doi:10.1093/pan/mpm005

Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data

Robert J. Franzese, Jr.
Department of Political Science, University of Michigan, Ann Arbor, MI 48109
e-mail: franzese@umich.edu (corresponding author)

Jude C. Hays
Department of Political Science, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL 61801
e-mail: jchays@uiuc.edu

In this paper, we demonstrate the econometric consequences of different specification and estimation choices in the analysis of spatially interdependent data and show how to calculate and present spatial effect estimates substantively. We consider four common estimators—nonspatial OLS, spatial OLS, spatial 2SLS, and spatial ML. We examine analytically the respective omitted-variable and simultaneity biases of nonspatial OLS and spatial OLS in the simplest case and then evaluate the performance of all four estimators in bias, efficiency, and SE accuracy terms under more realistic conditions using Monte Carlo experiments. We provide empirical illustration, showing how to calculate and present spatial effect estimates effectively, using data on European governments' active labor market expenditures. Our main conclusions are that spatial OLS, despite its simultaneity, performs acceptably under low-to-moderate interdependence strength and reasonable sample dimensions. Spatial 2SLS or spatial ML may be advised for other conditions, but, unless interdependence is truly absent or minuscule, any of the spatial estimators unambiguously, and often dramatically, dominates on all three criteria the nonspatial OLS commonly used currently in empirical work in political science.

1 Introduction

Empirical work in political science, while recognizing that panel and time-series-cross-section (TSCS) data usually correlate across both time and space, has tended

Authors' note: This research was supported in part by National Science Foundation grant no. 0318045. We thank Chris Achen, Jim Alt, Kenichi Ariga, Neal Beck, Jake Bowers, Kerwin Charles, Bryce Corrigan, Tom Cusack, David Darmofal, Jakob de Haan, John Dinardo, Zach Elkins, John Freeman, Fabrizio Gilardi, Kristian Gleditsch, Mark Hallerberg, John Jackson, Aya Kachi, Jonathan Katz, Mark Kayser, Achim Kemmerling, Gary King, Hasan Kirmanoglu, James Kuklinski, Tse-Min Lin, Xiaobo Lu, Walter Mebane, Covadonga Meseguer, Michael Peress, 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 other work in our broader project on spatial econometric models in political science. Bryce Corrigan, Aya Kachi, and Xiaobo Lu each provided excellent research assistance and Kristian Gleditsch, Mark Hallerberg, and Duane Swank also generously shared data. We alone are responsible for any errors.

to model temporal dependence directly. But spatial interdependence, if considered at all, is modeled solely as a nuisance to be "corrected" or to which standard error (SE) estimates should be made "robust" panel-corrected standard errors (PCSE). That is, current practice in political science (much of our own past work included) relies almost exclusively on nonspatial or, at most, "nuisance spatial" models. As argued and discussed elsewhere in this issue, direct modeling of temporal dynamics is methodologically and substantively laudable, and we follow that practice explicitly in our empirical example and at least implicitly in our analytic and simulation studies. This paper emphasizes cross-sectional interdependence, i.e., spatial dependence across units. In it, we determine and discuss the econometric consequences of alternative specification and estimation choices in empirical analyses of such spatial interdependence in TSCS data and show how to calculate and present substantively interesting spatial effects.

Relegating spatial interdependence to nuisance is often problematic on econometric grounds and, more often and importantly still, on substantive grounds because many political science contexts imply substantively meaningful and interesting interdependence across units of observation. The frequently ignored spatial relationships in TSCS data are an important part of the politics that political scientists aim to study. In comparative politics, for example, likelihoods and outcomes of demonstrations, riots, coups, and/or revolutions in one country almost certainly depend in substantively crucial ways on such occurrences in other countries (e.g., through *demonstration effects* or *snowballing*). In U.S. politics examples, election outcomes or candidate qualities or strategies in some contests surely depend on those in others, and individual legislators' votes certainly depend on others' votes or expected votes. In comparative and U.S. microbehavioral research, the recently surging interest in *contextual* or *network effects* usually refers to the effects on each individual's behavior or opinion of sets of other individuals' opinions or behavior; e.g., a respondent's opinion on some policy likely depends on the opinions of her state, district, community, or social group. In international relations, the interdependence of states' actions is almost the definition of the subject matter. States' entry decisions in wars, alliances, international organizations, e.g., heavily depend on how many and who enter and how. In comparative and international political economy (C&IPE), too, such spatial interdependence is very often substantively large and central. For example, *globalization*, i.e., international economic integration, implies strategic and nonstrategic interdependence across domestic politics and policies. Spatial interdependence is substantively central to the very concept of *integration* in general and to the fears and hopes surrounding *globalization* in particular. In short, spatial interdependence is ubiquitous, and often quite central, throughout the substance of political science.

We organize the paper thus. Section 2 presents and discusses our generic empirical model, which contains distinct unit-level, contextual, context-conditional, and spatial-interdependence components to establish the significance of spatial models to political science research and to separate conceptually several often-conflated sources of spatial correlation. We then discuss specification and estimation strategies, focusing on four common estimators—nonspatial OLS, spatial OLS, spatial 2SLS, and spatial ML. Section 3 examines analytically the respective omitted-variable and simultaneity biases of nonspatial OLS and spatial OLS for a simple but general case. Section 4 then evaluates the performance of all four estimators in bias, efficiency, and SE accuracy terms, using Monte Carlo simulations of more realistic contexts. We provide an empirical example and illustration of estimation and presentation of spatial econometric models in Section 5. Section 6 concludes.

2 Specifying and Estimating Spatial Econometric Models

We begin by specifying a generic spatial econometric model intended to reflect the basic structure of many political science arguments evaluated in TSCS data. Perhaps especially attuned to C&IPE, the model involves unit-level explanators (e.g., individual or domestic factors), $\mathbf{d}_{it}$, exogenous-external conditions or shocks (e.g., oil prices, technology), $\mathbf{s}_i$, and context-conditional responses to external shocks (i.e., unit responses to exogenous-external conditions may depend on unit characteristics), $\mathbf{d}'_{it} \otimes \mathbf{s}'_t$, as well as interdependence processes, $\rho \sum_{j \neq i} w_{ij} y_{jt}$:

$$y_{it} = \rho \sum_{j \neq i} w_{ij} y_{jt} + \mathbf{d}'_{it} \boldsymbol{\beta}_{\mathbf{d}} + \mathbf{s}'_t \boldsymbol{\beta}_{\mathbf{s}} + (\mathbf{d}'_{it} \otimes \mathbf{s}'_t) \boldsymbol{\beta}_{\mathbf{ds}} + \varepsilon_{it}. \tag{1}$$

Interdependence refers to processes by which outcomes in some units, y_j, affect outcomes in others, y_i.[1] We distinguish such interdependence processes, which will induce spatial correlation, from responses to spatially correlated outside influences—call these *exogenous-external conditions* or *common shocks*—and/or to spatially correlated unit level factors, both of which will also induce spatial correlation. Correlation across units can arise through any of these components of our generic model. For example, a country might respond to spatially correlated internal or exogenous-external political economic shocks by lowering its capital tax rate (the unit-level or contextual terms, $\mathbf{d}'_{it} \boldsymbol{\beta}_{\mathbf{d}}$ or $\mathbf{s}'_t \boldsymbol{\beta}_{\mathbf{s}}$), and its response to the external or internal shocks may depend on contextual or domestic conditions (the term reflecting context conditionality, $(\mathbf{d}'_{it} \otimes \mathbf{s}'_t) \boldsymbol{\beta}_{\mathbf{ds}}$), but the magnitude of its response may further depend on how its competitors respond and, conversely, its own response may affect the tax rates that policy makers in other countries choose (the term reflecting spatial interdependence, $\rho \sum_{j \neq i} w_{ij} y_{jt}$).

In this paper, we focus on C&IPE models like (1) that contain *domestic* (unit-level), *exogenous-external* (contextual), and *context-conditional* explanators as well as *interdependence* processes, and methods for estimating such models. We do not consider models with solely unit-level (domestic), contextual (exogenous-external), or context-conditional factors, although such models are also in common use, except to note that, if interdependence is important, estimates of coefficients for unit-level and contextual variables in such models will be inefficient in the best of circumstances and, more usually, biased and inconsistent as well. Indeed, a central challenge for empirical researchers, known as Galton's Problem, is the difficulty of distinguishing responses to spatially correlated contextual or unit-level conditions from *interdependence*. On one hand, to ignore or inadequately model interdependence processes when they are present will lead analysts to exaggerate the importance of common shocks and thus privilege contextual (exogenous-external) or unit-level (domestic) explanations. On the other hand, if simultaneity problems discussed below are insufficiently redressed, modeling interdependence with spatial lags will lead analysts to misestimate (usually overestimate) the importance of interdependence at the expense of common shocks, especially insofar as such common shocks are inadequately modeled. Instrumental-variables (IV) or maximum likelihood (ML) estimators (spatial two-stage least squares [S-2SLS] or S-ML) may provide effective resolutions (in somewhat different regards, under somewhat different conditions) to this dilemma.

[1]We do not consider models where interdependence arises unequally or solely through $\hat{y}_j$ and/or ε_j. For comprehensive textbook treatment of spatial econometrics, including such spatial error models and the like, see Anselin (1988) and, for newer developments, Anselin (2001).

2.1 *Estimating Models with Spatial Interdependence*

Much political science substance and theory will imply empirical models like (1). Econometrically, however, obtaining *good* (unbiased, consistent, efficient) coefficient estimates and accurate SEs in general, and, in particular, distinguishing effects of spatially correlated domestic, exogenous-external, and context-conditional factors from those of interdependence processes are not simple tasks. The first and primary consideration is the relative and absolute precision and power with which the empirical model specifies and measures its alternative nonspatial and spatial components, i.e., the interdependence part $\left(\rho\sum_{j\neq i} w_{ij}y_{jt}\right)$ and the parts reflecting common, correlated, or context-conditional responses to unit-level and exogenous-external factors ($\mathbf{d}'_{it}\boldsymbol{\beta}_{\mathbf{d}}$, $\mathbf{s}'_t\boldsymbol{\beta}_{\mathbf{s}}$, and $(\mathbf{d}'_{it}\otimes\mathbf{s}'_t)\boldsymbol{\beta}_{\mathbf{ds}}$). Insofar as unit level/domestic factors and/or the incidence and effects of exogenous-contextual/external factors correlate spatially (which is likely), and in patterns similar to the interdependence patterns (also likely), the two mechanisms produce similar effects, so inadequacies or omissions in specification of the one tend, quite intuitively, to induce overestimation of the other's impact.

Secondarily, however, even if the spatial and nonspatial components are modeled perfectly, the spatial lags in this empirical model will be endogenous (i.e., covary with residuals), so estimates of ρ will suffer simultaneity bias. As with the primary omitted variable or relative misspecification concern, these secondary simultaneity issues will tend to bias conclusions on the strengths of nonspatial and interdependence mechanisms in opposite directions. That is, relative failure to model either the nonspatial or spatial aspects adequately will tend to bias conclusions in favor of the other aspect; likewise, inadequate redress of the simultaneity involved in modeling interdependence will tend induce misestimation (often, but not necessarily, overestimation) of the strength of interdependence and thereby bias conclusions toward the one (usually interdependence) and against the other (usually nonspatial).

Spatial lagged–dependent variable, or just *spatial lag* models like (1) are an effective specification for estimating and testing the sign and strength of interdependence. In matrix notation,

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{y}$ is an $NT \times 1$ vector of observations (N units, T time periods per unit) on the dependent variable stacked by unit (i.e., unit 1, time 1 to T, then unit 2, time 1 to T);[2] ρ is the spatial autoregressive coefficient; $\mathbf{W}$ is an $NT \times NT$ block-diagonal spatial-weighting matrix, with elements w_{ij} reflecting the relative degree of connection from unit j to i. $\mathbf{Wy}$ is thus the spatial lag; i.e., for each observation on y_{it}, the corresponding element of $\mathbf{Wy}$ gives a weighted sum of the y_{jt}, with weights given by the relative connectivity from j to i. $\mathbf{X}$ is an $NT \times K$ matrix of observations on K independent variables; $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients on $\mathbf{X}$; $\boldsymbol{\varepsilon}$ is an $NT \times 1$ residual vector. Note that each $T \times T$ block along $\mathbf{W}$'s block diagonal,[3] which is the block multiplying $\mathbf{y}_i$ itself in the spatial lag–weighted

[2] As issues of temporal dependence are largely orthogonal to the spatial issues discussed here (see below and Franzese and Hays 2006a), and as other contributions to this volume emphasize temporal issues in TSCS data, we will assume for simplicity that $\mathbf{X}\boldsymbol{\beta}$ contains a full and effective model of the temporal dependence (e.g., time-lagged dependent variables) through most of our discussion. However, we will discuss substantive interpretation and presentation of estimated effects in models containing spatiotemporal dependence below, so we prefer to retain the i,t subscripts here.

[3] The methodological literature on spatial dependence mostly focuses on cross-section data (T = 1). In this case, each block referenced in the text and surrounding notes has just one element.

sum, is all zeros;[4] each of the off-diagonal $T \times T$ blocks has zero off-diagonal elements,[5] but nonzero diagonal elements reflecting the contemporaneous spatial correlation in **y**. Note also that, as the w_{ij} elements of **W** reflect the relative connectivity from unit j to i, **W** may not be symmetric.[6] Finally, ρ gives the impact on the outcome in i of the outcomes in all the other ($j \neq i$) spatial units, each weighted by w_{ij}.

Thus, ρ gauges overall interdependence strength, and the w_{ij} describe the relative magnitudes of specific interdependence paths between units. Typically, the set of w_{ij} is determined by theoretical and substantive argument as to which units will have greatest effect on outcomes in which others. ρ is the coefficient on **W**'s prespecified spatial weights, giving the strength of interdependence along those prespecified paths.[7] In C&IPE, e.g., the interdependence induced by international economic competition might be operationalized as a set of weights, w_{ij}, based on the trade- or capital-flow shares of countries j in country i's total. The inner product of that vector of weights with the stacked dependent variable **y** then gives as a regressor the weighted average (or sum) of **y** in the other countries j that time period. **Wy** gives the entire set of these vector inner products—here, the trade- or capital-flow–weighted averages—for all countries i and j. In other contexts (as well as in C&IPE), diffusion might alternatively occur via contiguity (borders), leader emulation or cultural connection mechanisms. Here, outcomes from some unit or set of units $\{j\}$, but not the outcomes from other units, would be expected to diffuse to the outcome in i. This implies that the weights are $(n_{\{j\}} - 1)^{-1}$ for those ij where i and j both belong to some group (e.g., share a border, language, or membership in an institution or any other group) and 0 for all others. Call this class of interdependence patterns *comembership*; our simulations below will reflect a special case of comembership interdependence where all sample units are comembers of the same group and affect each other equally, implying uniform weights of $(N - 1)^{-1}$.

To estimate equation (2), one could simply omit **Wy** and estimate $\boldsymbol{\beta}$ by ordinary least squares regression of **y** on **X**: *nonspatial OLS*. Despite obvious omitted-variable biases and inefficiency (and the lack of any estimate at all of ρ), this "strategy" of ignoring spatial interdependence is currently the most common (although often with spatial SE corrections, PCSE or certain types of *clustering*). A second strategy, almost as simple to implement and in increasing use, estimates $\boldsymbol{\beta}$ and ρ by OLS regression of **y** on both **X** and **Wy**: *Spatial OLS*. Unfortunately, because **Wy** is endogenous, S-OLS will suffer simultaneity biases. ML offers a third strategy of estimating ρ and $\boldsymbol{\beta}$ in a model that specifies the joint likelihood of **y** to reflect the spatial interdependence (Ord 1975). *Spatial ML* is computationally intense, especially as $N \times T$ dimensionality rises, but its parameter estimates will be consistent and asymptotically efficient if the model, including the interdependence pattern, is correctly specified. A fourth strategy instruments for **Wy** using **X** and **WX**. This *IV by S-2SLS* strategy also produces consistent and asymptotically efficient estimates, provided its necessary conditions are met: namely, that the **X** are indeed exogenously related to **y**.[8]

[4]If **y** also manifests temporal dynamics, only **W**'s prime diagonal is zero; the off-diagonal elements of the $T \times T$ block diagonal are nonzero and reflect these temporal dynamics.

[5]That is, unless **y** exhibits spatial, cross-temporal interdependence so y_{it} affects y_{js} for some $i \neq j$ and $t \neq s$.

[6]In fact, symmetric **W** is unlikely in most C&IPE contexts (at least), where it would imply, e.g., equal-strength effects United States → Belgium and Belgium → United States. Such asymmetry is one reason spatial approaches that exclusively stress error covariance have less useful applicability in political science than spatial lag models, although symmetric **W** may be more likely in some more homogenous contexts.

[7]The accuracy of **W**'s prespecification, both absolutely and relative to the nonspatial components of the model, is of crucial empirical, theoretical, and substantive importance. Strategies for parameterizing **W** and estimating such models are of great interest but as yet mostly remain for future work to develop.

[8]See Kelejian et al. (2003) and Kelejian and Robinson (1993) for more complete lists of estimators and more thorough coverage of the IV/method-of-moments class of estimators.

The two simplest estimators (OLS, S-OLS) are, as just noted, misspecified and so inconsistent, due to omitted variables (OLS) or simultaneity (S-OLS). This does not imply that their bias, inconsistency, and inefficiency will necessarily be large though, or that they will perform equally badly, or even that consistent estimators like S-2SLS or S-ML will necessarily outperform them in finite samples. Thus, the two "sophisticated" options asymptotically dominate the two simple ones, but we must still explore the relative performance of all four in samples and models reflecting typical political science TSCS contexts to assess which may be best strategies in practice, and by how much, under what conditions. Our experiments suggest that the omitted-variable biases of the current default practice of nonspatial OLS generally are large, whereas the simultaneity biases of S-OLS are typically smaller, especially as the strength of interdependence, ρ, remains modest in truth and domestic and exogenous-external factors are well-specified and powerful explanators. In fact, in some conditions (modest interdependence and small samples), S-OLS can perform adequately in mean-squared error in comparison to S-2SLS and S-ML even if IV instrumentation or ML joint-likelihood specification is perfect. At greater interdependence strength, however, the simultaneity bias grows large, and using one of the consistent estimators is more crucial. There we find that the (perhaps) simpler S-2SLS can perform acceptably well compared to S-ML under some conditions, notwithstanding IV estimators' known inefficiency relative to ML (and least squares [LS]), but also that S-ML seems to dominate more notably in other conditions and is rarely outperformed.

3 Analytical Results for OLS and S-OLS: Omitted-Variable and Simultaneity Biases

We now demonstrate analytically, in a case simplifying equation (2) to a single regressor, $\mathbf{x}$, that both nonspatial and spatial OLS estimates are biased and inconsistent in the presence of interdependence, and we specify those biases precisely. S-OLS estimation of equation (2) is inconsistent because the regressor $\mathbf{Wy}$, the spatial lag, covaries with the residual, $\boldsymbol{\varepsilon}$. The reason is simple, the spatial lag, $\mathbf{Wy}$, being a weighted average of outcomes in other units, places the left-hand side of some observations on the right-hand side of others: textbook simultaneity. To see the implications of this endogeneity, first rewrite equation (2) as

$$\mathbf{y} = \mathbf{Q}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad \text{where } \mathbf{Q} = [\mathbf{Wy} \quad \mathbf{x}] \text{ and } \boldsymbol{\delta} = [\rho \quad \beta]'. \tag{3}$$

The asymptotic simultaneity bias for the S-OLS estimator is then given by

$$\text{plim } \hat{\boldsymbol{\delta}}_{\text{S-OLS}} = \boldsymbol{\delta} + \text{plim}\left[\left(\frac{\mathbf{Q}'\mathbf{Q}}{n}\right)^{-1}\left(\frac{\mathbf{Q}'\boldsymbol{\varepsilon}}{n}\right)\right]. \tag{4}$$

In the case where $\mathbf{x}$ is exogenous, we can rewrite the biases expressed in equation (4) as

$$\text{plim } \hat{\boldsymbol{\delta}}_{\text{S-OLS}} = \begin{bmatrix} \rho \\ \beta \end{bmatrix} + \frac{1}{|\boldsymbol{\Psi}|}\begin{bmatrix} \text{cov}(\mathbf{Wy}, \boldsymbol{\varepsilon}) \times \text{var}(\mathbf{x}) \\ -\text{cov}(\mathbf{Wy}, \boldsymbol{\varepsilon}) \times \text{cov}(\mathbf{Wy}, \mathbf{x}) \end{bmatrix},$$
$$\text{where } \boldsymbol{\Psi} = \text{plim}\left(\frac{\mathbf{Q}'\mathbf{Q}}{n}\right). \tag{5}$$

So, e.g., in the likely common case of positive interdependence and positive covariance of the spatial lag and exogenous regressors, S-OLS would generally overestimate interdependence strength, $\hat{\rho}$, and correspondingly underestimate domestic, exogenous-external, and/or context-conditional effects, $\hat{\beta}$.

Conversely, the bias in $\hat{\beta}$ induced by omitting the spatial lag when interdependence exists—i.e., in nonspatial OLS—is simply omitted-variable bias, the formula for which is well known to be $\mathbf{F\beta}$ where $\mathbf{F}$ is the matrix of coefficients obtained by regressing the omitted on the included variables and $\boldsymbol{\beta}$ is the vector of (true) coefficients on the omitted variables. In this case,

$$\text{plim}\ \hat{\beta}_{\text{OLS}} = \beta + \rho \times \frac{\text{cov}(\mathbf{Wy}, \mathbf{x})}{\text{var}(\mathbf{x})}. \tag{6}$$

$\hat{\rho}_{\text{OLS}} \equiv 0$, of course, which is biased by $-\rho$. Thus, in the same likely case as above, OLS overestimates domestic, exogenous-external, or context-conditional effects while ignoring spatial interdependence.

We find further intuitions by simplifying even more radically to just two units (with one x each):

$$\begin{aligned} y_1 &= \rho_{12}y_2 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \rho_{21}y_1 + \beta_2 x_2 + \varepsilon_2. \end{aligned} \tag{7}$$

In equation (7), $\text{Cov}(y_1, \varepsilon_2) = \left(\frac{\rho_{12}}{1-\rho_{21}\rho_{12}}\right)\text{Var}(\varepsilon_2)$. Using $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \left(\frac{\mathbf{Q'Q}}{n}\right)^{-1}\frac{\mathbf{Q'\varepsilon}}{n}$ for OLS coefficients then reveals that:

$$\text{S-OLS bias of } \hat{\rho}_{21} = \frac{1}{n}(\mathbf{Q'Q})^{-1}_{22} \times \left(\frac{\rho_{12}}{1-\rho_{12}\rho_{21}}\right) \times \text{Var}(\varepsilon_2). \tag{8}$$

The {2, 2}th element of $\mathbf{Q'Q}/n$ is $V(y_1)$ and so is positive as is the corresponding element of the matrix inverse. Thus, the simultaneity bias in $\hat{\rho}_{21}$has the same sign as the true ρ_{12}.[9] Note the subscript reversal; it is intuitive. If Japan affects the United States negatively and the United States affects Japan positively, e.g., then S-OLS's simultaneity-biased estimates of U.S. → Japan interdependence would, by ignoring dampening feedback from the negative Japan → U.S. effect, induce underestimation of the positive U.S. → Japan dependence. Trying to estimate the negative Japan → U.S. interdependence by S-OLS would conversely incur positive bias by ignoring the countervailing U.S. → Japan feedback. Thus, oppositely signed interdependencies induce S-OLS simultaneity biases favoring interdependence-strength underestimation. We suspect that actual diffusion mechanisms more usually involve same-signed interdependence and so conclude that the simultaneity biases in S-OLS will usually inflate interdependence-strength estimates.

From equation (8), we also see that S-OLS's simultaneity bias will be large only as the function of ρ_{12} and ρ_{21} in the second term times the ratio of $V(\varepsilon_2)$ to $V(y_1)$ is large. Thus, simultaneity-induced overestimation of ρ will be large only insofar as mutual interdependence is relatively strong and outcomes are relatively inexplicable by the exogenous (nonspatial) model factors. To restate this substantively, S-OLS will suffer sizable simultaneity biases only if and insofar as spatial interdependence is strong and unit-level (domestic) and contextual (exogenous-external) factors account for relatively little systematic variation.

Finally, the omitted-variable biases in nonspatial OLS are larger the more domestic or exogenous-external factors correlate spatially. With fully common exogenous-external factors (e.g., time period fixed effects, as in our simulations), $x_1 = x_2$ and OLS

[9]Note that we assume throughout that $0 \le \rho < 1$. Spatial unit roots, $0 \le \rho < 1$, are highly problematic statistically but, thankfully, also highly improbable substantively. This implies that $0 \le \rho_{12}\rho_{21} < 1$ also.

yields $\hat{\beta}_1 = \beta_1 + \frac{\rho_{21}\beta_2 + \rho_{12}\rho_{21}\beta_1}{1-\rho_{12}\rho_{21}}$, *greatly* overestimating the importance of common shocks. Conversely, omitting common shocks when such are present induces great simultaneity-induced overestimation of ρ; S-OLS yields $\hat{\rho}_{12} = \rho_{12} + \frac{\beta_1\beta_2 + \rho_{21}\beta_1^2}{(1-\rho_{12}\rho_{21})[\beta_2^2 + \beta_1^2\rho_{21}^2 + \rho_{21}^2 \operatorname{var}(\varepsilon_1) + \operatorname{var}(\varepsilon_2)]}$.

In sum, S-OLS estimates of equation (1) or equation (2) suffer simultaneity bias; OLS estimates of equation (1) or equation (2) exclude the spatial lag and so suffer omitted-variable bias. S-OLS estimates of (net) interdependence from j to i will have bias of the same sign as the (net) interdependence from i to j[10]. If feedback from j to i and i to j reinforce (have the same sign), then S-OLS estimates will exaggerate interdependence. With countervailing feedback (opposite sign), which is probably less common, S-OLS attenuates $\hat{\rho}$. Moreover, simultaneity-induced inflation bias in $\hat{\rho}_{S-OLS}$ induces attenuation bias in $\hat{\beta}_{S-OLS}$, the coefficient on nonspatial factors. The $\hat{\rho}_{OLS} \equiv 0$ imposed by OLS, in turn, induces inflation bias in $\hat{\beta}_{OLS}$. These conclusions hold in degree as well: *insofar as* one specifies interdependence inadequately—absolutely, and relatively to the model's nonspatial components—one tends to underestimate the former and overestimate the latter, and vice versa. When spatial lags are generated with arbitrary and so likely inaccurate weights, e.g., interdependence-strength estimates will likely be biased downwards and unit-level, contextual, or context-conditional explanations conversely privileged. All these problems' magnitudes, intuitively, increase with the general strength of interdependence, ρ, and with the spatial correlation of domestic (unit-level), exogenous-external (contextual), and context-conditional regressors. Researchers interested in evaluating the relative strengths of unit-level, contextual, and interdependence effects thus especially need to weigh carefully these specification and estimation-strategy considerations.

Fortunately, IVs and ML methods for redressing the simultaneity problems of S-OLS exist. However, S-2SLS may suffer *quasi instrument* (Bartels 1991) and efficiency problems characteristic of all IV estimators, and S-ML is computationally demanding and not yet well implemented in software packages commonly used by political scientists. Furthermore, S-2SLS and S-ML, like all IV and ML, have only asymptotic properties (consistency and asymptotic efficiency, asymptotic normality), so their performance in realistic, limited samples demands further exploration. Moreover, even in these simple cases that we explored analytically, determining whether OLS omitted-variable or S-OLS simultaneity biases will typically be appreciable, which might be the larger concern typically, and how either OLS or S-OLS compare with S-2SLS or S-ML short of "asymptopia" is difficult. We turn therefore to Monte Carlo simulation to compare these estimators in richer, more realistic scenarios.

4 Experimental Evaluations of the Four Estimators

4.1 *Design of the Simulation Exercises (Monte Carlo Experiments)*

The true model generating the data for our experiments is the reduced-form solution of equation (2):[11]

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}, \tag{9}$$

with matrix $\mathbf{X}$ comprising $\mathbf{d} = \{d_{it}\}$, $\mathbf{s} = \{s_t\}$, $\mathbf{ds} = \{d_{it}s_t\}$, as in our generic context-conditional model (1). Residuals, $\boldsymbol{\varepsilon}$, are an $NT \times 1$ vector of i.i.d. draws from a standard normal distribution.

[10]"Net" here means including all the indirect feedback through third parties.

[11]We assume that the *spatial multiplier*, $(\mathbf{I} - \rho\mathbf{W})$, is invertible, which serves to debar spatial unit-roots.

d is also an $NT \times 1$ vector of i.i.d. draws from a standard normal, which, being unit-i and time-t specific, represent unit-level (domestic or individual) factors unique to each unit time. **s** is an $NT \times 1$ stack of T vectors of size $N \times 1$. Each subvector has N identical elements, but each of the T subvectors differs from others. **s** thus represents a set of common shocks (contextual factors), one per time period, each hitting each cross-sectional unit equally, and each also drawn i.i.d. standard normal.[12] The interaction term, **ds**, reflects the notion that unit-level variables condition the effects of these common shocks. Finally, the true model involves interdependence, with average magnitude $\boldsymbol{\rho}$ and relative connectivities from units j to i of w_{ij}.

We then generate **y** by equation (9) using coefficients $(\beta_d, \beta_s, \beta_{ds}, \rho) = (1, 1, 1, \rho)$, varying ρ from .1 to .9 and sample dimensions over subsets of $N = \{5, 10, 25, 40, 50\} \times T = \{20, 35, 40, 50\}$ across experiments. Recall that the spatial weights, w_{ij}, give the relative impact of each j on each i in the spatial interdependence pattern given by **W**, whereas ρ gives the general strength of interdependence following this pattern. Thus, larger ρ (for a given $\bar{w}_{ij}$)[13] implies stronger interdependence. We assume that the spatial dependence is time invariant—i.e., the w_{ij} connecting j to i persists in all T periods without change—so each diagonal element of each $T \times T$ off-diagonal block in **W** is equal. We also assume spatial connectivity to manifest without time lag, i.e., within an observation period. As Beck et al. (2006) note, time-lagged spatial lags need not covary with contemporaneous residuals, which alleviates estimation issues from endogeneity. However, many substantive contexts and data set structures will combine to suggest that interdependence across units occurs within the time span of one observational period, and, even if not, time-lagged spatial lags only alleviate the endogeneity insofar as temporal and *spatiotemporal* dynamics are modeled adequately.[14]

Finally, all data in our experiments are temporally uncorrelated; effectively, we are assuming that temporal dependence is successfully modeled elsewise (e.g., by time-lagged dependent variables) or absent (as likely, e.g., in pooled independent surveys).[15] We then set the interdependence pattern between spatial units by fixing all these w_{ij} elements of **W** to $(N-1)^{-1}$. That is, every unit affects every other unit equally, so the spatial lag is just an unweighted average of the dependent variable for the other units that year.[16]

To estimate spatial lag models, researchers must prespecify this spatial-weights matrix, **W**. This prespecification is a crucial theoretical and empirical step in studying interdependence.[17] As already stressed, distinguishing between and evaluating the relative strength of interdependence and that of unit-level or contextual effects rely firstly upon the relative precision with which these alternative sources of spatial correlation are specified. Our simulations, on the other hand, simply set all nonzero elements (i.e., the off-block-diagonal

[12]Note: **s** is one regressor, not a set of time period dummies. This assumes that the researcher effectively models exogenous-external conditions with some covariate (e.g., terms of trade). Franzese and Hays (2004) explore the implications of measurement/specification error in this crucial step.

[13]Following the spatial econometrics literature, we row normalize the w_{ij} to sum to 1 for each row i of **W**, so the parenthesis is unnecessary. Although little is discussed, row normalization is not necessarily substantively neutral. It implies that per-unit connectivity declines with numbers of connected units, or, e.g., that connectivity via trade depends only on shares of trade and not total trade exposure.

[14]As Beck et al. (2006) also stress, one can and should test the sufficiency of the modeled temporal dynamics as usual. Franzese and Hays (2004) discuss some tests to explore residual spatiotemporal dynamics.

[15]Franzese and Hays (2006a) consider temporal and spatial dependence jointly and find confirmation that temporal and spatial estimation issues are largely orthogonal, so we can safely discuss estimation issues regarding spatial interdependence while assuming temporal dependence to have been modeled otherwise. We do discuss, however, substantive interpretation of estimates from models containing both temporal and spatial dependence because spatiotemporal interactions are important there.

[16]This is similar, especially for binary outcomes and relatives like durations, to the (unweighted) counts or proportions of the other units with $y = 1$ or $y = 0$ often used in those contexts.

[17]Strategies for parameterizing and estimating **W** according to theoretical/substantive expectations would be of great interest but are, as yet, undeveloped. See also note 7.

diagonals) of the $\mathbf{W}$ specified for estimation to $(N-1)^{-1}$. That is, the hypothetical researcher estimates models with spatial lags given as unweighted averages of the dependent variable in the other cross-sectional units each period on the right-hand side, with instrumentation (S-2SLS) or without (S-OLS). S-ML likelihoods specified for empirical estimation likewise reflect this *flat* pattern of interdependence. In these experiments, in other words, interdependence is truly homogenous, and the hypothetical researcher has correctly specified the estimation-weighting matrix to equal the true one exactly.[18]

We evaluate nonspatial OLS, and S-OLS, S-2SLS, and S-ML, the LS estimators with and without PCSEs (i.e., estimates of the variance-covariance matrix of the coefficient estimates that are "robust to," i.e., consistent in the presence of, spatial correlation). We report some subset of experiments conducted over sample dimensions $N = \{5, 10, 25, 40, 50\}$, $T = \{20, 35, 40, 50\}$, $\rho = \{.1, \ldots, .9\}$.[19] Each table reports 1000 trial experiments; the figures show parameter sweeps in 100 trial experiments. We report means of coefficient estimates and the LS and ML SEs and PCSEs; actual SDs of coefficient estimates; and root mean-squared errors (RMSEs) of coefficient estimates. Comparing mean parameter estimates to their true values gives their (small sample) biases. By comparing mean reported SEs to true SDs of coefficient estimates, we observe SE accuracy and how well PCSEs may redress inaccuracies. RMSE is the square root of the sum of the square of the bias plus the variance of the estimated coefficients, and thus combines bias/consistency and efficiency concerns (with squared bias and variance weighted equally).

4.2 *The Estimators: OLS, S-OLS, S-2SLS, and S-ML*

To detail the estimators' mechanics: nonspatial OLS simply regresses $\mathbf{y}$ on $\mathbf{X}$, omitting $\mathbf{Wy}$, i.e., ignoring spatial interdependence; S-OLS regresses $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{Wy}$. Both are inconsistent given spatial interdependence, OLS suffering omitted-variable and S-OLS simultaneity biases.

S-2SLS also implements relatively easy. The $\mathbf{W}$ matrix already constructed to generate spatial lag $\mathbf{y}$ is also used to generate spatial lags of $\mathbf{X}$, $\mathbf{WX}$, which then serve as instruments for $\mathbf{Wy}$. To elaborate the standard IVs "solution" to endogeneity—i.e., covariance of regressors, here $\mathbf{Wy}$, with residuals—is to find some variables, $\mathbf{Z}$, that covary with the endogenous regressors but do not covary with the dependent variable (i.e., $\boldsymbol{\varepsilon}$) except insofar as they relate to those regressors. Given such a $\mathbf{Z}$, the IV estimator, $\mathbf{b_{iv}} = (\mathbf{X}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$, is consistent and asymptotically efficient (but usually inefficient in small samples relative to noninstrumented estimators like LS or ML). Practically, 2SLS implements this by, first, regressing the full set of $\mathbf{X}$, including endogenous regressors, on $\mathbf{Z}$ and the exogenous $\mathbf{X}$, and, second, regressing $\mathbf{y}$ on (all) the fitted $\mathbf{X}$'s from this first stage. If the instruments $\mathbf{Z}$ are indeed perfectly exogenous, i.e., their covariance with $\boldsymbol{\varepsilon}$ is *exactly* zero, then IV estimators will be consistent and asymptotically efficient regardless of how strongly the instruments covary with the endogenous regressors for which they instrument. If not, i.e, if the instruments are even slightly correlated with $\boldsymbol{\varepsilon}$, then the $\mathbf{Z}$ are only *quasi instruments* (Bartels 1991), and mean-squared error benefits or costs of instrumentation will depend on the ratio of the covariance of instruments with endogenous regressors to the (inestimable) covariance of $\mathbf{Z}$ with $\boldsymbol{\varepsilon}$. In our experiments, the $\mathbf{X}'$s (i.e., $\mathbf{d}$ and $\mathbf{s}$) are i.i.d., and in particular independent of the draws for $\boldsymbol{\varepsilon}$, so our $\mathbf{WX}$ are perfect instruments by

[18] Franzese and Hays (2004, 2006a) explore the consequences of measurement/specification error in this step and of patterns of interdependence other than this *flat* or *homogenous* one.

[19] These subsets suffice to demonstrate our major experimental findings. A much larger set of results are available (as is software to conduct further simulations oneself) from the authors upon request.

construction. More commonly in practice, researchers will often confront right-hand side $\mathbf{X}_i$ that are endogenous to left-hand side $\mathbf{y}_i$—i.e., the standard endogeneity concern that $\mathbf{y}$ causes $\mathbf{X}$ as well as $\mathbf{X}$ causes $\mathbf{y}$. If so, and if $\mathbf{X}$ exhibits spatial correlation also, then $\mathbf{WX}$ will offer only imperfect *quasi* instruments at best (intuitively, because *j*'s $\mathbf{X}$ will also contain some of *i*'s $\mathbf{y}$). In principle, researchers should be able to combine a standard 2SLS estimation strategy to address $\mathbf{X}$-$\mathbf{y}$ simultaneity with the S-2SLS estimation strategy just described to address the spatial simultaneity. Failing that (e.g., if perfectly valid instruments for the standard endogeneity problem prove difficult to discover, as they usually do), we expect that the utility of the available $\mathbf{WX}$ as *quasi instruments* will depend on the relative magnitudes of the intra-ε interdependence, call it γ, of the intra-$\mathbf{X}$ interdependency, call it ρ, the causal mechanism from $\mathbf{y}$ to $\mathbf{X}$, call that magnitude α, and the causal mechanism from $\mathbf{X}$ to $\mathbf{y}$, call it β. For now, we advise researchers either to employ only strictly exogenous $\mathbf{x}$ (i.e., those $\mathbf{x}$ whose exogeneity is most certain) in generating the spatial instruments $\mathbf{WX}$, or, if they trust our conjecture, to explain why the $\mathbf{x}$ used in $\mathbf{WX}$ have good "Bartels Ratios," which in this case translates to high $\rho\beta/\gamma\alpha$.[20]

Implementing S-ML is not much more complicated, although the maximization is computationally intense even while relying on a simplifying approximation to the determinant of ($\mathbf{I}$-$\rho\mathbf{W}$). To see the minor complication, start by isolating the stochastic component of the spatial lag model:

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \Rightarrow \boldsymbol{\varepsilon} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \equiv \mathbf{Ay} - \mathbf{X}\boldsymbol{\beta}. \tag{10}$$

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

$$L(\boldsymbol{\varepsilon}) = \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{NT}{2}} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right), \tag{11}$$

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

$$L(\mathbf{y}) = \mid \mathbf{A} \mid \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{NT}{2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})\right]. \tag{12}$$

This still resembles the typical linear normal likelihood, except that the transformation from $\boldsymbol{\varepsilon}$ to $\mathbf{y}$, is not by the usual factor of 1, but by $|\mathbf{A}| = |\mathbf{I} - \rho\mathbf{W}|$. Since $|\mathbf{A}|$ depends on ρ, each recalculation of the likelihood in maximization routine must recalculate this determinant for the updated ρ values. Ord's (1975) solution to this computational intensity issue was to approximate $|\mathbf{W}|$ by $\boldsymbol{\Pi}_i\,\lambda_i$ because the eigenvector $\boldsymbol{\lambda}$ in this approximation does not depend on ρ. Then, $|\mathbf{I} - \rho\mathbf{W}| = \boldsymbol{\Pi}_i(1 - \lambda_i)$, which requires the estimation routine only to recalculate a product, not a determinant, as it updates.[21] The estimated variance covariances

[20]We have begun to consider this conjecture in some simulations (Franzese and Hays 2004), but we have not yet explored it fully analytically or experimentally, nor have we yet considered the practical details of combining 2SLS for endogeneity of $\mathbf{X}$ and $\mathbf{y}$ with S-2SLS for spatial simultaneity of $\mathbf{y}$. Note: ρ, β, γ, α magnitudes cannot be estimated without a model whose identification conditions must assume them. That is, as Bartels stressed, the parameters that determine the quality of *quasi instruments* cannot be estimated; one can only offer theoretical/ substantive arguments about their likely relative magnitudes.

[21]Unfortunately, the approximation may be numerically unstable (Anselin 1988, 2001). On the other hand, S-ML may enjoy a practical advantage over S-2SLS in multiple $\mathbf{W}$ models in that S-ML does not require differentiated instrumentation for each $\mathbf{W}$ to gain distinct leverage on its corresponding ρ. The instruments, $\mathbf{WX}$, would differ by virtue of $\mathbf{W}$ differing for the alternative interdependence processes, so S-2SLS is estimable for multiple $\mathbf{W}$ models even with identical $\mathbf{X}$ in the $\mathbf{WX}$ instruments, but we harbor doubts about the practical identification leverage obtainable thereby.

Table 1 Average coefficient estimates across 1000 trials (bias)

			Coefficient	*OLS*	*S-OLS*	*S-2SLS*	*S-ML*
$\rho = .1$	$N = 5$	$T = 20$	$\hat{\beta}_s$	*1.112*	1.027	***1.003***	1.048
			$\hat{\rho}$	—	0.078	***0.097***	*0.063*
		$T = 40$	$\hat{\beta}_s$	*1.112*	0.991	***1.001***	1.021
			$\hat{\rho}$	—	0.108	***0.099***	*0.082*
	$N = 40$	$T = 20$	$\hat{\beta}_s$	*1.112*	*1.049*	***0.994***	*1.050*
			$\hat{\rho}$	—	*0.055*	***0.105***	*0.054*
		$T = 40$	$\hat{\beta}_s$	*1.112*	***0.999***	1.003	1.026
			$\hat{\rho}$	—	***0.101***	.098	*0.077*
$\rho = .5$	$N = 5$	$T = 20$	$\hat{\beta}_s$	*1.999*	*0.837*	***0.998***	*1.050*
			$\hat{\rho}$	—	*0.579*	***0.499***	*0.475*
		$T = 40$	$\hat{\beta}_s$	*2.001*	*0.826*	***1.000***	1.029
			$\hat{\rho}$	—	*0.587*	***0.500***	0.487
	$N = 40$	$T = 20$	$\hat{\beta}_s$	*2.004*	*0.861*	***1.008***	*1.050*
			$\hat{\rho}$	—	*0.570*	***0.497***	*0.474*
		$T = 40$	$\hat{\beta}_s$	*2.000*	*0.844*	***1.002***	1.025
			$\hat{\rho}$	—	*0.578*	***0.499***	0.487
Number of unbiased wins/ties				0	2	14	0
Number of times noticeable ($\geq$5%) bias				16	10	0	10

Note. Unbiasedness winner in bold italics and notable or appreciable biases in italics.

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 same approximation serves there.

4.3 *Simulation Results*

We note first that noticeable estimation problems arise virtually exclusively in $\hat{\rho}$, of course, and in $\hat{\beta}_s$, intuitively because omission or misestimation of $\hat{\rho}$ will induce biases primarily in the factors that correlate most closely with the missing/misestimated interdependence. In our experimental design of homogenous interdependence and fully common shocks, that is decidedly **s**. To conserve space, therefore, we report and discuss estimates related to those **s** and **Wy** regressors only.

Table 1 reports coefficient estimates for the subset sample dimensions and parameter values of $N = \{5, 40\}$, $T = \{20, 40\}$, and $\rho = \{.1, .5\}$. We see first that, as analytically shown for the simplest case, any bias in $\hat{\rho}$ induces a bias in $\hat{\beta}_s$ in approximate proportion and of opposite sign. Furthermore, as claimed, nonspatial OLS's erroneously imposed zero interdependence induces overestimates of $\hat{\beta}_s$, noticeably so even at weaker interdependence and radically so at stronger ρ. Simple S-OLS dramatically improves over this badly misspecified nonspatial OLS, replacing the latter's sizable omitted-variable biases with an actual estimate of ρ. The estimate does suffer a simultaneity bias, but that bias remains modest at lesser interdependence strength and virtually vanishes in larger samples with low ρ. However, the S-OLS simultaneity biases do become appreciable at greater ρ. In Table 1, S-2SLS emerges clearly dominant by an unbiasedness criterion, hardly erring at all on average, across any sample dimensions or parameter values. S-ML performs mostly acceptably also, with biases mostly below 5%, except in smaller samples at lower ρ, where it misses rather badly in percentage (although not in absolute) terms.

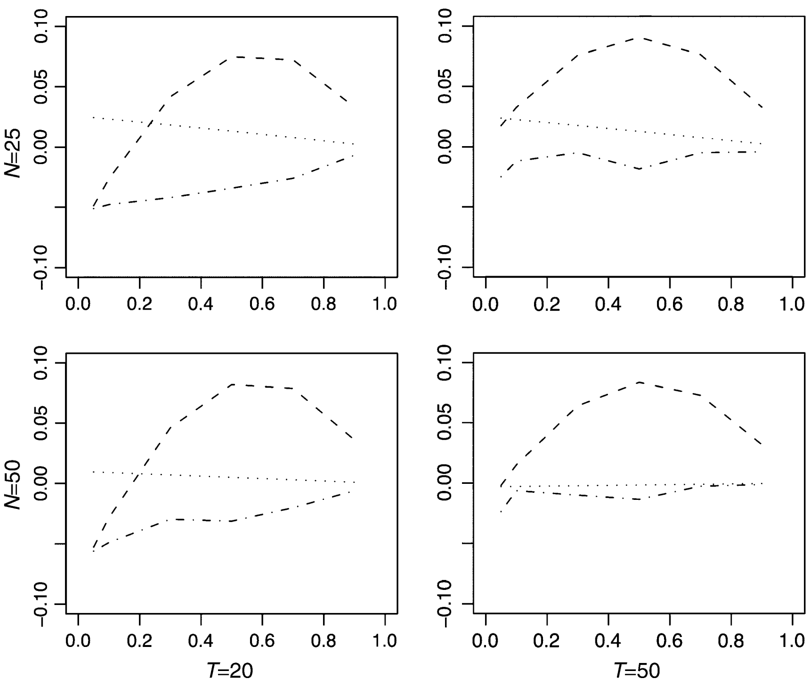

Fig. 1 Estimated bias in $\hat{\rho}$ plotted against true ρ across representative $N \times T$ sample dimensions. Dashed line, S-OLS; dotted line, S-2SLS-IV; dashed-dotted line, S-ML.

We also plot the percentage biases for each of the estimators across a range of interdependence strengths, $\rho = \{.1, \ldots, .9\}$, in grids of graphs arrayed by sample dimensions—we show $N = \{25, 50\}$, $T = \{20, 50\}$—such as in Fig. 1 for $\hat{\rho}$ and Fig. 2 for $\hat{\beta}_s$. The x axes in both figures are the true ρ. The y-axis in Fig. 1 is the estimated bias in $\hat{\rho}$ (i.e., the average $\hat{\rho}$ minus the true ρ) and in Fig. 2 just average $\hat{\beta}_s$ directly. The figures reinforce the conclusion that S-2SLS dominates on unbiasedness grounds, as the dotted line is always closest to zero in Fig. 1 and to one in Fig. 2. S-ML performs a little worse in the smaller T samples, but effectively closes the gap by the larger T sample sizes. Figure 1 shows clearly the asymptotic unbiasedness (consistency) of both estimators' $\hat{\rho}$ "kicking in" as either N or T increases; their $\hat{\beta}_s$ estimates, on the other hand, show this convergence mostly in N and barely if at all in T. Interestingly, S-2SLS has a (smaller) positive small sample bias to $\hat{\rho}$, whereas S-ML has a (perceptibly larger) negative one; i.e., their convergence is from above and from below, respectively.[22] Finally, their induced biases in $\hat{\beta}_s$ are (smaller and) oppositely signed, as expected. S-OLS, meanwhile, we can now see, actually begins with a negative bias to $\hat{\rho}$ at very low ρ, crossing an unbiased point and turning positive somewhere between the $\rho = .1$ value reported in Table 1 and $\rho = .2$ or so. Over most of the range, $\hat{\rho}_{S-OLS}$ suffers positive (i.e., inflation) bias, as we would generally expect. Interestingly, the bias peaks in absolute terms at around the $\rho = .5$ value reported in Table 1, but the magnitudes of these biases do not seem to depend in any intuitive way on sample dimensions.[23] As seen in Fig. 2, once again, the induced bias in $\hat{\beta}_s$ maintains opposite sign

[22] We have no intuition to impart for this intriguing finding.

[23] The proportionate bias, not shown, peaks at around $\rho = .4$. Also interestingly, all three $\hat{\rho}$ estimators approach (S-2SLS: stays) unbiased as $\rho \rightarrow 1$. Our intuition for this finding is underdeveloped, but we suspect it may be related to a result proven in Lin et al. (2006).

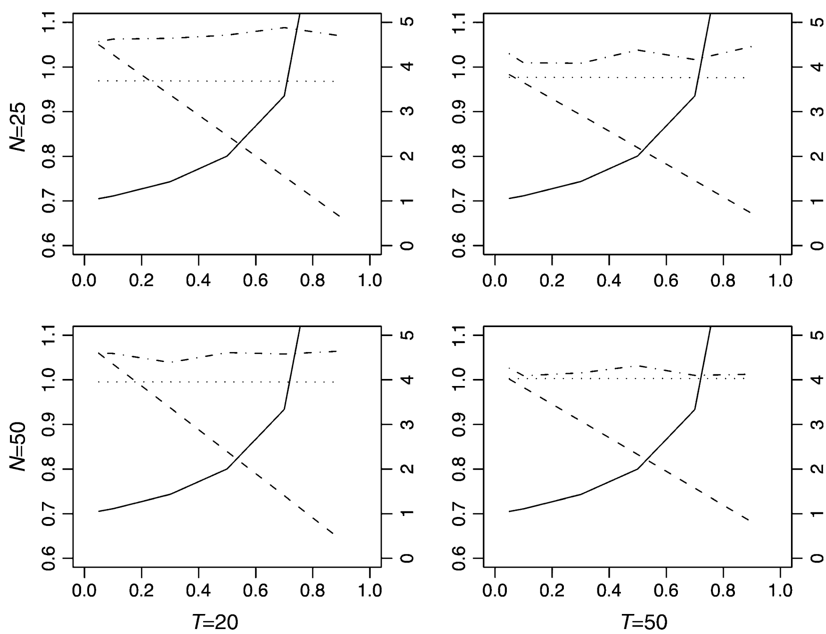

Fig. 2 Estimated $\hat{\beta}_s$ plotted against true ρ across representative $N \times T$ sample dimensions. Solid line, nonspatial OLS; dashed line, S-OLS; dotted line, S-2SLS-IV; dashed-dotted line, S-ML; nonspatial OLS results plotted against the larger-scale second *y*-axis on the right.

of the simultaneity bias in $\hat{\rho}_{\text{S-OLS}}$, with the S-OLS estimated importance of exogenous-external conditions falsely trending downward *linearly*. The bias of the nonspatial OLS "estimate" of ρ is, of course, -100%. (Plotted, it would be a $-45°$ line from 0 to -1.) The induced biases in the nonspatial OLS $\hat{\beta}_s$ are the omitted-variable biases considered analytically above. Their magnitudes are a function of ρ and not of the sample dimensions. No omitted-variable bias if $\rho = 0$, of course, but bias grows dramatically with ρ. In fact, the bias magnitude crosses 100% at $\rho = .5$, echoing the bottom of Table 1. Bias continues to rocket from there, formulaically by $\left(\frac{1}{1-\rho} - 1\right) \times 100\%$, requiring an alternative axis and truncation thereof at 400% to keep the other estimators' biases, some of which are themselves appreciable, in view. The disastrous implications of failing to model spatial interdependence when it occurs even moderately should now be abundantly clear.

Unbiasedness, being correct on average, is only one desirable property. We also prefer estimators that are generally close to true parameters, thus adding efficiency (sampling variation) concerns to bias ones. RMSE summarizes these concerns by adding the square of bias to the variance (then taking the square root), thus weighing those terms equally. Table 2 reports the RMSE of each parameter estimate under the same $2 \times 2 \times 2$ subset combinations of ρ, N, and T values as in Table 1.

From the perspective of RMSE, which combines bias and efficiency considerations in its specific way, summary of the results is much simpler: S-ML weakly dominates. Nonspatial OLS is very nearly (weakly) dominated by any spatial estimator and, in all sample dimensions, performs poorly at lower interdependence and abysmally at greater. S-OLS and S-2SLS both perform intermediately by this summary measure, but for different reasons. S-OLS suffers more bias but is relatively more efficient than S-2SLS, which has the opposite debilities. By the RMSE weighting, the net of these concerns may slightly favor S-2SLS over S-OLS, but S-ML weakly dominates either (although usually by less

Table 2 RMSE across 1000 trials (efficiency and bias)

			Coefficient	*OLS*	*S-OLS*	*S-2SLS*	*S-ML*
$\rho = .1$	$N = 5$	$T = 20$	$\hat{\beta}_s$	0.167	.216	.235	***.162***
			$\hat{\rho}$	—	.115	*.177*	***.106***
		$T = 40$	$\hat{\beta}_s$	0.139	.125	.139	***.107***
			$\hat{\rho}$	—	*.093*	*.103*	***.070***
	$N = 40$	$T = 20$	$\hat{\beta}_s$	0.119	*.215*	*.211*	***.104***
			$\hat{\rho}$	—	*.191*	*.188*	***.096***
		$T = 40$	$\hat{\beta}_s$	*0.116*	*.119*	*.136*	***.062***
			$\hat{\rho}$	—	*.105*	*.120*	***.057***
$\rho = .5$	$N = 5$	$T = 20$	$\hat{\beta}_s$	*1.04*	.242	.253	***.171***
			$\hat{\rho}$	—	*.110*	*.108*	***.066***
		$T = 40$	$\hat{\beta}_s$	*1.021*	*.211*	.146	***.116***
			$\hat{\rho}$	—	*.100*	.061	***.044***
	$N = 40$	$T = 20$	$\hat{\beta}_s$	*1.01*	.216	.270	***.213***
			$\hat{\rho}$	—	.107	.131	***.105***
		$T = 40$	$\hat{\beta}_s$	*1.00*	.185	.135	***.129***
			$\hat{\rho}$	—	.092	.065	***.063***
Number of RMSE wins/ties				0	0	0	16
Number of clearly (>50%) dominated				13	8	7	0

Note. RMSE winner in bold italics and clearly dominated RMSEs in italics.

than 25%). Again, we can see results across additional sample dimensions over more widely varying interdependence strengths in two grids of graphs paralleling those above. In Fig. 3 we can easily see that (a) all three spatial estimators yield smaller RMSE of $\hat{\rho}$ as N, T, and/or true ρ increase, (b) S-OLS's efficiency tends to outweigh S-2SLS's unbiasedness in RMSE terms at smaller true ρ and sample sizes (i.e., dashed below dotted) and the opposite at greater true ρ and sample sizes (dotted below dashed), and (c) S-ML (weakly) dominates in RMSE terms across all ρ and sample sizes. Figure 4 shows much the same pattern for RMSE of $\hat{\beta}_s$, although here S-OLS comes closer to being weakly dominated by S-2SLS and the near universal domination of nonspatial OLS by any spatial estimator is underscored dramatically by the necessity of the larger-scale second y-axis.

A third property of interest to empirical researchers is the accuracy of the estimators' reported SEs of their estimates. Table 3 reports, for the $N \in \{5, 40\}$, $T \in \{20, 40\}$, $\rho \in \{.1, .5\}$ subset of scenarios, ratios of the average reported SE (or PCSE) across the 1000 trials to the actual SD of the estimated parameters across those trials. SE accuracy ratios equal to (less than, greater than) one imply honest reporting (understatement/overconfidence, overstatement) on average. For the most part, nonspatial OLS reports inaccurate SEs for $\hat{\beta}_s$ except at small ρ. That is, at stronger interdependence, not only do omitted-variable biases favor overstating nonspatial factors' importance but also inaccurate SEs foster egregious overconfidence (64% or so) in that erroneous conclusion! PCSE only partially ameliorates this aspect of nonspatial OLS's flaws, still leaving 16%–22% overconfidence over these sample dimensions. As we have seen, S-OLS generally redresses the biases and inefficiencies of nonspatial OLS adequately at lower ρ, but we now see that it tends also to report the uncertainty of those better estimates with some overconfidence (7%–28.5%) in $\hat{\rho}$ and $\hat{\beta}_s$ SEs. Here, PCSE near uniformly *worsens* matters, yielding 12%–32% overconfidence. Intuitively, PCSE improves OLS SE estimates because the pattern of unmodeled spatial correlation there will correlate with **s**'s parts of the $\mathbf{X}'\mathbf{X}$ matrix, which is precisely

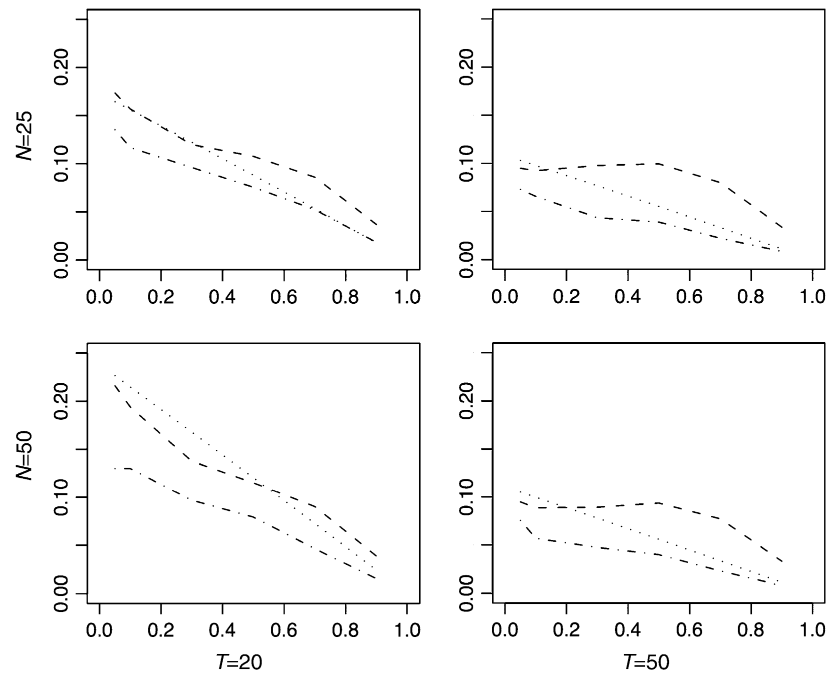

Fig. 3 RMSE $\hat{\rho}$ plotted against true ρ across representative $N \times T$ sample dimensions. Dashed line, S-OLS; dotted line, S-2SLS-IV; dashed-dotted line, S-ML.

that to which PCSE is consistent. It worsens S-OLS SE estimation because simultaneity bias in $\hat{\rho}$ is all that induces the spatial correlation that remains in the $\mathbf{ee}'$ matrix. This pattern will also correlate with $\mathbf{Wy}$'s and $\mathbf{s}$'s parts of the $\mathbf{X}'\mathbf{X}$ matrix, thereby leading PCSE to "correct" unnecessarily. S-2SLS produces SE estimates within 11% of true estimation variation except in the experiment with strong interdependence, $N = 40$ and $T = 20$ sample dimensions, where reported SEs are 22%–23% overconfident. S-ML consistently produces SE estimates within 11% of true estimation variation without exception, and so emerges as nearly (weakly) dominant on this desiderata also.

Figures 5 and 6 plot SE accuracy ratios in the now-familiar grid of broader experimental conditions, generally confirming the conclusions drawn from the (different) tabulated subset. Nonspatial OLS is not only greatly and increasingly biased at moderate to strong interdependence but wildly overconfident regarding $\hat{\beta}_s$ over that range as well. S-OLS offers acceptably honest SE reporting for $\hat{\beta}_s$ and $\hat{\rho}$ in longer T samples and across most of the middle ranges of ρ values in shorter T ones, but the latter, unfortunately, are exactly the ranges where its bias is worst. S-ML and S-2SLS SEs are both acceptably accurate over most ρ strengths and sample dimensions, and, although S-2SLS has some difficulty with the $N = 50 \gg T = 20$ case as it did with the $N = 40 \gg T = 20$, neither emerges as uniformly dominant.[24] Indeed, reminiscent of Beck and Katz's (1995, 1996) critique of Parks-Kmenta,

[24] LeSage's S-ML *MatLab* code uses $\mathbf{ee}'$ in the variance-covariance formula, whereas the maintained assumptions, most directly the sphericity of $V(\varepsilon)$, would allow the tighter $\mathbf{e}'\mathbf{e}/n$. Using $\mathbf{ee}'$ is roughly equivalent to calculating PCSEs for the ML estimates. Our tables impose $\mathbf{e}'\mathbf{e}/n$; our figures use LeSage's original $\mathbf{ee}'$. Both produce the correct calculation in expectation under current assumptions, but the former is more efficient (whereas the latter should be more consistent or robust to deviations from error sphericity that relate to the patterns in the regressors' variance-covariance matrix. This relative inefficiency and the smaller number of the graphed trials probably account for the greater variability in the S-ML SEs performance there.

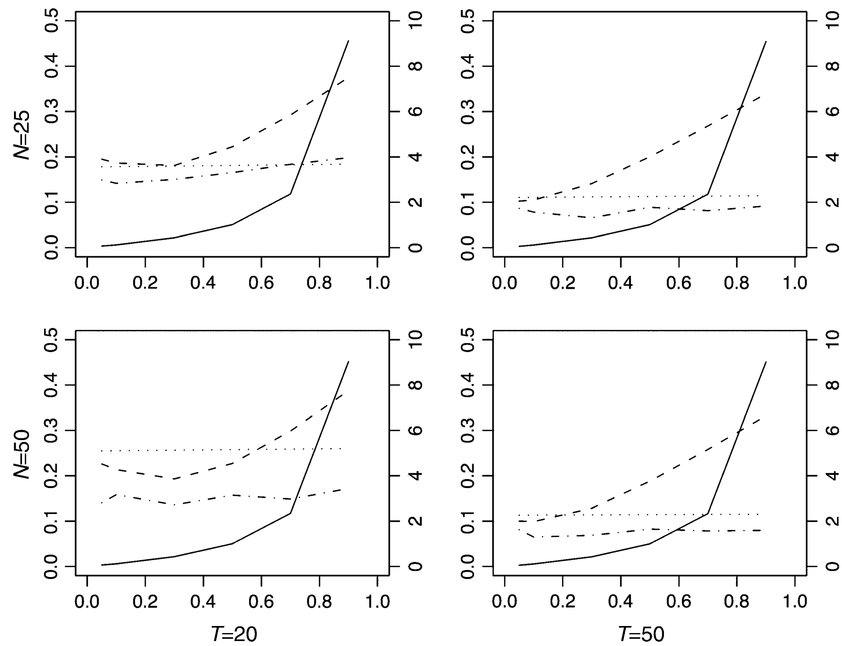

Fig. 4 RMSE $\hat{\beta}_s$ plotted against true ρ across representative $N \times T$ sample dimensions. Solid line, nonspatial OLS; dashed line, S-OLS; dotted line, S-2SLS-IV; dashed-dotted line, S-ML; nonspatial OLS results plotted against the larger-scale second y-axis on the right.

notice that all the spatial estimators, especially the LS ones, seem to require longer T length relative to N width samples (even more so than just larger samples) for their best SE accuracy. S-2SLS and S-ML thus emerge as the clear winners on SE accuracy.

4.4 *Conclusions from Analytic and Simulation Results*

In summary, we have shown analytically that nonspatial OLS ignores spatial interdependence and suffers omitted-variable biases fostering overestimation of nonspatial unit-level (domestic, individual) and contextual (exogenous-external) effects as a result. In simulations, we demonstrated that these biases quickly become substantively sizeable at even very modest interdependence strength ($\rho > .1$ or so) and gargantuan at greater ρ. The biases concentrate in common shocks for homogenous diffusion processes. SEs for these overestimated effects are also dramatically underestimated in these ranges, and PCSE offers only limited amelioration. In short, given any noticeable interdependence, nonspatial OLS is an unmitigated disaster. We have also shown analytically that S-OLS suffers converse simultaneity biases that tend toward overestimation of interdependence strength, inducing underestimation of nonspatial factors' roles. However, our simulations demonstrated that these simultaneity biases generally remain mild over a small-to-moderate interdependence range ($\rho < .3$). S-OLS is also rather efficient, so only its problems with SE accuracy in smaller T samples would argue against it as a simple and effective strategy for mildly spatially interdependent contexts. Programs and instruments to implement S-ML and S-2SLS, each consistent and asymptotically (normal and) efficient under its assumptions,

Table 3 Mean reported SE divided by SD across 1000 trials (SE accuracy)

				OLS		*S-OLS*			
			Coefficient	*SE*	*PCSE*	*SE*	*PCSE*	*S-2SLS*	*S-ML*
$\rho = .1$	$N = 5$	$T = 20$	$\hat{\beta}_s$	*.871*	.895	*0.814*	*.758*	0.901	***0.938***
			$\hat{\rho}$	—	—	*0.773*	***.727***	0.893	***0.943***
		$T = 40$	$\hat{\beta}_s$	.914	***.975***	0.928	.880	0.971	0.965
			$\hat{\rho}$	—	—	*0.859*	*.826*	***0.971***	0.955
	$N = 40$	$T = 20$	$\hat{\beta}_s$	.927	***.951***	**0.727**	**.689**	0.943	0.911
			$\hat{\rho}$	—	—	**0.715**	**.677**	***0.931***	0.890
		$T = 40$	$\hat{\beta}_s$	*.867*	.933	*0.849*	*.807*	0.926	***0.956***
			$\hat{\rho}$	—	—	*0.829*	*.790*	0.925	***0.942***
$\rho = .5$	$N = 5$	$T = 20$	$\hat{\beta}_s$	**.491**	*.799*	***1.034***	*.831*	0.901	0.936
			$\hat{\rho}$	—	—	***0.961***	*.776*	0.907	0.916
		$T = 40$	$\hat{\beta}_s$	**.485**	*.822*	1.017	*.857*	***0.993***	.979
			$\hat{\rho}$	—	—	0.980	*.837*	1.016	***1.010***
	$N = 40$	$T = 20$	$\hat{\beta}_s$	**.364**	*.782*	0.933	**.745**	*0.770*	***0.938***
			$\hat{\rho}$	—	—	0.914	**.728**	*0.779*	***0.936***
		$T = 40$	$\hat{\beta}_s$	**.370**	*.836*	***1.010***	*.840*	0.941	0.924
			$\hat{\rho}$	—	—	***1.000***	*.837*	0.954	0.914
Number of SE accuracy wins/ties				0	2	4	0	3	7
Number of notable/massive (>12.5%/ >25%) inaccuracies				6/4	4/0	7/2	15/5	2/0	0/0

Note. SE accuracy winner in bold italics, notable (>12.5%) inaccuracies in italics, and massive (>25%) inaccuracies in bold.

are available to handle other contexts.[25] Our simulations show S-2SLS to be admirably unbiased in both parameter and SE estimation,[26] especially as sample sizes increase, but efficiency can be an issue.[27] S-ML, for its part, weakly dominates or nearly does so across sample-dimension and parameter-value conditions in mean-squared error terms, especially and importantly so in smaller sample size, smaller ρ conditions. It does suffer some negative bias at lower interdependence strengths, and possibly some mildly erratic SE estimation under some conditions, but neither problem seems very large in magnitude. Therefore, comparing the two consistent estimators, S-ML and S-2SLS, against the simpler alternative S-OLS, we conclude that modest interdependence strength and imperfect exogeneity of instruments[28]—common conditions, we suspect—favor adequacy of the simpler LS over the IV or ML spatial estimators. Conversely, when interdependence is stronger, a consistent estimator should be chosen over S-OLS, whose simultaneity bias grows. S-ML has efficiency advantages over S-2SLS, although these come at some nonnegligible

[25]At the moment, S-2SLS is easier to implement than S-ML with the software packages commonly used by political scientists. We have not found existing S-ML code for *Stata*, e.g., to be reliable or efficient. LeSage's (1999) *MatLab* code, *sar.m* available from www.spatial-econometrics.com, is both *once one corrects a crucial error in line 183 of the code; the line references the incorrect element of the estimated coefficient variance-covariance matrix for returning the SEs of the spatial lag coefficient*. One might also want to tighten the SE formula as suggested in note 23.
[26]Whether PCSEs might further improve or worsen SE estimation awaits further research.
[27]Generalized method of moments extensions of S-2SLS (see, e.g., Kelejian 1993) might improve efficiency.
[28]Franzese and Hays (2004) show imperfect instruments add inconsistency to IV's inefficiency concerns.

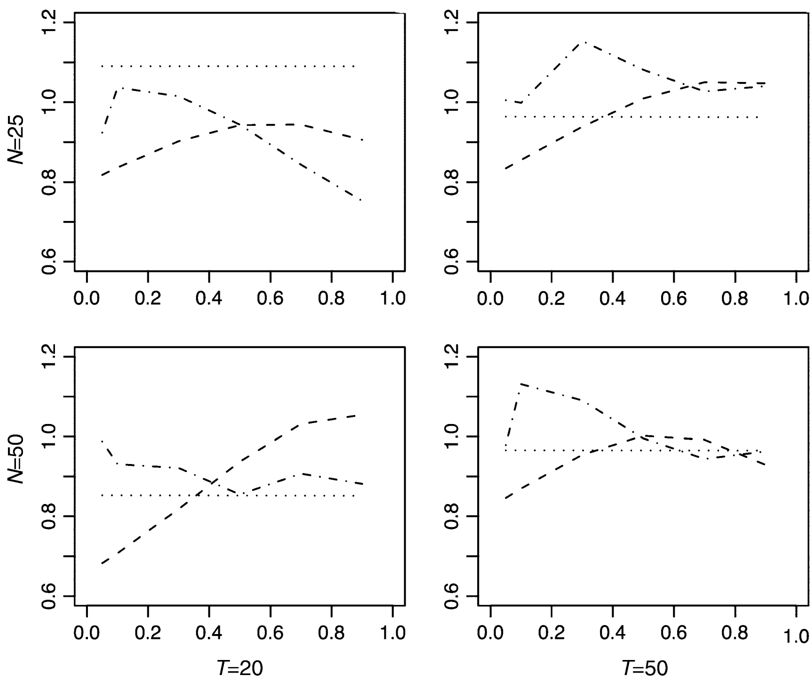

Fig. 5 SE $\hat{\rho}$ accuracy plotted against true ρ across representative $N \times T$ sample dimensions. Dashed line, S-OLS; dotted line, S-2SLS-IV; dashed-dotted line, S-ML. SE accuracy is gauged by the ratio of the average estimated SE to the true SD of the sampling distribution. Values less than one indicate overconfidence.

computational costs. The downward small sample bias of S-ML versus S-2SLS's upward one, plus their generally accurately reported SEs, suggests that S-ML might also be the more conservative hypothesis-testing option. On the other hand, unbiasedness and perhaps simplicity argue for S-2SLS, but either will typically work to redress S-OLS's simultaneity biases at stronger interdependence strengths. And, to repeat, nonspatial OLS is absolutely disastrous and to be avoided unless interdependence is known to be very weak or nonexistent.

5 Empirical Illustration: Estimating and Calculating Spatial Effects in TSCS Models

Calculation and presentation of *effects* in empirical models with spatial interdependence, as in any model beyond the purely linear additive, involve more than simply considering coefficient estimates.[29] In empirical models containing *spatial dynamics*, as in those with only temporal dynamics, coefficients on explanatory variables give only the *predynamic* impetuses to the outcome, which, incidentally, are actually unobservable if spatial dynamics are instantaneous (i.e., incur within an observation period). This section discusses the calculation of spatial multipliers, which allow expression of the *effects* across units of counterfactual shocks (Anselin 2003), and it applies the delta method to compute SEs for

[29]Even in models differing from the purely linear additive only in containing (time) dynamics (see, e.g., DeBoef and Keele 2005) or multiplicative interaction terms (see, e.g., Kam and Franzese 2007), *coefficients* and *effects* are different things.

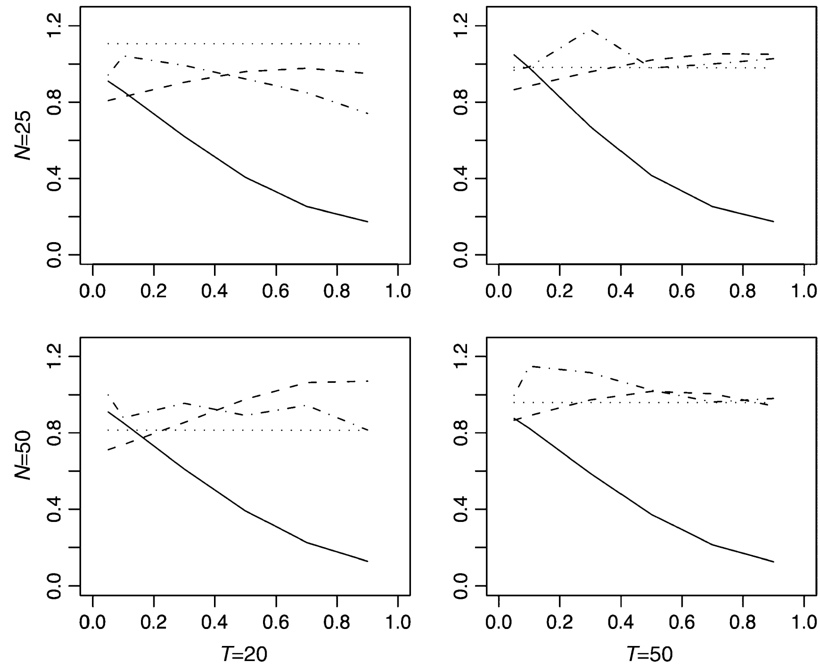

Fig. 6 SE $\hat{\beta}_s$ accuracy plotted against true ρ across representative $N \times T$ sample dimensions. Solid line, nonspatial OLS; dashed line, S-OLS; dotted line, S-2SLS-IV; dashed-dotted line, S-ML. SE accuracy is gauged by the ratio of the average estimated SE to the sampling distribution SD. Values less than one indicate overconfidence.

these *effects*. We illustrate these calculations with an analysis taken from Franzese and Hays (2006b) that estimated spatiotemporal lag models to assess empirically the strategic interdependence among European countries in active labor market policy, specifically labor market training (LMT) spending.[30] We also compare the four estimators here to show that the choice of estimator clearly influences the inferences one draws from the data regarding the determinants of LMT expenditures.

Spatiotemporal-lag models add temporally lagged dependent variables to spatially lagged ones in equation (2):

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \phi\mathbf{M}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{13}$$

where the dependent variable, **y**, is a vector of cross sections stacked by periods this time (i.e., the N first-period observations, then the N second-period ones, etc.). In this ordering, we can express **W** as the Kronecker product of a $T \times T$ identity matrix and an $N \times N$ spatial-weights matrix $(\mathbf{I}_T \otimes \mathbf{W}_N)$. **M** is an $NT \times NT$ matrix of zeros except for ones on the minor diagonal at coordinates $(N+1, 1)$, $(N+2, 2)$, ..., $(NT, NT-N)$, so **My** is the first-order temporal lag, making ϕ the temporal autoregressive coefficient.

[30]The conditional likelihood for the spatiotemporal-lag model, which assumes that the first observation is non-stochastic, is a straightforward extension of equation (12) (see Elhorst 2001, 2003, 2005). Our focus here is on presenting spatiotemporal dynamics, so we do not discuss it here (see Franzese and Hays 2006a).

Table 4 LMT expenditures in Europe (1987–1998)

	LMT/unemployed	*LMT/unemployed*	*LMT/unemployed*	*LMT/unemployed*
Temporal lag	.583*** (.078)	.439*** (.081)	.494*** (.085)	.488*** (.068)
Spatial lag	—	−.435*** (.101)	−.269** (.125)	−.288*** (.068)
Union density	.018* (.011)	.008 (.010)	.012 (.011)	.011 (.009)
Deindustrialization	.062* (.034)	.041 (.032)	.049 (.033)	.048* (.029)
Estimator	OLS	S-OLS	S-2SLS	S-ML

Note. All regressions include fixed period and unit effects; those coefficient estimates suppressed to conserve space. The spatial lags are generated with a binary contiguity–weighting matrix using shared territorial borders as the criterion, excepting that France, Belgium, and the Netherlands are coded as contiguous with Britain and Denmark as contiguous with Sweden. All the spatial weights matrices are row standardized.
Significant at *10%, **5%, ***1%.

With positive employment spillovers across borders, countries have incentives to free ride on neighbors' LMT expenditures. If so, the estimated spatial lag coefficient $\hat{\rho}$ should be negative. In a sample of annual data from 1987 to 1998 in 15 European countries,[31] the dependent variable $\mathbf{y}$ is the log of LMT expenditures per unemployed worker ($2000, purchasing-power parity [PPP]). In addition to its spatial and temporal lags, we include as regressors: macroeconomic performance (gross domestic product [GDP] per capita, unemployment), labor market characteristics (deindustrialization, union density), external conditions (trade openness, foreign direct investment), domestic politics (Left Party and Christian Democratic cabinet seats, left libertarian vote, government consumption), and country and year dummies to allow unit and period effects, a strategy that, given the tendency for inadequate modeling of such heterogeneity to induce overestimation of spatial interdependence, errs conservatively against our hypothesis of interdependent active labor-market (ALM) policy.

We calculated the spatial lag, $\mathbf{Wy}$, using a standardized *binary contiguity weights matrix* that first codes $w_{ij} = 1$ for countries i and j, which share borders, and $w_{ij} = 0$ for pairs ij, which do not.[32] Then, we row standardize the resulting matrix by dividing each cell in a row by that row's sum. We then estimate equation (13) by OLS, S-OLS, S-2SLS, and S-ML. Given the analytical and experimental results of previous sections, we expect the OLS estimates for $\boldsymbol{\beta}$ (and ϕ) to suffer inflation bias and the S-OLS estimates to suffer attenuation bias. The S-OLS estimate for ρ is likely to be inflated. We expect both S-2SLS and S-ML to give unbiased estimates with the latter's relative efficiency yielding smaller SEs.

Table 4 presents a subset of the estimates that are affected most consequentially by the choice of estimator. The OLS results show a marginally statistically significant positive relationship of union density and deindustrialization to government LMT spending. As expected, these estimates are larger than any of the spatial estimators' estimates of those coefficients, likely reflecting the omitted-variable bias of the nonspatial estimator, and favoring these (apparently spatially correlated) domestic explanators. Note also that the S-OLS estimates for $\boldsymbol{\beta}$ (and ϕ) are smaller than the S-2SLS and S-ML ones, whereas $\hat{\rho}_{\text{S-OLS}}$ is larger. This likely reflects S-OLS's simultaneity inflation bias in $\hat{\rho}$, and its corresponding induced attenuation bias in $\hat{\boldsymbol{\beta}}_{\text{S-OLS}}$(and ϕ). Penultimately, the SE estimates for the S-ML estimator are indeed smaller than the S-2SLS SE estimates as expected. Finally, notice also that the spatial estimates on both the union density and deindustrialization

[31]Austria, Belgium, Denmark, Germany, Greece, Finland, France, Ireland, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

[32]We code France, Belgium, and the Netherlands as "contiguous" with Britain, and Denmark with Sweden.

variables are statistically insignificant, which demonstrates that the inferences one draws about the relationship of domestic factors like labor market characteristics (union density and deindustrialization) to LMT spending depend on the estimator chosen. Given what we know about the relative performance of the estimators, we conclude that ignoring the spatial interdependence among European Union (EU) member states' LMT spending (nonspatial OLS) could lead to the mistaken conclusion that union density significantly increases LMT (a Type I error caused by omitted-variable bias), whereas estimating a spatial lag model by S-OLS could lead incorrectly to a failure to reject the null hypothesis that deindustrialization does not affect spending on LMT (a Type II error caused by an induced simultaneity bias). Moreover, the analyst who estimates equation (13) by S-OLS will overestimate the substantive magnitude of free riding in EU member states' LMT.

Note that the estimated effects of individual EU countries' ALM policies on their neighbors' are not understood immediately and fully from the $\hat{\rho}$ estimate alone. To see these *effects*, we need to calculate the *spatial multiplier* implied by model (13). Calculating the cumulative steady-state spatial effects is most convenient working with the spatiotemporal-lag model in $(N \times 1)$ vector form:

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t. \tag{14}$$

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 equation (14) and solve. This gives the following steady-state effect, assuming stationarity and that the exogenous right-hand side terms, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, remain permanently fixed to their counterfactual levels:

$$\begin{aligned}
\mathbf{y}_t &= \rho \mathbf{W} \mathbf{y}_t + \phi \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t = (\rho \mathbf{W} + \phi \mathbf{I}) \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\
&= [\mathbf{I}_N - \rho \mathbf{W} - \phi \mathbf{I}_N]^{-1} (\mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t) \\
&= \begin{bmatrix}
1-\phi & -\rho w_{1,2} & \cdots & \cdots & -\rho w_{1,N} \\
-\rho w_{2,1} & 1-\phi & & & \vdots \\
\vdots & & \ddots & & \vdots \\
\vdots & & & 1-\phi & -\rho w_{(N-1),N} \\
-\rho w_{N,1} & \cdots & \cdots & -\rho w_{N,(N-1)} & 1-\phi
\end{bmatrix}^{-1} (\mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t) \\
&\equiv \mathbf{S} \times (\mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t).
\end{aligned} \tag{15}$$

To get SE estimates for these steady-state estimates, we use the delta method; i.e., give a first-order Taylor series linear approximation to nonlinear equation (15) around the estimated parameter values and determine the asymptotic variance of that linear approximation. To do this, begin by denoting the ith column of $\mathbf{S}$ as $\mathbf{s}_i$ and its estimate as $\hat{\mathbf{s}}_i$. The steady-state spatial effects of a one-unit increase in explanatory variable k in country i are then $\mathbf{s}_i \beta_k$ giving delta method SEs of

$$\widehat{\mathbf{V}(\hat{\mathbf{s}}_i \hat{\beta}_k)} = \left[\frac{\partial \hat{\mathbf{s}}_i \hat{\beta}_k}{\partial \hat{\boldsymbol{\theta}}} \right] \widehat{\mathbf{V}(\hat{\boldsymbol{\theta}})} \left[\frac{\partial \hat{\mathbf{s}}_i \hat{\beta}_k}{\partial \hat{\boldsymbol{\theta}}} \right]', \tag{16}$$

Table 5 Steady-state spatial effects of deindustrialization on training

	Austria	*Germany*	*Switzerland*
Austria	.102* (.061)	−.025 (.016)	−.027 (.017)
Germany	−.01 (.006)	.102* (.061)	−.008* (.005)
Switzerland	−.018 (.011)	−.013* (.008)	.102* (.061)

Note. The cells report steady-state spatial effects of a 1% increase in the column country's degree of deindustrialization on its own LMT expenditures and the expenditures of its European counterparts (identified by the rows). SEs for this effect are in parentheses.
*Significant at 10%.

where $\hat{\boldsymbol{\theta}} \equiv [\hat{\rho} \quad \hat{\phi} \quad \hat{\beta}_k]'$, $\left[\frac{\partial \hat{\mathbf{s}}_i \hat{\beta}_k}{\partial \hat{\boldsymbol{\theta}}}\right] \equiv \left[\frac{\partial \hat{\mathbf{s}}_i \hat{\beta}_k}{\partial \hat{\rho}} \quad \frac{\partial \hat{\mathbf{s}}_i \hat{\beta}_k}{\partial \hat{\phi}} \quad \hat{\mathbf{s}}_i\right]$, 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 ith columns of $\hat{\beta}_k \hat{\mathbf{S}} \mathbf{W} \hat{\mathbf{S}}$ and $\hat{\beta}_k \hat{\mathbf{S}} \hat{\mathbf{S}}$, respectively.

Table 5 illustrates calculations of equations (15) and (16) for a subset of countries in our sample using the S-ML estimates of equation (14) from Table 4. The first number in each cell is the steady-state effect of a 1% increase in the column country's level of deindustrialization on the row country's LMT. The number in parentheses is the SE of this spatial effect. In Table 5, we see that a permanent 1% increase in the levels of Austrian, German, and Swiss deindustrialization increases their respective spending on LMT by 10.2% in the long run. Also, focusing on the statistically significant feedback, a permanent increase in Germany's level of deindustrialization reduces LMT expenditures in Switzerland by 1.3%, and the same increase in Switzerland decreases spending in Germany by .8%.

The spatiotemporal response path of the $N \times 1$ vector of unit outcomes, $\mathbf{y}_t$, to the exogenous right-hand side terms, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, could also emerge by rearranging equation (14) to isolate $\mathbf{y}_t$ on the left-hand side:

$$\mathbf{y}_t = [\mathbf{I}_N - \rho \mathbf{W}_N]^{-1} \{\phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\}. \tag{17}$$

This formula gives the response paths of all units $\{i\}$ to hypothetical shocks to $\mathbf{X}$ or $\boldsymbol{\varepsilon}$ in any unit(s) $\{j\}$, just by setting $(\mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)$ to one in the rows corresponding to $\{j\}$. This formulation is especially convenient for plotting estimated response paths in a spreadsheet, e.g. To calculate marginal spatiotemporal effects (noncumulative), i.e., the incremental change at some time $t + k$ in the overtime path resulting from a permanent one-unit change in an explanatory variable at time t, and their SEs, working with the entire $NT \times NT$ matrix is easier. Simply redefine $\mathbf{S}$ in equation (15) as $\mathbf{S} \equiv [\mathbf{I}_{NT} - \rho \mathbf{W} - \phi \mathbf{M}]^{-1}$ and follow the same steps. Franzese and Hays (2006a) illustrate calculation and presentation of such response paths with confidence intervals for a replication of Beck et al. (2006).

6 Conclusion

Our analytic and experimental explorations suggest, first, that the omitted-variable biases of excluding *interdependence* in nonspatial OLS pose far greater concerns than the simultaneity biases of including them in S-OLS, under a wide range of likely substantive conditions. Analysts who ignore interdependence will typically overestimate domestic (individual), exogenous-external, and context-conditional effects, with this overestimation concentrating in those factors most correlated with the omitted interdependence mechanism. Therefore, researchers always do better to include the spatial lags needed to specify the interdependence implied by theory than to ignore/omit that implication. However, we also showed that simultaneity biases from S-OLS regressions that include spatial lags to reflect interdependence can be appreciable when interdependence is more than moderate, that the biases tend toward overestimating interdependence strength (ρ), and that underestimation of its SE also occurs regularly.[33] Thus, hypothesis tests based on S-OLS estimations of *correctly specified* models like (2) or (13) would often be biased toward finding strong interdependence effects, with the relevant *t* statistics having inflated numerators and deflated denominators.

Fortunately, one can estimate models of interdependent processes like (2) or (13) by S-2SLS or by S-ML to obtain consistent estimates of ρ and $\boldsymbol{\beta}$. In fact, the former is not difficult to implement[34] because the spatial structure of the data itself suggests potential instruments. Valid instruments must satisfy that their (asymptotic) covariance with endogenous regressors—here, the spatially lagged outcomes in other units—are nonzero, and preferably large, whereas their (asymptotic) covariance with the residual is zero. Fortunately, too, our experiments show that such S-2SLS estimates produce not only consistent estimates but also essentially unbiased ones, even in relatively small samples under the experimental conditions considered here. Moreover, the accompanying S-2SLS SE estimates fairly accurately reflect the true sampling variability of the coefficient estimates across most sample size and parameter conditions explored. This suggests that S-2SLS, unlike S-OLS, will produce acceptably unbiased hypothesis tests. Unfortunately, S-2SLS is not very efficient and, indeed, is often outperformed in mean-squared error terms by simple S-OLS and almost always by S-ML. Therefore, these unbiased hypothesis tests may also be relatively weak.

S-ML seems to offer weakly dominant efficiency and generally solid performance in unbiasedness and SE accuracy, although it sometimes yields relatively little in reduced bias or enhanced efficiency relative to S-OLS and falls a little short of S-2SLS on unbiasedness grounds. Considering its computational intensity against this winning, but not universally dominant, performance, we do not see compelling reason to favor S-ML estimation too strenuously and generally over alternatives at this point. Nor do we necessarily see reason to push either of the consistent estimators, S-ML or S-2SLS, in all cases because S-OLS performs quite acceptably well under weak or mild interdependence-strength circumstances and, indeed, in some experimental conditions, even better than these more complicated alternatives. Unless one is *certain* that interdependence is zero or very small, however, the nonspatial models that are currently the common default are clearly inferior and to be avoided.

References

Anselin, L. 1988. *Spatial econometrics: Methods and models*. Boston: Kluwer Academic.

Anselin, L. 2001. Spatial econometrics. In *A companion to theoretical econometrics*, ed. B. Baltagi, 310–30. Oxford: Basil Blackwell.

Anselin, L. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26:153–66.

Bartels, L. 1991. Instrumental and 'quasi-instrumental' variables. *American Journal of Political Science* 35: 777–800.

Beck, N., Gleditsch, K. S., and Beardsley, K. 2006. Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly* 50:27–44.

[33]Furthermore, PCSEs did not seem to help much in this last regard.

[34]The latter is not so much difficult as computationally intensive. These two methods are not exhaustive of those potentially capable of returning consistent β and ρ estimates (see note 8).

Beck, N., and Katz, J. 1995. What to do (and not to do) with time series cross section data in comparative politics. *American Political Science Review* 89:634–47.

Beck, N., and Katz, J. 1996. Nuisance vs. substance: Specifying and estimating time series cross section models. *Political Analysis* 6:1–34.

De Boef, S., and Keele, L. 2006. Taking time seriously, dynamic regression models. Paper presented at the annual meeting of the Political Methodology Society, Tallahasee, FL.

Elhorst, J. P. 2001. Dynamic models in space and time. *Geographical Analysis* 33:119–140.

Elhorst, J. P. 2003. Specification and estimation of spatial panel data models. *International Regional Science Review* 26:244–68.

Elhorst, J. P. 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37:85–106.

Franzese, R., and Hays, J. 2004. Empirical modeling strategies for spatial interdependence: Omitted-variable vs. simultaneity biases. Summer meetings of the Political Methodology Society.

Franzese, R., and Hays, J. 2006a. Spatio-temporal models for political-science panel and time-series-cross-section data. Summer meetings of the Political Methodology Society.

Franzese, R., and Hays, J. 2006b. Strategic interaction among EU governments in active-labor-market policymaking: Subsidiarity and policy coordination under the European employment strategy. *European Union Politics* 7:167–89.

Kam, C., and Franzese, R. 2007. *Modeling and interpreting interactive hypotheses in regression analysis*. Ann Arbor: University of Michigan Press.

Kelejian, H. H., Prucha, I. R., and Yuzeforich, Y. 2004. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. *Advances in Econometrics* 18:163–198.

Kelejian, H. H., and Robinson, D. P. 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors and an application to a county expenditure model. *Papers in Regional Science* 72:297–312.

LeSage, J. P. 1999. *Spatial econometrics.* http://rri.wvu.edu/WebBook/LeSage/spatial/spatial.html.

Lin, T. M., Wu, C. E., and Lee, F. Y. 2006. ‘Neighborhood’ influence on the formation of national identity in Taiwan: Spatial regression with disjoint neighborhoods. *Political Research Quarterly* 59:35–46.

Ord, K. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70:120–6.