10-25-2012

Estimation of Spatial Models with Endogenous Weighting Matrices and an Application to a Demand Model for Cigarettes

Harry H. Kelejian

Gianfranco Piras
Gianfranco.Piras@mail.wvu.edu

Estimation of Spatial Models with Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes

Harry H. Kelejian
University of Maryland
College Park, MD

Gianfranco Piras
West Virginia University
Morgantown, WV

October 25, 2012

Abstract

Weighting matrices are typically assumed to be exogenous. However, in many cases this exogeneity assumption may not be reasonable. In these cases, typical model specifications and corresponding estimation procedures will no longer be valid. In this paper we specify a spatial panel data model which contains a spatially lagged dependent variable in terms of an endogenous weighting matrix.

We suggest an estimator for the regression parameters, and demonstrate its consistency and asymptotic normality. We also suggest an estimator for the large sample variance-covariance matrix of that distribution.

We then apply our results to a an interstate panel data cigarette demand model which contains an endogenous weighting matrix. Among other things, our results suggest that, if properly accounted for, the bootlegging effect of buyers, or "agents" for them, crossing state borders to purchase cigarette turns out to be positive and significant.

1 Introduction

A broad range of suggestions have been offered in the literature concerning the specification of the spatial weighting matrix. Historically, many of these suggested specifications are based on geographic criteria. For example, units are often defined to be "neighbors" if they are within a certain geographic distance of each other, or if they are in the same region, or if they have a common border, etc. In such cases, model estimation is simplified because the weighting matrix can be taken to be exogenous rather than endogenous.[1] Because of these simplifications, researchers rarely consider the endogeneity of their weighting matrix even when its endogeneity is evident! For example, after discussing a variety of issues associated with the elements of the weighting matrix, Anselin and Bera (1998) state that "in the standard estimation and testing approaches, the weights matrix is taken to be exogenous." They go on to suggest that if the elements of the weighting matrix involve socioeconomic indicators, the "weights should be chosen with great care to ensure their exogeneity, unless their endogeneity is considered explicitly in the model specification". Although not suggested by Anselin and Bera (1998), their statements might have lead some researchers to purposefully misspecify, and estimate, their model by either ignoring the (obvious) endogeneity of their weighting matrix, or by selecting an inappropriate matrix which can be comfortably viewed as exogenous.

There are many studies which involve weighting matrices for which the assumption of exogeneity may not be reasonable. For example, in estimating a cost-function model, Cohen and Morrison Paul (2004) specify the elements of the weighting matrix in terms of the share of the value of goods shipped from a state. In a study relating to the diffusion of knowledge spillover, Parent and LeSage (2008) specified a weighting matrix based on a technological proximity index. In the context of a growth model, Conley and Ligon (2002) use an economic definition based on transport costs. Similarly, Conley and Topa (2002) use a socio-economic distance based on social networks to study the spatial patterns of unemployment in Chicago. In a recent paper Behrens et al. (2012) point out that connectivity in terms of geographical proximity

[1] Among others, studies based on an exogenous weighting matrix are Baltagi et al. (2007), Kelejian and Prucha (1998), Kelejian and Prucha (1999), Kelejian and Prucha (2001), Kelejian and Prucha (2004), Kapoor et al. (2007), Lee (2004), Piras (2011), Rey and Boarnet (2004). For a spatial model involving a weighting matrix which is exogenous but in a nonparameteric framework, see Pinske et al. (2002).

is narrow. In deriving a quantity-based structural gravity equation system, they adopt a broader definition in terms of a similarity measure that is based on the relative size of regions as reflected by population share.[2] In a study relating to budget spillovers and fiscal policy interdependence, Case et al. (1993) stress that states that are economically similar, are more likely to have an effect on each other than states that simply share a common border. In the context of the above studies, many of the weighting matrices are likely to be endogenous.

As will become clear, if a weighting matrix in a spatial model is endogenous, typical model specifications will no longer be appropriate. Similarly, regression parameter estimators which do not account for the endogeneity of the weighting matrix will typically be inconsistent.

In this paper we specify a spatial panel data model which contains a spatially lagged dependent variable in terms of an endogenous weighting matrix. We consider model specification issues, and suggest an estimator for the regression parameters. We demonstrate the consistency and asymptotic normality of our estimator. We also point out a number of subtle issues that may not be obvious to all readers.

Our results are presented in terms of a fixed effects panel model. However, we demonstrate that minor modifications extend our results to a random effects panel data model.

We apply our theoretical results to a dynamic model for the demand for cigarettes. In our model, the elements of the weighting matrix depend upon cigarette price ratios which we view as endogenous. The model presented in this paper can be viewed as a variant of the model put forth by Baltagi and Levin (1986); however, their model did not contain spatial lags, and therefore did not contain an "endogeneity" problem.

As an overview of our results, we find that cigarette consumption per capita in a given state is negatively effected by the price of cigarettes in that

[2]Trade share variables were used to formulate a weighting matrix in a study on exchange market contagion by Kelejian et al. (2006) and in a study relating to GDP fluctuations by Mukerji (2009). Both the Mukerji (2009), and the Kelejian et al. (2006) studies recognized the endogeneity of the weighting matrix but they were applied, and did not contain formal estimation, or Monte Carlo results. Yet another study involving an endogenous weighting matrix was the one by Hondroyianis et al. (2012). Their study related to contagion issues in the bond markets of Europe. Their weighting matrix was formulated in terms of the differences of GPDs and national debts between the countries. However, in their study, the endogeneity of the weighting matrix was not accounted for.

state, and positively effected by the prices of cigarettes in neighboring states if those prices are lower than in the given state. This could be the result if consumers in a given state cross borders to capitalize on such lower prices, or if bootleggers purchase cigarettes in lower priced neighboring areas and then sell them to consumers in higher priced areas. As "somewhat" expected, we also find that cigarette consumption per capita in a given state is positively effected by the income level of that state.[3]

The paper is organized as follows. Section 2 presents the model and discusses some of the issues related to the error specification. In particular, we show that when the spatial weighting matrix is taken to be endogenous, typical models of spatially correlated disturbance terms would have a major shortcoming. In that section we specify our disturbance term in such a way that it does not have such "major shortcomings"; we also argue that, asymptotically, our disturbance specification is quite general. The model assumptions are presented and discussed in Section 3, while Section 4 describes the estimation procedure. The dynamic demand for cigarettes is introduced and estimated in Section 5. Conclusions and suggestions for further research are given in Section 6. Technical details are relegated to the appendix.

2 The Model

Consider the model

$$\begin{aligned} y_N(t) &= \mu + X_N(t)\beta + \lambda W_N(t) y_N(t) + Y_N(t)\gamma + u_N(t) \qquad (1) \\ &\equiv Z_N(t)\delta + \mu + u_N(t) \\ Z_N(t) &= [X_N(t), W_N(t)y_N(t), Y_N(t)];\ \delta' = (\beta', \lambda, \gamma') \\ t &= 1, ..., T;\ N > 1 \end{aligned}$$

where $y_N(t)$ is an $N \times 1$ vector of observations on the dependent variable at time t; μ is an $N \times 1$ vector of fixed effects, $X_N(t)$ is an $N \times k$ matrix of observations on k exogenous regressors at time t, β is an $k \times 1$ vector of parameters, λ is a scalar parameter, $W_N(t)$ is an $N \times N$ matrix of observations on an endogenous weighting matrix at time t, $Y_N(t)$ is an $N \times q$ matrix of

[3]We "somewhat" expected this income result because income levels positively correlate with educational levels, and the level of education could have a negative effect on cigarette consumption due to health issues.

observations on q endogenous variables at time t, γ is a corresponding $q \times 1$ vector of parameters, $\delta' = (\beta', \lambda, \gamma')$, and $u_N(t)$ is the $N \times 1$ disturbance vector which is specified below. The subscript N denotes triangular arrays.

Stacking the model over $t = 1, ..., T$ we have, using evident notation

$$\begin{aligned} y_N &= Z_N\delta + (i_T \otimes I_N)\mu + u_N \\ Z_N &= [X_N, diag_{t=1}^T[W_N(t)]y_N, Y_N] \end{aligned} \tag{2}$$

where i_T is a $T \times 1$ vector of unit elements, and

$$\begin{aligned} y_N &= [y_N(1)', ..., y_N(T)']';\ Z_N = [Z_N(1)', ..., Z_N(T)']'; \\ u_N &= [u_N(1)', ..., u_N(T)']';\ X_N = [X_N(1)', ..., X_N(T)']'; \\ Y_N &= [Y_N(1)', ..., Y_N(T)']'. \end{aligned}$$

We assume the following non-parametric specification for u_N

$$u_N = R_N\ \varepsilon_N \tag{3}$$

where R_N is an unknown $NT \times NT$ non-stochastic matrix, and ε_N is an $NT \times 1$ random vector whose mean is zero, and for which further details are defined below. At this point we note that the solution of the model for y_N in terms of $X_N, diag_{t=1}^T[W_N(t)]$, and Y_N is[4]

$$y_N = (I_{NT} - \lambda diag_{t=1}^T[W_N(t)])^{-1}[X_N\beta + (i_T \otimes I_N)\mu + Y_N\gamma + u_N] \tag{4}$$

Our large sample results below are based on the assumption that $N \to \infty$ with T fixed. Because of this the model in (2) is transformed to eliminate the fixed effects. Specifically, consider the standard transformation matrix $Q_{0,N} = (I_T - \frac{1}{T}J_T) \otimes I_N$, where $J_T = i_T i_T'$ and note that

$$Q_{0,N}(i_T \otimes I_N)\mu = 0 \tag{5}$$

Therefore, pre-multiplying (2) by $Q_{0,N}$ yields

$$\begin{aligned} Q_{0,N}y_N &= Q_{0,N}Z_N\delta + Q_{0,N}R_N\varepsilon_N \\ &= (Q_{0,N}X_N, Q_{0,N}diag_{t=1}^T[W_N(t)]y_N, Q_{0,N}Y_N)\delta + Q_{0,N}R_N\varepsilon_N \\ &= \bar{Z}_N\delta + Q_{0,N}R_N\varepsilon_N \end{aligned} \tag{6}$$

[4]Formal assumptions relating to the weights are given in Section 3.

where $\bar{Z}_N = Q_{0,N} Z_N.$

Unless the weighting matrix is time invariant, $Q_{0,N} diag_{t=1}^T [W_N(t)] \neq diag_{t=1}^T [W_N(t)] Q_{0,N}$ and therefore, in general, the term in (6)

$$Q_{0,N} diag_{t=1}^T [W_N(t)] y_N \neq diag_{t=1}^T [W_N(t)] Q_{0,N} y_N$$

2.1 Comments about the error specification

An issue relating to an endogenous weighting matrix

Before continuing to formal specifications, certain comments about the error term specification should be noted. First, the disturbance term u_N is being specified in a nonparametric way. One reason for this is that typical models of spatially correlated disturbance terms would have a major shortcoming because of the endogeneity of $W_N(t)$. To see this, let $\mathbf{W}_N = diag_{t=1}^T [W_N(t)]$. In a typical Cliff-Ord setting, the specification of the error term would be

$$u_N = \rho \mathbf{W}_N u_N + \varepsilon_N \tag{7}$$

where $(I_{NT} - \rho \mathbf{W}_N)$ would be assumed to be nonsingular and $E(\varepsilon_N) = 0$. However, in this setting,

$$u_N = (I_{NT} - \rho \mathbf{W}_N)^{-1} \varepsilon_N \tag{8}$$

and since $\mathbf{W}_N$ is endogenous

$$\begin{aligned} E(u_N) &= E[(I_{NT} - \rho \mathbf{W}_N)^{-1} \varepsilon_N] \\ &\neq 0 \end{aligned} \tag{9}$$

because the elements of W_N, and therefore those of $\mathbf{W}_N$ would not be independent with those of ε_N. Therefore, the specification in (7) would imply a disturbance term whose mean is not zero! Perhaps more significantly, under typical specifications, this mean would involve some or all of the exogenous variables which partially determine the elements of W_N,[5] as well as unknown parameters and so the mean would not just be a constant vector. Indeed,

[5]We say "partially" because if the elements of $W_N(t)$ are endogenous, then the determination of $W_N(t)$ would also involve endogenous variables, and perhaps, error terms. To simplify the presentation, we did not indicate that the expectation of $W_N(t)$ is conditional upon the exogenous variables.

unless extremely strong assumptions are made, even the functional forms of these means would not be known.

The comments above apply to any spatial error model which involves an endogenous weighting matrix -e.g., to spatial moving average models, etc. Obviously, the above comments would not apply if the weighting matrix in the error process is exogenous, and therefore different from the endogenous weighting matrix defining the spatially lagged dependent variable. However, such a specification might be difficult to justify since it would involve different measures of "meaningful" distance - e.g., a trade model with endogenous trade shares as weights defining a spatially lagged dependent variable, and a row normalized weighting matrix based on contiguous areas with equal weights in each row.

Asymptotic generality of the specification $u_N = R_N\ \varepsilon_N$

At first, this specification may seem to be restrictive in that some commonly used error term specifications in a panel framework can not be expressed as in (3). As one example, consider the somewhat typical error term specification in Kapoor et al. (2007).[6] They assumed a nonstochastic and time invariant weighting matrix, say W_N, spatial correlation in a Cliff-Ord setting, and an innovation term which has an error component structure with random effects. Using evident notation, their model is

$$\begin{aligned} \psi_N &= \rho(I_T \otimes W_N)\psi_N + \eta_N \\ \eta_N &= (i_T \otimes I_N)\mu_N + \nu_N \end{aligned} \tag{10}$$

where ψ_N is the $NT \times 1$ model error term, η_N is the corresponding $NT \times 1$ innovation vector, μ_N is an $N \times 1$ vector of *i.i.d.* random effects which have mean and variance $(0, \sigma_\mu^2)$, and v_N is an $NT \times 1$ vector of *i.i.d.* stochastic terms which have mean and variance $(0, \sigma_v^2)$. The error term ψ_N in (10) can not be expressed in a manner comparable to (3), and so in this sense the specification in (3) may seem "overly" restrictive.

On the other hand, given the specifications in Kapoor et al. (2007), the error term ψ_N in (10) is such that

$$E(\psi_N) = 0; \quad E(\psi_N \psi_N') = \Omega_{\psi_N} \tag{11}$$

where Ω_{ψ_N} is positive definite. Our assumptions below imply that R_N is nonsingular, and ε_N has a zero mean, and its VC matrix is I_{NT}. This implies

[6] This was pointed out to us by Ingmar Prucha.

that

$$E(u_N) = 0; \quad E(u_N u_N') = R_N R_N' \tag{12}$$

Clearly there exists an R_N such that

$$R_N R_N' = \Omega_{\psi_N} \tag{13}$$

For example, since Ω_{ψ_N} is positive definite and symmetric it can be expressed as $\Omega_{\psi_N} = \Omega_{\psi_N}^{1/2}\Omega_{\psi_N}^{1/2}$ [7] and so if $R_N = \Omega_{\psi_N}^{1/2}$ (13) will hold. The implication is that every panel data framework specification of the error term which implies a zero mean vector, and a VC matrix which is positive definite is consistent with the first two moments of our specification $u_N = R_N \varepsilon_N$ in (3). Furthermore, given that the mean of the error vector is zero, our assumptions in Section 3 imply that the large sample distribution of the IV estimators of the regression parameters given below do not depend upon any characteristic of the disturbance distribution other than its VC matrix. The implication of this is that in a large sample IV framework, it is of no consequence whether or not a "true" disturbance vector can be expressed as in (2). In this sense, our error term specification in (3) is not "overly" restrictive.[8]

3 Model assumptions

The mean of $W_N(t)$

Using evident notation, denote the elements of $W_N(t)$ as $w_{ij,N}(t)$, $t = 1, ..., T$. Some of these elements will be specified, apriori, to be zero, while others will be positive.[9] Denote the subset of elements of $W_N(t)$ which are not specified to be zero as $w^*_{ij,N}(t)$.[10] We assume that the mean of $w^*_{ij,N}(t)$ exists and is an unknown function of two sets of exogenous variables. One set, say $p_{ij,N}(t)$ is observable, and the other set, say $q_{ij,N}(t)$ is not observable, or perhaps not even known. The regressor matrix X_N and $p_{ij,N}(t)$ may have elements in common.

[7] See, e.g., equations (7) and (11) in Kapoor et al. (2007).

[8] Of course, other characteristics of the error distribution could have an effect on the corresponding small sample distributions.

[9] For example, among others, the diagonal elements of $W_N(t)$ would be specified to be zero.

[10] In the specification of $w^*_{ij,N}(t)$, the subscripts i, j only take on those values which correspond to the non-zero elements of $W_N(t)$.

Denote this mean relationship as

$$E[w^*_{ij,N}(t)] = f[p_{ij,N}(t), q_{ij,N}(t)] \tag{14}$$

The result in (14) implies

$$w^*_{ij,N}(t) = f[p_{ij,N}(t), q_{ij,N}(t)] + \zeta_{ij,N}(t) \tag{15}$$

where $E(\zeta_{ij,N}) = 0$.

Suppose $p_{ij,N}(t)$ is an $1 \times r$ vector, say $p_{ij,N}(t) = (p_{1,ij,N}(t), ..., p_{r,ij,N}(t))$, and consider the linear approximation to $f[p_{ij,N}(t), q_{ij,N}(t)]$ in terms of the elements of $p_{ij,N}(t)$, namely[11]

$$f[p_{ij,N}(t), q_{ij,N}(t)] \approx p_{1,ij,N}(t)a_1 + ... + p_{r,ij,N}(t)a_r \tag{16}$$

so that via (15) we have

$$w^*_{ij,N}(t) \approx p_{1,ij}(t)a_1 + ... + p_{r,ij,N}(t)a_r + \zeta_{ij,N}(t) \tag{17}$$

where $a_1, ..., a_r$ are constants. If an intercept is assumed in the approximation in (16), the values of $p_{1,ij,N}(t)$ can be taken as 1.0 .

Let $\hat{a}_1, ..., \hat{a}_r$ be the least squares estimates of the parameters, $a_1, ..., a_r$ based on (17) using all the nonzero values in $W_N(t)$, $t = 1, ..., T$. Let the approximation to $f[p_{ij,N}(t), q_{ij,N}(t)]$ be

$$\hat{f}[p_{ij,N}(t), q_{ij,N}(t)] = p_{1,ij,N}(t)\hat{a}_1 + ... + p_{r,ij,N}(t)\hat{a}_r \tag{18}$$

Let $P_{s,N}(t)$ be the $N \times N$ matrix whose zero elements are exactly in the same positions as those of $W_N(t)$, and whose nonzero elements are obtained by replacing $w^*_{ij,N}(t)$ with $p_{s,ij,N}(t)$, $s = 1, ..., r$. Then we assume the approximation to $E[W_N(t)]$ to be $\hat{W}^e_N(t)$ where

$$\hat{W}^e_N(t) = P_{1,N}(t)\hat{a}_1 + ... + P_{r,N}(t)\hat{a}_r \tag{19}$$

and, correspondingly, $diag^T_{t=1}[\hat{W}^e_N(t)]$ is

$$diag^T_{t=1}[\hat{W}^e_N(t)] = diag^T_{t=1}[P_{1,N}(t)\hat{a}_1 + ... + P_{r,N}(t)\hat{a}_r] \tag{20}$$

[11]We are assuming a linear approximation for ease of presentation; it will become clear that our results will hold if the approximation is in terms of a higher order polynomial which might be considered if r is small.

The mean of $Y_N(t)$

Let $Y_{ij,N}(t)$ be the ij^{th} element of $Y_N(t)$, and let

$$E[Y_{ij,N}(t)] = \phi(m_{ij,N}(t), \xi_{ij,N}(t)) \tag{21}$$

where $m_{ij,N}(t)$ is a $1 \times g$ vector of exogenous variables at time t that are observed, and $\xi_{ij,N}(t)$ is a vector of exogenous variables that are not observed. Let $M_N(t)$ be the $N \times g$ matrix of observations on $m_{ij,N}(t)$, and let M_N be the $NT \times g$ matrix of observations: $M_N = [M_N(1)', ..., M_N(T)']'$.

The instruments

Because the model in (6) contains endogenous regressors, as well as an endogenous weighting matrix we suggest an IV procedure for its estimation. We specify the form of these instruments before giving the assumptions of the model. We do this because some of the assumptions of the model are given in terms of these instruments. We also need to introduce some additional notation.

Let

$$\begin{aligned} S_{i,N}(t) &= P_{i,N}(t)X_N(t), i = 1, ..., r \\ V_{i,N}(t) &= P_{i,N}(t)M_N(t), i = 1, ..., r \end{aligned} \tag{22}$$

Stacking the data on these matrices over $t = 1, ..., T$ we define

$$\begin{aligned} S_{i,N} &= [S_{i,N}(1)', ..., S_{i,N}(T)']', i = 1, ..., r \\ V_{i,N} &= [V_{i,N}(1)', ..., V_{i,N}(T)']'i = 1, ..., r \end{aligned} \tag{23}$$

Finally, we define

$$\begin{aligned} S_N &= (S_{1,N}, ..., S_{r,N}) \\ V_N &= (V_{1,N}, ..., V_{r,N}) \end{aligned} \tag{24}$$

and note that S_N and V_N are, respectively, $NT \times rk$ and $NT \times rg$ matrices.

Our suggested instruments are $\hat{H}_N^*$ where[12]

$$\begin{aligned}
\hat{H}_N^* &= (Q_{0,N}X_N, Q_{0,N}M_N, Q_{0,N}diag_{t=1}^T[\hat{W}_N^e(t)]\ X_N, Q_{0,N}diag_{t=1}^T[\hat{W}_N^e(t)]\ M_N) \\
&= (Q_{0,N}X_N, Q_{0,N}M_N, Q_{0,N}S_N\hat{A}_1, Q_{0,N}V_N\hat{A}_2]) \qquad (25) \\
&\equiv Q_{0,N}(X_N, M_N, S_N\hat{A}_1, V_N\hat{A}_2)
\end{aligned}$$

where $\hat{A}_1$ and $\hat{A}_2$ are, respectively, the $rk \times k$ and $rg \times g$ matrices

$$\begin{aligned}
\hat{A}_1 &= [I_k\hat{a}_1, ..., I_k\hat{a}_r]' \qquad (26) \\
\hat{A}_2 &= [I_g\hat{a}_1, ..., I_g\hat{a}_r]'
\end{aligned}$$

Note that the instrument matrix $\hat{H}_N^*$ is $NT \times 2k + 2g$ and the regression parameter vector δ is $(k+1+q) \times 1$. We therefore assume that $2k + 2g \geq 1 + k + q$, or

$$k + 2g \geq 1 + q \qquad (27)$$

so that there are at least as many instruments as there are regression parameters. Note that even if $g = 0$ so that none of the exogenous variables that partially determine the mean of the endogenous matrix Y_N is observed, the condition in (27) will still hold if $k \geq 1 + q$. Note also that in this case because Y_N is endogenous, it partially depends upon y_N and therefore should depend upon X_N.

Another expression of $\hat{H}_N^*$ is useful for our analysis below. Specifically, $\hat{H}_N^*$ can be expressed as

$$\hat{H}_N^* = H_N\ \tilde{A} \qquad (28)$$

where

$$H_N = Q_{0,N}(X_N, M_N, S_N, V_N) \qquad (29)$$

$$\tilde{A} = \begin{bmatrix} I_k & 0_{k\times g} & 0_{k\times k} & 0_{k\times g} \\ 0_{g\times k} & I_g & 0_{g\times k} & 0_{g\times g} \\ 0_{rk\times k} & 0_{rk\times g} & \hat{A}_1 & 0_{rk\times g} \\ 0_{rg\times k} & 0_{rg\times g} & 0_{rg\times k} & \hat{A}_2 \end{bmatrix} \qquad (30)$$

[12] Obviously, we are assuming that if X_N and M_N have columns in common, these columns are not included in M_N. Also, our set of instruments could be extended to include products of the square of $\hat{W}_N^e(t)$ -e.g. terms such as $Q_{0,N}diag_{t=1}^T[(\hat{W}_N^e(t))^2]\ X_N$. Such an extended set of instruments could improve estimation efficiency. We have not included these additional instruments for ease of presentation. Our large sample results given below would still hold with obvious modifications.

Adjustments for the case of a random effects model

If the model has random effects instead of fixed effects, there are only two changes that need to be made in order for our results given below to hold. First, there would be an intercept term in the original model (1), and the vector of random effects would be incorporated into the disturbance specification in (3). In this case, there would be no reason to multiply the model across by $Q_{0,N}$ and therefore the list of instruments would be the same as those given in (25) except that none of the variables would be pre-multiplied by $Q_{0,N}$. Second, the list of instruments would contain the intercept. Of course, the assumptions given below would need to be correspondingly adjusted.

3.1 Assumptions

In this section we give detailed assumptions corresponding to the fixed effects model; their interpretations and justifications are given in the next section.

Assumption 1 *T is a given finite integer.*

Assumption 2 $\underset{N\to\infty}{p\lim}\hat{a}_i = c_i,\ i = 1, ..., r$ *where* $0 \leq c_i < \infty$ *and at least one of* $(c_1, ..., c_r)$ *is not zero.*

Assumption 3 *(a) (X_N) is nonstochastic with rank k for N large enough, and its elements are uniformly bounded for all $t = 1, ..., T$ and $N > 1$. (b) Let A be identical to $\tilde{A}$ except that $\hat{a}_i$ is replaced by $c_i, i = 1, ..., r$ and let $H_N^* = H_N A$. Then, we assume that H_N^* is nonstochastic and has full column rank, namely, $2k + 2g$ for N large enough, and its elements are uniformly bounded for all $t = 1, ..., T$ and $N > 1$.*

Assumption 4 *(a) The diagonal elements of $diag_{t=1}^{T}[W_N(t)]$ are zero (b) Each row and column of $diag_{t=1}^{T}[W_N(t)]$ has, at most, a finite number of non-zero elements. (c) The elements of $diag_{t=1}^{T}[W_N(t)]$ are uniformly bound in absolute value by a constant. (d) $(I_{NT} - \lambda diag_{t=1}^{T}[W_N(t)])$ is nonsingular for all $|\lambda| < \xi$, where ξ is a finite constant.*

Assumption 5 *The matrix R_N is nonsingular, and the row and column sums of R_N and R_N^{-1} are uniformly bound in absolute value.*

Assumption 6 *For all $1 \leq t \leq T$ and $1 \leq i \leq N$, $N \geq 1$ the elements of ε_N, namely $\varepsilon_{it,N}$, are identically distributed with zero mean and unit variance. In addition for each $N \geq 1$ and $1 \leq t \leq T$, $1 \leq i \leq N$ the error terms $\varepsilon_{it,N}$ are identically and independently distributed.*

Assumption 7 *We also assume*

$$\begin{aligned}
(a) &: p \lim_{N\to\infty} (NT)^{-1}\tilde{A}'[H_N' H_N]\tilde{A} = A'Q_{HH}A \equiv Q_{H^*H^*} \\
(b) &: p \lim_{N\to\infty} (NT)^{-1}\tilde{A}'H_N'\bar{Z}_N = A'Q_{H\bar{Z}} \equiv Q_{H^*\bar{Z}} \\
(c) &: p \lim_{N\to\infty} (NT)^{-1}\tilde{A}'H_N'R_N R_N' H_N\tilde{A} = A'Q_{HRRH}A \equiv Q_{H^*RRH^*}
\end{aligned}$$

where $Q_{HH}, Q_{H\bar{Z}}$, and Q_{HRRH} are finite matrices and $Q_{H^*H^*}, Q_{H^*\bar{Z}}$, and $Q_{H^*RRH^*}$ are finite full column rank matrices.

3.2 Assumption Interpretations

Given our model, Assumption 1 essentially indicates that our large sample results relate to $N \to \infty$. Assumption 2 requires that the linear approximation in (17), namely $[p_{1,ij}(t)a_1 + ... + p_{r,ij,N}(t)a_r]$, is at least correlated with $w^*_{ij,N}(t)$. This seems reasonable given the general correlation of most economic variables. Assumption 3 part (a) rules out perfect multicollinearity, and therefore an identification problem; it also rules out regressors whose values increase beyond limit. Part (b) of this assumption effectively rules out linearly dependent instruments. Note that a necessary condition for H^*_N to have full column rank is that A has full column rank. A necessary and sufficient condition for A to have full column rank is given in Assumption 2, namely that at least one $c_i \neq 0$. Obviously, part (b) of Assumption 3 would be violated if certain columns of M_N are linearly dependent upon those in X_N.

Assumption 4 (a) is a standard specification for all weighting matrices and is, effectively, a normalization of the model. In many spatial models it is typically assumed that each unit has a finite number of neighbors (typically bordering units) and so Assumption 4 (b) is reasonable. However the part of this assumptions that relates to the number of non-zero elements in each column rules out a model containing a central unit to which all units are related. Such a situation could arise if one unit is dominant either financially, militarily, or otherwise. In Assumption 4 (c) we are effectively assuming that the specification of the weights, the structure of the weighting matrix, and the stochastics involved are such that the weights are uniformly bound in absolute value by a constant. One example of this would be a model in which the elements of the weighting matrix are trade shares.[13] In this case each weight would be in the interval $[0, 1]$. Assumption 4 (d) implies that the model is complete in that it can be solved for the dependent variable, y_N, in terms of μ, X_N, Y_N, and u_N for all $|\lambda| < \xi$, which could be taken as the parameter space for λ. Given Gershgorin's Theorem, and Assumptions 4 (b) and (c), the researcher should be able to place an upper limit on the maximum row sum of the weighting matrix which can then be taken as the value of ξ. As one example, if the weighting matrix is based on trade shares, that sum would be 1.0.

Assumption 5 is also somewhat standard, see Kelejian and Prucha (2007), and along with Assumption 6, implies a well behaved VC matrix of the disturbance vector. These assumptions are consistent with a disturbance term which is both spatially correlated and heteroskedastic in a general way. They also account for triangular arrays. In Assumption 6 the innovation term is assumed to have a zero mean and unit variance. The variance assumption is not restrictive because the matrix R_N can always be defined to make this so- e.g., scaled by the inverse of a standard deviation!

Assumption 7 gives conditions which are typically made in determining large sample distributions - see e.g., Kelejian and Prucha (1999, 2004) and Kapoor et al. (2007).

4 Estimation and Inference

Let $P_{\hat{H}^*_N} = \hat{H}^*_N(\hat{H}^{*\prime}_N\hat{H}^*_N)^{-1}\hat{H}^{*\prime}_N$ and let $\hat{Z}_N = P_{\hat{H}^*_N}\bar{Z}_N$. For future reference note that

$$\begin{aligned} \hat{H}^{\prime *}_N Q_{0,N} &= \hat{H}^*_N \\ H^{\prime}_N Q_{0,N} &= H^{\prime}_N \end{aligned} \tag{31}$$

[13]In Assumption 4 we are assuming that each row and column of $diag_{t=1}^{T}[W_N(t)]$ contains, at most, a finte number of non-zero elements. Therefore, in a trade share model one must assume that no country trades (or has significant trades) with all other countries.

This follows from (28) and (29) because $Q_{0,N}$ is symmetric and idempotent.

Since we are assuming a non-parametric specification of the error term our estimator of δ is just the 2SLS estimator:

$$\hat{\delta} = (\hat{Z}'_N \hat{Z}_N)^{-1} \hat{Z}'_N y_N \tag{32}$$

The proof of Theorem 1 is given in the appendix.

Theorem 1 *Given our model in (1), and Assumptions 1-7:*

$$(NT)^{1/2}(\hat{\delta} - \delta) \xrightarrow{D} N(0, VC_{\hat{\delta}}) \tag{33}$$

where

$$\begin{aligned} VC_{\hat{\delta}} &= \Psi Q_{H^*RRH^*} \Psi' \qquad (34) \\ \Psi &= [p \lim_{N\to\infty} NT(\hat{Z}'_N \hat{Z}_N)^{-1}] \; Q'_{H^*\bar{Z}} \; (Q_{H^*H^*})^{-1} \end{aligned}$$

and where

$$p \lim_{N\to\infty} (NT)^{-1} \hat{Z}'_N \hat{Z}_N = Q'_{H^*\bar{Z}} (Q_{H^*H^*})^{-1} Q_{H^*\bar{Z}} \tag{35}$$

Small sample inferences can be based on the approximation

$$\begin{aligned} \hat{\delta} &\simeq N[\delta, \hat{\Omega}_{\hat{\delta}}) \\ \hat{\Omega}_{\hat{\delta}} &= NT \; G_N[\hat{Q}_{H^*RRH^*}]G'_N \\ G_N &= (\hat{Z}'_N \hat{Z}_N)^{-1} \bar{Z}'_N \hat{H}^*_N (\hat{H}^{*\prime}_N \hat{H}^*_N)^{-1} \end{aligned}$$

where $\hat{Q}_{H^*RRH^*} = \tilde{A}' \hat{Q}_{HRRH} \tilde{A}$ and where $\hat{Q}_{HRRH}$ is the HAC estimator of Q_{HRRH} – see, e.g., Kelejian and Prucha (2007) and Kim and Sun (2011).[14]

Let c' be a constant $1 \times 1 + k + q$ vector. Then, as an example, the hypothesis $H_0 : c'\delta = 0$ would be rejected at the two tail 5% level if

$$\frac{(c'\hat{\delta})^2}{c'\hat{V}C_{\hat{\delta}}c} > 1.96$$

[14] Neither Kelejian and Prucha (2007), nor Kim and Sun (2011) considered a panel framework. However their results can easily be applied in a panel framework given an evident condition if T is finite. Specifically, let $(q'_1, ..., q'_T)' = Q_0 u$ where $q_t, t = 1, ..., T$ is the t^{th} $N \times 1$ block of $Q_0 u$. Let q_{ti} be the i^{th} element of q_t.Then, using obvious notation, the "evident" condition is

$$\begin{aligned} |cov(q_{ti}, q_{sj})| &< |cov(q_{ti}, q_{tj})| \\ for \text{ all:} & \quad t, s = 1, ..., T; i, j = 1, ..., N \end{aligned}$$

Given this condition, the distance measure in the HAC estimator between q_{ti} and q_{sj}, say $d(q_{ti}, q_{sj})$, can simply be taken as $d(q_i, q_j)$.

4.1 Testing degenerate models

In this section we demonstrate in the context of our model that if the weighting matrix is endogenous the researcher can, under reasonable conditions, test for degenerate forms of the model. If, however, the weighting matrix is exogenous and, as usual, time invariant, the researcher can not test for degenerate models in which $E(Q_{0,N}y_N) = 0$.

The case when $W_N(t)$ is endogenous

First note that the only assumption that could depend upon nonzero values of the model parameters is Assumption 7, part (b). We will show that part (b) of Assumption 7 does not rule out the degenerate case of the model in which $\lambda = 0, \beta = 0, \gamma = 0$ as long as $W_N(t)$ is endogenous.[15] The implication of this is that one can test the hypothesis $\lambda = 0, \beta = 0, \gamma = 0$, or any subset of this hypothesis - e.g., the hypothesis $\beta = 0, \gamma = 0$ within the context of our model **if** $W_N(t)$ is endogenous.

To see this note that if $\lambda = 0, \beta = 0, \gamma = 0$, the data generating process for y_N would be

$$y_N = (i_T \otimes I_N)\mu + u_N \tag{36}$$

and so the true form of the considered transformed model would be

$$Q_{0,N}y_N = Q_{0,N}u_N \tag{37}$$

The transformed matrix $\bar{Z}_N = Q_{0,N}Z_N$ would still be

$$\bar{Z}_N = Q_{0,N}[X_N, diag_{t=1}^T[W_N(t)]y_N, Y_N] \tag{38}$$

Therefore

$$\begin{aligned} p\lim_{N\to\infty}(NT)^{-1}H_N'\bar{Z}_N &= p\lim_{N\to\infty}(NT)^{-1}H_N' * \\ &\quad [Q_{0,N}X_N, Q_{0,N}diag_{t=1}^T[W_N(t)]y_N, Q_{0,N}Y_N] \end{aligned} \tag{39}$$

First note that Assumption 2 part (b) does not, in any way, depend upon, or involve, non-zero values of any of the regression parameters. Given this,

[15] Note that we are only assuming the disturbance specification in (3) and Assumptions 5 and 6 so that maximum likelihood is not an option. Also, even if normality were assumed, one would have to extend the model by specifying a distribution for Y_N.

and assuming that $E(Y_N) \neq 0$, consider the limit matrix of a subset of the matrix in (39), namely

$$p \lim_{N\to\infty} (NT)^{-1} H_N' [Q_{0,N} X_N, Q_{0,N} Y_N] = \Phi_1 \tag{40}$$

It is quite reasonable to assume that Φ_1 has full column rank since the elements of Y_N involve variables all of which are not contained in X_N.

Now consider the remaining limit in (39), namely

$$p \lim_{N\to\infty} (NT)^{-1} H_N' Q_{0,N} diag_{t=1}^T [W_N(t)] y_N$$

Since $W_N(t)$ is endogenous, its elements would be correlated with those of u_N, and therefore with those of y_N. Thus,[16]

$$E(diag_{t=1}^T [W_N(t)] y_N) \neq 0 \tag{41}$$

The result in (41) does not involve the assumption that $\lambda \neq 0$. Therefore, the assumption that

$$p \lim_{N\to\infty} (NT)^{-1} H_N' Q_{0,N} diag_{t=1}^T [W_N(t)] y_N = \Phi_2 \tag{42}$$

where Φ_2 is a finite non-zero vector is reasonable unless one considers peculiar cases of unbounded moments, etc. It is also reasonable to assume that (Φ_1, Φ_2) has full column rank since the q columns of $p \lim_{N\to\infty} (NT)^{-1} H_N' Q_{0,N} Y_N$, in Φ_1 involve the endogeneity of Y_N while the vector Φ_2 involves the endogeneity of $diag_{t=1}^T [W_N(t)] y_N$.

The case when $W_N(t)$ is exogenous

Now consider the case in which $W_N(t)$ is exogenous and time invariant: $W_N(t) = W_N, t = 1, ..., T$. For this case we demonstrate, again in the context of our model, that one can not test the null hypothesis $\lambda = 0, \beta = 0, \gamma = 0$ if $E(Q_{0,N} y_N) = 0$. We show this by demonstrating that Assumption 7 part (b) would not hold because $Q_{H\tilde{Z}}$ would not have full column rank. It will also become clear from these results that any hypothesis which implies that $E(Q_{0,N} y_N) = 0$ can not be tested.

[16] Note that the expectation in (41) should involve $p_{ij,N}$ and $q_{ij,N}$ - see (14).

To see this first note that if (36) is the data generating process, $E(Q_{0,N}y_N) = E(Q_{0,N}u_N) = 0$ and so

$$\begin{aligned} E(H_N' Q_{0,n} diag_{t=1}^{T}[W_N(t)]y_N) &= E[H_N' Q_{0,N}(I_T \otimes W_N)\ y_N] \\ &= E[H_N'(I_T \otimes W_N) Q_{0,N}\ y_N] \\ &= H_N'(I_T \otimes W_N) E(Q_{0,N}\ y_N) \\ &= 0 \end{aligned} \tag{43}$$

Let VC_Ξ be the VC matrix of $\Xi = (NT)^{-1} H_N'(I_T \otimes W_N) Q_{0,N} y_N$. Then in this case, since $H_N, Q_{0,N}$, and W_N are exogenous, and recalling that $Q_{0,N} = (I_T - \frac{J_T}{T}) \otimes I_N$

$$\begin{aligned} VC_\Xi &= (NT)^{-2} H_N'(I_T \otimes W_N)\ [VC_{Q_0 y}]\ (I_T \otimes W_N') H_N \\ &= (NT)^{-2} H_N'(I_T \otimes W_N)[Q_{0,N} R_N R_N' Q_{0,N}]\ (I_T \otimes W_N') H_N \end{aligned} \tag{44}$$

Consistent with Assumption 4 (b) and (c), assume that the row and column sums of W_N are uniformly bound in absolute value by, say c_w; by Assumption 5 the row and column sums of R_N are also uniformly bound in absolute value, as are the row and column sums of $(I_T \otimes W_N)$ and $Q_{0,N} = (I_T - \frac{1}{T} J_T) \otimes I_N$, respectively by c_w and 1.0. Since the product of matrices whose row and column sums are uniformly bound in absolute value, also has row and column sums which are so bounded, the row and column sums of the matrix $(I_T \otimes W_N)[Q_{0,N} R_N R_N' Q_{0,N}]\ (I_T \otimes W_N')$ in (44) are uniformly bound in absolute value. By Assumption 3 part (b) the elements of H_N are uniformly bound in absolute value, and thus the elements of VC_Ξ in (44) are $0[(NT)^{-1}]$. It follows from (44) that $VC_\Xi \to 0$ and therefore by Chebyshev's inequality

$$(NT)^{-1} H_N'(I_T \otimes W_N) Q_{0,N} y_N \xrightarrow{P} 0$$

and therefore $p\lim_{N\to\infty}(NT)^{-1} H_N' \bar{Z}_N$ is not a full column rank matrix because one of its columns is a column of zeroes.

5 Empirical Application

In our empirical application, we consider a dynamic demand for cigarettes based on a panel data from 46 US states over the period 1963-1992 (Baltagi and Levin, 1986, 1992; Baltagi and Li, 2004). This data set has also been used for illustrative purposes in a number of spatial econometric studies (see e.g. Elhorst, 2005; Debarsy et al., 2010, among others). Our model is a variation of the one considered in Baltagi and Levin (1986). They considered a state level dynamic demand for cigarettes using the same data set over a limited period (1963-1980) to address several policy issues (such as the impositions of warning labels and the application of the Fairness Doctrine Act to cigarette advertising). They found a significant effect of the average retail price on cigarette consumption with a price elasticity of -0.2, but an insignificant income elasticity. A distinctive characteristic of their model is that consumer cigarette demand in each state is assumed to depend upon, among other things, the lowest cigarette price in neighboring states. This is meant to capture a bootlegging effect where buyers of cigarettes near state borders are tempted to buy from neighboring states if there is a price advantage in doing so. This bootlegging effect is found to be positive and statistically significant. However, the specification in Baltagi and Levin (1986) is only a first attempt at dealing with the issue of how to capture the bootlegging effect. As Baltagi and Levin (1986) point out, their model does not account for the fact that cross-border shopping may relate to more than one state and not just the one neighboring state with the minimum price. As will become clear later, our model is able to account for this.[17]

In this paper, we assume the following model:

$$\ln C_{it} = \beta_1 \ln C_{it-1} + \beta_2 \ln p_{it} + \beta_3 \ln I_{it} + \lambda \left[\sum_{j=1}^{46} \frac{p_{jt}}{p_{it}} d_{ijt} \ln C_{jt} \right] + \mu_i + \delta_t + u_{it} \tag{45}$$

where $i = 1, \ldots, 46$ denotes states, $t = 1, \ldots, 29$ denotes time periods, and the disturbance term u_{it} has the non-parametric specification in (3). C_{it} is cigarette sales to persons of smoking age in packs per capita in state i at time t and, therefore, is a measure of real per capita cigarette sales. p_{it} is the average retail price per pack of cigarettes in state i at time t. I_{it} is per capita disposable income in state i at time t. All values are measured in real terms. We expect $\beta_1 > 0, \beta_2 < 0, \beta_3 > 0$, and $\lambda > 0$. Our expectations concerning

[17]A number of variations on the Baltagi and Levin (1986) study have been considered but none of these studies considered the possibility of an endogenous weighting matrix. See, e.g., Baltagi (2008), Elhorst (2005), and Debarsy et al. (2010).

As a point of information, the data are available as supplementary material to Baltagi (2008) on the website: www.wileyeurope.com/college/baltagi.

β_1, β_2, and β_3 relate, respectively, to habit effects, the usual negative price effect on demand, and the positive effect of income. We expect λ to be positive for a number of reasons. First, the higher is the price ratio (p_{jt}/p_{it}), the less attractive is cross state shopping and so, given smoking habits of in-state consumers, the higher should be the in-state sales of cigarettes. Second, and somewhat indirect, values of $(p_{jt}/p_{it})d_{ijt} > 0$ imply that prices in neighboring states are lower than prices in the home state. These lower prices could lead to more intense smoking habits of in-state residents, as well as a broadening of the smoking in-state population, both of which would lead to an increase of in-state cigarette sales.[18]

The model in (45) also has state specific effects, μ_i, as well as time specific effects, δ_t. Baltagi and Levin (1986) and Baltagi (2008) give several motivations for the inclusions of time and state specific effects. The time effects can be justified by the various policy interventions and health warnings that occurred during the period under analysis. As for the state specific effects, they can represent any state-specific characteristics such as states that host military bases (that are tax-exempt), touristic states (where cigarette consumption is exceptionally high), or states with high percentage of Mormon population (a religion whose members are forbidden to smoke).

The spatial lag in brackets in (45) accounts for cross border cigarette shopping, or bootlegging. Specifically, d_{ijt} is a dummy variable which indicates the desirability of cross border shopping. In particular, $d_{ijt} = 1$ if i and j are border states **and** $p_{jt} < p_{it}$; $d_{ijt} = 0$ if i and j are not border states, **or** if $p_{jt} > p_{it}$. The multiplication of d_{ijt} by the price ratio in (45) effectively indicates that the price ratio p_{jt}/p_{it} will only be considered by cigarette consumers in state i at time t if $p_{jt} < p_{it}$. As will become clear, the first part of the term in brackets in (45), namely $(p_{jt}/p_{it})d_{ijt}$ is the i, j^{th} element of our weighting matrix at time t. Since the price of cigarettes is endogenous in a demand model for cigarettes, our weighting matrix is endogenous. The entire term in (45) is the spatial lag of the dependent variable.

Stacking (45) for each time period t, over $i = 1, \ldots, 46$, our model can be

[18]Our paper is not the first example of a demand equations that includes a spatial lag of the dependent variable. A previous influential application can be found in Case (1991) that, as an example of spatial modeling, estimates a demand for rice in Indonesia. However, the model in Case (1991) is not specified in terms of an endogenous weighting matrix.

written in the usual spatial form:

$$y_t = X_t\beta + \lambda W_t y_t + Y_t\gamma + \mu + e_{46}\delta_t + u_t; t = 1963, ..., 1992 \quad (46)$$

where y_t is the 46×1 vector of observations on $\ln C_{it}$ at time t; X_t is the corresponding vector of observations on the income variable $\ln I_{it}$; W_t is the 46×46 weighting matrix whose i, j^{th} element is $(p_{jt}/p_{it})d_{ijt}$; Y_t is the 46×2 matrix of observations on $\ln C_{it-1}$ and $\ln p_{it}$; μ is the vector of fixed effects; e_{46} is a 46×1 vector of unit elements; and u_t is the corresponding vector of disturbance terms.

To estimate (46) we took X_t as exogenous, and $W_t y_t$ and Y_t as endogenous. We then stacked the data over $t = 1963, ..., 1992$, multiplied by $Q_{0,N} = (I_T - \frac{1}{T}J_T) \otimes I_N$, and then used the IV procedure described in Section 4. Using evident notation, in doing this we took the variables in (17), to be the intercept, the cigarette tax rate in state i relative to that in state j, the relative populations of these two states, the relative compensation per employees in these two states, and the distance between them measured in hundreds of miles.[19] We took the variables defined by M_N as the cigarette tax rate, population, compensation per-employee, as well as the one period time lag of these three variables.[20] Our complete set of instruments also included the time dummy variables, as well as the product of the square of the estimated weighting matrix and the exogenous variables- i.e., via (25): $Q_{0,N} diag_{t=1}^T[(\hat{W}_N^e(t))^2]X_N$ and $Q_{0,N} diag_{t=1}^T[(\hat{W}_N^e(t))^2]M_N$.

Table 1 reports the estimation results of the cigarette demand equation. Standard errors are produced using the spatial HAC estimator of Kelejian and Prucha (2007) with a Parzen kernel.[21] The variables all have the expected signs and are strongly significant. As expected, the coefficient of income per capita points at a positive relationship with consumption, whereas price negatively affects consumption (with a coefficient equal to -0.80). The coefficient of lagged consumption is also positive and significant, thus pointing to a certain persistence in smoking behavior. The coefficient of the spatially lagged

[19]Data on the state cigarette tax rates were taken from the 58^{th} version of the annual compendium on tobacco revenue and industry statistics known as The Tax Burden on Tobacco. Data on compensation per-employee were available from the Bureau of Economic Analysis.

[20]All variables are in logarithms.

[21]Following Anselin and Lozano-Gracia (2008), we specify a variable bandwidth based on the distance to the seven nearest neighbors.

dependent variable is positive thus confirming the assumption of the bootlegging effect. Smokers, or agents for them,[22] living near the state border will purchase their cigarettes in near-by states when there is a price advantage for doing so.

Table 1: Estimation results of the cigarette demand equation.

	Coefficient	Std. Error	t-stat	$P > \|t\|$
$\ln(I_{it})$	0.1592	0.0360	4.4227	9.75e-06
$W\ln(C_{it})$	0.0026	0.0005	4.8824	1.05e-06
$\ln(C_{it-1})$	0.5941	0.0491	12.0883	1.22e-33
$\ln(p_{it})$	-0.8041	0.1029	-7.8145	5.52e-15

As a validation of our model, we tried to include the minimum price variable considered in Baltagi and Levin (1986). In fact, if this minimum price variable is the proper way to account for bootlegging effects, it will be significant and our spatially lagged dependent variable would no longer be significant. However, when we added the minimum price variable to our model our spatially lagged "bootlegging" variable remained significant, while the minimum price variable was not significant.[23]

A final point should be noted. In general, if a model contains a spatially lagged dependent variable, and no other endogenous regressors, the elasticities should be calculated with respect to the solution of the model for that dependent variable - e.g., its reduced form. If, however, that model also contains additional endogenous variables, and equations for these additional endogenous variables are not available, elasticities can not be calculated unless **very strong** assumptions are made. In fact, those elasticities should be calculated in terms of the solution for the corresponding system of equations. If, in addition, that model contains nonlinearities, further complications in the calculation of elasticities arise. In our case, in addition to a spatially

[22]If price advantages exist, boottleggers may purchase cigarettes in nearby states and sell them to cigarette consumers in higher priced states.

[23]Results for this specification can be obtained by writing to the authors.

lagged dependent variable, the model has an endogenous weighting matrix, endogenous regressors such as prices, and nonlinearities! It is for these reasons that we have not calculated elasticities.

6 Summary and Further Research Suggestions

In this paper we have specified a spatial panel data model that contains a spatially lagged dependent variable in terms of an endogenous weighting matrix. We suggested an estimator for the regression parameters and demonstrate its consistency and asymptotic normality. To the best of our knowledge, the present paper is the first attempt to formally establish the properties of a model that contains a spatially lagged dependent variable in terms of an endogenous weighting matrix.

We then apply our model to the estimation of a panel data model for the demand for cigarettes. We specify a spatially lagged dependent variable in terms of an endogenous weighting matrix that accounts for cross state bootlegging in cigarettes. We find that the bootlegging effect is positive and significant, suggesting that buyers cross state boards to purchase cigarettes when there is a price advantage in doing so. As expected, the coefficient of income per capita points to a positive relationship with consumption, whereas price negatively affects consumption.

One suggestion for further research would be a Monte Carlo study which focuses on the efficiency of our suggested estimator. Another would be a study which focuses on approximations to the calculation of elasticities in models such as ours.

References

Anselin, L. and Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A. and Giles, D., editors, *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, New York.

Anselin, L. and Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, 34(1):5–34.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data, 4th edition.* Wiley, New York.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2007). Estimating models of complex fdi: are there third-country effects? *Journal of Econometrics*, 140:260–281.

Baltagi, B. H. and 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. and Levin, D. (1992). Cigarette taxation: raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, 3:321–335.

Baltagi, B. H. and Li, Q. (2004). Prediction in panel data model with spatial correlation. In Anselin, L., Florax, R. J. G. M., and Rey, S. J., editors, *Advances in Spatial Econometrics: Methodology, Tools and Applications*, pages 283–295. Springer-Verlag, Berlin.

Behrens, K., Ertur, C., and Koch, W. (2012). "Dual" gravity: using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 27:773–794.

Case, A., Rosen, H., and Hines, J. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52:285–307.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica*, 59:953–965.

Cohen, J. and Morrison Paul, C. (2004). Public infrastructure investment, interstate spatial spillovers, and manufacturing costs. *The Review of Economics and Statistics*, 86(2):551–560.

Conley, T. and Ligon, E. (2002). Economic distance, spillovers and cross country comparison. *Journal of Economic Growth*, 7:157–187.

Conley, T. and Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17:303–327.

Debarsy, N., Ertur, C., and LeSage, J. P. (2010). Interpreting dynamic space-time panel data models. Document de Recherche n 2010-22, Laboratoire d'Economie d'Orléans.

Elhorst, J. P. (2005). Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis*, 37:62–83.

Hondroyianis, G., Kelejian, H. H., Mukerji, P., and Tavlas, G. S. (2012). Government bond spreads contagion among euro area countries. Manuscript.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data model with spatially correlated error components. *Journal of Econometrics*, 140(1):97–130.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two stages least square procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

Kelejian, H. H. and Prucha, I. R. (2001). On the Asymptotic Distribution of the Moran I test Statistic with Applications. *Journal of Econometrics*, 104:219–257.

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118:27–50.

Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140:131–154.

Kelejian, H. H., Tavlas, G. S., and Hondroyiannis, G. (2006). A spatial modeling approach to contagion among emerging economies. *Open Economies Review*, 17:423–441.

Kim, M. S. and Sun, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160(2):349 – 371.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72:1899–1926.

Mukerji, P. (2009). Ready for capital account convertibility? *Journal of International Money and Finance*, 28:1006–1021.

Parent, O. and LeSage, J. (2008). Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. *Journal of Applied Econometrics*, 23:235–256.

Pinske, J., Slade, M., and Brett, C. (2002). Spatial price competition: a semiparametric approach. *Econometrica*, 70:1111–1153.

Piras, G. (2011). Estimation of random effects spatial panel data models: some additional evidence. Manuscript.

Pötscher, B. M. and Prucha, I. R. (2000). Basic elements of asymptotic theory. In Baltagi, B. H., editor, *A Companion to Theoretical Econometrics*, pages 201 – 229. Basil Blackwell.

Rey, S. J. and Boarnet, M. G. (2004). A taxonomy of spatial econometric models for simultaneous equations systems. In Anselin, L., Florax, R. J. G. M., and Rey, S. J., editors, *Advances in Spatial Econometrics: Methodology, Tools and Applications*, pages 99–119. Springer-Verlag, Berlin.

Appendix

Since $P_{\hat{H}_N^*}$ is symmetric idempotent, $\hat{Z}_N'\bar{Z}_N = \hat{Z}_N'\hat{Z}_N$. It then follows from (6) and (32) that

$$\begin{aligned}
(NT)^{1/2}(\hat{\delta}-\delta) &= NT(\hat{Z}_N'\hat{Z}_N)^{-1}(NT)^{-1/2}\hat{Z}_N' Q_{0,N} R_N \varepsilon_N \qquad (47)\\
&= [NT(\hat{Z}_N'\hat{Z}_N)^{-1}][(NT)^{-1}\bar{Z}_N'\hat{H}_N^*][NT(\hat{H}_N^{*\prime}\hat{H}_N^*)^{-1}][(NT)^{-1/2}\hat{H}_N^{*\prime} Q_{0,N} R_N \varepsilon_N]\\
&\quad [NT(\hat{Z}_N'\hat{Z}_N)^{-1}][(NT)^{-1}\bar{Z}_N'\hat{H}_N^*][NT(\hat{H}_N^{*\prime}\hat{H}_N^*)^{-1}][(NT)^{-1/2}\tilde{A}' H_N' R_N \varepsilon_N]
\end{aligned}$$

where via (31) we have used $H_N' Q_{0,N} = H_N'$.

Consider the first term in brackets on the third line of (47). Since $\hat{Z}_N = P_{\hat{H}_N^*}\bar{Z}_N$ and $\hat{H}_N^* = H_N\tilde{A}$

$$\begin{aligned}
(NT)^{-1}\hat{Z}_N\hat{Z}_N &= [(NT)^{-1}\bar{Z}_N' H_N\tilde{A}][NT(\tilde{A}'H_N'H_N\tilde{A})^{-1}][(NT)^{-1}\tilde{A}'H_N'\bar{Z}_N] \qquad (48)\\
&= [(NT)^{-1}(\bar{Z}_N' H_N)\tilde{A}][(NT)^{-1}\tilde{A}'H_N'H_N\tilde{A}]^{-1}[(NT)^{-1}\tilde{A}'H_N'\bar{Z}_N]\\
&\xrightarrow{P} [Q_{H\bar{Z}}' A][(A'Q_{HH}A)^{-1}][A'Q_{H\bar{Z}}]\\
&\equiv [Q_{H^*\bar{Z}}'][(Q_{H^*H^*})^{-1}][Q_{H^*\bar{Z}}]
\end{aligned}$$

where the last line in (48) follows Assumptions 2, 3, and 7. Note that Assumption 7 implies that $Q_{H^*H^*}$ is positive definite and Q_{H^*Z} has full column rank, and so the quadratic form on the last line of (48) is positive definite and therefore nonsingular. Therefore the inverse of the probability limit of $(NT)^{-1}\hat{Z}_N\hat{Z}_N$ exists. Note that the expression on the last line of (48) is the same as the expression given in (35) of Theorem 1.

Consider now the second and third terms in brackets on the third line of (47). Assumptions 2, 3 and 7 imply

$$[(NT)^{-1}\bar{Z}_N' H_N\tilde{A}][NT(\tilde{A}'H_N'H_N\tilde{A})^{-1}] \xrightarrow{P} Q_{H^*\bar{Z}}' Q_{H^*H^*}^{-1} \qquad (49)$$

Let K equal the product of the limiting forms of the first three bracketed terms on the third line of (47):

$$K = [Q_{H^*\bar{Z}}' Q_{H^*H^*}^{-1} Q_{H^*\bar{Z}}]^{-1} Q_{H^*\bar{Z}}' Q_{H^*H^*}^{-1} \qquad (50)$$

Now let $\hat{\Psi}_N$ be the last bracketed term on the third line of (47), namely

$$\hat{\Psi}_N = (NT)^{-1/2}\tilde{A}'H_N'R_N\varepsilon_N \tag{51}$$

Let

$$\begin{aligned}\Psi_N &= (NT)^{-1/2}A'H_N'R_N\varepsilon_N \\ &= (NT)^{-1/2}H_N^{*\prime}R_N\varepsilon_N\end{aligned} \tag{52}$$

By Assumption 3 the elements of H_N^* are uniformly bounded in absolute value. By Assumption 7 part (c)

$$(NT)^{-1}H_N^{*\prime}R_NR_N'H_N^* \to Q_{H^*RRH^*} \tag{53}$$

where $Q_{H^*RRH^*}$ is a finite nonsingular matrix. Assumption 6, (52), (53) and the central limit theorem (30) in Pötscher and Prucha (2000) imply

$$(NT)^{-1/2}H_N^{*\prime}R_N\varepsilon_N \xrightarrow{D} N(0, Q_{H^*RRH^*}) \tag{54}$$

By Assumption 2, $\tilde{A} \xrightarrow{P} A$. It then follows from (51) - (54), and Assumption 7 part (a) that $\hat{\Psi}_N - \Psi_N \xrightarrow{P} 0$. Thus, Assumptions 3, 6 and 7 part (a), and the continuous mapping theorem,

$$(NT)^{-1/2}\tilde{A}'H_N'R_N\varepsilon_N \xrightarrow{D} N(0, Q_{H^*RRH^*}) \tag{55}$$

Theorem 1 follows from (47) - (50), and (55).