Interpreting dynamic space-time panel data models

Nicolas Debarsy*
CERPE, University of Namur,
Rempart de la vierge, 8, 5000 Namur, Belgium
E-mail: ndebarsy@fundp.ac.be

Cem Ertur
LEO, Université d'Orléans
Rue de Blois - BP 6739
45067 Orléans Cedex 2, France
E-mail: cem.ertur@univ-orleans.fr

James P. LeSage†
Fields Endowed Chair
Texas State University-San Marcos
Department of Finance & Economics
San Marcos, TX 78666
E-mail: jlesage@spatial-econometrics.com

Abstract

The literature on spatial econometrics has mainly focused its attention on the establishment of the asymptotic properties for several spatial autoregressive models. However, little attention has been paid to how the model estimates should be interpreted. The first goal of this paper is thus to summarize the impacts measures coming from the spatial autoregressive cross section and static panel data models. There is a great deal of literature regarding the asymptotic properties of various approaches to estimating simultaneous space-time panel models, but little attention has been paid to how the model estimates should be interpreted. The motivation for use of space-time panel models is that they can provide us with information not available from cross-sectional spatial regressions. LeSage & Pace (2009) show that cross-sectional simultaneous spatial autoregressive models can be viewed as a limiting outcome of a dynamic space-time autoregressive process. A valuable

*Nicolas Debarsy is research fellow at the F.R.S - FNRS and gratefully acknowledges their financial support.

†Corresponding author

aspect of dynamic space-time panel data models is that the own- and cross-partial derivatives that relate changes in the explanatory variables to those that arise in the dependent variable are explicit. This allows us to employ parameter estimates from these models to quantify dynamic responses over time and space as well as space-time diffusion impacts. We illustrate our approach using the demand for cigarettes over a 30 year period from 1963-1992, where the motivation for spatial dependence is a bootlegging effect where buyers of cigarettes near state borders purchase in neighboring states if there is a price advantage to doing so.
Keywords: Dynamic space-time panel data model, Markov Chain Monte Carlo estimation, dynamic responses over time and space.

1 Introduction

Great progress has been made in the past decade on the theoretical aspects of spatial econometrics. For cross-sectional data, Lee (2004) proved consistency and asymptotic normality of the quasi-maximum likelihood estimator while Kelejian & Prucha (1998, 1999) developed a generalized Method of Moments (GMM) estimator and its asymptotic properties for a regression model that includes spatial autoregressive disturbances. Besides, Lee (2007) derived the best GMM estimator and showed that it has the same limiting distribution as the maximum likelihood estimator (assuming normal disturbances). Kelejian & Prucha (2010) and Lin & Lee (2010) also developed GMM estimators robust to the presence of heteroskedasticity.

Besides, several authors focused their research on the development of the theoretical properties of both static and dynamic spatial panel data models estimators. Kapoor et al. (2007) were the first to derive a GMM estimator for a one-way spatially autocorrelated error panel data model. This paper has moreover been extended to the presence of an endogenous spatial lag by Mutl & Pfaffermayr (2011) who also derived a Hausman statistic to test the adequacy of the random effects specification in this GMM approach. Furthermore, Lee & Yu (2010*a*) developed an efficient GMM estimation for the spatial dynamic panel data model with fixed effects. In addition, several papers studied the properties of the quasi-maximum likelihood estimators (QMLE) in spatial panel data models. Yu et al. (2008) studied its asymptotic behavior in the spatial dynamic panel data with fixed effects when both the number of individual and time periods tends to infinity while Lee & Yu (2010*d*) further considered the presence of time fixed effects when establishing the asymptotic properties of QMLE. Lee & Yu (2010*b*) derived the asymptotic properties of the static fixed effects spatial autoregressive panel data model. Finally, Elhorst (2003) and Anselin et al. (2007) were the first to present the specifications an applied researcher could face when working on static spatial panel data models. These works were followed by two others papers (Lee & Yu 2010*c*,*e*) that integrated the recent advances

in dynamic space-time models in their reviews and further derived a Hausman and Lagrange Multiplier statistic to test the relevance of the random effects specification in the static spatial panel data model in the maximum likelihood framework.

In addition to the aforementioned papers devoted to the so-called frequentist framework, several articles developed estimation procedure for spatial models in a bayesian perspective. Lesage (1997) proposes a Markov Chain Monte Carlo (MCMC) method for cross-sectional spatial autoregressive models and Kakamu (2009) examines the small sample properties of a class of spatial models. Also, Kakamu & Wago (2008) study the small sample properties of spatial static panel data specifications. Finally, Parent & LeSage (2010*a*) propose a space-time filter to capture serial and spatial autocorrelation in the error term of a random effects panel data model while Parent & LeSage (2010*c*) derive a space-time filter for a spatial autoregressive random effects specification.

Notwithstanding the improvement of knowledges in theoretical foundations of spatial econometrics, there still remains confusion on the interpretation of impacts of a change in an explanatory variable on the dependent variable in the spatial autoregressive model. This confusion mainly comes from the fact that the spatial autoregressive model is estimated in implicit form and researchers should derive the associated reduced form and compute the associated matrix of partial derivatives used to evaluate these impacts. For the cross-sectional case, LeSage & Pace (2009, chap. 2) contains an excellent review of these interpretations. With the development of spatial panel data models, and more particularly, spatio-temporal specifications, where time and spatial dimensions interact, the need for clear interpretations of impacts of change in explanatory variables is stronger than ever. Indeed, such spatio-temporal models are of the highest interest for policy makers and constitute a complete tool to analyze diffusion effects (in time, space and spatio-temporal dimensions) of a shock given in a specific spatial unit on this specific unit as well as on all others locations considered in the sample. These specifications are also useful in many other economics fields, like public economics where the decisions taken by a government are partly affected by current and past decisions of neighboring government, due to mimicking behaviors of competition effects. Also, for researchers interested in political economics, these spatio-temporal models are of the highest interest. The most striking current example to illustrate these diffusions effects concerns the demand for political change in North Africa which started in Tunisia in December 2010 and spreads now over Egypt, Libya, Syria and Yemen.

The objective of the paper is thus to provide a complete overview of the computation of impacts of a change in a explanatory variable on the dependent variable for all spatial autoregressive models. We initially present the impacts for the simple spatial autoregressive cross-sectional and the static

panel model. We then describe the dynamic space-time panel data model along with a Bayesian Markov Chain Monte Carlo (MCMC) estimation procedure and develop analytical expressions for the partial derivatives that allows to compute spatial and spatio-temporal impacts of changes in explanatory variables. In this spatio-temporal panel data framework, we discuss interpretative considerations related to whether one is interested in responses to one-period change or permanent change in the level of these variables.

The rest of the paper is organized as follows. Section 2 presents impacts for the cross-sectional and static panel data models. Section 3 presents the spatio-temporal model estimation method and impacts computation. Section 4 illustrates the calculation of the impacts in this spatio-temporal model using a panel dataset from Baltagi & Li (2004) that relates state-level cigarette sales to prices and income over time. A final section contains our conclusions.

2 Impacts of spatial autoregressive cross-sectional and static panel data models

The cross-sectional spatial autoregressive model (SAR) presented in (1) for a sample of N observations serves as the workhorse of spatial regression modeling.

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \iota_N \alpha + \mathbf{X}\beta + \varepsilon \tag{1}$$

$\mathbf{y}$ is the vector of the dependent variable while $\mathbf{W}\mathbf{y}$ is its spatial lag allowing to capture the presence of spillovers. Also, ρ is the spatial autoregressive parameter representing the intensity of spatial autocorrelation. The $N \times N$ matrix $\mathbf{W}$ is an interaction matrix whose (i, j)th element takes some finite positive non-stochastic value if regions i and j interact and zero otherwise. This matrix is also assumed to be row-normalized such that its rows sum to unity. The matrix $\mathbf{X}$ contains the K exogenous explanatory variables while β is the K-dimensional vector of associated coefficients. The constant term is represented by α and ι_N is the N-dimensional unitary vector. The vector of theoretical disturbances, ε, is assumed to be *iid* with zero mean and constant variance σ^2. Let us also note that even though the interaction matrix $\mathbf{W}$ and all variables depend on the sample size N, forming triangular arrays, as the paper does not focus on asymptotic properties, we do not explicitly write this dependence to keep some clarity in notation.

To interpret the effect of a change in an explanatory variable in this SAR specification, one has to compute the reduced form of (1), which is presented in (2).

$$\mathbf{y} = (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\iota_N \alpha + \mathbf{X}\beta + \varepsilon) \tag{2}$$

To ensure that $\mathbf{I}_N - \rho\mathbf{W}$ is invertible, one needs to impose some restrictions on the parameter space ρ, which for a row-normalized interaction matrix $\mathbf{W}$, correspond to take use a compact set of $(-1, 1)$. The impacts of a change in the rth explanatory variable corresponds to the $N \times N$ matrix of partial derivatives of the dependent variable $\mathbf{y}$ with respect to the concerned covariate, $\mathbf{x}_r$, are shown in (3).

$$\Xi_y^{x_r} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}_r} = (\mathbf{I}_N - \rho\mathbf{W})^{-1} \beta_r \tag{3}$$

By contrast to the traditional classical linear model, diagonal elements of (3) are different from each others, off-diagonal elements differ from zero and the matrix itself is not symmetric. Spatial autoregressive models thus provide much richer information about impacts of a variation in a explanatory variable.

Diagonal elements represent own-partial derivatives, meaning the impact of a change in the rth explanatory variable in spatial unit i on the dependent variable in this ith spatial unit. They are formally written as

$$\frac{\partial y_i}{\partial x_{ri}} = \left[\Xi_y^{x_r}\right]_{ii} \quad i = 1, \ldots, N \tag{4}$$

These own-partial derivatives are labeled direct impacts and include feedback loop effects that arise as a result of impacts passing through neighboring spatial units j and back to spatial unit i. As the neighborhood is different for each spatial units, the feedback will be heterogeneous by nature, giving birth to the notion of interactive heterogeneity. This interactive heterogeneity is inherent to spatial autoregressive models and should not be confused with what the literature calls spatial heterogeneity, which refers to spatial instability of parameters (structural breaks, clubs) or heteroskedasticity.[1]

As mentioned above, off-diagonal elements of (3), representing cross-partial derivatives, differ from zero. This means that a change in the rth explanatory variable in location j will affect the dependent variable in location i. As matrix (3) is asymmetric this further imply that this impact will not be the same as the one caused by a change in location i on location j. Formally,

$$\frac{\partial y_i}{\partial x_{rj}} = \left[\Xi_y^{x_r}\right]_{ij} \neq \frac{\partial y_j}{\partial x_{ri}} = \left[\Xi_y^{x_r}\right]_{ji} \tag{5}$$

These cross-derivative elements are thus labeled indirect effects and show the response of the dependent variable in location i to a change in explanatory variables in any of the other locations.[2]

[1] This interactive heterogeneity is deeply discussed in the next chapter of the dissertation.

[2] A slightly different terminology has been developed in Kelejian et al. (2006) who replace indirect effects by emanating effects and direct effects by own-spillover effects. However, the underlying concepts are identical.

As one has N direct effects and $N(N-1)$ indirect effects, it can be problematic to provide sound interpretations and summarizing measures of this information are useful. As first summarizing step, one can compute the sum of indirect effects for the ith row in (3). One gets the cumulative indirect effect for location i derived from a change of the same magnitude in the rth explanatory variable in all but the ith location. Alternatively, the sum of indirects effects can be computed for column i. The obtained cumulative indirect effect is thus interpreted as the aggregated effect for all locations except i due to a change in the rth covariate in location i. One can also define total effect for location i either as the sums of all elements (including $\left[\Xi_y^{x_r}\right]_{ii}$) of column i or row i. In the former case, one gets the total impact for all locations (including i) of a change in the rth explanatory variable in location i while in the latter, one gets the total impact for location i of a change in the same magnitude of the rth covariate in *all* locations. LeSage & Pace (2009) propose to go one step ahead by using only three scalars measures to summarize information contained in the matrix (3). They suggest to use the average of the N diagonal elements as a measure of direct effects. Likewise, they define the average cumulative indirect effect as the average of the cumulative indirect effects and the average total effect as the mean of total effects.[3]

The notion that feedback and spillovers effects can exist in a cross-sectional setting where time is not involved requires that we think of these models as reflecting the outcome of a long-run equilibrium or steady state. Changes in the explanatory variables are then interpreted as setting in motion forces that lead to a new long-run equilibrium, which provides an intuitive motivation for spillovers and feedback effects in these cross-sectional models.

In addition to proposing scalar summary measures for the cross-sectional SAR model effects, LeSage & Pace (2009) provide a computationally efficient approach to determining measures of dispersion for these scalar summary effects estimates. These can be used to draw inference regarding the statistical significance of the average direct and cumulative indirect effects estimates for the explanatory variables in the model. These measures of dispersion are computed using Monte Carlo simulations based on the maximum likelihood multivariate normal distribution of the parameters.

To empirically compute these impacts, we first need estimates for ρ and β. Lee (2004) shows that under the set of assumptions reported below, the (quasi-)maximum likelihood estimator is $\sqrt{N}$ convergent and asymptotically normally distributed. Before stating the set of assumptions underlying this model, let us define $\mathbf{S}(\rho) = I_N - \rho W$ and $\mathbf{S}(\rho)^{-1}$ as its inverse.

Assumption 1. *The ε_i, $i = 1, \ldots, N$ are iid with zero mean and constant*

[3]For the scalar measures proposed by LeSage & Pace (2009), it does not matter whether cumulative indirect effects and total effects are computed by row or column sums.

variance σ^2. Its moment $E(|\varepsilon|^{4+\gamma})$ for some $\gamma > 0$ exists.

Assumption 2. *The elements of $\mathbf{X}$ are uniformly bounded constants, $\mathbf{X}$ is of full rank K and $\lim_{N\to\infty}(1/N)\mathbf{X}'\mathbf{X}$ exists and is nonsingular.*

Assumption 3. *$\mathbf{W}$ is a non-stochastic spatial weight matrix with zero diagonals.*

Assumption 4. *The matrix $\mathbf{S}(\rho)$ is invertible for all $\rho \in P$, where P is a compact interval. Besides, the true value of the parameter, ρ_0, is in the interior of P.*

Assumption 5. *$\mathbf{W}$ is uniformly bounded in both row and column sums and $\mathbf{S}(\rho)^{-1}$ is uniformly bounded in both row and column sums, uniformly in $\rho \in P$.*

Assumption 1 provides *iid* regularity conditions for ε_i. Assumption 2 rules out the presence of multicollinearity among covariates and suppose they are bounded. The zero diagonal elements in Assumption 3 helps the interpretation of the spatial effect, as self-influence will be excluded in practice. The parameter space for ρ defined in Assumption 4 ensures that (2) exists. In this paper, as the interaction matrix is assumed row-normalized, the parameter space for ρ is a compact subset of $(-1, 1)$. This statement comes from Lemma 2 of Kelejian & Prucha (2010, p.56). Let us also note that row-normalization is not the unique way to standardize $\mathbf{W}$. As such, some matrix norms like the spectral radius or the minimum between the maximum of row and column sums in absolute value can also be used. Kelejian & Prucha (2010) show that the parameter space for ρ can always be transformed to belong to $(-1, 1)$. However, Assumption 4 further requires the compactness of the parameter space P due to the nonlinearity of ρ in equation (2).

With these assumptions, the log-likelihood function for equation (1) can be written as:

$$\ln L(\beta, \rho, \sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln|\mathbf{S}(\rho)| - \frac{1}{2\sigma^2}\mathbf{V}'\mathbf{V} \tag{6}$$

where $\mathbf{V} = \mathbf{S}(\rho)\mathbf{y} - \iota_N\alpha - \mathbf{X}\beta$ is the disturbance vector of equation (1).

Even though the spatial autoregressive model captures contagion effects, LeSage & Pace (2009) advocate the use of the spatial Durbin model (SDM) since it allows for spatial correlation in the regressors. This model, presented in (7), consists in a SAR specification where the spatially lagged exogenous regressors are added to the original set of regressors.

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \iota_N\alpha + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\gamma + \varepsilon \tag{7}$$

From a pure econometric point of view, this model does not involve any further difficulties compared to the SAR specification. However, the matrix

of partial derivatives of $\mathbf{y}$ with respect to the rth explanatory variable, presented in (8) and computed from the reduced form of model (7), contains the additional term $\mathbf{W}\gamma_r$.

$$\Xi_y^{x_r} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}_r} = (\mathbf{I}_N - \rho\mathbf{W})^{-1} (\mathbf{I}_N\beta_r + \mathbf{W}\gamma_r) \tag{8}$$

Let us finally note that the impact definitions for this spatial Durbin specification are identical to those used in the SAR model above.

The last type of specification considered in this section is the spatial autoregressive static panel data model. Even though it can seem odd to mix panel data and cross-section models together, impacts measures implied by a spatial static panel data model are the same as those in a spatial autoregressive cross-sectional model, as soon as the interaction matrix and the parameters of interest of the former are assumed constant across time. In other words, in the specification (9), which represent a spatial Durbin static panel data model, neither $\mathbf{W}$, neither ρ, β and γ are indexed by t.

$$\mathbf{y}_t = \rho\mathbf{W}\mathbf{y}_t + \mathbf{X}_t\beta + \mathbf{W}\mathbf{X}_t\gamma + \mu + \iota_N\phi_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{9}$$

In this model, which is an extension of (7) to the case where the N individuals are observed during T periods of time, we allow for the presence of a N-dimensional vector of individual effects, μ, that are assumed to be time-invariant and a time effect, ϕ_t, common to all individuals in the model but peculiar to period t. Depending on the assumptions about individual and time effects, model (9) will be estimated using random or fixed effects. The former, more efficient, is adequate when the effects (individual and temporal) are independent from all regressors included in the specification and are traditionally assumed normally distributed. When this hypothesis of independence is rejected, either on the basis of a test statistic (Hausman, Lagrange multiplier (LM) or likelihood ratio (LR)) or from economic insights, the fixed effects specification should be preferred.[4] Even though these two estimation procedures are different, they both consist in first transforming the data (either applying the within operator for the fixed effects or a quasi-within transformation when the random effects estimation is used) and then applying "standard" spatial econometrics techniques on these transformed data to obtain the estimated parameters.

Under the assumption that both $\mathbf{W}$ and all parameters of interest are constant across time, formulas for impacts' computation in these spatial panel data models are identical to those used in the SDM above. The only difference with respect to the cross-sectional SDM is that parameters account

[4] Lee & Yu (2010*e*) derive a LM statistic to test the relevance of the random effects specification based on the between equation while the third chapter of this thesis develops a LR statistic that uses an auxiliary regression to assess the importance of the correlation between individual effects and regressors.

for the presence of individual effects (either as random or fixed effects). The chosen estimation procedure (fixed versus random effects), even though affecting the estimates, thus does not play any role in the computation of the impacts. The second and third chapters of this dissertation propose empirical applications where impacts are computed for a fixed effects and a random effects spatial autoregressive panel data model respectively. We will thus not discuss this type of specification any further but instead turn our attention to the dynamic space-time panel data model.

3 The dynamic space-time panel data model

3.1 Introduction

Our focus in this section is on extending this cross-sectional impacts analysis to the case of dynamic space-time panel data models. These panel data specifications allow us to compute own- and cross-partial derivatives that trace the effects (own-region and other-region) through time and space. Space-time dynamic models produce a situation where a change in the ith observation of the rth explanatory variable at time t will produce contemporaneous and future responses in all regions' dependent variables y_{it+T}, as well as other region future responses y_{jt+T}. This is due to the presence of an time lag (capturing time dependence), a spatial lag (that accounts for spatial dependence) and a cross-product term reflecting the space-time diffusion.

To the best of our knowledge, Parent & LeSage (2010*c*) is the only study dealing with impact coefficients for both space and time. They consider a time-space dynamic model that relates commuting times to highway expenditures. It seems clear that expenditures for an improvement in a single highway segment at time t (say segment i) will improve commuting times for those traveling on this highway segment (say y_{it}).[5] Improvements in the segment i will also produce future benefits of improved travel times to those using segment i ($y_{it+T}, T = 1, \ldots$). Equally important is the fact that commuting times on neighboring roadways will also improve in current and future time periods, which we might denote as: y_{jt} and y_{jt+T} where $j \neq i$. This is because less congestion on one highway segment will improve traffic flow on neighboring segments. It might also be the case that commuters adjust commuting patterns over time to take advantage of the improvements made in highway segment i and their impact on lessening congestion of nearby arteries.

Dynamic space-time panel data models have the ability to quantify these changes which should prove extremely useful in numerous applied modeling

[5] We abstract from the issue of time scale here and assume that measurements are taken over a sufficient period of time (say one year) to allow the improvements to be made in time t and for commuters to travel on the highway segment during some part of the year (time t).

situations. We show that the partial derivatives $\partial \mathbf{y}_t / \partial \mathbf{x}'_{rt}$ for these models take the form of an $N \times N$ matrix for time t and those for the cumulative effects of a change taking place in time t at future time horizon T take the form of a sum of T different $N \times N$ matrices. We derive explicit forms for these as a function of the dynamic space-time panel data model parameter estimates. This allows us to calculate the dynamic responses over time and space that arise from changes in the explanatory variables. In addition to setting forth expressions for the partial derivatives we also propose scalar summary measures for these and take up the issue of efficient calculation of measures of dispersion.

3.2 The specification

Anselin (2001) and Yu et al. (2008) consider a dynamic spatial autoregressive panel model that allows for both time and spatial dependence as well as a cross-product term reflecting spatial dependence at a one-period time lag. We add spatially lagged exogenous variables to the set of covariates, leading to a dynamic spatial Durbin model shown in (10).

$$\begin{aligned} &\mathbf{y}_t = \phi \mathbf{y}_{t-1} + \rho \mathbf{W} \mathbf{y}_t + \theta \mathbf{W} \mathbf{y}_{t-1} + \iota_N \alpha + \mathbf{x}_t \beta + \mathbf{W} \mathbf{x}_t \gamma + \eta_t, \\ &\eta_t = \mu + \varepsilon_t \quad t = 1, \ldots, T. \end{aligned} \tag{10}$$

Notations for this model are identical to those used before with furthermore ϕ defined as the autoregressive time dependence parameter and θ the spatio-temporal diffusion parameter. We assume ε_t is *i.i.d.* across i and t with zero mean and variance $\sigma_\varepsilon^2 \mathbf{I}_N$. The $N \times 1$ column vector μ represents individual effects with $\mu_i \sim N(0, \sigma_\mu^2)$, and it is typically assumed that μ is uncorrelated with ε_t.

In this paper, we use a one way error component to model individual heterogeneity. However, our results would also apply to a (time-space dynamic panel) model with fixed effects such as that from Yu et al. (2008). Parent & LeSage (2010*a*,*c*) propose a general framework for specifying space-time dependence that involves applying space and time filter expressions to the dependent variable vector or the disturbances.

Let $\mathbf{Y}_a = (\mathbf{y}'_0, \ldots, \mathbf{y}'_T)'$, and $\mathbf{A}$ be the $T+1 \times T+1$ time filter matrix shown in (11), which includes the term ψ from the Prais-Winsten transformation for the initial period.

$$\mathbf{A} = \begin{pmatrix} \psi & 0 & \ldots & 0 \\ -\phi & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & -\phi & 1 \end{pmatrix} \tag{11}$$

Specification of ψ, the (1,1) element in $\mathbf{A}$ depends on whether the first period is modeled or assumed to be known. We will not model this but

rather condition on the initial period, since our focus is on interpretation not estimation of these models.[6] Assuming the process is stationary, ψ is given by:

$$\psi = \sqrt{1-\phi^2}, \quad |\phi| < 1. \tag{12}$$

The filter for spatial dependence is defined as a nonsingular matrix $\mathbf{B} = (\mathbf{I}_N - \rho\mathbf{W})$. As already noted, $\mathbf{W}$ defines dependence between the cross-sectional (spatial) observations. We will also assume that $\mathbf{W}$ is row-normalized from a symmetric matrix, so that all eigenvalues (which we denote as $\varpi_i, i = 1, \ldots, N$) are real and less than or equal to one.

The two filter expressions are combined using the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{I}_{N,T+1} - \rho\mathbf{I}_{T+1} \otimes \mathbf{W} - \phi\mathbf{L} \otimes \mathbf{I}_N + (\rho \times \phi)\mathbf{L} \otimes \mathbf{W}. \tag{13}$$

where $\mathbf{L}$ is the $(T+1) \times (T+1)$ matrix time-lag operator. This filter implies a restriction that θ, the parameter associated with spatial effects from the previous period ($\mathbf{L} \otimes \mathbf{W}$) is equal to $-\rho \times \phi$. Parent & LeSage (2010*a*) show that applying this space-time filter to the error terms greatly simplifies estimation and Parent & LeSage (2010*c*) illustrate that interpretation of these models is also simplified by this restriction. The restriction produces a situation where space and time are separable, leading to simplifications in the space-time covariance structure as well as the own- and cross-partial derivatives used to interpret the model. We will have more to say about this later.

We consider the more general case shown in (14), where the simplifying restriction is not imposed, leading to three parameters ϕ, ρ, θ which will be estimated.

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{I}_{N,T+1} - \rho\mathbf{I}_{T+1} \otimes \mathbf{W} - \phi\mathbf{L} \otimes \mathbf{I}_N - \theta\mathbf{L} \otimes \mathbf{W}, \tag{14}$$

Applying the filter to the dependent variable results in a model specification:

$$\begin{aligned}
(\mathbf{A} \otimes \mathbf{B})\mathbf{Y}_a &= \iota_{N,T+1}\alpha + \mathbf{Z}\beta + (\mathbf{I}_{T+1} \otimes \mathbf{W})\mathbf{Z}\gamma + \eta, \\
\eta &\sim N(\mathbf{0}, \tilde{\mathbf{\Omega}}), \\
\tilde{\mathbf{\Omega}} &= \sigma_\mu^2(\mathbf{J}_{T+1} \otimes \mathbf{I}_N) + \sigma_\varepsilon^2\mathbf{I}_{N,T+1}, \\
\mathbf{J}_{T+1} &= \iota_{T+1}\iota'_{T+1},
\end{aligned} \tag{15}$$

where $\mathbf{Z} = (\mathbf{x}'_0, \ldots, \mathbf{x}'_T)'$, and we note that all model parameters are assumed to be constant across time and spatial units.

[6] See Parent & LeSage (2010*b*) for a discussion of issues pertaining to this.

For the case we deal with here where we condition on initial period observations, we work with the new filter $\mathbf{P}$ shown in (16), which corresponds to the filter in (15) where explanatory variable observations for the first time period are deleted:

$$\mathbf{P}_{NT,N(T+1)} = \begin{pmatrix} -(\phi\mathbf{I}_N + \theta\mathbf{W}) & \mathbf{B} & & \mathbf{0} \\ & \ddots & \ddots & \\ \mathbf{0} & & -(\phi\mathbf{I}_N + \theta\mathbf{W}) & \mathbf{B} \end{pmatrix}, \tag{16}$$

which allows us to rewrite the model in terms of: $\mathbf{e} = (\mathbf{PY_a} - \mathbf{X}\beta - (\mathbf{I}_T \otimes \mathbf{W})\mathbf{X}\gamma - \iota_{NT}\alpha)$, with $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_T')'$ so the log-likelihood function of the complete sample size (NT) is given by:[7]

$$\begin{aligned} \ln L_T(\upsilon) &= -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}| + T\sum_{i=1}^{N}\ln[(1-\rho\varpi_i)] - \frac{1}{2}\mathbf{e}'\mathbf{\Omega}^{-1}\mathbf{e}, \\ \mathbf{\Omega} &= (T\sigma_\mu^2 + \sigma_\varepsilon^2)(\bar{\mathbf{J}}_T \otimes \mathbf{I_N}) + \sigma_\varepsilon^{\mathbf{2}}\left[(\mathbf{I_T} - \bar{\mathbf{J}}_\mathbf{T}) \otimes \mathbf{I_N}\right], \\ \bar{\mathbf{J}}_T &= \mathbf{J}_T/T, \end{aligned} \tag{17}$$

where $\upsilon = (\beta', \gamma', \alpha, \sigma_\varepsilon^2, \sigma_\mu^2, \phi, \rho, \theta)$, and $\varpi_i, i = 1, \ldots, N$ represents eigenvalues of the matrix $\mathbf{W}$ which are real and less than or equal to one given our assumptions regarding the row-normalized matrix $\mathbf{W}$.

For this specification, stationary conditions are satisfied only if $|\mathbf{AB}^{-1}| < 1$, which requires the following properties, developed by Parent & LeSage (2010*a*), to be fulfilled:

$$\begin{array}{lll} \phi + (\rho+\theta)\varpi_{max} < 1 & \text{if} & \rho+\theta \geq 0, \\ \phi + (\rho+\theta)\varpi_{min} < 1 & \text{if} & \rho+\theta < 0, \\ \phi - (\rho-\theta)\varpi_{max} > -1 & \text{if} & \rho-\theta \geq 0, \\ \phi - (\rho-\theta)\varpi_{min} > -1 & \text{if} & \rho-\theta < 0, \end{array} \tag{18}$$

where $\varpi_{\min}$ and $\varpi_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{W}$ respectively.

As indicated, we rely on a Bayesian Markov Chain Monte Carlo estimation scheme to produce estimates of the parameters in the model. Complete details can be found in Parent & LeSage (2010*b*), but we make note of one issue that arises here. The priors for the space-time parameters ϕ, ρ and θ should be defined over the stationary interval in (18). A uniform joint prior distribution over this interval does not produce vague marginal priors. Sun & Berger (1998) propose different approaches to define priors

[7] The random effects parameters have been integrated out and we use the decomposition proposed by Wansbeek & Kapteyn (1982) to replace $\mathbf{J}_T$ by its idempotent counterpart.

on a constrained parameter space. Since we are concerned with the parameter vector (ρ, ϕ, θ), a prior can be constructed that takes the form $p(\rho,\ \phi,\ \theta) = p(\rho)\, p(\phi|\rho,\theta)\, p(\theta|\rho)$.

Assuming that the parameter space for ρ is a compact subset of (-1,1), we can define the following conditional prior $p(\phi|\rho,\theta) \sim U(-1+|\rho-\theta|, 1-|\rho+\theta|)$ based on the stationary interval defined in (18).[8] Then focusing only on the parameters θ and ρ it is easy to show that the conditional prior $p(\theta|\rho) \sim U(-1+|\rho|, 1-|\rho|)$. The last prior is therefore $p(\rho) \sim U(-1,1)$. Note that the joint prior is a uniform distribution and equal to 1/2 over the parameter space define by stationary interval (18).

For estimation purposes, we assign a prior distribution $p(\alpha, \theta', \sigma_u^2, \sigma_\varepsilon^2)$ with $\theta' = [\beta', \gamma']$ such that these parameters are a priori independent. Concerning the parameters (α, θ'), we estimate separately the intercept term α and the parameters θ assuming a non-hierarchical prior of the independent Normal-Gamma variety. Thus,

$$\alpha \sim N(\alpha_0, M_\alpha^{-1}), \tag{19}$$

$$\theta \sim N(\theta_0, \mathbf{M}_\theta^{-1}),$$

$$\sigma_\varepsilon^{-2} \sim G(v_0/2, S_0/2), \tag{20}$$

$$\sigma_\mu^{-2} \sim G(v_1/2, S_1/2).$$

We use diffuse priors with prior means α_0 and θ_0 set to zero, the variance parameter M_α^{-1} set to 10^{12} and $\mathbf{M}_\theta^{-1}$ to $10^{12} I_{2K}$. Parameters for the Gamma priors are all set to 0.001. Having the posterior distribution of the explanatory variables θ conditional on the random effects μ is not desirable because these two sets of parameters tend to be highly correlated which can create problems with mixing for the Markov Chain estimation procedure. We use the method proposed by Chib & Carlin (1999) who suggest first sampling β marginalized over μ and then sampling μ conditioned on β. Posterior distributions are standard and can be found in Koop (2003).

3.3 Interpreting the model estimates

Our focus here is on the partial derivative effects associated with a change in the explanatory variables in model (10). This model has own- and cross-partial derivatives that measure the impact on y_{it} that arises from changing the value of the rth explanatory variable at time t in region i. Specifically, $\partial y_{it}/\partial x_{it}^r$, represents the contemporaneous direct effect on region i's dependent variable arising from a change in the rth explanatory variable in region i. There is also a cross-partial derivative $\partial y_{jt}/\partial x_{it}^r$ that measures the contemporaneous spatial spillover effect on region j, $j \neq i$. We reserve the term

[8] This assumption regarding the parameter space is also used in Yu et al. (2008) despite the theoretical possibility that $\varpi_{\min}$ could be less than -1.

spillover to refer to contemporaneous cross-partial derivatives, those that involve the same time period.

We are most interested in partial derivatives that measure how region i's dependent variable responds over time to changes in a given time period of the explanatory variables. These cross-partial derivatives involving different time periods are referred to as *diffusion* effects, since diffusion takes time. The model allows us to calculate partial derivatives that can quantify the magnitude and timing of dependent variable responses in each region at various time horizons $t+T$ to changes in the explanatory variables at time t. Expressions for these are presented and discussed in what follows. We simply note here that we are referring to $\partial y_{it+T}/\partial x_{it}^r$ which measures the $T-$horizon own-region i dependent variable response to changes in own-region explanatory variable r, and $\partial y_{jt+T}/\partial x_{it}^r$, that reflects diffusion effects over time that impact the dependent variable in region $j \neq i$ when region i's explanatory variable r at period t is changed. We distinguish between two different interpretative scenarios, one where the change in explanatory variables represents a permanent or sustained change in the level and the other where we have a transitory (or one-period) change.

We condition on the initial period observation and assume that this period is only subject to spatial dependence. This implies that the dependent variable for the whole sample is written as $\mathbf{Y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$. In this case, the data generating process (DGP) for our model can be expressed by replacing the $NT \times N(T+1)$ space-time filter $\mathbf{P}$ by the $NT \times NT$ matrix $\mathbf{Q}$ as in (21), with $\mathbf{H} = \iota_T \otimes \mathbf{I}_N$, a matrix that assigns the same N random effects to each region for all time periods.

$$\mathbf{Y} = \mathbf{Q}^{-1}[\iota_{NT}\alpha + \mathbf{X}\beta + (\mathbf{I}_T \otimes \mathbf{W})\mathbf{X}\gamma + \mathbf{H}\mu + \varepsilon], \tag{21}$$

$$\mathbf{Y} = \sum_{r=1}^{K} \mathbf{Q}^{-1}(\mathbf{I}_{NT}\beta_r + (\mathbf{I}_T \otimes \mathbf{W})\gamma_r)\mathbf{X}^{(r)} + \mathbf{Q}^{-1}[\iota_{NT}\alpha + \mathbf{H}\mu + \varepsilon], \tag{22}$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{B} & \mathbf{0} & \ldots & & \mathbf{0} \\ \mathbf{C} & \mathbf{B} & & & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \ddots & & \vdots \\ \vdots & & \ddots & & \\ \mathbf{0} & \ldots & & \mathbf{C} & \mathbf{B} \end{pmatrix}, \tag{23}$$

$$\mathbf{C} = -(\phi\mathbf{I}_N + \theta\mathbf{W}),$$

$$\mathbf{B} = (\mathbf{I}_N - \rho\mathbf{W}).$$

In (22) we let $\mathbf{X}^r$ denote the rth column from the $NT \times K$ matrix $\mathbf{X}$, allowing us to express this DGP in a form suitable for considering the partial derivative impacts that arise from changes in the rth explanatory variable.

For future reference we note that the matrix $\mathbf{Q}^{-1}$ takes the form of a lower-triangular block matrix, containing blocks with $N \times N$ matrices.

$$\begin{aligned}
\mathbf{Q}^{-1} &= \begin{pmatrix} \mathbf{B}^{-1} & \mathbf{0} & \dots & & \mathbf{0} \\ \mathbf{D}_1 & & & & \vdots \\ \mathbf{D}_2 & \mathbf{D}_1 & \ddots & & \\ \vdots & & \ddots & & \mathbf{0} \\ \mathbf{D}_{T-1} & \mathbf{D}_{T-2} & \dots & \mathbf{D}_1 & \mathbf{B}^{-1} \end{pmatrix}, \\
\mathbf{D}_s &= (-1)^s (\mathbf{B}^{-1}\mathbf{C})^s \mathbf{B}^{-1}, \quad s = 0, \dots, T-1.
\end{aligned} \tag{24}$$

One implication of this is that we need only calculate $\mathbf{C}$ and $\mathbf{B}^{-1}$ to analyze the partial derivative impacts for any time horizon T. This means we can use a panel involving say 10 years to analyze the cumulative impacts arising from a permanent (or transitory) change in explanatory variables at any time t extending to future horizons $t + T$.

The one-period-ahead impact of a *permanent* change in the rth variable at time t is:

$$\partial \mathbf{Y}_{t+1} / \partial \mathbf{X}^{r'} = \left(\mathbf{D}_1 + \mathbf{B}^{-1}\right) \left[\mathbf{I}_N \beta_r + \mathbf{W} \gamma_r\right] \tag{25}$$

By a permanent change at time t we mean that: $\partial \mathbf{X}^{r'} = (\mathbf{x}_t + \delta, \mathbf{x}_{t+1} + \delta, \dots, \mathbf{x}_T + \delta)$, so the values increase to a new level and remain there in future time periods. More generally, the T-period-ahead (cumulative) impact arising from a permanent change at time t in the rth variable takes the form in (26). Note that we are cumulating down the columns (or rows) of the matrix in (24).

$$\partial \mathbf{Y}_{t+T} / \partial \mathbf{X}^{r'} = \sum_{s=0}^{T} \mathbf{D}_s [\mathbf{I}_N \beta_r + \mathbf{W} \gamma_r]. \tag{26}$$

By analogy to LeSage & Pace (2009), the main diagonal elements of the $N \times N$ matrix sums in (26) for time horizon T represent (cumulative) own-region impacts that arise from both time and spatial dependence. The sum of off-diagonal elements of this matrix reflect both spillovers measuring contemporaneous cross-partial derivatives and diffusion measuring cross-partial derivatives that involve different time periods. We note that it is not possible to separate out the time dependence from spillover and diffusion effects in this model. By this we mean that the matrix product involving the time filter $\mathbf{C}$ and space filter $\mathbf{B}$ are not separable in the expression for the cross-partial derivatives.

Of course, the $T-$horizon impulse response to a *transitory* change in the rth explanatory variable at time t would be given by the main- and off-diagonal elements of:

$$\begin{aligned} \partial \mathbf{Y}_{t+T}/\partial \mathbf{x}_t^{r'} &= \mathbf{D}_T[\mathbf{I}_N \beta_r + \mathbf{W}\gamma_r], \\ \mathbf{D}_T &= (-1)^T (\mathbf{B}^{-1}\mathbf{C})^T \mathbf{B}^{-1}. \end{aligned} \tag{27}$$

We note that (27) also corresponds to the marginal effect in period $t+T$ of a permanent change in the rth explanatory variable in time t.

A special case of the model and associated effects estimates was considered by Parent & LeSage (2010*c*) where the restriction $\theta = -\phi\rho$ holds. This allows the matrix $\mathbf{Q}^{-1}$ to be expressed as:

$$\begin{aligned} \mathbf{R}^{-1} &= \begin{pmatrix} \mathbf{B}^{-1} & \mathbf{0} & \dots & & \mathbf{0} \\ \mathbf{E}_1 & \ddots & & & \vdots \\ \mathbf{E}_2 & \mathbf{E}_1 & \ddots & & \\ \vdots & & \ddots & \ddots & \mathbf{0} \\ \mathbf{E}_{T-1} & \mathbf{E}_{T-2} & \dots & \mathbf{E}_1 & \mathbf{B}^{-1} \end{pmatrix}, \\ \mathbf{E}_s &= \phi^s \times \mathbf{B}^{-1}, s = 0, \dots, T-1. \end{aligned} \tag{28}$$

In this case, we have simple geometric decay over time periods of the spatial spillover (contemporaneous) effects captured by the matrix $\mathbf{B}^{-1}$. The computationally efficient approach to calculating the effects for cross-sectional spatial regression models described in LeSage & Pace (2009) can be used in conjunction with a scalar weighting term: $\phi^s, s = 0, \dots, T-1$. It should be clear that the time filter matrix $\mathbf{C}$ from the unrestricted model collapses to the scalar expression ϕ^s, which allows us to separate out the time dependence and spatial dependence contributions to the own- and cross-partial derivatives for this model. We say that space and time are separable in this model specification.

In any application of the model it is possible to test if the restriction $\theta = -\phi\rho$ holds, which suggests that the sample data is consistent with a model based on space-time separability. We illustrate this in our application in the next section.

4 Application to state-level smoking behavior

We use a panel consisting of 45 (of the lower 48) states plus the District of Columbia covering 30 years from 1963-1992 taken from Baltagi & Li (2004). The model is a simple (logged) demand equation for (packs of) cigarettes

as a function of the (logged) cigarette prices (per pack) and (logged) state-level income per capita.[9] We have observations for 30 years on (logged) real per capita sales of cigarettes measured in packs per person aged 14 years or older (the dependent variable). The two explanatory variables are the (logged) average retail price of a pack of cigarettes and (logged) real per capita disposable income in each state and time period.

Their motivation for spatial dependence (in their model disturbances) was a bootlegging effect where buyers of cigarettes near state borders purchase in neighboring states if there is a price advantage to doing so. They however did not allow for time dependence in the model disturbances. Baltagi & Levin (1986) use a panel covering the period from 1963 to 1980 to estimate a non-spatial dynamic demand equation for cigarettes and find a significant negative price elasticity of -0.2 but no significant income elasticity. This model accounted for the bootlegging effect by incorporating the lowest price for cigarettes from neighboring states as an explanatory variable. Bootlegging was found to be statistically significant.

We constructed a spatial weight matrix based on the state border miles in common between our sample of states. This was row-normalized to produce spatial lags $\mathbf{Wy}$ reflecting a linear combination of cigarettes sales from neighboring states weighted by the length of common borders. Given the cross-border shopping (bootlegging) motivation for spatial spillovers, this type of spatial weight matrix seemed intuitively appealing. It is interesting to note that the estimates and inferences did not change when a first-order contiguity weight matrix was used that assigned equal weight to all contiguous states (those with borders touching).

We report estimates for the model parameters in Table 1 based on 200,000 MCMC draws with the first 100,000 discarded to account for burn-in of the sampler. The table reports the posterior mean as well as lower 0.01 and 0.05 and upper 0.95 and 0.99 percentiles constructed using the retained draws.[10] Large variances were assigned to the prior distributions so these estimates should reflect mostly sample data information and be roughly equivalent to those from maximum likelihood estimation.[11]

The estimates for the parameters of the space-time filter indicate strong time dependence using the 0.01 and 0.99 intervals and weaker spatial dependence whose 0.01 and 0.99 intervals point to positive dependence. The cross product term θ is negative and the 0.01 and 0.99 intervals point to a difference from zero. We report the posterior distribution for the product $-\phi\rho$ constructed using the draws from the MCMC sampler. This distribution appears consistent with the restriction that can be used to simplify the model along with the effects estimates. It appears the sample data and model are

[9] Colorado, North Carolina and Oregon are the three missing states.

[10] Every tenth draw from the 100,000 retained draws was used to construct the posterior estimates reported in the tables to reduce serial dependence in the sampled values.

[11] This was checked and found to be the case.

Table 1: Dynamic space-time model parameter estimates

Parameters	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
ϕ	0.7940	0.7994	0.8287	0.8589	0.8632
ρ	0.2017	0.2048	0.2425	0.2697	0.2747
θ	-0.2214	-0.2180	-0.1913	-0.1637	-0.1600
$-\phi\rho$	-0.2267	-0.2243	-0.2009	-0.1695	-0.1679
σ^2_μ	0.0007	0.0008	0.0013	0.0020	0.0023
σ^2_ε	0.0012	0.0012	0.0013	0.0014	0.0015
Variables	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
price	-0.3582	-0.3445	-0.2988	-0.2537	-0.2400
income	0.0257	0.0416	0.0912	0.1416	0.1579
$W\times$ price	0.1160	0.1305	0.1777	0.2258	0.2402
$W\times$ income	-0.0753	-0.0584	-0.0061	0.0454	0.0617

consistent with space-time separability.

The coefficients associated with price, income and their spatial lags cannot be directly interpreted as if they were partial derivatives that measure the response of the dependent variable to changes in the regressors. As already shown, the partial derivatives take the form of $N \times N$ matrices for each time horizon and are non-linear functions of these coefficient estimates and the space-time filter parameters.

Table 2 shows the average direct effect estimates for the contemporaneous time period out to a time horizon T of 29 years.[12]

Except for the first row of both panels that show pure feedbacks effects, these effects should capture mostly impacts arising from time dependence of region i on changes in its own explanatory variables plus some of the feedback loop (spatial) effects, which will be fed forward in time. Since all variables in the model have been log-transformed, we can interpret our direct, indirect and total effects estimates in elasticity terms. The table reports the posterior mean of the period-by-period effects along with credible intervals for these constructed from the MCMC draws. The column labeled 'Cumulative' shows the cumulation of these period-by-period effects that would reflect the time horizon $t + T$ response to a permanent change in the explanatory variables at time t. Since our estimates for the the space-time filter parameters are consistent with model stability (the sum of the spatial filter parameters being less than one), we will see the (period-by-period) direct effects die down to

[12]The general expressions in (25) were used to produce these effects estimates despite the fact that the space-time separability restriction appears consistent with the model and data. These expressions collapse (approximately) to the simpler expressions in (28) in this case.

Table 2: Space-time average direct effect estimates

	Price (elasticity)					
Horizon T	Cumulative	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
0	-0.2906	-0.3467	-0.3337	-0.2906	-0.2478	-0.2348
1	-0.5311	-0.2811	-0.2716	-0.2405	-0.2087	-0.1985
2	-0.7302	-0.2313	-0.2237	-0.1991	-0.1744	-0.1665
3	-0.8952	-0.1936	-0.1866	-0.1649	-0.1442	-0.1380
4	-1.0319	-0.1636	-0.1571	-0.1367	-0.1180	-0.1127
5	-1.1453	-0.1393	-0.1330	-0.1133	-0.0958	-0.0914
6	-1.2393	-0.1188	-0.1130	-0.0940	-0.0775	-0.0735
7	-1.3174	-0.1016	-0.0962	-0.0780	-0.0624	-0.0590
8	-1.3822	-0.0871	-0.0820	-0.0648	-0.0502	-0.0472
9	-1.4361	-0.0748	-0.0700	-0.0538	-0.0404	-0.0378
10	-1.4808	-0.0643	-0.0598	-0.0447	-0.0325	-0.0303
15	-1.6142	-0.0305	-0.0275	-0.0178	-0.0108	-0.0099
20	-1.6679	-0.0145	-0.0127	-0.0072	-0.0036	-0.0032
25	-1.6897	-0.0069	-0.0059	-0.0029	-0.0012	-0.0010
29	-1.6975	-0.0038	-0.0032	-0.0014	-0.0005	-0.0004
	Income (elasticity)					
Horizon T	Cumulative	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
0	0.0924	0.0305	0.0453	0.0924	0.1401	0.1551
1	0.1689	0.0257	0.0381	0.0765	0.1148	0.1266
2	0.2324	0.0216	0.0319	0.0634	0.0946	0.1042
3	0.2850	0.0182	0.0267	0.0526	0.0782	0.0859
4	0.3287	0.0153	0.0224	0.0436	0.0649	0.0714
5	0.3650	0.0128	0.0187	0.0362	0.0541	0.0598
6	0.3951	0.0108	0.0155	0.0301	0.0452	0.0503
7	0.4201	0.0090	0.0129	0.0250	0.0378	0.0424
8	0.4409	0.0075	0.0107	0.0208	0.0318	0.0359
9	0.4582	0.0063	0.0089	0.0173	0.0268	0.0305
10	0.4726	0.0053	0.0074	0.0144	0.0227	0.0259
15	0.5157	0.0020	0.0028	0.0057	0.0099	0.0116
20	0.5331	0.0008	0.0010	0.0023	0.0044	0.0053
25	0.5402	0.0002	0.0003	0.0009	0.0020	0.0025
29	0.5427	0.0001	0.0001	0.0004	0.0010	0.0013

zero over time.

Consistent with microeconomic theory we see a greater long-run elasticity response of cigarettes sales to both price and income.[13] The direct effect period 0 price elasticity estimate of -0.29 is consistent with the estimate of -0.2 from Baltagi & Levin (1986). The high level of time dependence in the estimate for ϕ leads to a much more responsive long-run price elasticity of -1.69. This would be close to the long-run value, since at a time horizon

[13]Since it takes time for people to adjust behavior in response to price and income changes, the long-run elasticity is larger than short-run.

of 29 years, the period-by-period effects appear to have nearly died down to zero (the upper 0.99 interval value is -0.0004). Similarly for the income elasticity we see a period zero value of 0.09 and a thirty-year horizon value of 0.54, where again this is close to the long-run elasticity (since the lower 0.01 interval value is 0.0001 at the $T = 29$ horizon).

These results suggest that a 10 percent increase in (per pack) cigarette prices would lead to a short-run decrease in sales (of packs per capita) by 3 percent, but a long-run decrease in sales of 17 percent. A price elasticity of demand less than one would lead to an increase in tax revenue for states that raised cigarette taxes, and this appears to be the case for the initial three year time horizon. Beginning in the fourth and subsequent years following a tax increase, the elastic response of cigarettes sales would lead to a decrease in state tax revenue from cigarettes.

Since the income elasticity is positive, increases in state-level per capita income leads to increased sales of cigarettes.[14] In the short-run a 10 percent increase in income leads to a 1 percent increase in cigarette sales, whereas in the long-run sales are more responsive showing a 5 to 6 percent increase.

Table 3 shows the average total indirect effects in a format identical to that of Table 2. These effects represent contemporaneous spatial spillovers plus diffusion that takes place over time. The magnitude of these effects is likely to be small since the estimate for the spatial dependence parameter ρ was small. Following Baltagi & Levin (1986), one motivation for the presence of spatial spillover and diffusion effects is the bootlegging phenomena where buyers of cigarettes living near state borders purchase these at lower prices when possible.

The average total indirect effects for price are positive and different from zero up to a time horizon of 9 years using the 0.01 and 0.99 intervals. The positive sign is consistent with bootlegging since the (scalar summary) indirect effects estimates tell us that a positive change in own-state prices will lead to increased cigarette sales in all other states. Since the marginal or period-by-period positive spillover effects die down to zero by year $T = 10$, where the cumulative effects take a value of 0.62, we can conclude that bootlegging serves to offset a substantial portion of the cumulative negative own-price elasticity effect of -1.48 that we see for year 10. A 10 percent permanent increase in own-state cigarette prices would lead to around 7 percent long-run increase in bootleg sales from neighboring states. The cumulative spillover/bootlegging impact is around 0.7 which in conjunction with the negative cumulative direct price impact of -1.7 suggests a long-run total impact from price changes that would be close to unit-elastic. This means a 10 percent increase in price would lead to a 10 percent decrease in cigarette sales. Ignoring the spatial spillover/bootlegging impact would lead to an overestimate of the sensitivity of sales to price changes.

[14] Economists label commodities having positive income elasticities *normal goods*.

Table 3: Space-time average total indirect effect estimates

	Price spillover (elasticity)					
Horizon T	Cumulative	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
0	0.1308	0.0723	0.0859	0.1308	0.1757	0.1893
1	0.2370	0.0613	0.0723	0.1062	0.1398	0.1500
2	0.3232	0.0508	0.0595	0.0861	0.1133	0.1216
3	0.3932	0.0405	0.0472	0.0699	0.0934	0.1007
4	0.4499	0.0303	0.0362	0.0567	0.0781	0.0847
5	0.4959	0.0215	0.0268	0.0460	0.0658	0.0718
6	0.5333	0.0144	0.0190	0.0373	0.0557	0.0612
7	0.5635	0.0090	0.0129	0.0302	0.0474	0.0524
8	0.5881	0.0049	0.0083	0.0245	0.0404	0.0451
9	0.6079	0.0019	0.0048	0.0198	0.0346	0.0388
10	0.6240	-0.0001	0.0022	0.0160	0.0297	0.0334
15	0.6684	-0.0037	-0.0026	0.0055	0.0140	0.0163
20	0.6834	-0.0037	-0.0026	0.0018	0.0067	0.0081
25	0.6883	-0.0031	-0.0019	0.0005	0.0032	0.0040
29	0.6896	-0.0024	-0.0013	0.0002	0.0017	0.0023
	Income spillover (elasticity)					
Horizon T	Cumulative	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
0	0.0198	-0.0461	-0.0299	0.0198	0.0692	0.0850
1	0.0377	-0.0355	-0.0223	0.0178	0.0579	0.0708
2	0.0535	-0.0273	-0.0167	0.0158	0.0485	0.0593
3	0.0677	-0.0213	-0.0126	0.0141	0.0409	0.0499
4	0.0802	-0.0169	-0.0097	0.0125	0.0347	0.0422
5	0.0912	-0.0136	-0.0077	0.0110	0.0297	0.0359
6	0.1009	-0.0113	-0.0062	0.0097	0.0255	0.0306
7	0.1095	-0.0094	-0.0050	0.0085	0.0220	0.0263
8	0.1170	-0.0080	-0.0042	0.0074	0.0191	0.0227
9	0.1235	-0.0068	-0.0035	0.0065	0.0166	0.0200
10	0.1293	-0.0058	-0.0030	0.0057	0.0145	0.0175
15	0.1487	-0.0028	-0.0015	0.0028	0.0079	0.0101
20	0.1584	-0.0014	-0.0007	0.0014	0.0045	0.0061
25	0.1631	-0.0006	-0.0003	0.0007	0.0025	0.0036
29	0.1651	-0.0003	-0.0002	0.0003	0.0016	0.0024

Turning to the indirect effects for the income variable, these are small and not different from zero based on the 0.01 and 0.99 credible intervals. This suggests that increases in state-level income do not exert an influence on bootlegging behavior.

One point to note regarding our dynamic space-time model compared to models that deal with space and time dependence in the disturbances is that we have an explicit measure of spatio-temporal spillovers. The scalar summary effects estimates we propose here can be used to produce a quantitative assessment of the magnitude, timing and statistical significance of these spillovers.

A second point is that in the general space-time dynamic model considered here, the restriction $-\phi\rho = \theta$ is not imposed. This implies that except from the contemporaneous effects that represent pure spatial effects, future time horizons contain both time and space diffusion effects, which cannot be distinguished from each other. As noted by Parent & LeSage (2010*c*) when this restriction is consistent with the model and sample data, it is possible to separate out spatial, temporal and diffusion impact magnitudes.

Table 4 reports the average total effect/impact estimates in a format identical to that used for Tables 2 and 3. These effects are the sum of the direct and indirect effects, so they reflect the long-run elasticity associated with the price and income variables from the broader perspective of society at large. Individual state leaders or policy makers would be interested in the direct effects on cigarette sales from changes in own-state prices and incomes. The bootlegging spillovers impacting individual states are likely to be small and of little consequence. However, from the broader perspective of national policy makers the (cumulative) total effects estimates would be the relevant estimates for national policy purposes.

The total effects for both price and income are different from zero at all 29 time horizons reported in the table. However, the marginal effects die down to nearly zero based on an examination of the 0.01 and 0.99 interval magnitudes for the horizon $T = 29$.

The negative direct effect (elasticity) of -1.7 from changes in price are offset somewhat by the positive effect of 0.7 on cigarette sales from bootlegging, leading to a total effect long-run elasticity of -1.0. Of course, this represents a much more elastic long-run relationship relative to the short-run elasticity around -0.16. A similar result occurs for the income elasticity where we see the short-run elasticity of 0.11 increased to 0.70 over time.

5 Conclusion

This paper first presented an overview of interpretation of effects of changes in explanatory variables in cross-sectional and static panel data models. We then have extended the approach taken by LeSage & Pace (2009) for mea-

Table 4: Space-time average total effect estimates

	Price total (elasticity)					
Horizon T	Cumulative	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
0	-0.1597	-0.1964	-0.1887	-0.1597	-0.1312	-0.1236
1	-0.2940	-0.1597	-0.1542	-0.1343	-0.1137	-0.1072
2	-0.4070	-0.1310	-0.1273	-0.1129	-0.0977	-0.0925
3	-0.5020	-0.1100	-0.1065	-0.0950	-0.0828	-0.0790
4	-0.5819	-0.0942	-0.0906	-0.0799	-0.0693	-0.0665
5	-0.6493	-0.0815	-0.0780	-0.0673	-0.0573	-0.0548
6	-0.7060	-0.0709	-0.0678	-0.0567	-0.0470	-0.0447
7	-0.7538	-0.0623	-0.0591	-0.0478	-0.0384	-0.0362
8	-0.7941	-0.0550	-0.0517	-0.0403	-0.0312	-0.0291
9	-0.8281	-0.0487	-0.0454	-0.0340	-0.0254	-0.0234
10	-0.8568	-0.0432	-0.0399	-0.0286	-0.0206	-0.0188
15	-0.9458	-0.0242	-0.0210	-0.0123	-0.0071	-0.0062
20	-0.9844	-0.0136	-0.0112	-0.0053	-0.0024	-0.0020
25	-1.0013	-0.0077	-0.0059	-0.0023	-0.0008	-0.0006
29	-1.0078	-0.0049	-0.0036	-0.0012	-0.0003	-0.0002
	Income total (elasticity)					
Horizon T	Cumulative	lower 0.01	lower 0.05	mean	upper 0.95	upper 0.99
0	0.1123	0.0824	0.0892	0.1123	0.1361	0.1420
1	0.2066	0.0713	0.0771	0.0943	0.1111	0.1155
2	0.2860	0.0614	0.0662	0.0793	0.0918	0.0947
3	0.3528	0.0526	0.0562	0.0667	0.0765	0.0794
4	0.4089	0.0448	0.0471	0.0561	0.0647	0.0679
5	0.4562	0.0374	0.0392	0.0472	0.0555	0.0585
6	0.4961	0.0307	0.0324	0.0398	0.0482	0.0506
7	0.5296	0.0250	0.0266	0.0335	0.0419	0.0441
8	0.5579	0.0202	0.0217	0.0283	0.0366	0.0387
9	0.5818	0.0163	0.0177	0.0238	0.0320	0.0341
10	0.6020	0.0131	0.0144	0.0201	0.0280	0.0302
15	0.6644	0.0043	0.0050	0.0086	0.0147	0.0167
20	0.6915	0.0014	0.0017	0.0037	0.0078	0.0094
25	0.7034	0.0004	0.0006	0.0016	0.0041	0.0053
29	0.7079	0.0001	0.0002	0.0008	0.0025	0.0034

suring own- and cross-partial derivative impacts for (cross-sectional) spatial regression models to the case of dynamic space-time panel data models. They propose scalar summary measures along with measures of dispersion for these that allow the $N \times N$ matrices of impacts for each explanatory variable in the model to be summarized. Their approach is consistent with treatment of regression coefficient estimates where we view these as reflecting how changes in the explanatory variables impact the dependent variable on average over the sample. The extension results in a series of $N \times N$ matrix products for future horizons that can be cumulated to measure the dependent variable response over any time horizon. We follow LeSage & Pace (2009) and produce scalar summary measures using averages of the main diagonal elements of the sequence of $N \times N$ matrices for direct or own-partial derivatives and averages of the cumulated off-diagonal elements for the cross-partials.

A re-examination of the 30 year space-time panel data set on state-level cigarette sales, prices and income from Baltagi & Levin (1986) demonstrated the usefulness of our dynamic space-time elasticities/responses. In particular, we are able to capture spillovers attributed to bootlegging as part of the model. We found that over the period 1963 to 1992 positive spatial spillovers attributed to bootlegging reduced the short-run price elasticity of sales response from -0.29 to -0.13, and the long-run price elasticity of sales response from -1.7 to around -1.0. Spatial spillovers played no significant role in affecting the income elasticity, which exhibited a short-run elasticity of 0.11 and a long-run elasticity of 0.70.

References

Anselin, L. (2001), Spatial econometrics, *in* B. H. Baltagi, ed., 'A companion to theoretical econometrics', Blackwell Publishers Ltd, pp. 310–330.

Anselin, L., Le Gallo, J. & Jayet, H. (2007), Spatial panel econometrics, *in* L. Matyas & P. Sevestre, eds, 'The Econometrics of Panel Data', Kluwer Academic Publishers, Dordrecht.

Baltagi, B. H. & Levin, D. (1986), 'Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered', *The Review of Economics and Statistics* **68**, 148–155.

Baltagi, B. H. & Li, D. (2004), Predicttion in the panel data model with spatial correlation, *in* L. Anselin, R. J. Florax & S. Rey, eds, 'Advances in spatial econometrics: Methodology, tools and applications', Springer, pp. 283–295.

Chib, S. & Carlin, B. P. (1999), 'On mcmc sampling in hierarchical longitunal models', *Statistics and Computing* **9**, 17–26.

Elhorst, P. J. (2003), 'Specification and estimation of spatial panel data models', *International Regional Science Review* **26**, 244–268.

Kakamu, K. (2009), 'Small sample properties and model choice in spatial models: a bayesian approach', *Far East Journal of Applied Mathematics* **34**, 31–56.

Kakamu, K. & Wago, H. (2008), 'Small-sample properties of panel spatial autoregressive models: Comparison of the bayesian and maximum likelihood methods', *Spatial Economic Analysis* **3**, 305–319.

Kapoor, M., Kelejian, H. H. & Prucha, I. R. (2007), 'Panel data models with spatially correlated error components', *Journal of Econometrics* **140**, 97–130.

Kelejian, H. H. & Prucha, I. R. (1998), 'A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics* **17**, 99–121.

Kelejian, H. H. & Prucha, I. R. (1999), 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review* **40**, 509–533.

Kelejian, H. H. & Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.

Kelejian, H. H., Tavlas, G. & Hondroyiannis, G. (2006), 'A spatial modelling approach to contagion among emerging economies', *Open Economies Review* **17**, 423–441.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley and Sons, West Sussex.

Lee, L.-F. (2004), 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models', *Econometrica* **72**, 1899–1925.

Lee, L.-F. (2007), 'Gmm and 2sls estimation of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **137**, 489–514.

Lee, L.-F. & Yu, J. (2010*a*), Efficient gmm estimation of spatial dynamic panel data models with fixed effects, Working paper.

Lee, L.-F. & Yu, J. (2010*b*), 'Estimation of spatial autoregressive panel data models with fixed effects', *Journal of Econometrics* **154**, 165–185.

Lee, L.-F. & Yu, J. (2010*c*), 'Some recent developments in spatial panel data models', *Regional Science and Urban Economics* **40**, 255–271.

Lee, L.-F. & Yu, J. (2010*d*), 'A spatial dynamic panel data model with both time and individual fixed effects', *Econometric Theory* **26**, 564–597.

Lee, L.-F. & Yu, J. (2010*e*), Spatial panels: random components vs. fixed effects, Working paper.

Lesage, J. (1997), 'Bayesian estimation of spatial autoregressive models', *International Regional Science Review* **20**, 113–129.

LeSage, J. & Pace, K. R. (2009), *Introduction to Spatial Econometrics*, CRC Press Taylor and Francis Group, New York.

Lin, X. & Lee, L.-F. (2010), 'Gmm estimation of spatial autoregressive models with unknown heteroskedasticity', *Journal of Econometrics* **157**, 34–52.

Mutl, J. & Pfaffermayr, M. (2011), 'The hausman test in a cliff and ord panel model', *Econometrics Journal* **14**, 48–76.

Parent, O. & LeSage, J. P. (2010*a*), 'A space-time filter for panel data models containing random effects', *Computational Statistics and Data Analysis* . Forthcoming.

Parent, O. & LeSage, J. P. (2010*b*), Spatial dynamic panel data models with random effects, Technical report.

Parent, O. & LeSage, J. P. (2010*c*), 'A spatial dynamic panel model with random effects applied to commuting times', *Transportation Research Part B* **44**, 633–645.

Sun, D. & Berger, J. (1998), 'Reference priors with partial information', *Biometrika* **85**, 55–71.

Wansbeek, T. & Kapteyn, A. (1982), 'A simple way to obtain the spectral decomposition of variance components models for balanced data', *Communications in Statistics Part A -Theory Methods* **11**, 2105–2112.

Yu, J., de Jong, R. & Lee, L.-F. (2008), 'Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large', *Journal of Econometrics* **146**, 118–134.