3.916***
		(0.718)	(0.336)	(0.307)	(1.263)
Gqp		−0.478**	0.087		−1.560**
		(0.210)	(0.204)		(0.630)
Paa		0.221*	−0.010		.640**
		(0.113)	(0.080)		(0.302)
Hdl		−0.286	−0.069		−0.599
		(0.735)	(0.473)		(1.353)
Bdh		0.088	0.736***	0.692***	−1.789***
		(0.248)	(0.217)	(0.093)	(0.475)
Rhu		0.289	−0.101		1.069**
		(0.206)	(0.143)		(0.487)
Sh1		0.015	−0.054		.191
		(0.422)	(0.278)		(0.777)
Fhh		−0.045	−0.722**	−0.738**	2.116**
		(0.511)	(0.284)	(0.302)	(0.920)
Hvr		0.947***	−0.023		2.679***
		(0.270)	(0.222)		(0.656)
R^2	0.504	0.438	0.710	0.710	0.695
R^2-adjusted	0.500	0.360	0.667	0.698	0.652

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 levels.

to equation (4). The results are reported in the right-hand panel of Table I. The estimated coefficients of the additional covariates are consistent with those obtained in the auxiliary $\hat{\tau}_i$-regression in column 4 of Table II but are less significant since they are now unconditional on $\{\hat{\tau}_i\}_{i=1}^{N}$. As confirmed further in the next section, the substantial log-likelihood gain of 86.8 achieved by the additional covariates is essentially due to the inclusion of *Alp-II* and *Bdh*, whose coefficients are both highly significant. We already discussed the interpretation of the *Alp-II* and *Bdh* coefficient. Consistent with the broken-windows hypothesis, we see that the extension of the list of covariates has no major effect on the estimated coefficient of lagged log Part II crime, which remains statistically highly significant. Note that the latter is much smaller than that of *Alp-II* (0.044 versus 0.324). This suggests that policies aiming at a permanent reduction of the number of Part II crimes, whether directly through increased enforcement or indirectly, through action on the relevant socio-economic covariates, will have a much greater impact than policies aiming at transitory reductions of Part II crimes.

Furthermore, we find that $\hat{\sigma}_\tau$ has reduced from 0.388 to 0.151 by the additional socio-economic variables, corresponding to an 85% reduction in the variance of the time-invariant unobserved

DOI: 10.1002/jae

heterogeneity. However, it has not notably affected the estimated spatial lag coefficient ρ, which remains statistically significantly positive. This indicates that socio-economic similarities between census tracts cannot explain the contemporaneous spatial clustering in severe crime rates and confirms the spatially diffusive and contagious nature of severe crime (Tolnay *et al.*, 1996; Morenoff and Sampson, 1997; Loftin, 1986). The estimates of the temporal lag coefficient κ has slightly increased compared to the baseline model, while that of the spatial-temporal lag θ has decreased. Accounting for the fact that the estimate of ρ is virtually unchanged, it follows that the full model remains stable, with a maximum eigenvalue of the estimated K matrix well below 1. The strength of the temporal dependence predicted by the full model is roughly the same as that of spatial dependence. As for the time–space separability hypothesis $\theta = -\kappa\rho$, the likelihood-ratio test result reported in Table I confirms that the data are consistent with this restriction at the 1% level.

We conclude that the full model provides a sound statistical basis to analyze the spatio-temporal distribution of Part I crimes. Pending the time–space separability simplification together with the elimination of insignificant monthly dummies and covariates, it provides a solid basis for Part I crime predictions and policy implementation linked to the broken-windows hypothesis. In order to illustrate the flexibility of our model in this respect, we compute the marginal effects of Part I crimes w.r.t. transitory as well as permanent changes in lagged Part II crimes. In order to do so, we rely upon the results of Section 4.3, but have to account for the facts that lagged Part II crimes enter the matrix of covariates X_t in logs and that, in the case of long-run effects, a permanent change in lagged Part II crimes changes also accordingly the covariate *Alp-II*, defined as the time average of log Part II crimes. Let $z_{\cdot t} = (z_{it})$ denote the vector of lagged Part II crimes with a fixed-scenario value $z = (z_i)$ and assume that the $\ln(z_{it})$'s and their time averages are stacked in columns 1 and 2 of X_t, respectively, with slope coefficients γ_1 and γ_2. Then the L-step-ahead effects w.r.t. a transitory change in the vector of lagged Part II crimes, are given by

$$\frac{\partial \mu_{t+L}}{\partial z'_{\cdot t}} = \left(K^L D\gamma_1\right) \odot \left(\frac{\mu_{t+L}}{z_{1t}}, \ldots, \frac{\mu_{t+L}}{z_{Nt}}\right) \tag{19}$$

and the corresponding long-term effects w.r.t. a permanent change (see the online Appendix) by

$$\frac{\partial \mu}{\partial z'} = \left[(I_N - K)^{-1} D(\gamma_1 + \gamma_2)\right] \odot \left(\frac{\mu}{z_1}, \ldots, \frac{\mu}{z_N}\right) \tag{20}$$

Table III provides the estimates of the period-by-period marginal effects of a transitory change in lagged Part II crimes in downtown Pittsburgh on Part I crimes in downtown itself (direct effects) and the entire city (total effects), together with the corresponding direct and total elasticities. They are computed from equation (19) for $L = 0$ out to a horizon of $L = 13$ months. Also reported are the associated long-term marginal effects and elasticities according to equation (20). Consistent with the model stability we see that the period-by-period effects die out rapidly over time and essentially disappear after 8–10 months. The results further suggest that a 1% reduction of downtown Part II crimes would decrease in the subsequent month ($L = 0$) the Part I crimes by 0.05% in downtown and by 0.004% across the entire city. The corresponding long-term reductions of Part I crimes resulting from a permanent decrease of downtown Part II crimes add up to 0.87% (downtown) and 0.05% (entire city).[3] The fact that the long-term effects are significantly larger than the short-term ones follows from our finding that the estimate of γ_2 is much larger than that of γ_1 (0.324 versus 0.044). An important implication of that result is that policies aimed at reducing Part I crimes through reductions in Part II crimes (whether directly or indirectly through actions on socio-economics covariates) require long-term commitments to be fully effective.

[3] The direct and total short-term ($L = 0$) and long-term elasticities for all the 138 Pittsburgh census tracts are provided in the online Appendix.

DOI: 10.1002/jae

Table III. Marginal effects and elasticities for lagged Part II crimes in downtown ($i = 2$)

Horizon (L)	Direct marginal effects	Direct elasticities	Total marginal effects	Total elasticities
Transitory change in lagged Part II crimes				
0	0.0487	0.0468	0.0620	0.0038
1	0.0254	0.0254	0.0297	0.0018
2	0.0134	0.0140	0.0143	0.0009
3	0.0072	0.0077	0.0070	0.0004
4	0.0039	0.0043	0.0035	0.0002
5	0.0021	0.0024	0.0018	0.0001
6	0.0012	0.0013	0.0009	0.0001
7	0.0007	0.0008	0.0005	0.0000
8	0.0004	0.0004	0.0003	0.0000
9	0.0002	0.0002	0.0001	0.0000
10	0.0001	0.0001	0.0001	0.0000
11	0.0001	0.0001	0.0000	0.0000
12	0.0000	0.0000	0.0000	0.0000
13	0.0000	0.0000	0.0000	0.0000
Permanent change in lagged Part II crimes				
∞	0.8184	0.8657	0.9932	0.0522

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

Variable	Estimate	Asy. SE	MC SD
Temporal lag (κ)	0.506***	0.024	0.0004
Spatial lag (ρ)	0.511***	0.033	0.0007
σ_τ	0.192***	0.017	0.0002
σ_e	0.191***	0.007	0.0002
Constant	−0.686***	0.071	0.0004
Mar–Aug	0.154***	0.016	0.0002
Sep–Jan	0.104***	0.013	0.0002
Lagged log Part II	0.044***	0.008	0.0001
Ave. log Part II	0.361***	0.028	0.0003
Bdh	0.634***	0.091	0.0005
Missing data dummy	0.170**	0.072	0.0003
Max. eigenvalue of K	0.506		
Log-likelihood		−24,173.90	
LR-stat. H_0: predictive model vs. H_1: full model		15.66	[0.268]

Note: The ML-EIS estimates are based on an MC sample size of $S = 500$. The asymptotic standard errors (Asy. SE) are obtained from a numerical approximation of the Hessian and numerical MC standard deviations (MC SD) from 50 ML-EIS estimations conducted under 50 different random number seeds. Values are statistically significant at the *10%, **5% and ***1% significance levels. Probability value for the LR tests is given in brackets.

In order to further illustrate the global spatio-temporal effects of Part II crimes we display in Figure 1(b) a color map of the long-term Part I crime 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 such elasticities could play an important role for efficient allocation of scarce law enforcement resources.

DOI: 10.1002/jae

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

Period	Normalized PIT residuals				Point predictions							
	Mean	SD	JB stat.	JB P-value	Predictive model		Exponential smoothing (0.7)		Exponential smoothing (0.8)		Poisson regression	
					MSFE	MAFE	MSFE	MAFE	MSFE	MAFE	MSFE	MAFE
01/13	−0.04	0.98	5.02	0.08	8.80$^+$	2.26$^+$	12.45	2.72	12.25	2.70	18.98	3.13
02/13	0.01	0.97	1.43	0.49	6.84$^+$	2.01$^+$	14.24	2.85	14.88	2.87	13.06	2.56
03/13	−0.08	1.07	1.10	0.58	11.46$^+$	2.45$^+$	14.02	2.71	14.07	2.70	18.91	3.25
04/13	−0.11	1.05	0.53	0.77	13.70$^+$	2.68	14.02	2.67	13.95	2.67$^+$	25.85	3.42
05/13	−0.12	1.04	0.86	0.65	15.70	2.83	14.00	2.73	13.65$^+$	2.71$^+$	19.47	3.32
06/13	−0.19	1.11	1.84	0.40	15.26	2.91	14.44	2.89	13.59$^+$	2.81$^+$	21.21	3.11
07/13	−0.10	0.89	6.56	0.04	8.54$^+$	2.31$^+$	10.82	2.44	10.43	2.41	18.71	3.82
08/13	−0.12	0.98	3.79	0.15	12.58$^+$	2.51$^+$	17.12	2.66	17.55	2.68	31.30	3.21
09/13	0.09	0.93	2.48	0.29	13.91$^+$	2.42$^+$	17.38	2.65	16.62	2.60	21.26	2.93
10/13	−0.12	0.99	0.24	0.89	10.34$^+$	2.37$^+$	11.93	2.55	11.19	2.49	18.24	2.90
11/13	0.02	0.89	1.90	0.39	9.77$^+$	2.33$^+$	12.06	2.52	11.75	2.51	13.35	2.73
12/13	−0.03	1.02	3.27	0.19	12.83$^+$	2.58$^+$	15.72	2.85	15.06	2.78	16.25	2.99
01/13–12/13	−0.07	1.00	4.22	0.12	133.46$^+$	29.66$^+$	158.39	32.26	157.45	31.93	233.97	36.35

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

6.4. Predictive Model

For prediction accuracy, it is important that we simplify the model further by imposing the time-separability restriction $\theta = -\rho\kappa$ as supported by the data and, additionally, 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 I) suggest that we can safely capture seasonality with only two dummies (March to August and September to January). Moreover, we can also eliminate the insignificant covariates *Lmi*, *U18* and *Fhh*. This leaves us with a parsimonious total of seven variables in X_t (constant, two seasonal dummies, three covariates and one missing data dummy). The ML-EIS results for the simplified model reported in Table IV indicate that this elimination of 11 variables together with the time-separability restriction produces a log-likelihood ratio test statistic of 15.66, which is insignificant at the 0.27 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 % 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. We conducted an 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 1-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 (12) and (14). This provides us with a total of $138 \cdot 12 = 1656$ point predictions which we then compare with three sets of benchmark values. The first two are 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). The third set results from a simple Poisson regression model obtained by imposing the restrictions $\kappa = \rho = \theta = \sigma_\tau = \sigma_e = 0$ in the predictive model, which can then be estimated by standard ML. For all four methods we provide mean squared forecast errors (MSFE) and mean absolute forecast error (MAFE) for each month in 2013 as well as for the entire year of 2013. The results are reported in the right-hand panel of Table V. We find that, monthly as well as yearly, our predictive model clearly outperforms the Poisson regression model. Its predictive accuracy is also higher than that of 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 (Gorr *et al.*, 2003; Cohen and Gorr, 2005). It also indicates that for crime prediction purposes it is important to account for spatial and temporal dependence as well as for unobserved heterogeneity.

The left-hand panel of Table V summarizes the results of diagnostic checks on the normalized predictive PIT residuals $\xi^*_{iT'+1}$ in equation (17). Both sample means and standard deviations are close to their respective benchmark values of 0 and 1 under a correctly specified model. Moreover, the P-values of the Jarque-Bera statistic (Lütkepohl, 2007) indicate that the null hypothesis of normality expected for a correctly specified predictive model is never rejected at the 1% level.

As shown by equation (15) 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. This is important since full predictive cdfs are obviously far more informative than point predictions.

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

DOI: 10.1002/jae

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 enable 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 direct 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 1-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.* (2007).

ACKNOWLEDGEMENTS

We thank two anonymous referees for their helpful and constructive comments. We are also particularly grateful to Chris Briem (UCSUR, University of Pittsburgh) for discussions, as well as for having provided us with the socio-economic data from the Census 2000. We thank participants at the 2015 Pittsburgh Economics Medley Conference (University of Pittsburgh), IAAE 2015 Conference (Thessaloniki), Spatial Statistics 2015 Conference (Avignon), the 2015 Advances in Econometrics Conference (Louisiana State University), the 2015 Conference of the Verein für Socialpolitik (Münster) and the Statistische Woche 2015 (Hamburg) for helpful comments and suggestions. R. Liesenfeld and J. Vogler acknowledge support by the Deutsche Forschungsgemeinschaft (DFG) (grant LI 901/3-1). The estimation results of this paper were obtained using computing resources of the Institute of Statistics and Econometrics at the CAU Kiel, and CHEOPS, a scientific high-performance computer at the Regional Computing Center of the University of Cologne (RRZK) funded by the DFG.

REFERENCES

Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic: Boston, MA.

Anselin L, Cohen J, Cook D, Gorr W, Tita G. 2000. Spatial analyses of crime. In *Measurement and Analysis of Crime and Justice*, Duffee D (ed). National Institute of Justice: Washington, DC; 213–262.

Baller RD, Anselin L, Messner SF, Deane G, Hawkins DF. 2001. Structural covariates of U.S. county homicide rates: incorporating spatial effects. *Criminology* **39**: 561–590.

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

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

Cliff A, Ord JK. 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**: 29–37.

Cohen J, Gorr WL. 2005. Development of crime forecasting and mapping systems of use by police. Report, Carnegie Mellon University, Pittsburgh, PA.

Cohen J, Gorr WL, Olligschlaeger AM. 2007. Leading indicators and spatial interactions: a crime-forecasting model for proactive police deployment. *Geographical Analysis* **39**: 105–127.

Cohen LE, Felson M. 1979. Social change and crime rate trends: a routine activity approach. *American Sociological Review* **44**: 588–608.

Cornish DB, Clarke RVG. 1986. *The Reasoning Criminal: Rational Choice Perspectives on Offending*. Springer: New York.

Czado C, Gneiting T, Held L. 2009. Predictive model assessment for count data. *Biometrics* **65**: 1254–1261.
Debarsy N, Ertur C, LeSage JP. 2012. Interpreting dynamic space-time panel data models. *Statistical Methodology* **9**: 158–171.
Elhorst JP. 2012. Dynamic spatial panels: models, methods, and inferences. *Journal of Geographical Systems* **14**: 5–28.
Gilbert JR, 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, UK.
Helbich M, Arsanjani JJ. 2014. Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science* **42**: 134–148.
Jung RC, Kukuk M, Liesenfeld R. 2006. Time series of count data: modeling, estimation and diagnostics. *Computational Statistics and Data Analysis* **51**: 2350–2364.
Kubrin CE. 2003. Structural covariates of homicide rates: does type of homicide matter? *Journal of Research in Crime and Delinquency* **40**: 139–170.
Land KC, McCall PL, Cohen LE. 1990. Structural covariates of homicide rates: are there any invariances across time and social space? *American Journal of Sociology* **95**: 922–963.
LeSage JP, Pace RK. 2009. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC: Boca Raton, FL.
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 JF, Vogler J. 2016. Likelihood evaluation of high-dimensional spatial latent Gaussian models with non-Gaussian response variables. In *Spatial and Spatiotemporal Econometrics*, Baltagi BH, LeSage JP, Pace RK (eds), Advances in Econometrics, vol. 37. Emerald: Bingley, UK. (forthcoming).
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 PAP. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society B* **10**: 243–251.
Morenoff JD, Sampson RJ. 1997. Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. *Social Forces* **76**: 31–64.
Morenoff JD, Sampson RJ, Raudenbush SW. 2001. Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology* **39**: 517–558.
Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46**: 69–85.
Pace RK. 2014. Maximum likelihood estimation. In *Handbook of Regional Science*, Fischer MM, Nijkamp P (eds). Springer: Berlin; 1553–1570.
Pace RK, LeSage JP. 2016. Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. In *Spatial and Spatiotemporal Econometrics*, Baltagi BH, LeSage JP, Pace RK (eds), Advances in Econometrics, vol. 37. Emerald: Bingley, UK. (forthcoming).
Parent O, LeSage JP. 2012. Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics* **42**: 727–738.
Ratcliffe JH. 2013. Crime mapping: spatial and temporal challenges. In *Handbook of Quantitative Criminology*, Piquero AR, Weisburd D (eds). Springer: New York; 5–24.
Richard JF, Zhang W. 2007. Efficient high-dimensional importance sampling. *Journal of Econometrics* **141**: 1385–1411.
Roth RE, Ross KS, Finch BG, Luo W, MacEachren AM. 2013. Spatiotemporal crime analysis in U.S. law enforcement agencies: current practices and unmet needs. *Government Information Quarterly* **30**: 226–240.
Shaw CR, McKay HD. 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: Chicago, IL.
Tita GE, Greenbaum RT. 2009. Crime, neighborhoods, and units of analysis: putting space in its place. In *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, Weisburd D, Bernasco W, Bruinsma G (eds). Springer: New York; 145–170.
Tita GE, Radil SM. 2013. Spatial regression models in criminology: modeling social processes in the spatial weights matrix. In *Handbook of Quantitative Criminology*, Piquero AR, Weisburd D (eds). Springer: New York; 101–121.
Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46**: 234–240.

DOI: 10.1002/jae

Tolnay SE, Deane G, Beck EM. 1996. Vicarious violence: spatial effects on southern lynchings, 1890–1919. *American Journal of Sociology* **102**: 788–815.
US Department of Justice, Federal Bureau of Investigation. 2004. *Uniform Crime Reporting Handbook*.
Wang X. 2014. Limited and censored dependent variable models. In *Handbook of Regional Science*, Fischer MM, Nijkamp P (eds). Springer: Berlin; 1619–1635.
Willits D, Broidy L, Denman K. 2013. Schools, neighborhood risk factors, and crime. *Crime and Delinquency* **59**: 292–315.
Wilson JQ, Kelling GL. 1982. Broken windows: the police and neighborhood safety. *Atlantic Monthly* **249**: 29–38.
Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.