A spatial panel data model for estimating the impact of social and economic determinants on opioid mortality rates in the US

Gopal, Sucharita; Fischer, Manfred M.

DOI:
10.57938/d15da340-d792-4556-9c26-f017a198b285

Published: 01/08/2022

Document Version:
Publisher's PDF, also known as Version of record

Document License:
Unspecified

A Spatial Panel Data Model for Estimating the Impact of Social and Economic Determinants on Opioid Mortality Rates in the US

Sucharita Gopal*[1] , Manfred M. Fischer[2]

Abstract: This paper employs a spatial Durbin panel data model, an extension of the cross-sectional spatial Durbin model to a panel data framework, to estimate the impact of a set of social and economic determinants on opioid-induced mortality in the US. The empirical model uses a pool of US states over six years from 2014 to 2019 and a k=8 neighbor matrix that represents the topological structure between the states. Calculation of direct (own state) and indirect (cross-state spillovers) effects estimates – based on Bayesian estimation and inference – reflects a proper interpretation of the marginal effects for our nonlinear model that involves lags of the dependent variable vector. The study provides evidence for the existence of spatial effects working through the dependent variable vector and points to the importance of larger indirect effects of *Asian* and *Hispanic/Latino* minorities on the one side and the population age groups *35-44 years* and *65 years and older* on the other. This finding echoes the first law of geography that *"everything is related to everything else, but near things are more related than distant things"* (Tobler 1970). Space – largely neglected in previous research – matters for gaining a valid and better understanding of why and how neighboring states contribute to opioid-induced mortality in the states.

JEL: C23, I19, O51

Keywords: Spatial Durbin panel data model, Bayesian econometrics, Markov Chain Monte Carlo, direct (own state) effects, indirect (cross-state spatial spillover) effects, inferential statistics

August 10, 2022

*Corresponding author: Sucharita Gopal, Boston University, 675 Commonwealth Avenue, Boston University, Boston, MA 02215; email: suchi@bu.edu

[1] Department of Earth and Environment, Boston University, [2] Department of SocioEconomics, Vienna University of Economics and Business

.

1 Introduction

Opioid-induced deaths (hereafter, briefly opioid mortality) increased sharply in the United States (US) during the past two decades, and opioid overdose mortality has become a pressing health problem today (Kopel 2019).[1] Commonly, the opioid epidemic is perceived as a rural, white problem (Sobotka and Stewart 2020). This perception stems from opioid epidemics in rural Appalachia (Schalkoff et al. 2020) and the substantial increase in overdose mortality rates in rural counties (see, for example, Rossen et al. 2013). But opioid rates tend to be similar or even higher in urban counties (Monnat 2019) and have started to increase in recent years for Blacks (Alexander et al. 2018). Nationally, from 2016 to 2017, the opioid death rate for Blacks increased by 25 percent compared to 11 percent for Whites (Keturah and Ayana 2018).

Indeed, there is a substantial degree of heterogeneity across space in the US (Monnat 2019), and this heterogeneity is largely neglected in regression-based studies when analysing the impact of explanatory variables on opioid mortality. One prominent exception relates to the study by Yang et al. (2015) that explains geographic variation in US counties, using a spatial Durbin model approach, a state-of-the-art spatial econometric model, to analyse geographic variation across the US. The model approach used, however, is cross-sectional and lacks a longitudinal perspective.

To overcome this shortcoming, we employ a spatial Durbin panel data model, an extension of the cross-sectional model to a panel data framework, for estimating direct and indirect (cross-state spillovers) effects of a set of key factors (including race/ethnicity, age, opioid prescribing among other variables) on opioid mortality. The model is applied to 49 states over six years from 2014 to 2019. A k=8 nearest-neighbor matrix is chosen to represent the topological structure between the states. [2] Calculation of effects estimates is based on Bayesian estimation and inference, and reflects a proper interpretation of the marginal effects for our nonlinear model that involves lags

[1] Opioids include natural opioids such as morphine and codeine, semisynthetic opioids such as oxycodone, hydrocodone, hydromorphone and oxymorphone, and synthetic opioids such as methadone, tramadole and fentanyl. Heroine is an illicit opioid synthesized from morphine. Prescription opioids include natural and semisynthetic opioids and methadone (Scholl et al. 2019).

[2] This weight matrix constrains the neighbor structure to the 8-nearest neighbors. For a definition of k-nearest neighbor matrices see, for example, Fischer and Wang (2011)

of the dependent variable vector. The study provides a rich picture on how determinants affect mortality. The findings clearly point to the important role indirect effects play. A neglection of the indirect effects would cause incorrect inferences.

The remainder of the paper is structured as follows. Section 2 presents the econometric framework, outlines the proposed model, discusses prior selection for Bayesian estimation of the model, and shows how to calculate direct and indirect effects estimates necessary for proper interpretation of the model estimates. Section 3 outlines model specification issues, provides a brief summary of the data set used, and then presents the direct and indirect effects estimates of the explanatory variables on the opioid mortality rates. The last section summarises and concludes the paper.

2 Econometric Framework

2.1 The model

The econometric approach we employ in this study is a spatial econometric panel data model that is known as spatial Durbin panel data model in the spatial econometrics literature (see, e.g., LeSage 2014) and can be written in matrix form as

$$y = \rho W y + X\beta + WX\theta + \iota_T \otimes \mu + \nu \otimes \iota_N + \epsilon \tag{1}$$

where y is the $NT \times 1$ dependent variable vector of opioid mortality rates in state i ($i = 1, ..., N$) at time t ($t = 1, ..., T$), organized with t being the slow index for elements y_{it} in the vector y. X is an $NT \times K$ matrix of K explanatory variables that is organized in the same way. β is the associated $K \times 1$ parameter vector.

W is an $NT \times NT$ weight matrix that represents connections between the states and has a block diagonal form: $I_T \otimes w$, with $\otimes$ being a Kronecker product. w is a $N \times N$ spatial weight matrix reflecting spatial proximity of the N states that make up the panel of states over $t = 1, ..., T$ time periods. The diagonal elements $w_{ii} = 0$ ($i = 1, ..., N$), and the matrix is row-normalized to have row-sums of unity. The $NT \times 1$ matrix-vector product Wy is referred to as spatial lag and reflects a linear combination of neighboring state values for the dependent variable. The scalar parameter ρ

denotes the spatial dependence parameter that indicates the strength of spatial dependence between the vector y and the vector Wy.[3] The $NT \times K$ matrix-matrix product WX is used to create spatial lags of the K explanatory variables of the model, and represents a linear combination of characteristics from neighboring states, with the associated parameters θ.

$\iota_T \otimes \mu$ represents an N-vector of state-specific fixed effects μ, with $\otimes$ being a Kronecker product that repeats the vector μ for each time period. $\nu \otimes \iota_N$ reflects a Kronecker product of the T-vector of time-specific effects ν, one for each time period. The $NT \times 1$ vector ϵ is a stochastic disturbance term, assumed to be normally distributed with zero mean and a scalar variance, σ^2.

$$\epsilon \sim \mathcal{N}(0_{NT}, \sigma^2 I_{NT}) \tag{2}$$

2.2 Bayesian estimation

Estimation of the spatial Durban panel data model outlined in Eq.(1) is based on Markov Chain Monte Carlo (MCMC) estimation, with prior distributions assigned to the parameters ρ, $\delta = (\beta, \theta)'$ and σ^2 (see Mills and Parent 2021 for a recent survey of Bayesian inference methods in the regional science).[4] Parameter restrictions are imposed on the dependence parameter ρ during MCMC sampling, using methods described in LeSage (2020), and LeSage and Fischer (2020). MCMC estimation involves sequentially sampling each parameter from their conditional distributions. One of the main advantages of MCMC sampling is that conditional distributions for each parameter, given values of all other model parameters, take a form that is computationally simple to sample from (LeSage 2020).

The prior setup we use is standard. We rely on a normal prior for the parameters $\delta = (\beta, \theta)'$, associated with the X- and WX-variables, as in LeSage and Fischer (2020):

$$p(\delta) \sim \mathcal{N}(\bar{\delta}, \bar{\Sigma}_\delta) \tag{3}$$

[3]Note that the dependence parameter ρ is well defined over a limited interval that safeguards the existence of the matrix inverse $R^{-1} = (I_{NT} - \rho W)^{-1}$. This interval is $(\lambda_{min}, \lambda_{max})$, where λ_{min}, λ_{max} are the minimum and maximum eigenvalues of the matrix $(I_{NT} - \rho W)$, respectively (LeSage 2021). λ_{min} will be negative, and λ_{max} will be one, provided the matrix W is row-normalized. In this study we use a lower bound of -1 to avoid the need to calculate the minimum and maximum eigenvalues of the large matrix R.

[4]For μ and ν no priors, just demeaning transformations are used (see LeSage 2021).

where $\bar{\delta}$ is a $2K \times 1$ vector of prior means, and $\bar{\Sigma}_\delta$ a $(2K) \times (2K)$ prior variance-covariance matrix. We set the prior means to a value of 0.5 and the prior variance to 0.001.

For the dependence parameter ρ we employ a uniform prior

$$p(\rho) \sim U(-1, 1) \tag{4}$$

because this scalar parameter is constrained to lie in the open interval $(-1, 1)$. The constraint is imposed during MCMC estimation using a Metropolis-Hastings approach with rejection sampling, using methods described in LeSage and Fischer (2020), and LeSage (2020).

An informative prior is placed on the noise variance parameter σ^2, because we use an inverse gamma $(\bar{a}, \bar{b})$ distribution shown in Eq. (5).

$$p(\sigma^2) = \frac{\bar{b}^{\bar{a}}}{gamma(\bar{a})} \; (\sigma^2)^{-(\bar{a}+1)} \; exp(-\bar{b}/\sigma^2) \tag{5}$$

where $\bar{a} \to 0$, $\bar{b} \to 0$ reflect no prior information for this parameter.

As is traditional in the literature, we assume that priors for ρ, δ and σ^2 are independent. Given these priors we use the conditional distributions described in LeSage and Fischer (2020). A set of 4,000 draws were taken for determining posterior estimates of the parameters, with the first 500 omitted for burn-in of the sampler

2.3 Direct and indirect impact estimates

The estimates for the coefficients β and θ, unfortunately, are not meaningful as LeSage and Pace (2009) point out, if one is interested how changes in the explanatory variables impact the dependent variable outcome, which is the object of interest in this paper.[5] We follow the suggestion of LeSage and Pace (2009) to calculate direct and indirect effects estimates that reflect a proper interpretation of the marginal effects for our nonlinear model that involves spatial lags of the dependent variable.[6]

[5] One implication of the nonlinear relationship between y and X in our model is that changes in the kth variable $(k = 1, ..., K)$ for a single observation i can potentially impact all observations in the dependent vector y.

[6] See LeSage and Pace (2021) for a discussion of issues surrounding proper interpretation of the estimates from a variety of spatial regression models.

The partial derivatives used to calculate the direct and indirect effects estimates take the form of an $NT \times NT$ matrix. Using matrix notation, the partial derivative for the kth explanatory variable X^k can be written as shown in Eq. (6).

$$\partial y/\partial X^k = (I_{NT} - \hat{\rho}W)^{-1} \ (\hat{\beta} + W\hat{\theta}) \tag{6}$$

where $\hat{\rho}, \hat{\beta}$ and $\hat{\theta}$ denote parameter estimates. Diagonal elements of this matrix represent own-partial derivatives, while off-diagonal elements reflect cross-partial derivatives.

Using the mean of the main diagonal elements of the matrix of partial derivatives in Eq. (6) produces a scalar summary of the *direct effects*. Direct effects show how changes in the kth explanatory variable for the ith state impact the ith state's dependent variable for $i = 1, ..., N$. Using the mean of these NT different values generates a scalar summary that may be interpreted as representing how a change in the kth explanatory variable in the typical state impacts outcome y for the typical state (LeSage and Pace 2021).[7]

Indirect effects represent the impact of the ith state outcomes y_i from a change in the kth explanatory variable from the jth state, and capture the off-diagonal elements of the $NT \times NT$ matrix. Specifically, the elements in the ith row of the matrix show

$$\partial y_i/\partial X_j^k \text{ for } i \neq j; \ i = 1, ..., N$$

reflecting how changes in each of the other states' kth explanatory variable impact outcomes in the jth state. LeSage and Pace (2009) suggest using the mean of the sum of the off-diagonal elements from each row to produce a scalar summary measure of *cumulative indirect effects* or *spatial spillovers*.

Note that our model can have positive (or negative) direct effects associated with negative (or positive) indirect effects for the kth variable, so that spillover impacts might work in the opposite direction of direct impacts, arising from changes in each explanatory variable. *Total effects* represent

[7]Since the panel model includes time-specific fixed effects that capture variation over time periods, we can reasonably average the scalar summary estimates. But these scalar summary measures might ignore impact information over all time periods regarding the impact of changing values X^k at the observation level of direct (indirect) effects of the N states.

the sum of direct and indirect effects. In addition to calculating scalar summary measures of the effects, there is a need to calculate an empirical distribution from which measures of dispersion for the effects can be constructed and used for inference regarding the statistical significance of the effects (LeSage and Pace 2021). This is done using a sequence of say 1,000 MCMC draws for the parameters ρ, β, θ to calculate a sequence of 1,000 scalar summary measures of these effects that reflect the posterior distributions.

Given an estimate of the standard deviation for the scalar summary point estimates we can test hypotheses regarding the significance of the various types of effects for each of the explanatory variables used in the model. t-statistics, t-probabilities, and lower 0.05 and upper 0.95 credible intervals can be computed using the retained MCMC draws. The mean of the draws for direct effects (for each of the K explanatory variables), for example, and the standard deviation of the draws is taken to construct the t-statistic, which is then used to find the associated t-probability. The lower 0.05 and upper 0.95 are based on the MCMC draws. Given a set of 10,000 MCMC draws, the lower 0.05 interval would be determined by the 500th value from the lowest set of sorted values, and the upper 0.95 by the 9,500th value from this sorted set (LeSage 2021).

3 Application of the Model

This section applies the model. The empirical model estimated uses a pool of the 49 contiguous US states ($N = 49$) over six years from 2014 to 2019 ($T = 6$). The dependent variable vector y represents the annual opioid mortality rates and we consider a suit of K=17 explanatory variables. The choice of the time was based on the availability of data.

3.1 Model specification issues

The definition of a spatial lag in spatial regression models in general and our spatial Durbin panel data model in particular depends on the choice of a spatial weight matrix that summarizes the topology of the data set. Clearly, a large number of weight matrices can be derived for the same spatial layout (see Fischer and Wang 2011 for a review). In this study we use a nearest-neighbor-based weight matrix that constrains the neighbor structure to the k-nearest neighbors. The number

of neighbors, k, is the parameter of this weighting scheme. Its choice is an empirical issue (see LeSage and Fischer 2008) yielding k=8 neighbors in this study.[8]

W is an 234×234 weight matrix that represents connections between the states and has a block diagonal $I_T \otimes w$, with $\otimes$ being a Kronecker product. w is a 39×39 spatial weight matrix constructed based on 8-nearest neighboring states and identical during each time period in our model formulation. The diagonal elements $w_{ii} = 0$ $(i = 1, ..., 39)$, and the matrix is row-normalized to have row-sums of unity. The 234×1 matrix-vector product Wy is referred to as spatial lag and reflects a linear-combination of neighboring states values for the dependent variable opioid-induced death rates. The $234 \times K$ matrix product WX is used to create spatial lags of the $K = 17$ explanatory variables described in the following subsection.

3.2 Data and data sources

Data for the dependent variable come from the US Centers for Disease Control and Prevention (CDC) Wonder Online Databases of Multiple Cause of Death, filtered for drug-induced causes of death including opioids, synthetic opioids and psychostimulants, estimated per 10,000 population per each year for each state.[9] Based on previous literature (see, for example, Sun 2022, Wilson et al. 2020, Bohnert et al. 2019, Marotta et al. 2019) we consider the following set of X-variables to explain variation in state-specific opioid mortality rates over the observation period.

(a) The first variable relates to one of the opioid supply factors to US opioid mortality rates, measured in terms of the *opioid prescribing rate* per 100 person in each state every year. The data come from the US Opioid Dispensing Rate Maps provided by IQVIA Xponent from the CDC in 2020. The prescribing rate is calculated by dividing the total number of opioid prescriptions dispensed annually at the state-level by annual resident population and multiplying the number by 100.

(b) To uncover the role of race/ethnicity we consider its composition in terms of the percentages of

[8]Comparing ten different k-neighbor matrices running from k=1 to k=10, we find that k=8 leads to a log-marginal likelihood of -174.8609 and an associated model probability of 0.9999, close to one.

[9]Note that more than 80 percent of drug overdose death certificates named at least one drug in 2017 and 2018. Drug-specific overdose numbers and rates might have changed substantially from 2017 to 2018 as a result of changes in reporting in some states (see Wilson et al. 2020).

White, *Black or African American*, *American Indian and Alaska Native*(=*Native Americans*), *Asian*, *Hispanic or Latino* (of any race/ethnicity) out of the total population in each state every year. The data was derived from the US Census Bureau's Population Estimates Program called American Community Survey (ACS).

(c) Opioid mortality has occurred across all age groups. In this study we consider *age* in five categories (*18-24 years*, *25-34 years*, *35-44 years*, *45-64 years*, *65 years and more*) representing percentage of total population in each age group in each state every year (Source: US Census Bureau ACS).

(d) The *education* level is measured by two variables, the percentage of people who hold a *high school's degree or higher*, and the percentage of people having a *bachelor's degree or higher* in each state every year (Source: US Census Bureau ACS).

(e) The *poverty rate*, the *unemployment rate* and the variable *lower income* represent the socioeconomic context in each state and year. The *poverty rate* is calculated as the percentage of people whose income falls below the state-specific level per year (Source: American Community Survey), while *lower income* is measured in terms of the number of people with income of less than $ 34,999 per 10,000 persons (Source: US Census Bureau ACS). The *unemployment rate* is given as percentage of people aged 16 years and older who are in the resident labor force, but unemployed (Source: US Census Bureau ACS).

(f) Finally, we also consider insurance coverage, measured in terms of the percentage of people with *health insurance* in each state and year, as a potential explanatory variable (Source: US Census Bureau ACS).

3.3 Estimation results and interpretation of the coefficient estimates

This subsection presents the estimation results of the model and quantifies the impact of the explanatory variables on the opioid mortality rates, using the scalar summary direct and indirect impact measures described in Section 2.3.

Table 1: Posterior parameter estimates of the spatial Durbin panel data model

Variable (Parameter)	Posterior Mean	Asymp t-stat	z-prob
Opioid Prescribing Rate (β_1)	−0.0064	−2.3763	0.0175
Percent White (β_2)	−0.0163*	−0.7389	0.4600
Percent Black/African American (β_3)	0.0078*	0.1706	0.8645
Percent Native (β_4)	−0.2221*	−1.8458	0.0649
Percent Asian (β_5)	−0.2059	−2.3068	0.0211
Percent Hispanic/Latino (β_6)	0.0360*	0.7658	0.4438
Percent Age 18-24 (β_7)	−0.1057*	−1.2978	0.1944
Percent Age 25-34 (β_8)	0.1372*	1.4731	0.1407
Percent Age 35-44 (β_9)	0.0610*	0.5755	0.5650
Percent Age 45-64 (β_{10})	−0.0594*	−0.8160	0.4145
Percent Age $\geq$ 65 (β_{11})	−0.1694	−2.2128	0.0269
Percent High School and above (β_{12})	−0.0807	−2.4654	0.0137
Percent Bachelor and higher (β_{13})	0.0466*	1.6510	0.0987
Poverty Rate (β_{14})	−0.0282*	−1.9677	0.0491
Lower Income(β_{15})	0.0238*	1.3788	0.1680
Unemployment Rate (β_{16})	0.0210*	0.6234	0.5330
Percent Health Insurance (β_{17})	−0.0145*	−1.6671	0.0955
$W*$ Opioid Prescribing Rate (θ_1)	−0.0222	−2.5986	0.0094
$W*$ Percent White (θ_2)	0.2098	2.5127	0.0120
$W*$ Percent Black/African American (θ_3)	−0.3091*	−1.8291	0.0674
$W*$ Percent Native (θ_4)	−0.7714*	−1.7840	0.0744
$W*$ Percent Asian (θ_5)	0.7766	2.4292	0.0151
$W*$ Percent Hispanic/Latino (θ_6)	0.6277	3.1032	0.0019
$W*$ Percent Age 18-24 (θ_7)	0.0487*	0.2115	0.8325
$W*$ Percent Age 25-34 (θ_8)	−0.3329*	−1.1208	0.2624
$W*$ Percent Age 35-44 (θ_9)	−0.7836	−2.9728	0.0030
$W*$ Percent Age 45-64 (θ_{10})	−0.2515*	−1.0777	0.2812
$W*$ Percent Age $\geq$ 65 (θ_{11})	−0.8490	−4.7356	0.0000
$W*$ Percent High School and above (θ_{12})	−0.0220*	−0.2096	0.8340
$W*$ Percent Bachelor and higher (θ_{13})	0.0962*	1.0993	0.2716
$W*$ Poverty Rate (θ_{14})	−0.1078	−2.1844	0.0289
$W*$ Lower Income (θ_{15})	0.1852	3.7613	0.0002
$W*$ Unemployment Rate (θ_{16})	−0.4948	−5.7702	0.0000
$W*$ Percent Health Insurance (θ_{17})	−0.0158*	−0.9359	0.3493
Rho (ρ)	0.2939	4.0535	0.0001

Notes: Results are based on a k-nearest neighbor matrix W with $k = 8$ nearest neighboring states; $N = 39$, $T = 6$ and $K = 17$; parameter estimates calculated using LeSage's (2021) Panel Data Toolbox for MATLAB; performance: log-marginal likelihood= −174.8609, R-square= 0.9353, corr-square= 0.6507 and sigma-square= 0.0502; * indicates not significantly from zero

In Table 1 we report the posterior parameter estimates for the spatial Durbin panel data model, although these are not directly interpretable in terms of the impacts associated with changes in the X- and WX-variables on the dependent variable. The table also summarizes the associated (asymptotic) t-statistics and the z-probabilities. t-statistics represent the posterior mean divided by the posterior standard deviation both of which are calculated from the retained MCMC draws. The t-statistics are then used to determine the associated z-probabilities. The table indicates and this is important to note that the parameter estimate of the spatial autoregressive parameter is positive ($\hat{\rho} = 0.2939$) and significantly different from zero, providing evidence for the existence of significant spatial effects working through the dependent variable.

The proper way to interpret the model results is in terms of the (summary) direct and indirect effects estimates of the X- and WX-variables. Table 2 outlines these (cumulative) effects estimates, along with inferential statistics. A positive (negative) mean with positive (negative) lower and upper credible intervals should be interpreted as a positive (negative) effect. Effects whose credible intervals span zero are not significant. The scalar summary effects estimates average over all the US states in the sample.[10]

Direct (own state) effects responses in Table 2 indicate four negative and significant effects. The own-state (direct) impact of changes in the percentage of the senior population (65 years and older) on the opioid mortality rate is negative (-0.2042) and significant at the 99 percent level. This result matches our intuition, as we should expect that a larger percentage of senior people in the typical state i would reduce the opioid mortality rate. The direct effect of *Native Americans* is negative (-0.2545) and significant at the 95 percent level while that of the *population with a high school degree or higher* (-0.0825), the *poverty rate* (-0.0327) and the *opioid prescribing rate* (-0.0073) are also negative and significant, but at a much lower magnitude. The variables *percent Hispanic/Latino*, *percent age 25-34* and *percent age 35-44* are positive, but not significantly different from zero. All other estimates reported in Table 2 are negative, but not significantly different from zero.

[10]Relying on the scalar summary effects estimates might obscure important variation over time in the direct and indirect effects that may be of substantive interest in spatial regression relationships being analysed. In such cases plots of the observation-level direct and indirect effects should be considered.

Table 2: Direct and indirect impact estimates for changes in X- and WX-variables

Impact Estimates	Posterior Mean	t-stat	t-prob	Lower 0.05	Upper 0.95
Direct Impact Estimates					
Opioid Prescribing Rate	−0.0073	−2.6367	0.0086	−0.0127	−0.0018
Percent White	−0.0085*	−0.3738	0.7087	−0.0524	0.0353
Percent Black/African American	−0.0041*	−0.0842	0.9330	−0.1000	0.0873
Percent Native	−0.2545	−2.0072	0.0453	−0.5028	−0.0083
Percent Asian	−0.1785*	−1.9235	0.0550	−0.3623	0.0016
Percent Hispanic/Latino	0.0606*	1.2011	0.2303	−0.0356	0.1582
Percent Age 18-24	−0.1050*	−1.2992	0.1945	−0.2639	0.0500
Percent Age 25-34	0.1262*	1.3160	0.1888	−0.0583	0.3103
Percent Age 35-44	0.0316*	0.2972	0.7664	−0.1774	0.2446
Percent Age 45-64	−0.0698*	−0.9563	0.3394	−0.2075	0.0750
Percent Age ≥65	−0.2042	−2.6293	0.0088	−0.3535	−0.0522
Percent High School and above	−0.0825	−2.4486	0.0147	−0.1488	−0.0167
Percent Bachelor and higher	0.0508*	1.7462	0.0814	−0.0075	0.1085
Poverty Rate	−0.0327	−2.2147	0.0272	−0.0610	−0.0037
Lower Income	0.0312*	1.7802	0.0757	−0.0025	0.0672
Unemployment Rate	0.0021*	0.0630	0.9498	−0.0654	0.0688
Percent Health Insurance	−0.0153*	−1.7790	0.0759	−0.0321	0.0012
Indirect Impact Estimates					
Opioid Prescribing Rate	−0.0336	−2.6377	0.0086	−0.0594	−0.0097
Percent White	0.2851	2.3991	0.0168	0.0569	0.5318
Percent Black/African American	−0.4279*	−1.7446	0.0817	−0.9173	0.0489
Percent Native	−1.1673*	−1.8483	0.0652	−2.4187	0.0171
Percent Asian	0.9952	2.1655	0.0308	0.1122	1.9368
Percent Hispanic/Latino	0.8889	2.9519	0.0033	0.3175	1.5079
Percent Age 18-24	0.0244*	0.0776	0.9382	−0.5871	0.6369
Percent Age 25-34	−0.4046*	−0.9602	0.3374	−1.2489	0.4292
Percent Age 35-44	−1.0646	−2.8531	0.0045	−1.8139	−0.3602
Percent Age 45-64	−0.3749*	−1.1445	0.2530	−1.0468	0.2564
Percent Age ≥65	−1.2530	−4.4361	0.0000	−1.8542	−0.7393
Percent High School and above	−0.0640*	−0.4300	0.6674	−0.3648	0.2361
Percent Bachelor and higher	0.1535*	1.2188	0.2235	−0.0835	0.4039
Poverty Rate	−0.1619	−2.2505	0.0249	−0.3084	−0.0266
Lower Income	0.2680	3.5939	0.0004	0.1306	0.4227
Unemployment Rate	−0.6805	−5.0752	0.0000	−0.9641	−0.4371
Percent Health Insurance	−0.0280*	−1.2301	0.2193	−0.0754	0.0158

Notes: Impact estimates are calculated using LeSage's (2021) Panel Data Toolbox; * indicates not significantly different from zero.

The impact estimates differ from the corresponding coefficient estimates outlined in Table 1. The difference is due to some feedback effect that comes into play in the direct effects estimates. The discrepancy is positive since the impact estimates exceed the coefficients, reflecting some positive feedback. Note that it would be a mistake to interpret the β-parameters as representing direct effects estimates. If we would incorrectly view, for example, the model coefficient β_4 on the *percent Native American* variable as representing the direct impact, this would lead to an effect not significantly different from zero. But the true direct impact estimate points to effects estimates that are negative and significant.

Table 2 also shows the cumulative indirect (cross-state spatial spillover) effects with a change in the WX-variables. The indirect impact of the percentage of the senior population (*65 years and older*) in neighboring states j ($j \neq i$) is negative and significant at the 99 percent level. The cumulative spillover magnitude of 1.2530 appears to be rather large when compared to the direct effects magnitude of 0.2042. It must be noted, however, that the indirect impact estimates are cumulative spillovers where cumulation takes place over all neighboring states $j \neq i$, neighbors to the neighboring states, and so on. While this may seem counterintuitive, the indirect effects falling on any single state would be generally much smaller, consistent with spillovers being a "second-order effect". Also, the largest indirect effects would fall on nearby states. It is the cumulation of the spatial spillovers over all states that leads to a relatively larger indirect than direct effects. Of course, one could investigate direct and indirect impacts for an individual state without averaging, but this would take the form of a 1-by-294 row vector for each state.

Moreover, it is worth noting that the indirect (spatial spillover) impact of the percentage of the age group *35-44 years* in neighboring states j ($j \neq i$) is negative (-1.0646) and significant at the 99 percent level, while the indirect impact estimates of the percentages of *Asian*, *Hispanic/Latino* and *White* populations in neighboring states j are significant and positive: 0.9952, 0.8889 and 0.2851, respectively. Finally, the following variables exert significant indirect effects on opioid mortality rates: the *opioid prescribing rate* (-0.0336), the *poverty rate* (-0.1619), *lower income* (0.2680), and the *unemployment rate* (-0.06805).

4 Closing Remarks

This paper shows that the spatial Durbin panel data model is a suitable model specification for quantifying the direct and indirect impact of a suit of explanatory variables on opioid mortality rates in a panel data framework. The analysis provides evidence for the existence of spatial externalities and interaction. Since the model involves spatial lags of dependent and independent variables, the traditional least-squares ceteris paribus interpretation of the regression parameters does not hold any longer. Thus, direct and indirect (spillover) effects estimates had to be calculated to quantify the impact of the explanatory variables on the variable vector that represents the state-specific opioid mortality rates over 2014 to 2019.

The main results of the analysis may be summarized in terms of the total effects estimates as follows.[11] *First*, *opioid prescribing* has a negative total effect, statistically different from zero but with a rather low magnitude of 0.0409. *Second*, more important are the *race/ethnicity* categories. The *Hispanic/Latino*, but also the *White American* category reveals a statistically significant positive total impact effect (0.9495 and 0.2767, respectively) while the impact of *Native Americans* is negative (-1.4218) and significant. *Third*, *age* plays the most important role evidenced by a total effects estimate of -1.4571 in the case of the senior population (*65 years and older*) and -1.0330 in the case of the age category *35-44 years*, both are highly significant. *Fourth*, the total impacts of the *poverty rate*, the *unemployment rate* and the *lower income* variable are significantly different from zero and much more important than the *opioid prescribing rate*. A *final* point to note is the important role that indirect effects play whose omission would lead to incorrect inferences. Space matters and may lead to a better understanding of why and how neighboring states contribute to the opioid-induced mortality in the typical US state.

The study provides a rich picture on how potential determinants affect opioid mortality in US states, yielding useful insights to motivate the adoption of regional public policies for tackling the opioid crisis. State-specific policies may be derived by calculating observation-level direct and indirect effects. This, however, would require to analyse 294x1 row vectors, one for each of the 49

[11] Note that total impact estimates are defined as sum of the direct and indirect estimates, but inference statistics has to based on the sum of the two coefficients $\hat{\beta} + \hat{\theta}$, and a proper measure for dispersion of this sum of coefficients. This can be constructed from the MCMC draws.

states, and this might be done in future studies.

Note that all the inferences were made conditional on the data and the specification of the spatial weight matrix. The assumption that a particular spatial weight specification is correct might be relaxed by explicitly incorporating model uncertainty in the statistical analysis. To accommodate this uncertainty issue one might use a convex combination of spatial weight matrices to be used for estimation of the model, in the spirit of Fischer and LeSage (2020).

Acknowledgements

References

Alexander MJ, Kiang MV, Barbieri M (2018) Trends in Black and White opioid mortality in the United States, 1979-2015, *Epidemiology*, vol. 29(5), 707-715

Blake-Gonzalez B, Cebula RJ, Koch JV (2021) Drug-overdose death rates: The economic misery explanation and its alternatives, *Applied Economics*, vol. 53(6), 730-741

Bohnert AS, Valenstein M, Bair MJ, Ganoczy D, McCarthy JF, Ilgen MA, Blow FC (2011) Association between opioid prescribing patterns and opioid overdose-related deaths, *Jama*, vol. 305(13), 1315-1321

Center for Disease Control and Prevention [Internet] https://www.cdc.gov/drugoverdose/rxrate-maps/state2014.html

Fischer MM, Wang J (2011) *Spatial Data Analysis. Methods and Techniques* [Springer Briefs in Regional Science Book Series]. Springer, Berlin, Heidelberg and New York

Fischer MM, LeSage JP (2020) Network dependence in multi-indexed data on international trade flows, *Journal of Spatial Econometrics*, vol 1(1): article 4, available online, doi: 10.1007/s43071-020-00005-w

Keturah J, Ayana J (2018) The opioid crisis in Black communities, *The Journal of Law, Medicine Ethics*, vol. 46(2), 404-421

Kopel J (2019). Opioid mortality in rural communities, *The Southwest Respiratory and Critical Care Chronicle*, vol. 7(31), 59-62

LeSage JP (2014) Spatial econometric panel data model specification: A Bayesian approach, *Spatial Statistics*, vol 9, 122-145

LeSage JP (2020) Fast MCMC estimation of multiple W-matrix spatial regression models and Metropolis-Hastings Monte Carlo log-marginal likelihoods, *Journal of Geographical Systems*, vol 22(1), 47-75

LeSage JP (2021) A panel data tool box for MATLAB, Department of Economics, University of Toledo

LeSage JP, Fischer MM (2008) Spatial growth regressions. Model specification, estimation and interpretation, *Spatial Economic Analysis*, vol 3(3), 275-304

LeSage JP, Fischer MM (2020) Cross-sectional dependence specifications in a static trade panel data setting, *Journal of Geographical Systems*, vol 22(1), 5-46

LeSage JP, Pace RK (2009) *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group, Boca Raton

LeSage JP, Pace RK (2021) Interpreting spatial econometric models. In Fischer MM and Njikamp P (eds) *Handbook of Regional Science*, 2nd and extended edition, pp. 2201-2218. Springer, Berlin and Heidelberg

Marotta PL, Hunt T, Gilbert L, Wu E, Goddard E, El-Bassel N (2019) Assessing spatial relationships between prescription drugs, race, and overdose in New York State from 2013 to 2015, *Journal of Psychoactive Drugs*, vol. 51(4), 360-370

Monnat SM (2019) THe contributions of socioeconomic and opioid supply factors to U.S. drug mortality rates: Urban-rural and within-rural differences, *Journal of Rural Studies*, vol. 68, 319-335

Mills JA, Parent O (2021) Bayesian Markov Chain Monte Carlo estimation. In Fischer MM and Njikamp P (eds) *Handbook of Regional Science*, 2nd and extended edition, pp. 2073-2096. Springer, Berlin and Heidelberg

Rossen LM, Khan D, Warner M (2013) Trends and geographic patterns in drug-poisoning death rates in the U.S., 1999-2009, *American Journal of Preventive Medicine*, vol. 45(6), e19-e25

SAMHSA [Internet] (2018), Opioid overdose prevention toolkit − Updated 2018. Available from: https://store.samhsa.gov/product/Opioid-Overdose-Prevention-Toolkit/SMA18-4742

Schalkoff CA, Lancaster KE, Gaynes BN, Wang V, Pence BW, Miller WC (2020) The opioid and related drug epidemics in rural Appalachia: A systematic review of populations affected, risk factors, and infectious diseases, *Substance Abuse*, vol.41(1), 35-69

Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G (2019) Drug and opioid-involved overdose deaths — United States, 2013–2017, *Morbidity and Mortality Weekly Report*, vol. 67, 1419–1427

Sobotka TC, Stewart SA (2020) Stereotyping and the opioid epidemic: A conjoint analysis, *Social Science & Medicine*, vol. 255, available online, doi: 10.1016/j.socscimed.2020.113018

Sun F (2022) Rurality and opioid prescribing rates in U.S. counties from 2006 to 2018: A spatiotemporal investigation, *Social Science & Medicine*, vol. 296, 114788, available online, doi: 10.1016/j.socscimed.2022.114788

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region, *Economic Geography* , vol. 46 (Supplement), 234-240

Wilson N, Kariisa M, Seth P, Smith H, Davis NL (2020) Drug and opioid-involved overdose deaths - United States, 2017-2018, *Morbidity and Mortality Weekly Report*, vol. 69, 290-297

Yang T-C, Noah A, Shoff C (2015) Exploring geographic variation in US mortality rates using a spatial Durbin approach, *Population, Spase and Place*, vol. 21(1), 18-37