Spatio-Temporal Models for Political-Science Panel and Time-Series-Cross-Section Data*

Robert J. Franzese, Jr.
Associate Professor of Political Science
The University of Michigan, Ann Arbor
franzese@umich.edu

Jude C. Hays
Assistant Professor of Political Science
University of Illinois at Urbana-Champaign
jchays@uiuc.edu

18 July 2006

* This research was supported in part by NSF grant #0318045. Earlier or related versions—this being a selection from the authors' broader book project on spatial econometric models for political science,were presented at the Tulane-Murphy/Nottingham-Leverhulme Institute 2003, Harvard University Center for International Affairs 2003, MPSA 2003, *Wissenschaftszentrum*-Berlin 2004, MPSA 2004, Political Methodology Society Summer Meetings 2004, APSA 2004, RC33 Sixth International Conference on Social Science Methodology, Amsterdam 2004, University of Texas–Austin 2004, Leitner Political Economy Workshop at Yale 2005, MPSA 2005, *Rijksuniversitet Groningen* 2005, APSA 2005, MPSA 2006, ECPR 2006, Jean Monnet Centre of the Freie Universität-Berlin 2006. We thank Chris Achen, James Alt, Kenichi Ariga, Neal Beck, Jake Bowers, Kerwin Charles, Bryce Corrigan, Thomas Cusack, Jakob de Haan, John Dinardo, Zachary Elkins, John Freeman, Fabrizio Gilardi, Kristian Gleditsch, Mark Hallerberg, Aya Kachi, Mark Kayser, Achim Kemmerling, Hasan Kirmanoglu, James Kuklinski, Tse-Min Lin, Xiaobo Lu, Covadonga Meseguer, Thomas Pluemper, Dennis Quinn, Megan Reif, Frances Rosenbluth, Ken Scheve, Phil Schrodt, Beth Simmons, Duane Swank, Wendy Tam-Cho, Craig Volden, Michael Ward, and Gregory J. Wawro for useful comments on this and/or previous work in our project on spatial-econometric models in political science. Xiaobo Lu, Bryce Corrigan, and Aya Kachi have each provided excellent research assistance, and Kristian Gleditsch, Mark Hallerberg, and Duane Swank also generously shared data. We alone are responsible for any errors.

Spatio-Temporal Models for Political-Science Panel and Time-Series-Cross-Section Data

ABSTRACT: Building from our broader project exploring spatial-econometric models for political science, this paper discusses estimation, interpretation, and presentation of spatio-temporal models. We first present a generic spatio-temporal-lag model and two methods, OLS and ML, for estimating parameters in such models. We briefly consider those estimators' properties analytically before showing next how to calculate and to present the spatio-temporal dynamic and long-run steady-state equilibrium effects—i.e., the spatio-temporal substance of the model—implied by the coefficient estimates. Then, we conduct Monte Carlo experiments to explore the properties of the OLS and ML estimators, and, finally, we conclude with a reanalysis of Beck, Gleditsch, and Beardsley's (2006) state-of-the-art study of directed export flows among major powers.

I. Introduction

Empirical work in political science has increasingly recognized the cross-time and cross-space interdependence of outcomes. Most current practice, however, models temporal dependence directly, which is laudable, but spatial interdependence as a nuisance to be "corrected" (FGLS) or against which standard-error estimates are to be made "robust" (PCSE), which is less so. Other spatial modeling strategies relegate interdependence to spatial-covariance matrices and the like or otherwise to stages of the empirical analyses separate from that considering the ultimate dependent variable of interest. Relegating spatial interdependence to mere nuisance is usually problematic on econometric grounds, and demoting it to empirical sidebar can be as well, but, ultimately, these strategies are unsatisfying on substantive grounds because these ignored, sidelined, or analytically isolated spatial relationships frequently play central roles in many of the substantive contexts political scientists study. Even those recent works that do grant a direct and central substantive role for spatial interdependence, however, generally fail to present the spatio-temporal dynamic and steady-state implications—i.e., the spatio-temporal substance—of their model estimates, confining themselves to reporting the estimated strength of interdependence (as parameterized by their specific model) and its statistical significance.

We organize the rest of this paper to advance this state of affairs thusly. Section II presents a generic spatio-temporal-lag model, which embeds the substantive spatio-temporal dynamics directly in the model of outcome. The section explains two methods—spatial ordinary-least-squares (S-OLS) and spatial maximum-likelihood (S-ML)—for estimating the parameters of the spatio-temporal-lag model and considers those estimators' properties analytically. (The section also discusses alternative instrumental-variable/method-of-moments estimation strategies briefly, but we do not explore them explicitly in this paper.[1]) Section III describes how to calculate and present the spatio-temporal effects implied by the coefficient estimates of such models. Section IV describes the design of our Monte Carlo experiments to explore the properties of the S-OLS and S-ML estimates of parameters and (of dynamic (not done yet) and) steady-state effects, and then it presents the results. Section V offers a reanalysis of Beck, Gleditsch, and Beardsley's (2006) state-of-the-art study of directed export flows among major powers to illustrate the estimation, calculation, and presentation of spatio-temporal dynamics. Section VI concludes summarily.

II. Estimating Spatio-Temporal Models

We can write the spatio-temporal lag model in matrix notation as:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \phi\mathbf{M}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2.1),$$

where $\mathbf{y}$, the dependent variable, is an $NT \times 1$ vector of cross sections stacked by periods (i.e., the N first-period observations, then the N second-period ones, and so on to the N in the last period, T).[2] The parameter ρ is the spatial autoregressive coefficient and $\mathbf{W}$ is an $NT \times NT$ block-diagonal spatial-weighting matrix. More specifically, we can express this $\mathbf{W}$ matrix as the Kronecker product of a $T \times T$ identity matrix and an $N \times N$ weights matrix $\left(\mathbf{I}_T \otimes \mathbf{W}_N\right)$, with elements w_{ij} of $\mathbf{W}_N$ reflecting the relative degree of connection from unit j to i. $\mathbf{W}\mathbf{y}$ is thus the spatial lag; i.e., for each observation y_{it}, $\mathbf{W}\mathbf{y}$ gives a weighted sum of the y_{jt}, with weights, w_{ij}, given by the relative connectivity from j

[1] Franzese and Hays (2004) explore the spatial 2SLS-IV estimator in a spatial, but not spatio-temporal, context.
[2] Nonrectangular panels and/or missing data are possible, but we assume rectangularity and completeness for simplicity.

to i. The parameter ϕ is the temporal autoregressive coefficient, and $\mathbf{M}$ is an $NT\times NT$ matrix with ones on the minor diagonal, i.e., at coordinates $(N+1,1)$, $(N+2,2)$, …, $(NT, NT\text{-}N)$, and zeros elsewhere, so $\mathbf{My}$ is the (first-order) temporal lag. The matrix $\mathbf{X}$ contains $NT\times k$ observations on k independent variables, and $\boldsymbol{\beta}$ is a $k\times 1$ vector of coefficients on them. The final term in (2.1), $\boldsymbol{\varepsilon}$, is an $NT\times 1$ vector of disturbances, assumed to be independent and identically distributed.[3]

A. Least Squares Estimation

OLS estimation of model (2.1), sometimes called spatial OLS or S-OLS, is inconsistent because the regressor $\mathbf{Wy}$, the spatial lag, covaries with the residual, $\boldsymbol{\varepsilon}$. The reason is simple; the spatial lag, $\mathbf{Wy}$, is a weighted average of the outcome in other units, thus placing the left-hand side (LHS) of some observations on the right-hand side (RHS) of others: textbook simultaneity. To see the implications of this endogeneity, first rewrite the spatial-lag model as

$$\mathbf{y} = \mathbf{Q}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \qquad (2.2),$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Wy} & \mathbf{My} & \mathbf{X} \end{bmatrix} \text{ and } \boldsymbol{\delta} = \begin{bmatrix} \rho & \phi & \boldsymbol{\beta} \end{bmatrix}' \qquad (2.3).$$

The matrices $\mathbf{Q}$ and $\boldsymbol{\delta}$ have dimensions $N\times(k+2)$ and $(k+2)\times 1$ respectively. The asymptotic simultaneity bias for the S-OLS estimator is given by

$$\text{plim}\ \hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \text{plim}\left[\left(\frac{\mathbf{Q}'\mathbf{Q}}{n}\right)^{-1}\frac{\mathbf{Q}'\boldsymbol{\varepsilon}}{n}\right] \qquad (2.4).$$

In the case where $\mathbf{Q}$ contains a single exogenous regressor $\mathbf{x}$ (i.e., $k=1, \text{cov}(\boldsymbol{\varepsilon},\mathbf{x})=0$) and the error term retains no serial dependence controlling for time-lagged $\mathbf{y}$ (i.e., $\text{cov}(\mathbf{My},\boldsymbol{\varepsilon})=0$), we can rewrite equation (2.4) as

[3] Alternative distributions of $\boldsymbol{\varepsilon}$ are possible but add complication without illumination.

$$\text{plim}\,\hat{\boldsymbol{\delta}} = \begin{bmatrix} \rho \\ \phi \\ \beta \end{bmatrix} + \frac{1}{|\boldsymbol{\Psi}|} \begin{bmatrix} \text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon}) \times \text{var}(\mathbf{My}) \times \text{var}(\mathbf{x}) \\ -\text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon}) \times \text{cov}(\mathbf{Wy},\mathbf{My}) \times \text{var}(\mathbf{x}) \\ -\text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon}) \times \text{cov}(\mathbf{Wy},\mathbf{x}) \times \text{var}(\mathbf{My}) \end{bmatrix} \quad (2.5),$$

where $\boldsymbol{\Psi} = \text{plim}\left(\frac{\mathbf{Q}'\mathbf{Q}}{n}\right)$.

Since $\boldsymbol{\Psi}$ is a variance-covariance matrix, its determinant is strictly positive. Thus, when the data exhibit positive (negative) spatial and temporal dependence, the covariances in equation (2.5) will be positive (negative), and so S-OLS will over- (under-) estimate ρ and under- (over-) estimate ϕ and β. To elaborate, assuming $\mathbf{W}$ positive definite, $\text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon})$ and $\text{cov}(\mathbf{Wy},\mathbf{My})$ have the same signs as ρ and ϕ, respectively, and $\text{cov}(\mathbf{Wy},\mathbf{x})$ is non-zero if $\mathbf{x}$ exhibits spatial interdependence, say $\mathbf{x} = \theta\mathbf{W}_x\mathbf{u}$, and, assuming both $\mathbf{W}$ are positive definite, has the same sign as $\rho\theta$. The simultaneity biases of S-OLS are then as given in Table 1 below.

<u>Table 1</u>: S-OLS Simultaneity Biases	$\text{cov}(\mathbf{Wy},\mathbf{My})$			$\text{cov}(\mathbf{Wy},\mathbf{x})$		
	Positive ($\phi > 0$)	Zero ($\phi = 0$)	Negative ($\phi < 0$)	Positive ($\rho\theta > 0$)	Zero ($\rho\theta = 0$)	Negative ($\rho\theta < 0$)
$\text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon}) > 0$ $\rho > 0$	$E(\hat{\rho}) > \rho$ $E(\hat{\phi}) < \phi$	$E(\hat{\rho}) > \rho$ $E(\hat{\phi}) = \phi$	$E(\hat{\rho}) > \rho$ $E(\hat{\phi}) > \phi$	$E(\hat{\rho}) > \rho$ $E(\hat{\beta}) < \beta$	$E(\hat{\rho}) > \rho$ $E(\hat{\beta}) = \beta$	$E(\hat{\rho}) > \rho$ $E(\hat{\beta}) > \beta$
$\text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon}) = 0$ $\rho = 0$	$E(\hat{\rho}) = \rho$	$E(\hat{\rho}) = \rho$	$E(\hat{\rho}) = \rho$	$E(\hat{\rho}) = \rho$	$E(\hat{\rho}) = \rho$	$E(\hat{\rho}) = \rho$
$\text{cov}(\mathbf{Wy},\boldsymbol{\varepsilon}) < 0$ $\rho < 0$	$E(\hat{\rho}) < \rho$ $E(\hat{\phi}) > \phi$	$E(\hat{\rho}) < \rho$ $E(\hat{\phi}) = \phi$	$E(\hat{\rho}) < \rho$ $E(\hat{\phi}) < \phi$	$E(\hat{\rho}) < \rho$ $E(\hat{\beta}) > \beta$	$E(\hat{\rho}) < \rho$ $E(\hat{\beta}) = \beta$	$E(\hat{\rho}) < \rho$ $E(\hat{\beta}) < \beta$

In short, assuming positive spatial and temporal dependence, the most common case in practice, S-OLS estimation of spatial-lag models tends to overestimate the strength of spatial interdependence at the expense of unit-level and exogenous-external explanatory factors, including the temporal dynamics, all of which will tend consequently to be underestimated in proportion to their relative correlation with the spatial lag. We showed elsewhere (2004, 2006ab), however, (a) that these S-OLS simultaneity biases remain small if the true strength of interdependence remains moderate but (b) that the converse estimation strategy of ignoring the interdependence (i.e., omitting the spatial lag) biases substantive conclusions much further in the opposite directions (against interdependence, for domestic/exogenous-external), unless true spatial-interdependence is very weak. That is, non-spatial OLS estimation in contexts with even moderate spatial interdependence tends to overestimate exogenous-external and unit-level effects dramatically, again in proportion to their correlation with the omitted spatial-lag, while it fixes interdependence strength erroneously to zero by definition.

One easy way to ease or even erase the simultaneity problem with S-OLS is to lag temporally the spatial lag. To the extent that this makes the spatial lag pre-determined—that is, to the extent spatial interdependence does not have instantaneous effect, where *instantaneous* here means within an observation period, given the model—the S-OLS bias disappears. In other words, provided that the spatial-interdependence process does not have effect within an observational period, and, of course, that the spatial and temporal dynamics are adequately correctly modeled to prevent that problem arising via measurement/specification error, OLS with a temporally lagged spatial-lag on the RHS is a simple and effective estimation strategy. However, even in this best-case scenario, *OLS with time-lagged spatial-lags only provides unbiased estimates when the first observation is non-stochastic*. Elhorst (2001:128) shows that the likelihood function for the spatio-temporal lag model retains the offending Jacobian even in this case if the first observation is stochastic (see Appendix A). On the other hand, testing for either or both of remaining temporal or spatial correlation in residuals given the time-lagged spatio-temporal-lag model is possible and so strongly recommended. Standard Lagrange-multiplier tests for remaining temporal correlation in regression residuals remain valid, and see Appendix B for introduction to several tests for/measures of spatial correlation, some of which retain validity when applied to estimated residuals from models containing spatial and temporal lags.

B. Instrumental-Variables, Two-Stage-Least-Squares, and Method-of-Moments Estimation

As always with covariance-of-regressors-with-residual problems (i.e., endogeneity), instrumental variables (IV) provides one strategy for obtaining consistent parameter estimates. Generically, the IV estimator relies on finding some variable(s), $\mathbf{Z}$, that covary with the endogenous regressor(s), $\mathbf{X}$, but not with the residual, $\boldsymbol{\varepsilon}$, i.e., not with the dependent variable, $\mathbf{y}$, except through $\mathbf{X}$. Formally, if

$$\begin{array}{l} \text{plim}\left(\tfrac{1}{n}\mathbf{Z'X}\right) = \mathbf{V_{ZX}}, \\ \quad \text{a finite matrix of full column rank, and} \\ \text{plim}\left(\tfrac{1}{n}\mathbf{Z'}\boldsymbol{\varepsilon}\right) = \mathbf{0} \end{array} \qquad (2.6),$$

then the instrumental-variable estimator,

$$\mathbf{b}_{\text{IV}} = \left(\mathbf{Z'X}\right)^{-1}\mathbf{Z'y}, \qquad (2.7),$$

is consistent. In the spatial-lag-model case, $\mathbf{WX}$ may often offer an effective and readily available set of such instruments, $\mathbf{Z}$.

This section introduces the spatial-2SLS and spatial-GMM estimators for the spatio-temporal model formally for future consideration, although we do not discuss them further in this paper (yet; and we have considered S-2SLS in several previous papers). As just suggested, the S-2SLS and S-GMM provide consistent estimates for the coefficients in the spatial-lag model by using spatially weighted values of the exogenous variables in other units as instruments. S-GMM extends S-2SLS to account the heteroskedasticity that the spatial lag typically induces in the quadratic form of the sample orthogonality conditions (see, e.g., Kelejian 1993). When such second-moment conditions apply, the S-GMM estimator yields a smaller asymptotic variance than does the S-2SLS estimator.

When they do not, the two estimators are equivalent.[4]

To estimate the spatial-lag model in equations (2.2) and (2.3) by S-2SLS, define the linear prediction of $\mathbf{Wy}$ as

$$\mathbf{Wy} = \mathbf{\Pi}[\mathbf{\Pi}'\mathbf{\Pi})]^{-1}\mathbf{\Pi}'\mathbf{Wy} \qquad (2.8),$$

where $\mathbf{\Pi}$ is the full set of *exogenous* variables including, at least, $\mathbf{X}$ and $\mathbf{WX}$, with $\mathbf{WX}$ providing the *spatial* instruments.[5] Thus, $\mathbf{\Pi}$ is an $N\times L$ matrix, where $L \geq 2k$. Formally, the orthogonality condition for the S-2SLS estimator is: $E\left[\mathbf{\Pi\varepsilon}\right] = 0$.

Next, define $\hat{\mathbf{Z}}$ as an $N\times(k+1)$ matrix of the predicted values of $\mathbf{Wy}$ and $\mathbf{X}$,

$$\hat{\mathbf{Z}} = \begin{bmatrix} \mathbf{Wy} & \mathbf{X} \end{bmatrix} \qquad (2.9).$$

Using this definition, the S-2SLS estimator is

$$\hat{\boldsymbol{\delta}}_{\text{S2SLS}} = (\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'\mathbf{y} \qquad (2.10),$$

and

$$\text{var}\left(\hat{\boldsymbol{\delta}}_{\text{S2SLS}}\right) = s^2(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1} \qquad (2.11),$$

where s^2 is calculated using residuals from the original structural model, equation (2.2), with $\hat{\boldsymbol{\delta}}_{\text{S2SLS}}$ substituted for $\boldsymbol{\delta}$. In other words, S-2SLS is simply the familiar 2SLS estimator with $\mathbf{WX}$ providing the instruments for the endogenous spatial lag. In previous work (Franzese and Hays 2004), the S-2SLS estimator performed (perhaps surprisingly) well in our simulations, provided the true data-generating function met its necessary conditions: the exogeneity of $\mathbf{WX}$ and large samples (an

[4] When the number of excluded exogenous variables equals the number of endogenous variables, the GMM, 2SLS, and ILS estimators are equivalent. Therefore, more fully accurate is to say that S-GMM improves on S-2SLS given this heteroskedasticity *and* overidentification of the coefficients in the system/equation. The case we consider below has only one endogenous variable, the spatial lag, in a single equation. Thus, provided the number of exogenous variables in **X** equals or exceeds two, the number of spatial instruments will also, so the spatial-lag coefficient will be overidentified.

[5] One can also include higher-order spatial instruments in $\mathbf{\Pi}$, i.e., $\left\{\mathbf{W}^2\mathbf{X}, \mathbf{W}^3\mathbf{X}, \mathbf{W}^4\mathbf{X}, \ldots\right\}$.

unsurprising proviso). Perhaps more interestingly, violation of the exogeneity of $\mathbf{WX}$ occurs, we noted, when the $\mathbf{y}$ in some unit(s) i correlate with the $\mathbf{X}$ in some other unit(s) j in some manner beyond the dependence of i's outcome, y_{it}, on y_{jt}, a condition we called *cross-spatial endogeneity*. While such "diagonal causal arrows", $\begin{pmatrix} y_{it} & \\ & \mathbf{X}_{jt} \end{pmatrix}$, might be unlikely to arise directly in many substantive contexts, a vertical arrow, $\begin{pmatrix} \mathbf{X}_{it} \\ \\ \mathbf{X}_{jt} \end{pmatrix}$, i.e., spatially interdependent $\mathbf{X}$, and a horizontal one $\left(y_{it} \quad \rightarrow \quad \mathbf{X}_{it}\right)$, i.e., the typical (non-spatial) simultaneity problem, combine to the equivalent of a diagonal one, so violation of the "no cross-spatial endogeneity" condition might not be so rare.

The S-GMM estimator minimizes a weighted quadratic form of the sample moment conditions derived from the orthogonality assumptions. More specifically, this criterion is

$$q = E[\mathbf{m}(\boldsymbol{\delta})\boldsymbol{\Sigma}^{-1}\mathbf{m}(\boldsymbol{\delta})'] \qquad (2.12),$$

where the moment conditions are

$$\mathbf{m}(\boldsymbol{\delta}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\pi}_i(y_i - \mathbf{z}_i'\boldsymbol{\delta}) \qquad (2.13)$$

and

$$\begin{aligned}\boldsymbol{\Sigma} &= E\left[\mathbf{m}(\delta)\mathbf{m}(\delta)'\right] \\ &= \frac{1}{N}E\left[\sum_{i=1}^{N}\boldsymbol{\pi}_i\boldsymbol{\pi}_i'(y_i - \mathbf{z}_i'\boldsymbol{\delta})^2\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\omega\boldsymbol{\pi}_i\boldsymbol{\pi}_i \\ &= \frac{1}{N}\left(\boldsymbol{\Pi}'\boldsymbol{\Omega}\boldsymbol{\Pi}\right)\end{aligned} \qquad (2.14).$$

In these equations, $\boldsymbol{\pi}_i$ is a column vector ($L\times 1$) transposing the i^{th} row of $\boldsymbol{\Pi}$ (representing the i^{th}

observation) and, similarly, $\mathbf{z}_i$ is a $(k+1)\times 1$ vector that transposes the i^{th} row of $\mathbf{Z}$. Equation (2.13) is the usual least-squares optimand, expressed as the geometric problem of rendering the information remaining in y minus its linear prediction, i.e., the estimated error term, orthogonal to the information in the observation. These are sometimes called the normal equations in the geometry of least-squares regression; they are the (first-)moment conditions in the method-of-moments approach. Equations (2.12) and (2.14) show, however, that the second moment is generally heteroskedastic in the spatial case even if the stochastic error term in spatial model is homoskedastic.

The S-GMM estimator calculates the weighting matrix needed to account this heteroskedasticity by inverting some consistent estimator of the variance-covariance matrix, $\mathbf{\Sigma}$, of these moment conditions.[6] White's estimator provides a consistent estimator of $\mathbf{\Sigma}$ provided we have a consistent estimator of $\boldsymbol{\delta}$. Fortunately, S-2SLS can provide these. Thus, the estimate for $\mathbf{\Sigma}$ is

$$\mathbf{S}_0 = \sum_{i=1}^{N} \boldsymbol{\pi}_i \boldsymbol{\pi}_i' \left(y_i - \mathbf{z}_i' \hat{\boldsymbol{\delta}}_{\text{S2SLS}} \right)^2 \tag{2.15}$$

so the GMM estimator for $\boldsymbol{\delta}$ is

$$\hat{\boldsymbol{\delta}}_{\text{SGMM}} = \left[\mathbf{Z}'\mathbf{\Pi}(\mathbf{S}_0)^{-1}\mathbf{\Pi}'\mathbf{Z} \right]^{-1} \left[\mathbf{Z}'\mathbf{\Pi}(\mathbf{S}_0)^{-1}\mathbf{\Pi}'\mathbf{y} \right] \tag{2.16}$$

and the variance-covariance estimator for $\hat{\boldsymbol{\delta}}$ is

$$\text{var}\left(\hat{\boldsymbol{\delta}}_{\text{SGMM}} \right) = \left[\mathbf{Z}'\mathbf{\Pi} \left(\hat{\mathbf{S}}_0^{-1} \right) \mathbf{\Pi}'\mathbf{Z} \right]^{-1} \tag{2.17}$$

C. Maximum-Likelihood Estimation

The conditional likelihood function for the spatio-temporal-lag model, which assumes the first observation non-stochastic, is a straightforward extension of the standard spatial-lag likelihood function, which, in turn, adds only one mathematically and conceptually small complication (albeit a

[6] The logic parallels that behind the weighted-least-squares estimator, WLS. OLS is consistent in the presence of heteroskedasticity, but WLS is more efficient. Likewise, 2SLS is consistent under heteroskedasticity, but GMM is asymptotically more efficient.

computationally intense one) to the likelihood function for the standard linear-normal model. To see this, start by rewriting the spatial-lag model with the stochastic component on the left:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{XB} + \boldsymbol{\varepsilon} \Rightarrow \boldsymbol{\varepsilon} = \left(\mathbf{I} - \rho\mathbf{W}\right)\mathbf{y} - \mathbf{XB} \equiv \mathbf{Ay} - \mathbf{XB} \qquad (2.18).$$

Assuming *i.i.d.* normality, the likelihood function for $\boldsymbol{\varepsilon}$ is then the typical linear one:

$$L(\boldsymbol{\varepsilon}) = \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{NT}{2}} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right) \qquad (2.19),$$

which, in this case, will produce a likelihood in terms of $\mathbf{y}$ as follows:

$$L(\mathbf{y}) = |\mathbf{A}| \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{NT}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{Ay} - \mathbf{XB}\right)'\left(\mathbf{Ay} - \mathbf{XB}\right)\right) \qquad (2.20).$$

This still resembles the typical linear-normal likelihood, except that the transformation from $\boldsymbol{\varepsilon}$ to $\mathbf{y}$ is not by the usual factor, 1, but by $|\mathbf{A}|=|\mathbf{I}-\rho\mathbf{W}|$. Since $|\mathbf{A}|$ depends on ρ, it seemed that each time the maximum-likelihood routine recalculates the likelihood with updated estimates of ρ, it would have to recalculate the determinant at these new ρ-values. Ord (1975) redressed this computational-intensity issue by using the approximation $\prod_i \lambda_i$ for $|\mathbf{W}|$ because the eigenvector $\boldsymbol{\lambda}$ does not depend on ρ. Using $|\mathbf{I}-\rho\mathbf{W}| = \prod_i \left(1-\lambda_i\right)$ for $|\mathbf{A}|$ requires the estimation routine only to recalculate a product, not a determinant, as it updates. The estimated variance-covariances of parameter estimates follow the usual ML formula (negative the inverse of Hessian of the likelihood) and so are also functions of $|\mathbf{A}|$. The analogous strategies may serve there.

In some of our earliest explorations (e.g., at PolMeth 2004), we found that direct maximization of this likelihood using this eigenvector approximation had speed, stability, and accuracy issues, at least as implemented in the third-party spatial-econometrics routines written for *Stata*[TM] on which we had then relied. Subsequent to Ord (1975), however, a quicker, more stable, and more accurate concentrated-likelihood approach has developed. First, unpack the spatial-lag likelihood function from equation (2.20) further to:

$$\ln L(\mathbf{y}) = \ln|\mathbf{A}| - \left(\frac{N}{2}\right)\ln\pi - \left(\frac{N}{2}\right)\ln\sigma^2 - \left(\frac{1}{2\sigma^2}(\mathbf{Ay}-\mathbf{X\beta})'(\mathbf{Ay}-\mathbf{X\beta})\right) \qquad (2.21).$$

Then, given an estimate of the spatial-lag coefficient, ρ, an analytic optimum estimate of the non-spatial coefficients can be found thusly:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Ay} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \rho(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Wy} = \hat{\boldsymbol{\beta}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\beta}}_{\mathbf{L}} \qquad (2.22).$$

Note that the first term in the final expression of (2.22) is just the OLS regression of $\mathbf{y}$ on $\mathbf{X}$, and the second term is just ρ times the OLS regression of $\mathbf{Wy}$ on $\mathbf{X}$. Both rely solely on observables, (except for ρ), and so are readily calculable given some ρ (estimate). Next, define these terms:

$$\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} \equiv \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{O}}$$
and
$$\hat{\boldsymbol{\varepsilon}}_{\mathbf{L}} \equiv \mathbf{Wy} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{L}} \qquad (2.23).$$

Then:

$$\hat{\sigma}^2 = (1/N)(\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\varepsilon}}_{\mathbf{L}})'(\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\varepsilon}}_{\mathbf{L}}) \qquad (2.24)$$

yields the S-ML estimate of the standard-error of the regression, and

$$\ln L_C(\mathbf{y}) = -\left(\frac{N}{2}\right)\ln\pi + \ln|\mathbf{A}| - \frac{N}{2}\ln\left(\frac{1}{N}(\boldsymbol{\varepsilon}_{\mathbf{O}} - \rho\boldsymbol{\varepsilon}_{\mathbf{L}})'(\boldsymbol{\varepsilon}_{\mathbf{O}} - \rho\boldsymbol{\varepsilon}_{\mathbf{L}})\right) \qquad (2.25)$$

yields the S-ML estimate of ρ. These latter substituted into (2.22) yields $\hat{\boldsymbol{\beta}}$. The procedure may be iterated, and estimated variance-covariances of parameter estimates derive from the information matrix as usual, although they could also be bootstrapped. Routines written in *MatLab*^{TM} to employ this concentrated-likelihood approach (LeSage XXXX, www.spatial-econometrics.com), still using the eigenvector approximation, have thusfar proven much more reliable.

Now, as mentioned, the likelihood for the spatio-temporal model is a straightforward extension of this spatial-lag likelihood. Written in ($N \times 1$) vector notation, spatio-temporal-model conditional-likelihood is mostly conveniently separable into parts, as seen here:

$$\operatorname{Log} f_{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_2 | \mathbf{y}_1} = -\frac{1}{2} N(T-1) \log\left(2\pi\sigma^2\right) + (T-1)\log\left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{1}{2\sigma^2}\sum_{t=2}^{T} \boldsymbol{\varepsilon}_t' \boldsymbol{\varepsilon}_t \qquad (2.26),$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \rho\mathbf{W}_N\mathbf{y}_t - \phi\mathbf{I}_N\mathbf{y}_{t-1} - \mathbf{X}_t\boldsymbol{\beta}$. This separation conveniently implies, for example, that the strategy that Beck and Katz (1995) suggested in a related context (in introducing and discussing PCSE's), the strategy of modeling the temporal dynamics separately and considering the properties of the spatial model remaining after conditioning on that temporal-dynamic model, will generally work effectively in the spatio-temporal-lag context also.

Penultimately, the issue of stationarity arises in more-complicated fashion in spatio-temporal dynamic models than in purely temporally dynamic ones. Nonetheless, the conditions and issues arising in the former are reminiscent although not identical to those arising in the latter. Define $\mathbf{B} = \phi\mathbf{I}$, $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$, and ω as a characteristic root of $\mathbf{A}$, the statio-temporal process generating the data is covariance stationary if

$$\left|\mathbf{B}\mathbf{A}^{-1}\right| < 1 \qquad (2.27),$$

or, equivalently, if

$$\begin{cases} |\phi| < 1 - \rho\omega_{\max}, & \text{if } \rho \geq 0 \\ |\phi| < 1 - \rho\omega_{\min}, & \text{if } \rho < 0 \end{cases} \qquad (2.28).$$

For example, in the contexts our simulations below consider, that of positive temporal dependence and positive, uniform spatial dependence ($\rho > 0$ and $w_{ij} = 1/(N-1)\ \forall\ i \neq j$), stationarity requires simply that $\phi + \rho < 1$. In fact, the maximum characteristic root is +1 for any row-standardized $\mathbf{W}$.

Finally, we note that he unconditional (exact) likelihood function, the one that retains the first time-period observations as non-predetermined, is more complicated (Elhorst 2001, 2003, 2005).[7]

$$\begin{aligned}\text{Log} f_{\mathbf{y}_t,\ldots,\mathbf{y}_1} &= -\frac{1}{2}NT\log\left(2\pi\sigma^2\right)+\frac{1}{2}\sum_{i=1}^{N}\log\left(\left(1-\rho\omega_i\right)^2-\phi^2\right)+\left(T-1\right)\sum_{i=1}^{N}\log\left(1-\rho\omega_i\right)\\ &\quad -\frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t'\varepsilon_t-\frac{1}{2\sigma^2}\varepsilon_1'\left(\left(\mathbf{A}-\mathbf{B}\right)'\right)^{-1}\left(\mathbf{A}'\mathbf{A}\text{ - }\mathbf{A}'\mathbf{B}\mathbf{A}^{-1}\left(\mathbf{A}'\mathbf{B}\mathbf{A}^{-1}\right)'\right)^{-1}\left(\mathbf{A}\text{ - }\mathbf{B}\right)^{-1}\varepsilon_1\end{aligned} \tag{2.29}$$

where $\boldsymbol{\varepsilon}_1=\mathbf{y}_1-\rho\mathbf{W}_N\mathbf{y}_1-\phi\mathbf{I}_N\mathbf{y}_1-\mathbf{X}_1\boldsymbol{\beta}$. When T is small, the first observation contributes greatly to the overall likelihood, and the unconditional likelihood should be used to estimate the model. In other cases, the more compact conditional likelihood is acceptable for estimation purposes.

III. Calculating and Presenting Spatio-Temporal Effects

Calculation, interpretation, and presentation of effects in empirical models with spatio-temporal interdependence, as in any model beyond the strictly linear-additive (in variables and parameters, explicitly and implicitly[8]), involve more than simply considering coefficient estimates. *Coefficients* do *not* generally equate to *effects* beyond that simplest strictly linear-additive world. In empirical models containing spatio-temporal dynamics, as in those with only temporal dynamics, for example, coefficients on explanatory variables give only the pre-dynamic impetuses to the outcome variable from changes in those variables. The coefficients represent only the (often inherently unobservable) pre-interdependence impetus to outcomes associated with each RHS variable.

This section discusses the calculation of spatio-temporal multipliers, which allow expression of the effects of counterfactual shocks of various kinds to some unit(s) on itself (themselves) and other

[7] Note that the same condition that complicates ML estimation of the spatio-temporal lag model, namely the first set of observations is stochastic, also invalidates the use of OLS to estimate a model with a temporally lagged spatial lag. (The Appendix presents this alternative likelihood.) Hence, asymptotically, this consideration offers no econometric reason to prefer S-OLS over S-ML estimation of spatio-temporal-lag models or the converse.

[8] For example, the familiar (a) linear-interaction models are explicitly nonlinear in variables although linear-additive in parameters; (b) logit/probit class of models are explicitly nonlinear in both variables and parameters; and (c) temporally dynamic models of all sorts are implicitly nonlinear in parameters and sometimes in variables too (via the presence of terms like $\rho\beta X_{t-s}$ implicitly in the right-hand-side lag terms). Spatial-lag models are likewise implicitly nonlinear-additive. In any of these cases, i.e., in all models beyond those with strictly linear-additively separable right-hand-side terms, like the introductory textbook linear-regression model, *coefficients* and *effects* are very different things.

units over time, accounting both the temporal and spatial dynamics. These multipliers also allow expression the long-run, steady-state, or equilibrium impact of permanent such shocks. In this section, we also apply the delta-method to derive analytically the asymptotic approximate standard errors for these response-path and long-run effect estimates.[9]

Calculating the cumulative, steady-state spatio-temporal effects is most convenient working with the spatio-temporal-lag model in (Nx1) vector form:

$$\mathbf{y}_t = \rho\mathbf{W}\mathbf{y}_t + \phi\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \qquad (3.1).$$

To find the long-run, steady-state, equilibrium (cumulative) level of $\mathbf{y}$, simply set $\mathbf{y}_{t-1}$ equal to $\mathbf{y}_t$ in (3.1) and solve. This gives the steady-state effect, assuming stationarity and that the exogenous RHS terms, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, remain permanently fixed to their hypothetical/counterfactual levels:[10]

$$\begin{aligned}
\mathbf{y}_t &= \rho\mathbf{W}\mathbf{y}_t + \phi\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\
&= \left(\rho\mathbf{W} + \phi\mathbf{I}\right)\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\
&= \left[\mathbf{I}_{\mathrm{N}} - \rho\mathbf{W} - \phi\mathbf{I}_{\mathrm{N}}\right]^{-1}\left(\mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right) \\
&= \begin{bmatrix} 1-\phi & -\rho w_{1,2} & & & -\rho w_{1,\mathrm{N}} \\ -\rho w_{2,1} & 1-\phi & & & \\ & & & & \\ & & & 1-\phi & -\rho w_{(\mathrm{N}-1),\mathrm{N}} \\ -\rho w_{\mathrm{N},1} & & & -\rho w_{\mathrm{N},(\mathrm{N}-1)} & 1-\phi \end{bmatrix}^{-1}\left(\mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right) \\
&\equiv \mathbf{S}\left(\mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right)
\end{aligned} \qquad (3.2).$$

To offer standard-error estimates for these steady-state estimates, one could use the delta method. I.e., give a first-order[11] Taylor-series linear-approximation to nonlinear (3.2) around the estimated

[9] For an excellent discussion of spatial multipliers, see Anselin (2003).
[10] The counterfactual addressed here is usually the steady-state effect of *permanent* shocks; since, given stationarity, the long-run steady-state effect of a temporary shock is zero.
[11] First-order approximations are almost universal in practice, although mounting simulation evidence supports second or higher orders in some specific modeling contexts. One suspects the second- or higher orders could be important here, given the highly nonlinear nature of **S**, perhaps compounded by the heteroskedasticity induced by spatio-temporal dynamics. Simulations below are oddly intriguing, however, in that S-ML standard errors for the steady-state-effect estimates outperform those for the coefficients themselves. We suspect counteracting biases at those stages.

parameter-values and determine the asymptotic variance of that linear approximation. To find the key elements needed for this, begin by denoting the i^{th} column of $\mathbf{S}$ as $\mathbf{s}_i$ and its estimate as $\hat{\mathbf{s}}_i$. The steady-state spatio-temporal effects of a one-unit increase in explanatory variable *k* in country *i* are $\mathbf{s}_i\beta_k$ giving delta-method standard-errors of

$$\mathbf{V}\left(\hat{\mathbf{s}}_i\hat{\beta}_k\right)=\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\mathbf{V}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]' \quad (3.3),$$

where $\hat{\boldsymbol{\theta}}\equiv\begin{bmatrix}\hat{\rho} & \hat{\phi} & \hat{\beta}_k\end{bmatrix}'$, $\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\equiv\begin{bmatrix}\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\rho}} & \frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\phi}} & \hat{\mathbf{s}}_i\end{bmatrix}$, and the vectors $\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\rho}}\right]$ and $\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\phi}}\right]$ are the i^{th} columns of $\hat{\beta}_k\hat{\mathbf{S}}\mathbf{W}\hat{\mathbf{S}}$ and $\hat{\beta}_k\hat{\mathbf{S}}\hat{\mathbf{S}}$ respectively. We will explore in simulation below how well these analytic, but asymptotic and approximate, delta-method standard-errors perform. Ultimately, we intend to do so both absolutely (as in our simulations) and relatively to boot-strapping, which is always an alternative and often a robust and effective one.

The spatio-temporal response path of the $N\times 1$ vector of unit outcomes, $\mathbf{y}_t$, to the exogenous RHS terms, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, could also emerge by rearranging (3.1) to isolate $\mathbf{y}_t$ on the LHS:

$$\mathbf{y}_t=\left[\mathbf{I}_N-\rho\mathbf{W}_N\right]^{-1}\left\{\phi\mathbf{y}_{t-1}+\mathbf{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t\right\} \quad (3.4).$$

This formula gives the response-paths of all unit(s) $\{i\}$ to hypothetical shocks to $\mathbf{X}$ or $\boldsymbol{\varepsilon}$ in any unit(s) $\{j\}$, including a shock in $\{i\}$ itself/themselves, just by setting $\left(\mathbf{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t\right)$ to one in the row(s) corresponding to $\{j\}$. This formulation may be especially convenient for plotting estimated response paths in a spreadsheet, for example. To calculate marginal spatio-temporal effects (non-cumulative), i.e., the incremental change at some time *t*+*k* in the over-time path resulting from a permanent one-unit change in an explanatory variable at time *t*, and their standard errors, working with the entire NT x NT matrix is easier. Simply redefine $\mathbf{S}$ in the (3.2) as $\mathbf{S}\equiv\left[\mathbf{I}_{\text{NT}}-\rho\mathbf{W}-\phi\mathbf{M}\right]^{-1}$ and follow the steps outlined above. We calculate these effects for the presentation of our empirical reanalysis below, for example.

IV. Monte Carlo Simulations

A. Experimental Design

We generate data using the reduced form of the spatio-temporal model and then estimate the structural model using both OLS and ML. The structural model is

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \phi\mathbf{M}\mathbf{y} + \beta\mathbf{x} + \boldsymbol{\varepsilon} \quad (4.1),$$

where

$$\mathbf{x} = (\mathbf{z} + \theta\mathbf{W}\mathbf{z}) \quad (4.2)$$

and

$$\boldsymbol{\varepsilon} = g \times (\mathbf{u} + \lambda\mathbf{M}\mathbf{u}) \quad (4.3).$$

The reduced form is thus

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{M})^{-1}\left[\beta(\mathbf{z} + \theta\mathbf{W}\mathbf{z}) + (\mathbf{u} + \lambda\mathbf{M}\mathbf{u})\right] \quad (4.4)$$

We draw $\mathbf{z}$ and $\mathbf{u}$ as standard normals to generate $\mathbf{y}$. g scales the variance of $\boldsymbol{\varepsilon}$ relative to $\mathbf{x}$. Non-zero values of θ will correlate $\mathbf{x}$ with $\mathbf{Wy}$. Non-zero values of λ will render $\mathbf{My}$ endogenous.

B. Experimental Results

We concentrated our experiments in sample-size and parameter ranges that our experiences have suggested is typical of panel datasets in comparative political economy: $N = (15, 25)$, $T = (35, 45)$, $\rho = \theta = .2$, $\phi = .4$, and $\lambda = (0, .1)$. The results of our experiments are presented in Tables 2-3. The results are easy enough to summarize. In terms of bias, for both the parameter estimates and the estimated long-run steady-state effect of permanent shocks, S-ML typically outperforms S-OLS, although usually not dramatically. In terms of efficiency too or, rather, mean-squared-error (not reported), S-ML also typically outperforms S-OLS for both types of estimates, usually a bit more noticeably. In terms of reported standard-error accuracy, *excepting the reported standard-errors for* $\hat{\rho}$*, the crucial strength-of-interdependence parameter*, S-ML and S-OLS perform equivalently.

TABLE 2: Homogenous Interdependence Pattern $\{(N\text{-}1)^{-1}\}$

				BIAS		STANDARD ERROR ACCURACY	
			Coeff.	**S-OLS**	**S-ML**	**S-OLS**	**S-ML**
ρ=.2 φ=.4 β=1 θ=.2 λ=0	**N=15**	**T=35**	$\hat{\rho}$	.258	.159	.774	.404
			$\hat{\phi}$	.391	.394	.980	.979
			$\hat{\beta}$	1.038	1.033	.988	.979
		T=45	$\hat{\rho}$	.272	.167	.822	.418
			$\hat{\phi}$	.392	.395	1.031	1.033
			$\hat{\beta}$	.927	.933	1.159	1.036
	N=25	**T=35**	$\hat{\rho}$	.269	.167	.835	.329
			$\hat{\phi}$	.394	.396	.983	.983
			$\hat{\beta}$	1.057	1.062	1.024	1.022
		T=45	$\hat{\rho}$	.274	.169	.866	.334
			$\hat{\phi}$	.395	.397	1.003	1.004
			$\hat{\beta}$	.960	.957	.991	.985
ρ=.2 φ=.4 β=1 θ=.2 λ=.1	**N=15**	**T=35**	$\hat{\rho}$	.243	.155	.823	.410
			$\hat{\phi}$	.468	.470	1.011	1.019
			$\hat{\beta}$	.901	.902	.983	.975
		T=45	$\hat{\rho}$	.268	.170	.822	.418
			$\hat{\phi}$	.465	.468	1.031	1.033
			$\hat{\beta}$	.625	.628	1.041	1.036
	N=25	**T=35**	$\hat{\rho}$	.230	.148	.752	.296
			$\hat{\phi}$	.471	.473	1.170	1.163
			$\hat{\beta}$	1.107	1.103	1.081	1.070
		T=45	$\hat{\rho}$	.258	.164	.866	.334
			$\hat{\phi}$	.475	.476	1.003	1.004
			$\hat{\beta}$	.819	.830	.991	.985

Notes: Standard error accuracy is the ratio of the mean estimated standard-error for each coefficient to the standard deviation of its (experimental) sampling distribution. Results < 1.0 imply overconfidence.

For $\hat{\rho}$, however, S-ML on average reports wildly over-confident standard errors (standard errors

60-70% too small) whereas S-OLS is more mildly over-confident in its reported uncertainty (standard errors 15-25% too small). This "feature" of S-ML estimation has emerged in many of our previous experiments as well, especially in those with relatively modest strength-of-interdependence like the $\rho = .2$ used here. Oddly, though, the relative overconfidence of the estimators reverses when we consider the estimated cumulative steady-state effect of permanent shocks (Table 3), with S-ML reporting relatively accurate 5-20% under-confident standard-errors and S-ML reporting standard errors a whopping 130-190% too large. Perhaps the upshot is to use S-ML for parameter and effect estimation but bootstrap standard errors for parameter estimates rather than rely on these asymptotic formulae (steady-state-effect standard-errors seem more reliably approximated by the delta method). We intend to explore this strategy in future work.

TABLE 3: Long-Run Steady-State Spatio-Temporal Effects.

			BIAS/CONSISTENCY		STANDARD ERROR ACCURACY	
			S-OLS	**S-ML**	**S-OLS**	**S-ML**
ρ=.2 φ=.4 β=1 θ=.2 λ=0	**N=15**	**T=35**	.907	.894	2.572	1.178
		T=45	.909	.914	2.501	1.177
	N=25	**T=35**	.927	.938	2.695	1.214
		T=45	.808	.804	2.473	1.160
ρ=.2 φ=.4 β=1 θ=.2 λ=.1	**N=15**	**T=35**	.909	.919	2.868	1.182
		T=45	.683	.706	2.395	1.113
	N=25	**T=35**	1.246	1.226	2.519	1.271
		T=45	.980	1.014	2.486	1.059

Notes: Results are expressed as ratios of the estimated coefficients/standard errors to the true coefficients/standard errors.

V. Reanalysis

In this section, we reanalyze the Beck, Gleditsch, and Beardsley (henceforth BG&B) model of directed export flows among major powers using a contemporaneous spatial lag and the conditional ML estimator. Our purpose is not to criticize BG&B's analysis but rather to build on what they recommend by illustrating how to calculate and present some of the spatio-temporal effects implied by their model. We agree with BG&B that theory should drive our spatio-temporal specification choices (with openness to being informed and refined by empirical results), and we find little in their empirical results and conclusions with which to disagree. Indeed, we choose this article for our re-analysis precisely because it represents the start of the art in our view. However, BG&B, citing the relative difficulty in implementing S-ML estimator for panel and TSCS data and the theretofore-apparent lack of an unambiguously superior estimator in such data,[12] estimate their spatio-temporal models exclusively by S-OLS with time-lagged spatial-lags. As we have shown above, though, the conditional S-ML estimator is relatively straightforward to specify. Programs are also available to calculate these estimates now (thanks to Lesage, Elhorst, and others). And, with relatively large T (as in BG&B's case), analysts should not hesitate to use conditional S-ML and these programs when theory suggests that initial spatial-effects are likely to occur quickly (i.e., within period, as in BG&B's case, in our opinion) or when they harbor suspicions that temporal or spatial-dynamic misspecification might induce contemporaneous interdependence (as in BG&B's case, by their own assessment[13]). Accordingly, we reformulate BG&B's empirical model to a contemporaneous rather than a time-lagged spatial-lag and re-estimate by conditional S-ML before proceeding to illustrate the calculation and presentation of spatio-temporal dynamic and steady-state effects.

There are seven major powers during the period BG&B examine, and their unit of analysis is the directed dyad, giving 42 total directed-dyads ($N\times(N\text{-}1)$). We are comfortable with conditioning on

[12] Indeed, an unambiguously superior estimator for all data conditions and spatial-dependence processes has not quite emerged yet, and may be unlikely to emerge, although S-ML has increasingly established itself as the leading candidate in our experiments to date.
[13] BG&B note that their models fail Lagrange-multiplier tests for remaining residual (temporal) autocorrelation, with residual (temporal) autocorrelation on the order of 0.1.

the first set of observations because the average number of observations for each dyad is 61 years. The first set of observations represents less than 2% of the full dataset and therefore contributes relatively little to the overall value of the likelihood function.

Our specification differs slightly from BG&B. We use a contemporaneous spatial lag (as noted above) and include the full sets of dyad and year indicators (space and time fixed effects). We see no reason to prefer on *a priori* theoretical grounds a time-lagged spatial-lag over a contemporaneous one, we rely on the data to help choose our specification.[14] Note that our estimates of the strength of spatio-temporal interdependence are highly conservative due to the inclusion of the dyad and year dummies. Omitted unit effects can inflate coefficient estimates for temporal lags (Judson and Owen 1999) and omitted period effects can inflate coefficient estimates for spatial lags (Franzese and Hays 2004). Including unit and period dummies is an extremely conservative way to control for these effects because the converses are also true in limited samples; that is, insofar as effects are not fixed but simply correlated across time and space, inclusion of those fixed effects will tend to deflate estimates of spatio-temporal dependence (Beck and Katz 2001 make a related point).

Our results are reported in Table 5. Several of our estimates differ from BG&B's in predictable ways (compare with Table 4 from Beck et al. (2006:41)). For example, our coefficient estimates on slowly changing variables (e.g., joint democracy, distance, alliance) are much smaller than BG&B's estimates because we include dyad dummies.[15] Our coefficient estimate for the temporal lag is also

---

[14] If we replace the time-lagged spatial-lag in the BG&B model with a contemporaneous one and estimate by conditional S-ML, the estimated coefficient is approximately three times larger than the estimate reported by BG&B in Table 4 (.06 vs. .02). (Likelihood and R^2 are also higher, but these models are non-nested and have the same degrees of freedom, so those quantities are not directly comparable and standard statistical tests would not apply.) On theoretical grounds, one might argue that, because the "spatially" connected dyads are dyads with a common member (e.g., US-Germany and US-Russia) the idea that the effects are instantaneous (i.e., occur within one year) is highly plausible. The same factors that cause the US to alter its trade with Germany would likely cause it to adjust trade with Russia more or less instantaneously and, in any event, largely within one year. Note too that with a spatio-temporal specification spatial effects are not entirely felt instantaneously but rather unfold over time given the temporal dynamics. The maximum marginal effect could even occur several years after the initial effect.

[15] Due to political division and reunification, distance, a seemingly time invariant measure, does change for the dyads

noticeably smaller, and our estimate for the spatial lag is almost five times larger. Both these changes most likely are attributable to our inclusion of a contemporaneous spatial-lag and the dyad dummies, and also to the estimate of quicker temporal dynamics.

<u>Table 4</u>: Reanalysis of BG&B. Directed Export Flows, Major Powers, 1907-1990.

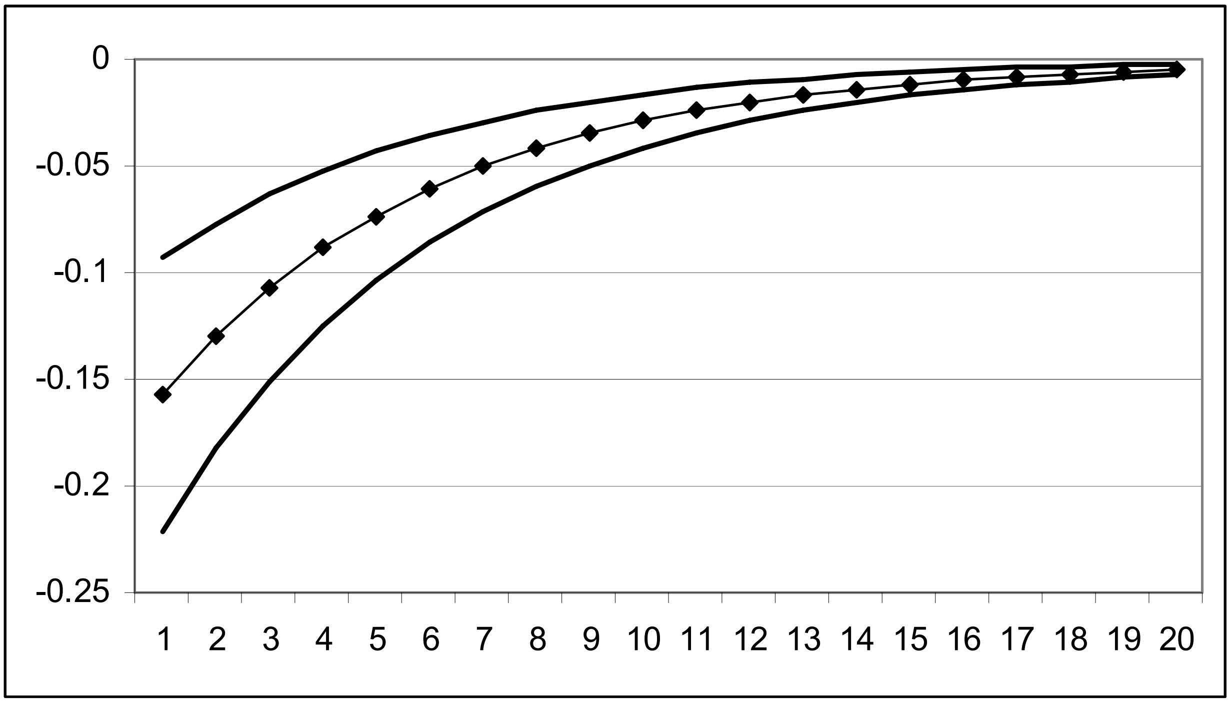
Spatio-Temporal Model Estimates					
$R^2 = 0.9702$				$\bar{R}^2 = 0.9687$	
$\hat{\sigma}^2 = 0.0939$				N, k = 2565, 120	
LL = 282.677				# of iterations = 12	
Variable	**BG&B Coeff**	**Our Coeff**	**BG&B s.e.**	**Our s.e.**	**p-level**
constant	0.17	-0.5888	(0.11)	0.6130	0.3367
Ln GNP A	0.02	*0.0276*	(0.01)	0.0161	*0.0861*
Ln GNP B	0.03	-0.0034	(0.01)	0.0156	0.8272
Ln Pop A	0.04	0.0561	(0.02)	0.0650	0.3879
Ln Pop B	0.03	0.0593	(0.02)	0.0594	0.3183
Ln Dist	-0.04	-0.0041	(0.01)	0.0728	0.9551
Ln tau-b	0.11	-0.0633	(0.06)	0.0598	0.2901
Ln Dem	0.14	*0.0883*	(0.03)	*0.0374*	*0.0181*
Ln MID	-0.20	*-0.1570*	(0.04)	0.0389	*0.0001*
Ln Multipolar	-0.28	-0.0553	(0.05)	0.0559	0.3226
Ln Bipolar	-0.04	0.0317	(0.05)	0.0546	0.5613
Temporal Lag	0.91	*0.8252*	(0.01)	*0.0100*	*0.0000*
Spatial Lag	0.02	*0.0980*	(0.01)	*0.0385*	*0.0110*
Coefficients for Fixed Dyad and Year Effects Suppressed to Save Space.					

Note that the sum of our coefficient estimates for the temporal and spatial lags is less than one, which suffices to show the process stationary in this case. Therefore, we can calculate the spatio-temporal effects along the lines described in Section III. In Figures 1-3, we present the over-time path of the marginal (i.e., the year-by-year incremental, not the cumulative) spatio-temporal effects from a permanent one-unit increase in the MID variable. These figures show three types of estimated responses to this counterfactual: Figure 1 shows the temporal effects with spatial feedback (effect of a US-Russia MID on US exports to Russia over time; Figure 2 gives the first-order spatio-temporal

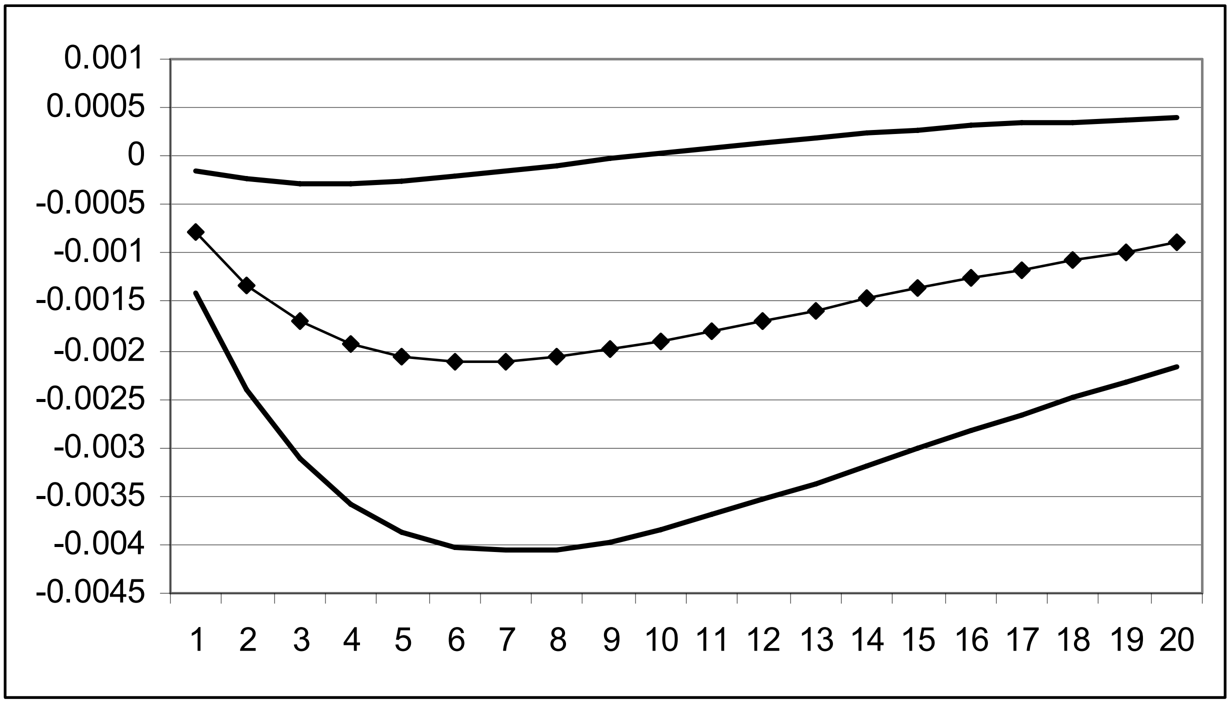
including Germany.

effects (effect of a US-Russia MID on US exports to Germany over time; and Figure 3 shows the second-order spatio-temporal effects (effect of a US-Russia MID on German exports to Russia over time). The cumulative (20-year) type 1 response to a permanent one-unit increase in the MID variable is to decrease the log of exports by almost -.90 (approximately 90%). The two other effects are smaller in size, take longer to reach their maximum, with increments fading more slowly.

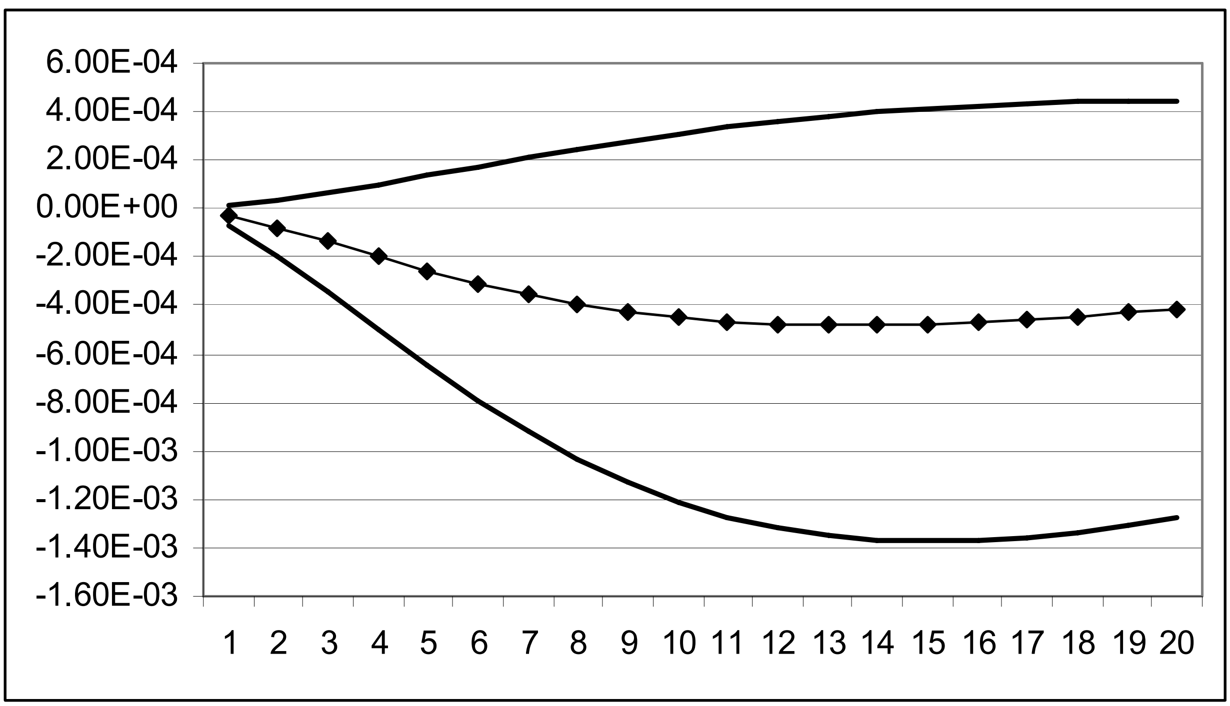
<u>Figure 1</u>: Temporal Effects with Spatial Feedback (E.g., US Exports to Russia response to US-Russia MID)



<u>Figure 2</u>: First Order Spatio-temporal Effects (E.g., US Exports to Germany response to US-Russia MID)



<u>Figure 3</u>: 2nd-Order Spatio-temporal Effects (E.g., German Exports to Russia response to US-Russia MID)



VI. Conclusion

Wasn't this fun? Thanks for your patience, and stay tuned for more introductions, analytics, simulations, conclusions, advice on interpretation and presentation and, eventually, some programs to implement all of this in Stata™ so practicing empirical political scientists might actually use it!

References

Anselin, Luc. 2003. "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics." *International Regional Science Review* 26(2);153-66.

Anselin, Luc, et al. 1996. "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics* 26:77-104.

Beck, Nathaniel N. and Jonathan Katz. 2001. "Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon," *International Organization* 55(2):487-95.

Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006. "Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy, *International Studies Quarterly* 50(1):27-44.

Darmofal. 2006. "Spatial Econometrics and Political Science." http://polmeth.wustl.edu/retrieve.php?id=575

Elhorst, J. Paul. 2001. "Dynamic Models in Space and Time." *Geographical Analysis* 33(2):119-140.

Elhorst, J. Paul. 2003. "Specification and Estimation of Spatial Panel Data Models." *International Regional Science Review* 26(3):244-268.

Elhorst, J. Paul. 2005. "Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels." *Geographical Analysis* 37(1):85-106.

Franzese & Hays. 2004. "Empirical Modeling Strategies for Spatial Interdependence: Omitted-Variable vs. Simultaneity Biases," presented to PolMeth 2004 & RC33 6th International Conference on Social Science Methodology 2004. http://www.umich.edu/~franzese/FranzeseHays.PolMeth.2004.pdf

Franzese & Hays. 2006a. "Spatial Econometric Models for the Analysis of TSCS Data in Political Science," *Political Analysis*, in review for TSCS special issue.

Franzese & Hays. 2006b. "Strategic Interaction among EU Governments in Active-Labor-Market Policymaking: Subsidiarity and Policy Coordination under the European Employment Strategy," *European Union Politics* 7(2):167-89.

Judson, Ruth A and Ann L Owen (1999): "Estimating dynamic panel data models: a guide for macroeconomists." *Economics Letters* 65:9-15.

Appendix A: Unconditional Likelihood for Time-Lagged Spatio-Temporal Model

The spatio-temporal model with time-lagged dependent variable and time-lagged spatial-lag is

$$\mathbf{y}_t = \eta \mathbf{W} \mathbf{y}_{t-1} + \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t .$$

If the first set of observations is stochastic, the unconditional (exact) log-likelihood is

$$\ln f_{\mathbf{y}_t,\ldots,\mathbf{y}_1} = -\frac{1}{2} NT \times \ln\left(2\pi\sigma^2\right) + \frac{1}{2}\sum_{i=1}^{N} \ln\left(1-\left(\phi+\eta\omega_i\right)^2\right) - \frac{1}{2\sigma^2}\sum_{t=2}^{T} \varepsilon_t' \varepsilon_t$$
$$-\frac{1}{2\sigma^2} \varepsilon_1' \left(\left(\mathbf{I}-\mathbf{B}\right)'\right)^{-1} \left(\mathbf{I}-\mathbf{B}\mathbf{B}'\right)^{-1} \left(\mathbf{I}-\mathbf{B}\right)^{-1} \varepsilon_1$$

where $\boldsymbol{\varepsilon}_1 = \mathbf{y}_1 - \left(\phi + \eta \mathbf{W}_N\right)\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \eta \mathbf{W}_N \mathbf{y}_{t-1} - \phi \mathbf{y}_{t-1} - \mathbf{X}_t \boldsymbol{\beta}$, and $A = \phi I + \eta W$. For the derivation of this likelihood function, see Elhorst (2001, 126-130). Note that the second term in the likelihood function causes the OLS estimator to be biased. Asymptotically $\left(T \to \infty\right)$, this bias goes to zero.

Appendix B: Some Tests for Contemporaneous Correlation in Regression Residuals

Moran's I,

$$I = \frac{N}{S}\frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}} \qquad (2.1),$$

where

$$S = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \qquad (2.2),$$

is anologous to the Durbin-Watson family of time-serial-correlation tests against the specific form of correlation specified in the test (here, by $\mathbf{W}$).

A family of *Spatial Lagrange-Multiplier Tests* exist as well. (This section draws heavily from (Anselin et al. 1996). A one-directional test against a spatial-error-process alternative is:

$$LM_{\lambda} = \frac{\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T} \qquad (2.3),$$

where

$$T = \text{tr}[(\mathbf{W}' + \mathbf{W})\mathbf{W}] \qquad (2.4).$$

A robust one-directional test against a spatial-error alternative is

$$LM_{\lambda}^{*} = \frac{\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2} - \left[\mathbf{T}\hat{\sigma}_{\varepsilon}^{2}\left(\mathbf{G} + \mathbf{T}\hat{\sigma}_{\varepsilon}^{2}\right)^{-1}\right]\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{\mathbf{T}\left[1 - \frac{1}{\hat{\sigma}_{\varepsilon}^{2}}\left(\mathbf{G} + \mathbf{T}\hat{\sigma}_{\varepsilon}^{2}\right)\right]^{-1}} \qquad (2.5),$$

where

$$\mathbf{G} = (\mathbf{WX}\hat{\boldsymbol{\beta}})'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{WX}\hat{\boldsymbol{\beta}}) \qquad (2.6).$$

A one-directional test against a spatial-lag (of $\mathbf{y}$) alternative is

$$LM_{\rho} = \frac{\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{\mathbf{G} + \mathbf{T}\hat{\sigma}_{\varepsilon}^{2}} \quad (2.7).$$

And a robust one-directional test against spatial-lag alternative is

$$LM_{\rho}^{*} = \frac{\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y} / \hat{\sigma}_{\varepsilon}^{2} - \hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{\mathbf{G} + \mathbf{T}(\hat{\sigma}_{\varepsilon}^{2} - 1)} \quad (2.8).$$

Note also that any of these global tests can be conducted locally as well (see, e.g., Darmofal 2006).

Unfortunately, none of these tests are valid with a spatial-lag included among the regressors in the model generating the estimated residuals to be tested. For those purposes, likelihood-ratio tests of the alternative models, with and without the spatial lag, or Anselin et al. (1996) LM_{a} and LM_{a}^{*} are appropriate. Other valid tests for such considerations may exist, the Kelejian-Robinson diagnostic may serve, for example, but these remain to be analyzed or tested for their utility toward such purposes. The relative power and accuracy of these valid tests for remaining spatial correlation has also not been well explored in the literature to our knowledge.