Spatial-Econometric Models of Interdependence

(*A Book Prospectus*)

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

Jude C. Hays (JCHays@uiuc.edu)
Assistant Professor of Political Science
The University of Illinois, Urbana-Champaign

1 June 2007

ABSTRACT

This book covers the specification, estimation, interpretation, and presentation of empirical models that reflect spatial and spatio-temporal interdependence among observations.

Chapter 1 provides overview of the substance and theory of interdependence in the social sciences, with an emphasis on political science and, especially therein, comparative and international politics and political economy. Part A surveys the broad substantive range in which, and the multifarious mechanisms by which, spatial interdependence arises across these subjects. Part B first presents and discusses a general theoretical model which demonstrates that interdependence arises whenever the marginal utility to unit *i* of its actions/policies depends on the actions/policies of some unit(s) *j*. The model illuminates the logical connections from positive externalities to strategic-substitute relations, free-rider incentives, first-/early-mover advantages, and strategic-delay/war-of-attrition dynamics on the one hand, and from negative externalities to strategic-complement relations, beggar-thy-neighbor incentives, last-/late-mover advantages, and competitive-race (e.g., *race-to-the-bottom*) dynamics on the other. It then presents, and derives empirically testable propositions from, a more-specific theoretical model in the substantive context of international capital-tax competition for foreign investment before offering briefer overview of theoretical models of interdependence in other substantive areas across the social sciences.

Chapter 2 introduces and explores the serious challenges for empirical analysis posed by interdependence. It first contrasts three possible sources of spatial correlation—correlated unit-specific explanators, correlated exogenous-external explanators, and true interdependence—distinguishing between which is the essence of *Galton's* (famous) *Problem*. Whether these spatial correlations or interdependencies arise in stochastic and/or systematic components creates a thorny additional complication of considerable importance because each implies substantively very different spatial dynamics. It then discusses in this vein several empirical observations commonly thought to be evidence of interdependence—e.g., S-shaped patterns of policy-adoption or temporal fixed-effects or regional dummies—and explains how they, unlike well-specified spatial-econometric models with spatial lags, are incapable of distinguishing among possible sources of spatial correlation. Next it introduces and discusses various measures and tests of spatial correlation and/or interdependence, some of which are capable of offering at least some judgment on their alterative possible sources.

Chapter 3 turns to the technical specification and estimation of empirical models of interdependence. It begins by presenting a generic empirical model with spatial lag, unit-specific and exogenous-external shocks, and the interaction of the latter two. Such a model corresponds to modern theoretical and substantive expectations in political science and highlights the empirical difficulty of distinguishing the possible sources of spatial correlation. It then introduces and discusses four common estimation strategies for such spatial-lag models: non-spatial OLS (or ML), OLS (or ML) with spatial lags, spatial method-of-moments (IV/2SLS/GMM), and fully specified spatial ML. It next identifies four broad classes of models with spatial interdependence—spatial-error, spatial-lag, spatio-temporal-lag models, and spatial- and spatio-temporal-lag qualitative-dependent-variable models. The final section discusses more technically than did Chapter 2 the challenges in estimating these sorts of models for these estimators: omitted-variable and simultaneity biases for the two simpler estimators, specification and measurement error for all of them, and the many serious issues surrounding the all-important spatial-connectivity matrix, **W**, that describes the relative strengths of interdependence from units *j* to *i*. This last part includes discussion of (1) the pre-specification of **W**, which is current practice and is perhaps the most crucial stage of empirical

spatial analysis (*inter alia*, this part explains and explores *Achen's Problem* with lagged-dependent-variable models, namely that insufficiently accurate dynamic specifications push the LDV toward "stealing explanatory power" from the other regressors, in the spatial context), (2) how estimators' properties tend to vary, often in complicated fashion, with the nature of **W** (this is what necessitates the many repeated sets of simulations in Chapter 4), (3) issues raised by multiple-**W** models (used in the literature, for example, to evaluate the strength of several alternative mechanisms of interdependence that may be operating), and (4) some inroads into the parameterization and estimation of **W** (widely agreed to be an important future direction for spatial-econometric modeling, but plagued by thorny complications and not far advanced as yet).

Chapter 4 considers the properties of these estimators. It first derives analytically (in simple cases, asymptotically) the respective omitted-variable and simultaneity biases of the non-spatial estimators and the LS or ML estimators for models with spatial lags, spatial and temporal lags, and multiple spatial lags. It then evaluates by simulation the bias, efficiency, and standard-error-accuracy performance of all the estimators in richer contexts and limited samples. These simulations explore, first, data-generating processes and empirical specifications matching the generic empirical model of Chapter 3, with common exogenous-external shocks and uniform interdependence among spatial units. The spatial lag is thus the simple average of the dependent variable in other units, representing a "rough & ready" spatial-lag proxy to which empirical researchers may often resort. *Galton's Problem* manifests severely here, given the close similarity of uniform interdependence and common exogenous-external shocks. Across simulations, the estimators are evaluated first with empirical specifications matching the true data-generating process; then with measurement-*cum*-specification error in the empirical model of the exogenous shock or the spatial-interdependence pattern; and then, for the moment (instrumentation) estimators, with imperfectly exogenous instruments under varying degrees of spatial correlation and endogeneity among right- and left-hand-side variables and, for the fully specified ML estimators, with varyingly sizable violations of their strong distributional assumptions. For the spatio-temporal models, the simulations also explore *Achen's Problem* with lagged-dependent-variable (LDV) models in the spatial context. All these simulation exercises are then repeated for several realistically non-homogenous patterns of spatial-interdependence literally drawn from substantive literatures; i.e., elements of the spatial-connectivity matrix used for experimental consideration are drawn directly from actual datasets used in the literature. One set of patterns reflects international economic relations among the units; another reflects "co-memberships" (e.g., in organizations, language or cultural groups, etc.; shared borders, dyads, etc.) among the units; a third reflects the geographic or Euclidean distance between the units; and the fourth pattern reflects strategic-complement (cooperative) and -substitute (competitive) relations between the units. The last two sections of the chapter conduct similar sets of simulations of multiple-**W** and spatial-lag qualitative-dependent-variable models.

Chapter 5 shows how to calculate (by the delta method and by simulation), and offers suggestions on presenting (via tabular feedback grids, spatio-temporal response-path graphs, and maps), estimates and standard errors of estimates of spatial and spatio-temporal multipliers, effects, dynamics (response paths), and long-run steady-state effects.

Chapter 6 illustrates all this via replications and extensions of existing work using spatial lags or failing to use them when theory and substance clearly imply them (Swank & Steinmo 2002, Hays 2003, Basinger and Hallerberg 2004, Beck, Gleditsch, & Beardsley 2006, Volden 2006, Franzese & Hays 2006).

Two appendices provide Stata™ code to implement all specification, estimation, testing, and presentation procedures described in the text, and all data and code to replicate the empirical work.

OUTLINE / TABLE OF CONTENTS

I. **(CHAPTER 1)** The Substance & Theory of Spatial Interdependence

A. The Broad Substantive Range of Spatial Interdependence

1. Topics and Literature Review

2. Mechanisms: Coercion; Competition, Cooperation, and Externalities; Learning and Emulation; Migration

3. => Tobler's First Law of Geography (*all things related, near things more so*), with Beck, Gleditsch, & Beardsley's addendum (*space is more than geography*).

B. Theoretical Models Strategic Interdependence:

1. A General Theoretical Model of Interdependence

a) Basic Principle: if marginal utility of *i*'s actions depend on *j*'s actions, then interdependence

b) General Properties:

(1) Positive Externalities => Strategic Substitutes wherein policies/actions move in opposite directions, i.e., negative interdependence, in which *free-rider* problems & late-mover advantages, & so strategic delay/wars of attrition dynamics, arise.

(2) Negative Externalities => Strategic Complements wherein policies/actions move in same directions, i.e., positive interdependence, in which *beggar-thy-neighbor* problems & early-mover advantages, & so *races-to-the-bottom* (or *top*, or, generally, competitive races), arise.

c) Examples & Discussion

2. A Specific Theoretical Model of Interdependence in One Example Context: The Political Economy of Globalization, Capital-Tax Competition, and Domestic Fiscal-Policy Autonomy

a) Review of theories in previous literature; all imply interdependence, but most previous empirical specifications fail to incorporate that implication.

b) A citizen-candidate model of capital-tax competition and foreign investment (from Persson & Tabellini 2000).

(1) Also implies interdependence & renders it formally explicit.

(2) Discuss model & some examples of derived comparative statics.

3. Briefer overview of several other specific theoretical models, in other contexts within and beyond comparative/international political economy.

II. **(CHAPTER 2)** The Empirical Challenge of Spatial Interdependence

A. Galton's Problem: 3 Sources of Spatial Correlation, and a modern complication

1. correlated unit-specific conditions/explanators (examples)

2. correlated exogenous-external conditions/explanators (examples)

3. interdependence (examples)

4. complication: distinguish error/stochastic-term correlation or interdependence from systematic-component/outcome interdependence: implied spatial-dynamics differ substantively importantly

5. Galton's Problem is that these things look very much alike empirically, and so the fact that any, all, or any combination may operate in any given context presents an extreme empirical challenge to distinguish them (by any estimation strategy or, for that matter, by any empirical methodology, quantitative or qualitative).

B. Empirical observations often taken as evidence of interdependence that not necessarily

1. Dummies for Groups/Regions

2. S-Shaped Adoption Curves

3. Measures & Tests of Spatial Correlation/Clustering (briefly introduced)

4. Parks FGLS, PCSE, & other estimation strategies based on error-covariance matrices with spatial patterns

C. Measures and tests of spatial correlation and interdependence:

1. Discusses how each of the above by itself and, for that matter, a spatial-lag model by itself fail to distinguish among the multiple possible sources of spatial correlation, including interdependence, and so are not necessarily evidence of the latter.

2. Elaborates some appropriate nested & non-nested test-strategies of fuller models

III. **(CHAPTER 3)** Specifying and Estimating Empirical Models of Spatial Interdependence

A. A Generic Empirical Model of Open Socio-Political-Economies with Context-Conditionality and Interdependence

B. Estimators for Spatial-Interdependence Models

1. Non-Spatial Least-Squares and Maximum-Likelihood models (inconsistent: omitted-variable bias)

2. Least-Squares and (Quasi) Maximum-Likelihood with Spatial Lags (inconsistent: simultaneity bias)

3. Spatial Method-of-Moments Estimators:

a) Spatial Instrumental-Variables / Two-Stage-Least-Squares (consistent but inefficient)

b) Spatial Generalized-Method-of-Moments (consistent and asymptotically efficient)

4. (Fully Specified) Spatial Maximum-Likelihood (consistent and asymptotically efficient and normal)

C. Classes of Spatial-Econometric Models

1. Spatial-Error Models (spatial interdependence in the stochastic component)

a) Utility & Limits of Consistent Variance-Covariance Estimation (PCSE's)

b) A Simple, Data-Driven Efficiency-Enhancement of PCSE: Utility, Limits

c) True Spatial-Error Models

2. Spatial-Lag Models (spatial interdependence in systematic component or *y*)

a) Can address Galton; key is the relative & absolute accuracy and power of the spatial and non-spatial aspects of the empirical model.

b) As in time-series, modern view is increasingly that this is way to proceed.

3. Spatio-Temporal-Lag Models

a) Time-lagged spatial interdependence: utility and limits. Spatial LS or ML consistent under some conditions. Give and discuss those conditions.

b) Contemporaneous spatial interdependence

4. Spatial- and Spatio-Temporal-Lag Qualitative-Dependent-Variable Models:

a) Spatial Probit

b) Spatial Over- and Under-dispersion in Count Models

D. Challenges for Estimation and Inference

1. Simultaneity of the simple estimation strategies (LS/ML, S-OLS/Quasi-ML)

2. Specification & Measurement Error

a) In the non-spatial components of the model

b) In the spatial-connectivity matrix

c) In instruments' exogeneity or in likelihood

3. **W**: The Spatial-Connectivity Matrix of Relative Interdependence

a) **W** commonly pre-specified; crucial stage of spatial analysis

(1) Examples given, discussion

(2) Achen's Problem with lagged-dependent-variable (LDV) models

(a) Review: Time-LDV v. AR(p) residuals

(b) Analogous issue in spatial-LDV v. spatial-error models

b) Estimators' properties tend to vary, often in complicated fashion, with **W**

c) Multiple-**W** Models: raise issues beyond the usual ones associated with controls (e.g., interpretation, colinearity) due to simultaneity and "overlap"

d) Inroads into parameterization and estimation of **W**

IV. **(CHAPTER 4)** Properties and Performance of Spatial Estimators

A. Analytic Results

1. In 2-unit, 2-regressor case: omitted-variable bias (OVB) in non-spatial OLS and simultaneity in Spatial OLS

2. General Derivation of Simultaneity Biases in Spatial and Spatio-Temporal Model

3. Simultaneity Biases in the multiple-**W** case.

B. Simulation Results: using generic empirical model of interdependence, and:

1. Interdependence Processes/Patterns of Interdependence:

a) Uniform/Homogenous Interdependence: The full set of experiments described are conducted first with this specification; then repeated for all of the following, on which the reports will be briefer.

b) Interdependence proportional to some economic/strategic connection; w_{ij} drawn from actual datasets in lit: trade/capital flows, trade/capital mobility,

c) Strategic complements/substitutes (cooperation/competition) relations between units; w_{ij} drawn from uniform-distribution {-1…1}.

d) Unit interdependence due to "co-membership" (borders, organizations, culturo-linguistic-religious groups, dyads, etc.); w_{ij} drawn from actual datasets in literature. Considerations in exploring this parameter space:

(1) Overlapping v. non-overlapping group-memberships.

(2) Complete group membership (no isolates)

(3) Size & number of these groups relative to totals

e) Unit interdependence proportional to geographic or Euclidean distance b/w units (drawn from actual datasets in the literature)

2. Specification and Measurement Error:

a) Fully accurate specification and measurement: The full set of experiments are conducted first thusly; then repeated for all of the following.

b) Measurement/specification error in exogenous-external factors

c) Measurement/specification error in **W**

d) Exogenous unit-specific and exogenous-external explanators vs. varying magnitudes of endogeneity from left- to right-hand-side as well as varying magnitudes spatial interdependence among left- and right-hand-side.

3. Spatio-Temporal Lag Models

4. Multiple-**W** models; Parameterization & Estimation of **W**

5. Spatial- and Spatio-Temporal-Lag Models of Qualitative Dependent-Variables

V. **(CHAPTER 5)** Interpretation & Presentation of Spatial & Spatio-Temporal Effects & Dynamics:

A. Calculating Spatial and Spatio-Temporal Multipliers, Effects, and Dynamics

1. The Spatial Multiplier and Contemporaneous Spatial Effects

2. The Spatio-Temporal Multiplier and Spatio-Temporal Response-Paths and Long-Run Spatio-Temporal Steady-States

B. Calculating Standard Errors for S and ST Multipliers, Effects, and Dynamics

1. Delta Method

2. Simulation Methods

C. Presenting Estimates of Spatial and Spatio-Temporal Effects and Dynamics

1. Spatial and Spatio-Temporal Effect-Grids
2. Plotting Spatio-Temporal Response-Paths
3. Maps (and Dynamic Maps) of Spatial and Spatio-Temporal Effects

VI. **(CHAPTER 6)** Empirical Applications: Replications & extensions (of existing work using spatial lags or failing to use them when theory & substance imply them, ordered by mechanism of interdependence)

A. Competition: Globalization, Capital-Tax Competition for Investment, & Domestic-Policy Autonomy (Swank&Steinmo APSR '02; Hays WP '03; Basinger&Hallerberg APSR '04)

B. Learning & Emulation: Volden (AJPS 2006) on US State Adoption of Policy Innovations

C. Competition and Negative Externalities: Beck et al. (ISQ 2006) on International Conflict and Trade

D. Cooperation & Positive Externalities: Franzese & Hays (EUP 2006) on ALMP in the EU

VII. (Web) Appendices:

A. Stata™ code to implement all of the procedures described in the text.

B. Data and code to replicate all of the estimation results given in the text.

CONTENTS

I. (CHAPTER 1) The Substance & Theory of Spatial Interdependence

A. The Broad Substantive Range of Spatial Interdependence

1. Topics and Literature Review

Until recently, empirical analyses of spatial interdependence (a.k.a., *diffusion*[1]) in the social sciences remained largely confined to specialized areas of applied economics (e.g., urban/regional, real-estate, environmental economics) and sociology (i.e., network analysis). However, social-scientific interest in and applications of spatial modeling have burgeoned lately, due partly to advances in theory that imply interdependence and in methodology for addressing it, partly to global substantive developments that have raised at least the perception of and attention to interconnectivity, and likely the actual degree and extent of it also, at all levels, from micro/personal to macro/international, and partly to advances in technology for obtaining and working with spatial data. In comparative politics also, spatial empirical analyses have become increasingly common of late. This is a very welcome development as many phenomena that comparativists study entail substantively important spatial interdependence.

Perhaps the most extensive classical and contemporary interest in spatial interdependence surrounds the diffusion of policy and/or institutions across national or sub-national governments. While the study of policy-innovation diffusion among U.S. States has deep roots and much contemporary interest, with sustained attention between,[2] similar policy-learning mechanisms underlie some comparative studies of policy diffusion (e.g., Schneider and Ingram 1988, Rose 1993, Meseguer 2004, 2005, Gilardi 2005). In

---

[1] In the conceptualizations and terminologies of the literature, *diffusion* and *interdependence* seem sometimes subtly distinct and sometimes perfectly synonymous. They are highly overlapping sets in any event, and care in making and maintaining possible distinctions is not always evident. For us here, *interdependence* will refer to contexts in which the outcome(s) of interest (i.e., dependent variable(s)) in some units of analysis (e.g., countries) directly affect the outcome(s) of interest in others. Such interdependence may arise by any of numerous different mechanisms (see section I.B. below), some of which must involve direct contact between the units involved (or of elements therein such as citizens of countries) but others do not necessarily involve such direct contact (e.g., in contexts of economic competition, direct exchange between units is not necessary for interdependence between them to arise). *Diffusion* here will refer to the patterns and paths across space (and time, usually) by which some outcome(s) spread across units of analysis. It, too, may occur by any number of mechanisms, but, additionally, patterns of spatial diffusion may appear without any direct effect from the outcomes in some units to the outcomes in others, perhaps via spatial correlation in domestic or external conditions to which the units are exposed. (Paradoxically, though, direct contact between units is necessary to some conceptions of diffusion.) Distinguishing these possibilities is the essence of *Galton's Problem* (see Section I.C.), an extremely challenging problem indeed empirically.

[2] See, e.g., Crain (1966), Walker (1969, 1973), Gray (1973), Knoke (1982), Caldiera (1985), Lutz (1987), Berry and Berry (1990), Case et al. (1993), Berry (1994), Rogers (1995), Mintrom (1997ab), Mintrom and Vergari (1998), Mossberger (1999), Berry and Berry (1999), Godwin and Schroedel (2000), Balla (2001), and Mooney (2001), Bailey and Rom (2004), Boehmke and Witmer (2004), Daley and Garand (2004), Grossback et al. (2004), Shipan and Volden (2006), Volden (2006).

comparative politics, perhaps the closer historical parallel is interest in institutional or regime diffusion, which is likewise long-standing and recently much reinvigorated. Dahl's (1971) classic *Polyarchy*, for instance, (implicitly) references international interdependence among the eight causes of democracy he lists; Starr's (1991) "Democratic Dominoes" and Huntington's (1991) *Third Wave* accord it a central role; Beissinger (2007) and Bunce and Wolchik (2006, 2007), *inter alia*, emphasize it in the context of post-communist democratic transitions in Eastern Europe, and Hagopian, Mainwaring, and colleagues (Hagopian and Mainwaring 2005) among others in the Latin American context; finally, O'Loughlin et al. (1998), Brinks and Coppedge (2006), and Gleditsch and Ward (2006, 2007) estimated empirically the extent, paths, and/or patterns of international diffusion of democracy.

The substantive range of important spatial-interdependence effects extends well beyond these more-obvious contexts of intergovernmental diffusion, however, spanning the subfields and substance of political science. Inside democratic legislatures, e.g., representatives' votes certainly depend on others' votes or expected votes; in electoral studies, election outcomes or candidate qualities or strategies in some contests surely depend on those in others. Outside legislative and electoral arenas, the probabilities and outcomes of coups (Li and Thompson 1975), riots (Govea and West 1981), and/or revolutions (Brinks and Coppedge 2006) in one unit depend in substantively crucial ways on those in others. In micro-behavioral work, too, some of the recently surging interest in *contextual effects* surrounds the effects on each respondent's behaviors or opinions of aggregates of others' behaviors and opinions—e.g., those of the respondent's region, community, or social network. Within the mammoth literature on contextual effects in political behavior (Huckfeldt and Sprague (1993) review), recent contributions that stress interdependence include Braybeck and Huckfeldt (2002ab), Cho (2003), Huckfeldt et al. (2005), Cho and Gimpel (2007), Cho and Rudolph (2007), Lin et al (2006). In international relations, meanwhile, the interdependence of states' actions essentially defines the subject. States' entry decisions in wars, alliances, or international organizations, e.g., heavily depend on how many and who (are expected to) enter.[3] Empirical attention to the inherent spatial interdependence of international relations has been greatest in the work of Ward, Gleditsch, and colleagues (e.g., Shin and Ward 1999, Gleditsch and Ward 2000, Gleditsch 2002, Ward and Gleditsch 2002, Hoff and Ward 2004, Gartzke and Gleditsch 2006, Salehyan and Gleditsch 2006, Gleditsch 2007). In comparative and international political economy also, interdependence is perhaps especially frequently substantively large and central. Simmons and Elkins (2004), Elkins et al. (2006), and Simmons et al. (2006) stress cross-national diffusion as the main force behind recent economic liberalizations, for examples, as do Eising (2002), Brune et al. (2004), Brooks

[3] Signorino (1999,2002) and Signorino & Tarar (2006) also stress the strategic interdependence of international relations in specifying empirical models, as do Signorino and Yilmaz (2003) and Signorino (2003) regarding strategic choice broadly.

(2005), and many others. In fact, globalization and international economic integration, arguably today's most-notable (and indisputably most-noted) political-economic phenomena, imply strategic and/or non-strategic interdependence of domestic politics, policymakers, and policies. Empirical work emphasizing such globalization-induced interdependencies includes Genschel (2002), Basinger and Hallerberg (2004), Knill (2005), Jahn (2006), Swank (2006), Franzese and Hays (2006b, 2007abc), and Kayser (2007).

2. Mechanisms: Coercion; Competition, Cooperation, and Externalities; Learning and Emulation; Migration

3. => Tobler's First Law of Geography (*all things related, near things more so*), with Beck, Gleditsch, & Beardsley's addendum (*space is more than geography*).

Spatial interdependence is, in summary, ubiquitous and often quite central throughout the substance of comparative politics. *Tobler's Law* (geographer Waldo Tobler: 1930-) aptly captures this ubiquity: "Everything is related to everything else, but near things are more related than distant things." What's more, as Beck et al.'s (2006) pithy title stresses: "Space is More than Geography." I.e., the substantive content of the *proximity* in Tobler's Law, and so the pathways along which interdependence between units may operate, extend well beyond simple physical distance and bordering (as several examples above illustrate). Elkins and Simmons (2005) and Simmons et al. (2006), e.g., define and discuss four mechanisms by which international interdependence may arise: coercion, competition, learning, and emulation. *Coercion*, which may be direct or indirect and hard (force) or soft (suasion), encompasses a generally "vertical" pathway by which the powerful induce actions among the weaker. *Competition* refers to interdependence stemming from economic pressures that the actions of each unit place upon others in competition with it or as substitutes for or complements to it. We might usefully broaden this category to include interdependencies induced by externalities and spillovers, more generally, and so broaden also its label for our purposes to *competition, cooperation, and externalities*. This mechanism would encompass also complement and substitute relations in economics (e.g., countries being complementary or substitute destinations for foreign investors). Next, *learning* entails situations where actors learn, in rational-Bayesian or some other fashion, from others' actions something regarding the attractiveness of their own alternative actions.[4] *Emulation*, finally, is ritualistic (i.e., neither coerced nor responsive to competition or to learning) following or doing oppositely of others (e.g., leaders, co-ethnics, co-

[4] What is learned affects actors' choices, but may be objective or subjective, true/correct or false/incorrect, and may regard the politics, economics, sociology, or any other aspect of those choices.

partisans). These last two are perhaps similar enough in their operation—emulation being distinct from learning only in degree of its rationality and enlightenment aspects— to be merged for our use as *learning and emulation*. Although Simmons et al. (2006) enumerated these specifically for the context of international diffusion, these four categories nicely span many of the possible channels of spatial interdependence across its broader substantive domain in the social sciences. We might only add a fifth channel, *migration*, wherein some components of some units move directly into other units, the most obvious example being human im-/emigration, which will tend to generate a direct, mechanical interdependence in addition to strategic or idea-dissemination, only some of which effects *competition, cooperation, and externalities* or *learning and emulation* could cover in our modified version of Simmons et al. (2006) schema.

B. Theoretical Models Strategic Interdependence:

1. A General Theoretical Model of Interdependence

a) Basic Principle: if marginal utility of *i*'s actions depend on *j*'s actions, then interdependence

b) General Properties:

(1) Positive Externalities => Strategic Substitutes wherein policies/actions move in opposite directions, i.e., negative interdependence, in which *free-rider* problems & late-mover advantages, & so strategic delay/wars of attrition dynamics, arise.

(2) Negative Externalities => Strategic Complements wherein policies/actions move in same directions, i.e., positive interdependence, in which *beggar-thy-neighbor* problems & early-mover advantages, & so *races-to-the-bottom* (or *top*, or, generally, competitive races), arise.

c) Examples & Discussion

More generally, following Brueckner (2003), one can show that strategic interdependence arises whenever the actions of some unit(s) affect the marginal utility of alternative actions for some other unit(s). Consider two units (i,j) with utilities, (W^i,W^j), from alternative actions or policies, (p_i,p_j), that they may choose. Due to externalities, *i*'s utility depends on its policy and that of *j*. E.g., imagine two countries with homogenous populations regarding, say, their economic and environmental preferences. Due to environmental externalities (e.g., those surrounding pollution) and economic ones (e.g., those surrounding the costs of environmental regulations), domestic welfare (i.e., net political-economic benefits/utilities to policymakers) in each country will depend on both countries' actions:

$$W^i \equiv W^i(p_i, p_j) \ \ ; \ \ W^j \equiv W^j(p_j, p_i) \tag{1}$$

When the government in country *i* chooses its policy, p_i, to maximize its own social welfare, this alters the optimal policy in country *j*, and *vice versa*. For example, *i* implementing more (less) effective anti-pollution policy reduces (increases) the need for effective anti-pollution policy in *j* due to environmental spillovers. We can express such strategic interdependence between countries *i* and *j* with a pair of best-response functions, giving *i*'s optimal policies, p_i^*, as a function of *j*'s chosen policies, and *vice versa*:[5]

$$p_i^* \equiv \text{Argmax}_{\text{p}_\text{i}} W^i(p_i, p_j) \equiv R^i(p_j) \ \ ; \ \ p_j^* \equiv \text{Argmax}_{\text{p}_\text{j}} W^j(p_j, p_i) \equiv R^j(p_i) \tag{2}$$

The slopes of these best-response functions indicate whether actions by *i* induce *j* to move in the same direction, in which case we call the actions of *i* and *j* *strategic complements*, or in the opposite direction, in which case they are *strategic substitutes*. For example, anti-pollution policies are strategic substitutes in terms of their environmental effects as described above. The slopes of these best-response functions depend on the following ratios of second cross-partial derivatives:

$$\frac{\partial p_i^*}{\partial p_j} = -W^i{}_{p_i p_j} / W^i{}_{p_i p_i} \ \ ; \ \ \frac{\partial p_j^*}{\partial p_i} = -W^j{}_{p_j p_i} / W^j{}_{p_j p_j} \tag{3}$$

If the government maximizes its utility, the second-order condition implies negative denominators in (3). Therefore, the slopes depend directly on the signs of the second cross-partial derivatives (i.e., the numerators). If $W^{i,j}_{p_i p_j} > 0$, i.e., if policies are strategic complements, reaction functions slope upward. Regarding the economic costs of anti-pollution regulation, e.g., increased (reduced) regulation in *i* lowers (raises) the costs of regulation in competitors *j*, and so spurs *j* to increase (reduce) regulation too. If $W^{i,j}_{p_i p_j} < 0$, policies are strategic substitutes, so reaction functions slope downward, as noted regarding in the environmental benefits of anti-pollution regulation. If the second cross-partial derivative is zero, strategic interdependence does not materialize and best-response functions are flat (Brueckner 2003).

Generally speaking, then, positive externalities induce strategic-substitute relations, and policies will move in opposite directions as *free-rider* dynamics obtain. Franzese and Hays (2006) argue and find such

[5] Explicitly, country *i*'s optimum policy is obtained by maximizing with respect to p_i, taking p_j as given (fixed); i.e., setting the first derivative of the welfare function with respect to p_i equal to zero and solving for the resulting p_i^* as a function of p_j (and then verifying that the second derivative is negative).

free-riding dynamics in EU active-labor-market policies, for example. Notice, furthermore, that free-rider advantages also confer late-mover advantages and so war-of-attrition (strategic delay and inaction) dynamics are likely. Conversely, negative externalities induce strategic complementarity, with policies moving in the same direction. The common example of tax-competition has these features. Tax cuts in one unit have negative externalities for competitors, who are thereby spurred to cut taxes as well. These situations advantage early movers, so *competitive races* can unfold.[6] Other good examples here are competitive currency-devaluations or trade-barriers. Economically, earlier movers in these contexts reap disproportionate benefits, so races to be first are likely. Thus, positive and negative externalities induce strategic-complement and -substitute relations, respectively, which spur competitive-races and free-riding, respectively, with their corresponding early- and late-mover advantages, and so strategic rush to go first on the one hand and delays and inaction on the other.

2. A Specific Theoretical Model of Interdependence in One Example Context: The Political Economy of Globalization, Capital-Tax Competition, and Domestic Fiscal-Policy Autonomy

a) Review of theories in previous literature; all imply interdependence, but most previous empirical specifications fail to incorporate that implication.

b) A citizen-candidate model of capital-tax competition and foreign investment (from Persson & Tabellini 2000).

(1) Also implies interdependence & renders it formally explicit.

(2) Discuss model & some examples of derived comparative statics.

Globalization and, specifically, capital-tax competition, which example we elaborate next and (re-)analyze empirically below, creates strategic policy-interdependence by increasing cross-border negative externalities (i.e., negative cross-partial derivatives), which steepens the positively sloped reaction functions, possibly thereby spawning competitive tax-policy races.

In theory, strong inter-jurisdictional competition for capital undermines the tax-policy autonomy of

[6] We intentionally eschew the labels *races to the bottom* (or *top*) and *convergence* because these competitive races need not foster convergence on any top, bottom, or mean, and could further divergence (see below and Plümper and Schneider 2006).

individual states, inducing tax rates to converge, especially those levied upon more-mobile assets. Such inter-jurisdictional competition intensifies as capital becomes more liquid and mobile across borders. Indeed, many scholars of domestic or international fiscal-competition (*e.g.*, Zodrow and Mieszkowski 1986; Wilson 1986, 1999; Wildasin 1989; Oates 2001) expect intense inter-jurisdiction competition to engender a virtually unmitigated race to some (*ill-defined*: see note 6 and below) bottom. As a central exemplar, most scholarly and casual observers see the striking post-1970s rise in international capital-mobility and steady postwar increase in trade integration as forcing welfare- and tax-state retrenchment and shifting tax-burden incidence from relatively mobile bases (e.g., capital, especially finance) toward more immobile (e.g., labor, especially the less-flexibly-specialized).[7] Rising capital-market integration and asset mobility across jurisdictions supposedly enhances these pressures by sharpening capital's threat against domestic governments to flee perceived excessive and inefficient welfare and tax systems.

Several notable recent studies of the comparative and international political economy of tax policy over this period challenge these claims. Quinn (1997), Swank (1998, 2002), Swank and Steinmo (2002), Garrett and Mitchell (2001), and others do not find rising international economic integration to have constrained governments' fiscal policies much or at all. The theoretical explanation for such results, occasionally implicit, is that other cross-national differences also importantly affect investment-location decisions, affording governments maneuvering room. Hines (1999), e.g., found commercial, regulatory, and other policies, labor-market institutions, intermediate-supply availability, and proximity to final markets, among other factors, to be critical in corporate investment-location decisions. Moreover, other factors than capital mobility affect governments' tax policies. For example, Swank (2002) argues that corporate and capital tax-rates depend on macroeconomic factors like inflation and economic growth, partisan politics, and funding requirements of programmatic outlays. Controlling for these domestic factors, he finds little relation of taxation to capital mobility, although including spatial lags to reflect tax-policy interdependence as in our re-analyses below and Swank's subsequent work (e.g., 2006) significantly modifies that conclusion. Still, a common finding in this area is that internal political-economic determinants of fiscal policy can be at least as important as external, globalization-related factors. As just presaged, our empirical reanalysis returns to this controversy.

On closer analysis, these recent challenges to simplistic *globalization-induces-welfare/tax-state-retrenchment* views have at least four distinct bases. Garrett (1998) argues that certain combinations of left government and social-welfare, active-labor-market, and related policies, coupled with coordinated-

[7] Unskilled labor is usually relatively mobile within (national) jurisdictions but highly immobile across jurisdictions, especially those borders delineating strongly differentiated ethnic, linguistic, religious, and cultural societies. Some types of skilled labor are highly specialized into specific productive activities, which may limit intra- and inter-jurisdictional mobility; other types, some *human capitalists* for example, may be relatively mobile across jurisdictions.

bargaining, can be as or more efficient than neoliberal state-minimalism and conservative government. Therefore, he argues, capital will not flee such efficient combinations. Boix (1998) argues that public human- and physical-capital investment-strategies comprise an alternative to neoliberal minimalism that is sufficiently efficient economically to retain capital, and perhaps even attract it, *and* politically effective enough to sustain electoral competitiveness for the left. Hall, Soskice, and colleagues (2001) argue that complex national networks of political-economic institutions confer *comparative* advantages in differing productive activities, which, as Mosher and Franzese (2002) stress, implies capital mobility and trade integration could (*if international tax-competition remains sufficiently muted*: see below) spur institutional and policy specialization, which would imply persistent welfare/tax-system variation or even divergence rather than convergence or global retrenchment. These three views—and we should add a fourth which emphasizes the possibility that globalization may increase populations economic insecurity and so foster greater welfare-state demand (Rodrik, Scheve, Iversen)—fundamentally question whether globalization actually creates net economic pressures to retreat from welfare/tax-state commitments (or at least whether all aspects of globalization do so, so strongly: see below).

Swank's (2002) argument that the institutional structures of the polity and the welfare system itself shape the domestic policy-response to integration represents a fifth basis (or augments the fourth) for challenge. He does not fundamentally challenge claims of the exclusively superior macroeconomic efficiency of neoliberal minimalism but rather stresses the primacy of domestic political conditions—the policymaking access, cohesion and organization, and relative power of contending pro- and anti-welfare/tax interests—in determining the direction and magnitude of welfare/tax-policy reactions to economic integration. Specifically, he argues and finds (i) inclusive electoral institutions, (ii) social-corporatist interest-representation and policymaking, (iii) centralized political authority, and (iv) universal welfare systems relatively favor the political access and potency of pro-welfare/public-policy interests and bolster supportive social norms in the domestic political struggle over the policy response to integration. The opposite conditions favor anti-tax/welfare interests and norms in this struggle. Capital mobility and globalization therefore induce increased welfare/tax-state largesse in previously generous states and retrenchment in tight ones: i.e., divergence for domestic reasons not convergence for interdependent ones. Swank's is, thus, the most directly and thoroughly political of these critiques. It may also be the most thoroughly explored empirically, offering comparative-historical statistical and qualitative analyses of six alternative versions of a *globalization-induces-retrenchment* thesis: a simple version (a regression including one of five capital-openness measures), and five other versions that he labels the *run-to-the-bottom* (capital openness times lagged welfare-policy), *convergence* (capital openness times the gap from own to cross-country mean welfare-policy), *nonlinear* (capital openness and its square), *trade-and-*

capital-openness (their product), *capital-openness-times-fiscal-stress* (deficits times capital openness), and *capital-flight* (net foreign direct investment). He finds little support for any globalization-induces-retrenchment argument, and, indeed, some indications that capital mobility tends on average to enhance welfare effort (perhaps supporting those stressing its effect in increasing popular demand for social insurance against global risks).[8]

Basinger and Hallerberg (2004), in a sense, take the implied next step of Swank's central point. Swank stresses the domestic political and political-economic institutions and structures of interest that shape governments' policy responses to economic integration. It then follows, however, as Basinger and Hallerberg (2004:261) stress, that "[if] countries with higher political costs are less likely themselves to enact reforms, [then this] also reduces competing countries' incentives to reform regardless of their own political costs." That is, the magnitude of the tax-competition pressures that economic integration places upon one government's fiscal policies depends upon the policy choices of its competitors, which is precisely the strategic interdependence that we emphasize here as well.

Such critiques underscore that the *bottom* toward which globalization and capital mobility may push tax-competing states may not be neoliberal minimalism. Insofar as alternative economic advantages allow some states to retain higher tax rates, or insofar as restraining political conditions prevent some from reaching neoliberal minimum, the competitive pressures on all states diminish, more so, of course, the more economically integrated and important are those states whose domestic political-economic conditions allow such maneuvering room or raise such constraints. Furthermore, if, as Mosher and Franzese (2002) suggest, national economic-policy differences contribute to *comparative* advantages—which, if they do, they do regardless of their *absolute* efficiency—then both trade and global *fixed-capital* integration would actually enhance economic pressures toward specialization, i.e., divergence, and not convergence. From this view, international *liquid-capital* mobility alone, through the tax-competition it engenders, produces whatever competitive races may occur. In this case, interestingly, such competitive races would occur regardless and independent of the efficiency of the tax systems in question or of the public policies they support. Furthermore, as both Hays (2003) and Basinger and Hallerberg (2004) stress, such races need not be to some *bottom*; rather, the competitiveness and the destination(s) of the race depend on the constellation of domestic political-economic conditions present in, and the economic integration of, the international system. Conversely, as Mosher and Franzese (2002) stress, zero offers no inherent *bottom* to such tax-cut races as may occur. In the competition for liquid portfolio-

[8] See Franzese (2003) for a more-complete review of Swank (2002). Again, however, his later work (2006) and our own (2004, 2006ab, 2007, and below) specifies the impact of globalization as properly based in interdependence and modifies these conclusions considerably, finding much stronger evidence for constraint on domestic autonomy from capital mobility.

capital specifically, governments always have incentives to cut taxes further, perhaps deep into subsidy; only their abilities to tax other less liquid and/or mobile assets and to borrow limit (in internationally interdependent manner, as just noted) those races.

Thus, international tax and fiscal-policy competition arguments, in any of their conventional forms and throughout each of these critiques, imply cross-national (i.e., spatial) interdependence in fiscal policymaking. Whatever pressures on domestic policymaking may derive from rising capital mobility, their nature and magnitude will depend on the constellation of tax (and broader economic) systems with which the domestic economy competes. As we have shown, race-to-the-bottom (RTB) dynamics may occur when policies are strategic complements across jurisdictions—i.e., when policy changes in one jurisdiction create incentives for other jurisdictions to adopt similar changes. Such arguments have been applied *inter alia* to capital taxation or costly environmental or labor regulatory standards. Cuts in taxes or regulatory standards in one jurisdiction increase the costs to others of maintaining high taxes and regulatory standards, inducing the affected jurisdictions to follow suit in their own policies. By contrast, free-riding occurs when policies are strategic substitutes—i.e., when policy changes in one jurisdiction create incentives for governments in others to adopt change in the opposite direction. For example, an increase in defense expenditures in one country might lower the marginal security benefit from defense spending in its military allies, creating an incentive for them to free ride (see, e.g., Redoano 2003).

We gave a general theoretical formulation of interdependence above, following Brueckner (2003); here we follow Persson and Tabellini (2000:ch. 12) to show formally how tax competition specifically implies spatial interdependence. The model's essentials are these. In two jurisdictions (e.g., countries), denote the domestic and foreign capital-tax rates τ_k and τ_k^*. Individuals can invest in either country, but foreign investment incurs *mobility costs*. Governments use revenues from taxes levied (by the source, not the residence, principle) on capital and on labor to fund a fixed spending-level.[9] Individuals differ in their relative labor-to-capital endowment, e^i, and make labor-leisure, l and x, and savings-investment, $s=k+f$ (k=domestic; f=foreign), decisions to maximize quasi-linear utility, $\omega=U(c_1)+c_2+V(x)$, over leisure and consumption in the model's two periods, c_1 and c_2, subject to a time constraint, $1+e^i=l+x$, and to period-1 and -2 budget constraints, $1\text{-}e^i=c_1+k+f+\equiv c_1+s$ and $c_2=(1\text{-}\tau_k)k+(1\text{-}\tau^*_k)f\text{-}M(f)+(1\text{-}\tau_l)l$.

The equilibrium economic choices of citizens i in this model are as follows:

$$s=S(\tau_k)=1-U_c^{-1}(1-\tau_k) \tag{4}$$

$$f=F(\tau_k,\tau_k^*)=M_f^{-1}(\tau_k-\tau_k^*) \tag{5}$$

[9] Government consumption is not only fixed but also entirely wasted; i.e., it enters no one's utility function.

$$k = K(\tau_k, \tau_k^*) = S(\tau_k) - F(\tau_k, \tau_k^*) \tag{6}$$

With labor, $L(\tau_l)$, leisure, x, and consumption, c_1, c_2, implicitly given by these conditions, this leaves individuals with indirect utility, W, defined over the policy variables, capital and labor tax-rates, of:

$$W(\tau_l, \tau_k) = U\left(1 - S(\tau_k)\right) + \left(1 - \tau_k\right) S(\tau_k) + \left(\tau_k - \tau_k^*\right) F(\tau_k, \tau_k^*)\ M\left(F(\tau_k, \tau_k^*)\right) + \left(1 - \tau_l\right) L(\tau_l) + V\left(1 - L(\tau_l)\right) \tag{7}$$

Facing an electorate with these preferences, using a Besley-Coate (1997) citizen-candidate model wherein running for office is costly and citizens choose whether to enter the race by an expected-utility calculation, some citizen candidate will win and set tax rates to maximize its own welfare. The model's stages are: 1) both countries hold elections, 2) elected citizen-candidates set their countries' tax rates, and 3) all private economic decisions are made. In this case, the candidate who enters and wins will be the one with endowment e^P such that s/he desires to implement the following *Modified Ramsey Rule*:

$$\frac{S(\tau_k^p) - e^p}{S(\tau_k^p)}\left[1 + \varepsilon_l(\tau_k^p)\right] = \frac{L(\tau_l^p) + e^p}{L(\tau_l^p)}\left[1 + \frac{S_\tau(\tau_k^p) + 2F_\tau^*(\tau_k^{p*}, \tau_k^p)\tau_k}{S_\tau(\tau_k^p)}\right] \tag{8}$$

This modified Ramsey rule gives the optimal capital-tax-rate for the domestic policymaker, which is a function of the capital tax-rate abroad, τ^*_k. The game is symmetric, so the optimal capital tax-rate for the foreign policymaker is analogous, including in its dependence on the domestic capital-tax-rate, τ^*. That is, this equation gives best-response functions of the sort we specified generically above, and specifically in this context of $\tau_k = T(e^P, \tau_k^*)$ and $\tau_k^* = T^*(e^{P*}, \tau_k)$ for the foreign and domestic policymaker, respectively. In words, the domestic (foreign) capital-tax rate depends on the domestic (foreign) policymaker's labor-capital endowment and the foreign (domestic) capital tax rate—i.e., capital taxes are strategically interdependent. Substantively, the slope of these functions, $\partial T/\partial \tau_k^*$ and $\partial T^*/\partial \tau_k$, can be either positive or negative in this case because an increase in foreign tax-rates induces capital flow into the domestic economy, but the domestic policymaker may use the increased tax-base to lower tax-rates or to raise them (the latter to seize the greater revenue opportunities created by the decreased elasticity of this base). Figure 1 (borrowed with permission) graphs the reaction functions assuming that both slope positively. The illustrated comparative static shows an increase in the domestic policymaker's labor-capital endowment (intended to reflect a leftward shift in government). This change shifts the function T outward, raising the equilibrium capital-tax rate *in both countries*.

Figure 1. Best Response Functions (Persson and Tabellini 2000, 334)

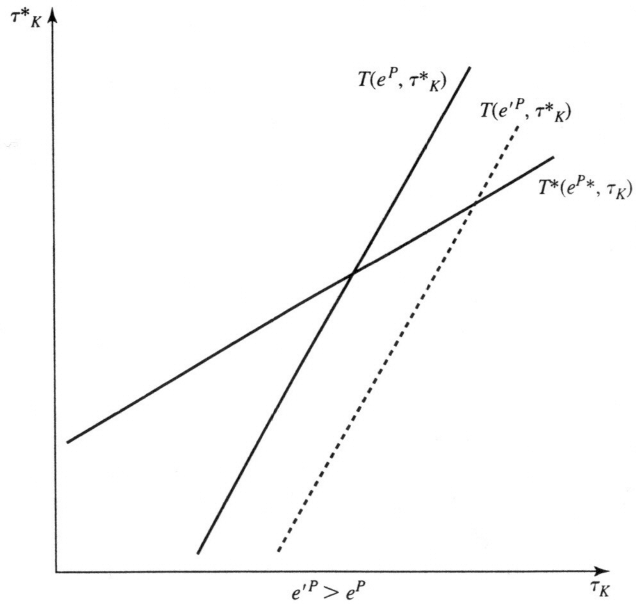


Figure 12.4

Although formal tax-competition models like this one or Hays (2003) or Basinger and Hallerberg (2004) clearly imply the strategic spatial-interdependence of capital taxes, as do any of the alternative arguments reviewed above, very few scholars have empirically modeled that interdependence directly. Not all tax/welfare-state retrenchment arguments, however, necessarily involve tax-competition. Iversen and Cusack (2000), e.g., argue that structural change in the labor force, specifically deindustrialization, is the primary force pushing welfare/tax-state retrenchment. Pierson (2001) concurs in part but also emphasizes *path dependence* (technically: state dependence, see Page 2006), namely the accumulation and entrenchment of interests (or their absence) behind welfare/tax-state policies and institutions. Rodrik (1998), and Cameron (1978) before him, stressed instead the added demand from some domestic interests for certain social policies that increased economic exposure would engender.

Such forces—labor-force structural-change, domestic-interest entrenchment and/or change—may be related to, or even partly caused by, aspects of globalization, but, ultimately, these are domestic-factor explanations, or arguments about domestic-factors modifying responses to *exogenous* external trends, and therefore do not by themselves imply a strategic interdependence among policy choices. Below, we term these sorts of *domestic-factors*, or *exogenous-external-factors*, or *domestic-factors-conditioned-responses-to-exogenous-external-factors* approaches: *Open-Economy Comparative-Political-Economy (OE-CPE).* Pure tax-competition arguments exemplify the (internationally) strategic-interdependence approaches that we term *International-Interdependence Political-Economy* (*II-PE*). Of course, the two are easily combined in *(Open-*

Economy) Comparative and International Political Economy (C&IPE) models that reflect domestic factors, exogenous-external factors, and domestically modified responses to exogenous-external conditions (*context-conditional*, for short) on the one hand and international interdependence on the other. We reanalyze such *C&IPE* empirical models of globalization and capital-tax competition below, but, first, we explain the severe empirical challenges to estimating and distinguishing these *OE-CPE* and *II-PE* alternatives or components of *C&IPE* models and some methodological approaches that may surmount these challenges.

3. **Briefer overview of several other specific theoretical models, in other contexts within and beyond comparative/international political economy.**

II. (CHAPTER 2) The Empirical Challenge of Spatial Interdependence

A. Galton's Problem: 3 Sources of Spatial Correlation, and a modern complication

1. **correlated unit-specific conditions/explanators (examples)**
2. **correlated exogenous-external conditions/explanators (examples)**
3. **interdependence (examples)**
4. **complication: distinguish error/stochastic-term correlation or interdependence from systematic-component/outcome interdependence: implied spatial-dynamics differ substantively importantly**
5. **Galton's Problem is that these things look very much alike empirically, and so the fact that any, all, or any combination may operate in any given context presents an extreme empirical challenge to distinguish them (by any estimation strategy or, for that matter, by any empirical methodology, quantitative or qualitative).**

Galton originally raised the issue in these terms: "[F]ull information should be given as to the degree in which the customs of the tribes and races which are compared together are independent. It might be that some of the tribes had derived them from a common source, so that they were duplicate copies of the same original. ...It would give a useful idea of the distribution of the several customs and of their relative prevalence in the world, if a map were so marked by shadings and colour as to present a picture of their geographical ranges" (Sir Francis Galton, *The Journal of the Anthropological Institute of Great Britain* *and Ireland* 18:270, as quoted in Darmofal (2007).) In http://en.wikipedia.org/wiki/Galton's_problem, we find further historical context: "In [1888], Galton was present when Sir Edward Tylor presented a paper at the Royal Anthropological Institute. Tylor had compiled information on institutions of marriage and descent for 350 cultures and examined the correlations between these institutions and measures of societal complexity. Tylor interpreted his results as indications of a general evolutionary sequence, in which institutions change focus from the maternal line to the paternal line as societies become increasingly complex. Galton disagreed, pointing out that similarity between cultures could be due to borrowing, could be due to common descent, or could be due to evolutionary development; he maintained that without controlling for borrowing and common descent one cannot make valid inferences regarding evolutionary development. Galton's critique has become the eponymous *Galton's Problem* (Stocking 1968: 175), as named by Raoul Naroll (1961, 1965), who proposed [some of] the first statistical solutions." That is, Galton was noting that some common behavior pattern across societies—more generally, of some correlation in outcomes across some set(s) of contexts—may arise because societies borrow from each other, which would be interdependence by the emulation, learning, or migration mechanism, depending on how one understands *borrowing*, or by common descent, which in our terms would be understood to mean because society-specific characteristics happened correlate (due to common origin in this case), or because of (common) evolutionary development, which more generally would mean by virtue of the unit responses to common or correlated exogenous-external conditions naturally being correlated as well. *Galton's Problem*, and it remains equally true, relevant, and problematic today and across the nearly universal substantive domain to which *Tobler's Law* may apply, is that these things look very much alike empirically, and so the fact that any, all, or any combination may operate in any given context presents an extreme empirical challenge to any attempt to distinguish them (by any estimation strategy or, for that matter, by any empirical methodology, quantitative or qualitative).

To this early recognition of the empirical challenge of spatial correlation and interdependence, we must add a further complication. Modern researchers will need for most purposes further to distinguish correlation or interdependence in the stochastic component (i.e., error term) of the outcomes of their study from interdependence in the systematic-component or of the entire outcome because the spatial-dynamics implied by these alternative possibilities differ substantively critically (as elaborated below).

B. Empirical observations often taken as evidence of interdependence that not necessarily

1. Dummies for Groups/Regions

2. S-Shaped Adoption Curves

3. Measures & Tests of Spatial Correlation/Clustering (briefly introduced)

4. Parks FGLS, PCSE, & other estimation strategies based on error-covariance matrices with spatial patterns

To clarify the importance of modeling interdependence, consider three sorts of empirical observations that researchers sometimes conclude imply diffusion: significant regional or other group dummies, spatially correlated observations or a spatial or clustering pattern in the variance-covariance of residuals, or an S-shaped accumulation of the share of a sample or population that adopts some action.

Convenient expedients like regional, cultural, organizational, linguistic group-membership dummies illustrate well the difficulties of distinguishing these alternatives. If EU membership, e.g., correlates, controlling for the model's other cofactors, with, say, foreign-aid, is this because values-based and/or utilitarian arguments for aid circulate among EU-member elites to diffuse among its members, or because EU institutions provide forums for policymaking elites in member countries to learn from each other the ethical or practical benefits of aid, or because EU membership lowers the cost of responding to some exogenous external shock like a famine, or because some domestic conditions that favor joining the EU also favor foreign aid? These are two different diffusion arguments, one common-shock-varying-response argument, and one domestic-factors argument, and a dummy for EU membership is incapable of distinguishing between them on its own. However, insofar as the researcher can model these alternative mechanisms more precisely than by the membership indicator, she should do so (of course). Also or alternatively, insofar as she can connect the time-invariant conditional-mean-shifting nature of the dummy's effect to one mechanism, substantively, theoretically, and empirically distinctly from specifications of alternative effect types, then the reader may attribute the coefficient on the dummy to this linked mechanism.

Likewise, approaches to modeling spatial interdependence that emphasize estimates of the variance-covariance structure of residuals are inherently limited and potentially misleading as means of distinguishing diffusion from alternative explanations for such correlation. In the first place, a finding that observations or residuals in some set of countries correlate or cluster in some manner can as easily support a common-exposure as a diffusion argument for the same reasons a dummy alone cannot distinguish these alternatives. In the second place, residual variance-covariance approaches tend to require strong multivariate normality and separability assumptions for tractability and, relegating all spatial correlation to a variance-covariance matrix, logically cannot reflect any asymmetry in the pattern of diffusion that may exist. If we estimate the variance-covariance or clustering of residuals and interpret that as diffusion, then we must conclude *a priori* that, in a model of tax competition estimated in a sample of OECD countries for example, taxes in Luxembourg affects the US as much as the US affects Luxembourg. In the third place, finally, insofar as the diffusion *cum* spatial interdependence lies at the center of the theoretical or substantive issues under exploration, their relegation to the stochastic properties of residuals rather obfuscates the point.

Finally, some refer to an S-shaped accumulation in a plot of the share of a population that has adopted some particular action over time as the characteristic pattern of diffusion: first one unit adopts some practice, then the idea or knowledge of its success diffuses to a few others, from which it diffuses more densely to many others, until enough of the population has adopted the course of action that most of the diffusion paths lead to units that have already adopted and the pace of adoption tapers toward everyone having adopted. That's certainly one story consistent with a sigmoidal accumulation of adopters, but many, many cumulative functions are sigmoidal or roughly sigmoidal. One very important one is the cumulative normal, for example. Suppose, then, that each unit would have to pay some cost (say, to overcome some domestic political opposition) to adopt some course of action. The central limit theorem gives every reason, absent some other theory, to suppose these costs are normally distributed across units. Now suppose globalization or some other exogenous trend that dampens the opposition to or enhances the benefit from adoption of this course of action is commonly experienced by the sample units. Then, cumulative share having adopted would follow the cumulative normal, which is S-shaped. Again, the ability to distinguish these alternative explanations for the adoption path relies in the first order on modeling both accurately, and observational or relatively non-parametric expedients like this one or the preceding two are wholly inadequate to that task.

C. Measures and tests of spatial correlation and interdependence:

1. **Discusses how each of the above by itself and, for that matter, a spatial-lag model by itself fail to distinguish among the multiple possible sources of spatial correlation, including interdependence, and so are not necessarily evidence of the latter.**

2. **Elaborates some appropriate nested & non-nested test-strategies of fuller models**

Regarding tests for and statistical measures of spatial correlation, so far we have assumed that analysts working with TSCS data know they correlate spatially. Such data certainly will typically exhibit spatial correlation, but analysts will still want to gauge the magnitude of this substantively interesting phenomenon (or diagnose this problem before treating it, if thought pure nuisance) especially if the bias-efficiency-complexity tradeoffs we discuss in later chapters manifest as frequently and prominently as we expect. For example, parallel to how including temporally lagged dependent-variables as regressors induces bias if temporal correlation is inadequately modeled (*Achen's Problem*), including spatially lagged dependent-variables as regressors will induce bias if spatial correlation is inadequately modeled. Accordingly, this section discusses several parametric, semi-parametric, and non-parametric tests for (gauges of) spatial correlation and, where necessary or pedagogically helpful, provides analytic and/or experimental results documenting their large- and small-sample properties.

[…Moran's *I*, Anselin's *LISA*, and related spatial-correlation and clustering statistics…]

Certain measures or tests derived by analogy to those from time-serial and cross-sectional regression literature may also prove useful. For example, Lagrange multiplier tests derived by analogy to White's heteroscedasticity (i.e., non-constant variance) test, which regress $e_{i,t} \times e_{j,t}$ (e estimated error) rather than White's $e_{i,t}^2$ on the corresponding products and, as feasible, cross products of the elements of $\mathbf{x}_{i,t}$ and $\mathbf{x}_{j,t}$ rather than White's products and, as feasible, cross products of the elements of $\mathbf{x}_{i,t}$, may prove useful. As with heteroskedasticity, the sample pattern of which $e_{i,t}^2$ estimates, the spatial correlation, the sample pattern of which $e_{i,t} \times e_{j,t}$ estimates, that most imperils standard regression statistics will have a pattern that relates to the corresponding products and cross-products (i.e., moments) of the independent-variable matrix. Moreover, we expect this test to retain validity with or without spatial lags in the model as does White's. However, the very large number of products and cross products of $\mathbf{x}_{i,t}$ and $\mathbf{x}_{j}$ ($1 \times k$ vectors) may severely limit feasibility (even ignoring cross-products, at some cost, to limit this concern). This issue plagues White's test also, but its magnitude increases exponentially in the spatial case because directionality *i-j* and *j-i* matters, effectively squaring the number of necessary regressors in the test.

Lagrange multiplier (LM) tests derived by analogy to the standard LM test for serial correlation, which would regress $e_{i,t}$ on $e_{j,t}$, rather than $e_{i,t}$ on $e_{i,t-1}$, controlling for the other variables of the model. Like that inspired by White, this test retains validity with or without spatial lags as its temporal analogue does.

It may also prove more-widely feasible, and its large- and small-sample properties easier to derive and to simulate. However, such tests cannot distinguish the more perilous forms of spatial correlation where the spatial-dependence pattern correlates with the moments of the **x** matrix from the lesser forms.

Breusch and Pagan's (1980) LM test, which is described in more detail below where we begin the Monte Carlo assessments of this test's size. However, although this test's asymptotic properties are known as its validity with and without spatial lags in the model, its small-sample properties are unknown, and it unfortunately cannot distinguish more from less perilous patterns of spatial correlation.

Some semi- and non-parametric tests and measures based on groupings of residuals, including ones based on the familiar Durbin-Watson statistic, various spatial versions of autocorrelation and partial-autocorrelation correlograms, and Ljeung-Box Q. Some of these will not be robust to inclusion of spatial lags in the model, and their generally weaker structural assumptions may weaken them as tests for specific patterns of spatial dependence, such as that termed most-perilous above. However, conversely, we expect that their looser structure could broaden the range of patterns that might register empirically.

Tests for and Measures of Spatial Correlation: The Size of Breusch and Pagan's LM Test

The size of a statistical test is its probability of a Type I Error, i.e., of falsely rejecting a null hypothesis that is, in fact, true. In this section, we conduct Monte Carlo experiments to examine the size of Breusch and Pagan's (1980) LM test in small samples. Each experiment is defined by its sample dimensions, i.e., the number of units and periods. The results presented derive from 1000 independent trials. We use the following model to create a sample for N units over T periods:

$$y_{i,t} = x_{i,t} + \varepsilon_{i,t} \tag{1.9}$$

Variables are indexed *i=1..N* and *t=1…T* to identify each of the sample's *NT* observations. The exogenous variables, $y_{i,t}$ and $\varepsilon_{i,t}$ are independent draws from a standard normal distribution. Hence, the data exhibit no spatial correlation. We estimate the following model in such a sample by OLS:

$$y_{i,t} = \beta x_{i,t} + \gamma \bar{y}_{\sim i,t} + \varepsilon_{i,t} \tag{1.10}$$

The variable $\bar{y}$ is the average *y*-value for unit *i*'s sample counterparts, $\{j\}$, at time *i*, this being one way some (e.g., Franzese 2002) have attempted to address (some of) the spatial correlation in their TSCS data. Breusch and Pagan (1980: 247) calculate their LM statistic as:

$$LM = T\sum_{i=2}^{N}\sum_{j=1}^{i-1} r_{ij}^2 \tag{1.11}$$

with the *ij*th residual correlation. Under the null, LM is distributed asymptotically chi-squared with (N•(N 1))/2 degrees of freedom. A distribution thus determines the critical values for our test.

For now, our primary interest is how this LM statistic performs under different sample dimensions

(NxT). We consider several NxT dimensions common in political economy research: T's of 20, 30, or 40 (years) by N's of 5, 10, 15, 20, or 30 (countries). We conduct three experiments for each T using different N's, giving nine in total: 5x20, 10x20, 15x20, 5x30, 10x30, 20x30, 10x40, 20x40, and 30x40.

Table 4 reports the results for each experiment. We focus on the 95th percentile of the relevant chi-squared distribution as the critical value for comparison. Not surprisingly, the experimental size of the LM test for spatial correlation in small samples always exceeds the asymptotic size of the test: 0.05. In some cases, it more than doubles this 5% size. The LM test may help diagnose spatial correlation, but, as always, analysts must use care in interpreting borderline results suggesting correlation, especially in small samples where it rejects appreciably more often than warranted. One result of the experiments is surprising: for a given T, smaller N does not always produce more-accurate test-sizes (e.g., test size seemed truer at 10x30 than at 5x30); nor does increasing T for a given N always yield truer test-size (e.g., test size is truer at 10x30 than at 10x40). This suggests an optimal NxT ratio for test accuracy may exist, but, generally, this LM statistic performs reasonably well, though p-values based may understate true probabilities of Type I Errors in small samples. As we shall see below, however, the test easily reveals strong spatial correlation where it is present, although whether this power is accurate (i.e., avoidance of Type II Errors) at smaller degrees of spatial correlation and/or in smaller samples remains untested (for now).

Table 1: **Size of Breusch and Pagan (1980) LM Test under True Null Hypothesis of No Spatial Correlation**

N:	***5***	***10***	***15***	***5***	***10***	***20***	***10***	***20***	***30***
T:	***20***	***20***	***20***	***30***	***30***	***30***	***40***	***40***	***40***
Trials:	*1000*	*1000*	*1000*	*1000*	*1000*	*1000*	*1000*	*1000*	*1000*
Percentiles									
1%	3.68	28.29	82.22	3.12	29.03	158.60	28.79	154.57	381.87
5%	5.04	34.59	91.22	4.70	32.88	168.14	33.04	164.82	399.21
10%	6.20	36.87	94.10	5.87	35.84	174.75	35.59	172.09	409.12
25%	8.19	41.20	100.90	8.27	40.64	184.67	40.86	182.10	424.64
50%	11.23	47.10	110.83	11.04	46.51	196.48	47.26	195.81	446.34
75%	14.60	53.15	121.48	14.19	53.30	211.30	54.23	209.05	466.71
90%	18.20	59.50	130.99	18.23	60.12	224.42	60.64	224.03	487.62
95%	**20.62**	**65.06**	**138.49**	**20.58**	**64.04**	**230.62**	**65.21**	**233.83**	**497.55**
99%	27.96	76.46	150.91	27.62	74.87	248.61	74.79	251.69	521.02
Degrees Freedom:	10	45	105	10	45	190	45	190	435
Chi-Squared (95%):	***18.31***	***61.66***	***129.92***	***18.31***	***61.66***	***223.16***	***61.66***	***223.16***	***484.63***
Size of LM Test:	***0.095***	***0.08***	***0.113***	***0.096***	***0.074***	***0.108***	***0.086***	***0.104***	***0.116***

The likelihood-ratio tests described in Greene (1997: 661), which compare the sums of logged diagonal elements of the residual variance-covariance matrix under a restriction holding all off-diagonal elements to zero with the log of the unrestricted variance-covariance matrix's determinant, may also prove useful as preliminary test of the null hypothesis that the unrestricted matrix's off-diagonals are all zero. Determinants of diagonal matrices are just the products of their diagonal elements, so, if restricted and unrestricted variance-covariance matrices are equal, the sum of the logged diagonal elements and the log of the determinant will be equal. The sum of the restricted matrix's logged diagonals will exceed the log of the unrestricted matrix's determinant by more the larger are the absolute values of the off-diagonal elements in the unrestricted matrix. Under the null hypothesis, T times this difference is distributed chi-squared with N(N 1)/2 degrees of freedom, which, notice, being equal to the number of sample covariances, depends only on N.

The following tests and measures may be used upon the residuals from non-spatial regression models to explore whether some pattern of spatial correlation manifests in the residuals, suggesting that closer consideration of Galton's possibilities, with complications, may be in order. We illustrate them below using the base case OLS regression without any spatial components of our replication of Volden (2006), the details of which model appear in that part of Chapter 6.

Among the most common and well-known diagnostics of spatial correlation is *Moran's I*:

$$I = \frac{N}{S}\frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}, \text{ where } S = \sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N} w_{ij}\,, \tag{12}$$

When **W** (see below) is row-standardized (so row elements sum to one), the expression simplifies to:

$$I = \frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}} \tag{13}$$

To test a null of no spatial correlation (in patterns given by **W)**, one can compare a properly standardized Moran's I to the standard normal distribution (Cliff and Ord 1973, Burridge 1980, Kelejian and Prucha 2001).

In addition to Moran's I, several Lagrange multiplier (LM) tests based on OLS residuals exist. The standard LM tests assume that the spatial autoregressive process is either a spatial lag or spatial error model. More precisely, in terms of **(51)**, the standard LM test for the null hypothesis $\rho = 0$ against the spatial lag alternative *assumes* $\lambda = 0$. Likewise, the LM test for $\lambda = 0$ *assumes* $\rho = 0$. The standard, one-directional test against spatial lag alternative is calculated as

$$LM_{\rho} = \frac{\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y}/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{G + T\hat{\sigma}_{\varepsilon}^{2}}, \tag{14}$$

where $G = (\mathbf{WX}\hat{\boldsymbol{\beta}})'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{WX}\hat{\boldsymbol{\beta}})$ and $T = \text{tr}[(\mathbf{W}' + \mathbf{W})\mathbf{W}]$. The standard, one-directional test against spatial error alternative is

$$LM_{\lambda} = \frac{\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T} \qquad (15).$$

The drawback with these tests is that they have power against the incorrect alternative, which means they are usually not helpful for making specification choices. Regardless of whether the true spatial autoregressive process is a lag or error process, both tests are likely to reject the null hypothesis. Anselin et al. (1996) present robust LM tests for spatial dependence that are less problematic in this regard. The robust, one-directional test against spatial error alternative treats ρ in the mixed SAR model, **(51)**, as a nuisance parameter and controls for its effect on the likelihood. The statistic is then calculated as

$$LM_{\lambda}^{*} = \frac{\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2} - \left[T\hat{\sigma}_{\varepsilon}^{2}\left(G + T\hat{\sigma}_{\varepsilon}^{2}\right)^{-1}\right]\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T\left[1 - \frac{1}{\hat{\sigma}_{\varepsilon}^{2}}\left(G + T\hat{\sigma}_{\varepsilon}^{2}\right)\right]^{-1}}. \qquad (16).$$

The robust, one-directional test against spatial lag alternative is

$$LM_{\rho}^{*} = \frac{\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y} / \hat{\sigma}_{\varepsilon}^{2} - \hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{G + T(\hat{\sigma}_{\varepsilon}^{2} - 1)}. \qquad (17).$$

The two-directional LM test, finally, can be decomposed into the robust LM test for one alternative (lag or error) and the standard LM test for the other:

$$LM_{\rho\lambda} = LM_{\lambda} + LM_{\rho}^{*} = LM_{\rho} + LM_{\lambda}^{*} \qquad (18).$$

The one-directional test statistics are distributed χ_1^2 while the two-directional statistic is distributed χ_2^2. Using Monte Carlo simulations, Anselin et al. (1996) show that all five tests have the correct size in small samples. I.e., they all reject the null hypothesis at the stated rate when the null is true. The robust LM tests have lower power compared with the standard ones against the correct alternative, but the loss is relatively small and the robust tests are less likely to reject the null against the wrong alternative.

So, for example, when the true data generating process is a spatial AR error model ($\lambda \neq 0, \rho = 0$), rejection rates for LM_{λ} are about 5 percentage points higher on average across the range of λ than for LM_{λ}^{*}. The robustness of LM_{ρ}^{*} relative to LM_{ρ} is clear in this experiment. At $\lambda = .9$, LM_{ρ} rejects in favor of the incorrect alternative 89.9% of the time whereas LM_{ρ}^{*} rejects 17.1% of the time. The power advantage of the standard LM test is smaller when the true data generating process is a spatial AR lag model ($\lambda = 0, \rho \neq 0$). Rejection rates for LM_ρ are less than 2 percentage points higher on average than for LM^*_ρ across the full range of ρ . At $\rho = .9$, LM_λ rejects in favor of the incorrect alternative 100% of the time whereas LM^*_λ rejects 0.6% of the time. It seems the reduced power for increased robustness tradeoff strongly favors that the robust LM tests be included in the set of diagnostics.[10]

To help illustrate how these tests can be used in empirical research, we present OLS estimates for a non-spatial model of welfare policy generosity in column 1 of Table 1. All variables in our illustrative analysis are states' averages over the five years 1986-1990. The dependent variable is the maximum monthly AFDC benefit, and the independent variables are the state's poverty rate, average monthly wage in the retail sector, government ideology (ranging from 0=conservative to 100=liberal), degree of interparty competition (ranging from .5=competitive to 1.0=non-competitive, tax effort (revenues as a percentage of tax "capacity"), and the portion of AFDC benefits paid by the federal government. We use a standardized binary contiguity-weights matrix, which begins by coding $w_{ij} = 1$ for states *i* and *j* that share a border and $w_{ij} = 0$ for states that do not border. Then, we row-standardize (as commonly done in spatial-econometrics) the resulting matrix by dividing each cell in a row by that row's sum. This gives the unweighted average of the dependent variable in "neighboring" (so-defined) states.

[10] See Anselin et al. (1996) Tables 3-6. These results are for the N=40 experiments.

Table 1. State Welfare Policy (Maximum AFDC Benefit)

Independent Variables	OLS	Spatial AR Lag (S-OLS)	Spatial AR Lag (S-2SLS)	Spatial AR Lag (S-GMM)	Spatial AR Lag (S-MLE)	Spatial AR Error (S-MLE)
Constant	54.519	-246.76	-422.09	-500.05	-156.282	676.120
	(531.830)	(450.75)	(437.74)	(413.02)	(429.130)	(471.965)
Poverty Rate	-6.560	8.04	13.205	7.29	3.657	3.239
	(11.262)	(10.022)	(9.977)	(8.452)	(8.917)	(10.062)
Retail Wage	-.121	.016	.089	-.008	-.025	-.344
	(.226)	(.193)	(.187)	(.201)	(.181)	(.243)
Government	1.513	1.397	1.359*	1.655**	1.432*	1.696**
Ideology	(1.030)	(.863)	(.825)	(.761)	(.806)	(.822)
Inter-party	621.799**	368.65	286.98	438.9**	444.677*	263.887
Competition	(290.871)	(250.55)	(243.72)	(197.47)	(226.911)	(238.419)
Tax Effort	3.357**	2.022	1.553	2.397	2.423*	2.936**
	(1.587)	(1.364)	(1.328)	(1.493)	(1.262)	(1.213)
Federal Share	-4.405	-5.818	-6.012	-3.654	-5.393	-6.882*
	(5.001)	(4.20)	(4.014)	(3.415)	(3.901)	(4.099)
Spatial AR		.767***	1.069***	.840***	.537***	.565***
		(.178)	(.232)	(.237)	(.122)	(.131)
Moran I-statistic	3.312***					
$LM_{\rho\lambda}$	12.322***					
LM_{ρ}	11.606***					
LM_{ρ}^{*}	6.477**					
LM_{λ}	5.845***					
LM_{λ}^{*}	.716					
Log-likelihood					-270.763	-272.728
Adj.-R^2	.461	.622	.595	.606	.510	.588
Obs.	48	48	48	48	48	48

Notes: The spatial lags are generated with a binary contiguity weighting matrix. All the spatial weights matrices are row-standardized. ***Significant at the 1% Level; **Significant at the 5% Level; *Significant at the 10% Level.

The results for our non-spatial model suggest that high tax effort and low party competition are associated with more generous AFDC benefit payments. This seems reasonable. However, if the data exhibit spatial dependence, we need to worry about validity of these inferences. To check this possibility, we implement the diagnostic tests outlined above starting with Moran's I. The value of the standardized Moran-I test statistic is 3.312, which is statistically significant. We can reject the null hypothesis of no spatial dependence. We also include the LM tests. The result of the two-directional test leads to the same conclusion. Both the standard one-directional tests seem, predictably, statistically significant, which, unfortunately, gives us little guidance for specification. As expected, the robust one-directional tests are more helpful in this regard. The robust test against the spatial lag alternative is statistically significant while the robust test against the spatial error alternative is not. This suggests a spatial lag specification. We conclude with a warning. Ignoring evidence of spatial dependence can be extremely problematic, especially if the data suggest the true source of dependence is a spatial-lag process. In this case, simple OLS is likely to provide inaccurate coefficient estimates, particularly for variables that happen to cluster spatially (e.g., Franzese and Hays 2004, 2006a, 2007b).

[Non-nested model-testing strategies]

III. (CHAPTER 3) Specifying and Estimating Empirical Models of Spatial Interdependence

A. A Generic Empirical Model of Open Socio-Political-Economies with Context-Conditionality and Interdependence

We begin this chapter on specification by following the development of a subfield of political science that we call *comparative and international political economy* (*C&IPE*), identifying therein four approaches that motivate distinct empirical models. These generic empirical models, in turn, help illuminate the possible sources of spatial correlation, and highlight one of the fundamental challenges surrounding interdependence for empirical researchers, namely that correlated responses to common external conditions or to correlated unit-specific/domestic conditions are very difficult to distinguish from interdependence (*Galton's Problem)*.

Closed-economy comparative political-economy (*CPE*) focuses on domestic variables and ignores external shocks and international-interdependence processes.[11] *Open-economy CPE*, contrarily, recognizes the importance of external conditions/shocks (e.g., oil prices) for the domestic political-economy. The domestic policy or outcome responses to these foreign shocks may be moderated by domestic variables (*context-conditional open-economy CPE*) or unconditioned thereby (*orthogonal open-economy CPE*), but, either way, external shocks are held *exogenous* and treated as isolated phenomena. That is, in *open-economy CPE*, exogenous external conditions affect domestic

[11] Following recent practice in political science, we refer to processes by which the outcomes in some units directly affect the outcomes in other units as *interdependence*. We distinguish such interdependence processes, which will induce *spatial correlation*, from *spatially correlated responses (outcomes) to spatially correlated exogenous shocks*, or *common shocks* for short, and for that matter from spatially correlated domestic conditions, both of which will also induce *spatial correlation*. For us, synonyms for *interdependence* include *contagion*, *diffusion*, *strategic interdependence, strategic dependence*; and synonyms for *spatial (auto-)correlation* include *spatial dependence.* We have noticed, however, no consistency within or across disciplines in how these terms are used. For example, *contagion* would be synonymous with *interdependence* specifically in much of biometrics whereas it is often synonymous with *spatial correlation* generally in much of econometrics, and it seems equally likely to mean either in sociology.

policies and outcomes, but these domestic policies and outcomes do not themselves affect other units' policies and outcomes and so do not reverberate throughout the global polity or political economy. Finally, *international political economy* (*IPE*) focuses explicitly on spatial linkages and mechanisms of *interdependence* in the global political economy whereby policies and outcomes in some units directly affect the policies and outcomes of other units, perhaps in addition to the possibility that multiple units are exposed to common (or correlated) exogenous-external shocks or correlated domestic conditions. For example, a country might respond to an exogenous domestic or global political or economic shock by lowering its capital tax-rate (*open-economy CPE*), but the magnitude of its response may further depend on how its competitors respond and, conversely, its own response may affect the capital tax-rates that policymakers in other countries choose (*IPE*).[12] This is the hallmark of spatially interdependent processes: outcomes (i.e., explanandums, left-hand sides, or dependent variables, y_i, or components thereof, $\hat{y}_i$ or ε_i) in some units are among the explanators (right-hand sides, "independent" variables) of outcomes in others.

In this paper, we focus on models representing *context-conditional open-economy CPE* and *IPE* and methods for estimating such models. We do not consider purely domestic models except to note that, if external influences are important, these models will, even in the best of circumstances, produce inefficient estimates of the coefficients for domestic variables and, more usually, biased and inconsistent estimates. A central empirical challenge for C&IPE researchers to be explored here is the difficulty distinguishing *common shocks* (correlated responses to correlated exogenous-external or domestic conditions) from *international interdependence* (i.e., Galton's Problem *plus*[13]). On one hand, ignoring interdependence processes when they are present will lead analysts to exaggerate the importance of external shocks and thus to privilege open-economy CPE explanations. On the other hand, if certain simultaneity problems discussed below are insufficiently addressed, modeling interdependence with spatial lags will lead analysts to misestimate (perhaps usually to overestimate) the importance of interdependence at the expense of common shocks, especially insofar as such common shocks are inadequately modeled. Spatial two-stage-least-squares instrumental-variables or spatial maximum-likelihood estimators (S-2SLS-IV or S-ML) seem to provide effective resolutions (in somewhat different regards, under somewhat different conditions) to this dilemma, at least under

[12] If these responses are competitive, the initiating country will likely lower its capital taxes by more than it would have in the absence of *tax competition*.

[13] The *plus* underscores that Galton had not considered the further complication introduced by modern context-conditional open-economy CPE that (potentially correlated) domestic factors might condition the responses to correlated exogenous-external conditions.

circumstances so far considered: namely, that domestic explanatory factors are not endogenous to foreign dependent variables or that just one, well-specified interdependence process operates.

Thus, four broad visions of substantive political-economy—closed-economy CPE, orthogonal open-economy CPE, context-conditional open-economy CPE, and (comparative and) international political economy (C&IPE)—have evolved the field's intellectual history. Each of these visions has a characteristic mathematical expression of its empirical implications, and these characteristic empirical specifications clarify the inherent theoretical stance (assumption) in each regarding the substantive roles of domestic factors, exogenous-external conditions, and interdependence.

A. Closed-Economy Comparative-Political-Economy:

In closed-polity-and-economy CPE, domestic political and/or economic institutions (*e.g.*, electoral systems, central-bank autonomy), structures (*e.g.*, socioeconomic-cleavage or economic-industrial structures), and conditions (*e.g.*, electoral competitiveness, business cycles) are the paramount explanators of domestic outcomes. Such domestic-primacy substantive stances imply theoretical and empirical models of this form:[14]

$$y_{it} = D_{it}\beta_D + \varepsilon_{it} \qquad \textbf{(1)}$$

where y_{it} are the policies or outcomes to be explained (dependent variables) and D_{it} are the *domestic* institutional, structural, and other conditions that explain them (independent variables), each of which may vary across time and/or space. Most early 'quantitative' empirical studies in comparative politics and political economy were of this form,[15] later beginning to allow the stochastic component, ε_{it}, to exhibit some spatial correlation, but still presently treating this correlation as nuisance either to be "corrected" by Parks procedure (FGLS) or, later, merely to require an adjustment to standard-errors (PCSE). Examples here include most of the early empirical literature on the political economy of fiscal and monetary policy (*e.g.*, Tufte 1978, Hibbs 1987), coordinated wage bargaining and corporatism (*e.g.*, Cameron 1984, Lange 1984, Lange and Garrett 1985), and the early central-bank-independence literature (*e.g.*, Cukierman 1992, Alesina and Summers 1993).

B. Orthogonal Open-Economy Comparative-Political-Economy:

As economies grew more open and interconnected by international trade and, later, finance,

[14] To avoid a potentially confusingly technical expression in the third of these models, we will be somewhat loose in notation and algebra, using scalars, y and D etc., interchangeably with vectors and matrices, $\mathbf{y}$ and $\mathbf{D}$. The fully correct model here, assuming there are multiple domestic factors (but still a single dependent variable) is: $y_{it} = \mathbf{D}'\boldsymbol{\beta} + \varepsilon_{it}$ **(1a)**.

[15] Many early 'qualitative' studies also tended to ignore the spatial interdependence of their subject(s), or, at most, to mention the international context as among explanatory factors but generally elaborating little. Moreover, many modern political-economy studies, of both 'quantitative' and 'qualitative' varieties continue to ignore the spatial interdependence of their data (see Persson and Tabellini 2004, e.g.).

through the postwar period, and as perhaps their geopolitical interconnectedness increased also, comparative political-economists began to consider controlling for the effects of global political and economic conditions on domestic policies and outcomes to be more crucial. At first, however, such global conditions were assumed to hit all domestic units equally and to induce equal responses from each domestic unit to that impact. This implies theoretical/empirical models of the following form:[16]

$$y_{it} = D_{it}\beta_D + S_t\beta_S + \varepsilon_{it} \qquad \textbf{(2)}$$

where S_t are the exogenous global shocks (e.g., oil crises), felt equally by all sample units (each i feels an identical S_t), each of whom respond equally (by amount β_I) thereto. Again, the stochastic term, ε_{it}, may exhibit spatial correlation—i.e., spatial correlation distinct from that induced by exposure to these common shocks—but any such correlation was treated as nuisance either to be "corrected" by Parks procedure (FGLS) or, later, merely to require a standard-error adjustment (PCSE). Examples of empirical models reflecting such stances (often implicit) include many post-oil-crisis political-economy studies, including later rounds of the above literatures wherein time-period indicators or controls for global economic conditions began to appear: e.g., Alvarez, Garrett, and Lange (1991) on partisanship and corporatism interactions; Alesina, Roubini, and Cohen (1997) on political and/or partisan cycles; Powell and Whitten (1993) on economic voting.

C. Context-Conditional Open-Economy Comparative-Political-Economy:

Modern approaches to CPE continue to recognize the large effects of global shocks and other conditions abroad on the domestic political economy, but tend now to emphasize also how domestic institutions, structure, and conditions *shape* the degree and nature of domestic exposure to such shocks/conditions and *moderate* the domestic policy or outcome responses to these differently felt foreign stimuli. This yields characteristic theoretical and empirical models of the following sort:[17]

$$y_{it} = D_{it}\beta_D + S_t\beta_S + (D_{it}\cdot S_t)\beta_{DS} + \varepsilon_{it} \qquad \textbf{(3)}$$

where the incidence, impact, and/or effects of global shocks, S_t, on domestic policies and outcomes, y_{it}, are conditioned by domestic institutional-structural-contextual factors, D_{it}, and so differ across spatial units. Examples here include much of modern CPE, including Franzese (2002)[18] and all of the

---

[16] The model in formally correct matrix notation would be $\mathbf{y}_{it} = \mathbf{D}'_{it}\boldsymbol{\beta}_0 + \mathbf{S}'_t\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_{it}$ **(2b)**.

[17] In correct matrix notation, the model would be $\mathbf{y}_{it} = \mathbf{D}'_{it}\boldsymbol{\beta}_0 + \mathbf{S}'_t\boldsymbol{\beta}_1 + (\mathbf{D}_{it}\otimes\mathbf{S}_t)'\boldsymbol{\beta}_3 + \boldsymbol{\varepsilon}_{it}$ **(3b)**, which involves a Kronecker product, $\otimes$, which operation—the entire k_Dx1 vector, **D**, multiplies each element of the k_Sx1 vector **S**, resulting in the k_Dk_Sx1 vector in parentheses—we suspect may be arcane to some and so have evaded in the text.

[18] That varying domestic institutions and structures condition domestic policy and outcome responses to external shocks is a core argument of Franzese (2002). Although the estimated empirical models in that book are actually of type **(4)**, the resulting IPE and interdependence aspects receive little mention, less discussion, and no emphasis.

contributions to an *International Organization* special issue (Bernhard, Broz, and Clark 2002) on the choice of exchange-rate regimes (a notable *international*, or, at least, *foreign-policy*, institution). Once more, any spatial correlation distinct from that induced by common or correlated responses to globally *common shocks* would be left to FGLS or PCSE "corrections".

D. (Comparative &) International Political Economy:

International political economy is distinct from comparative political economy (insofar as the distinction exists) in the emphasis IPE places on international *relations* in political economy. That is, whereas open-economy CPE recognizes a role for international conditions, global conditions there remain *exogenous* to domestic policies and outcomes. International factors may thus be part of CPE explanations, but they are usually not central to the analysis. IPE analyses, contrarily, directly stress the international relations, i.e. the *interdependence* of policies and outcomes across nations, implying that, in general, policies in outcomes in countries i and j affect each other:[19]

$$y_{it} = \rho\sum_{j\neq i} w_{ij} y_{jt} + D_{it}\beta_D + S_t\beta_S + \left(D_{it}\cdot S_t\right)\beta_{DS} + \varepsilon_{it} \quad \textbf{(4)}$$

where y_{jt} are the outcomes in the other ($j\neq i$) units, which in some manner (given by ρw_{ij}) directly affect the outcome in unit i. Note, importantly, that w_{ij} reflects the *relative degree* of connection from j to i, and ρ reflects the overall *strength of interdependence* of the outcome in i on the outcomes in the other ($j\neq i$) spatial units, as weighted by w_{ij}. The rest of the right-hand-side model reflects the domestic political economy and, in the literature, has been as simple as **(1)** or as complex as **(3)**. Examples of these sorts of models include the recent work of Elkins and Simmons (2004) on the global interdependence of liberalization policies and reforms or Hays (2003) or Basinger and Hallerberg (2004) on the interdependence implied by tax competition.[20]

B. Estimators for Spatial-Interdependence 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

[19] A formal expression for this C&IPE model would be $y_{it} = \rho\mathbf{w}_i'\mathbf{y}_{jt} + \mathbf{D}_{it}'\boldsymbol{\beta}_D + \mathbf{S}_t'\boldsymbol{\beta}_S + \left(\mathbf{D}_{it}\otimes\mathbf{S}_t\right)\boldsymbol{\beta}_{DS} + \varepsilon_{it}$, where $\mathbf{w}_i$ is a *(J-1)x1* vector (*J*=the number of units) expressing the relative strengths of the connectivities from $j\neq i$ to i.

[20] Franzese (2003) gives a nonlinear least-squares spatial-lag model where domestic inflation policy/outcomes depend upon foreign inflation-rates, weighted (w_{ij}) in a manner determined by patterns of international monetary exposure and exchange-rate commitments. Franzese (2002) also estimates simple, unweighted (*i.e.*, constant w_{ij}) OLS models of type **(4)**, but the discussion there emphasizes the context-conditional CPE aspects of such models (see note 18).

comparative and international political-economy (C&IPE), the model involves unit-level explanators (e.g., individual or ***d***omestic factors), $\mathbf{d}_{it}$, exogenous-external conditions or ***s***hocks (e.g., oil prices, technology), $\mathbf{s}_t$, and context-conditional responses to external shocks (i.e., unit responses to exogenous-external conditions may depend on unit characteristics), $\left(\mathbf{d}'_{it} \otimes \mathbf{s}'_t\right)$, as well as interdependence processes, $\rho\sum_{j\neq i} w_{ij}y_{jt}$:

$$y_{it} = \rho\sum_{j\neq i} w_{ij}y_{jt} + \mathbf{d}'_{it}\boldsymbol{\beta}_{\mathbf{d}} + \mathbf{s}'_t\boldsymbol{\beta}_{\mathbf{s}} + \left(\mathbf{d}'_{it} \otimes \mathbf{s}'_t\right)\boldsymbol{\beta}_{\mathbf{ds}} + \varepsilon_{it} \qquad (19).$$

Interdependence refers to processes by which outcomes in some units, y_j, directly affect outcomes in others, y_i.[21] 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, $\left(\mathbf{d}'_{it} \otimes \mathbf{s}'_t\right)\boldsymbol{\beta}_{\mathbf{ds}}$), 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 policymakers in other countries choose (the term reflecting spatial interdependence, $\rho\sum_{j\neq i} w_{ij}y_{jt}$).

Much political-science substance and theory will imply empirical models like (1). Econometrically, however, obtaining *good* (unbiased, consistent, efficient) coefficient estimates and accurate standard-errors 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 non-spatial and spatial components, i.e., the interdependence part ($\rho\sum_{j\neq i} w_{ij}y_{jt}$) and the parts reflecting

[21] We do not consider models where interdependence arises solely or unequally though $\hat{y}_j$ 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).

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 $\left(\mathbf{d}'_{it}\otimes\mathbf{s}'_t\right)\boldsymbol{\beta}_{\mathbf{ds}}$). 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 non-spatial 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 non-spatial and interdependence mechanisms in opposite directions. That is, relative failure to model either the non-spatial 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 mis-estimation (often, but not necessarily, over-estimation) of the strength of interdependence and thereby bias conclusions toward the one (usually interdependence) and against the other (usually non-spatial).

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

$$\mathbf{y}=\rho\mathbf{W}\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \qquad (20),$$

where $\mathbf{y}$ is a $NTx1$ 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, etc.);[22] ρ is the spatial-autoregressive coefficient; $\mathbf{W}$ is an $NTxNT$ 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 $NTxK$ matrix of observations on K independent variables; $\boldsymbol{\beta}$ is a $Kx1$ vector of coefficients on $\mathbf{X}$; $\boldsymbol{\varepsilon}$ is a $NTx1$ residual vector. Note that each TxT block along $\mathbf{W}$'s

[22] 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 spatio-temporal dependence below, so we prefer to retain the *i,t* subscripts here.

block-diagonal,[23] which is the block multiplying $\mathbf{y}_i$ itself in the spatial-lag weighted-sum, is all zeros;[24] each of the off-diagonal TxT blocks has zero off-diagonal elements,[25] but non-zero 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.[26] 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} are 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 pre-specified spatial-weights, giving the strength of interdependence along the pre-specified paths.[27] 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 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 *co-membership*; our simulations below will reflect a special case of co-membership interdependence where all sample units are co-members of the same group and affect each other equally, implying uniform weights of $(N-1)^{-1}$.

1. Non-Spatial Least-Squares and Maximum-Likelihood models (inconsistent:

[23] The methodological literature on spatial dependence mostly focuses on cross-section data ($T=1$). In this case, each block referenced here in the text and surrounding notes has just one element.
[24] If **y** also manifests temporal dynamics, only **W**'s prime diagonal is zero; the off-diagonal elements of the TxT block-diagonal are non-zero and reflect these temporal correlations.
[25] I.e., unless **y** exhibits spatial, cross-temporal interdependence so y_{it} affects y_{js} for some $i \neq j$ and $t \neq s$.
[26] In fact, symmetric **W** is unlikely in most C&IPE contexts (at least), where it would imply, e.g., equal-strength effects US→Belgium and Belgium→US. 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.
[27] The accuracy of **W**'s pre-specification, both absolutely and relative to the non-spatial 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.

omitted-variable bias)

To estimate (2), one could simply omit **Wy** and estimate **β** by ordinary least-squares regression of **y** on **X**: *non-spatial 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 standard-error corrections, PCSE or *clustering*). A second strategy, almost as simple to implement and in increasing use, estimates **β** 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. Maximum-likelihood offers a third strategy of estimating ρ and **β** in a model that specifies the joint likelihood of **y** to reflect the spatial interdependence (Ord 1975). *Spatial ML* is computationally intense, especially as *NxT*-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 *instrumental-variables by spatial two-stage-least-squares* (S-2SLS), also produces consistent and asymptotically efficient estimates, provided its necessary conditions are met: namely, that the **X** are indeed exogenously related to **y**.[28]

The two simplest estimators (OLS, S-OLS) are, as just noted, mis-specified 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 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 non-spatial 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 comparison to S-2SLS and S-ML even if IV instrumentation or ML joint-

[28] See Kelejian et al. (2003) and Kelejian and Robinson (1993) for more-complete lists of estimators and more-thorough coverage of the Instrumental-Variable/Method-of-Moments class of estimators.

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 LS), but also that S-ML seems to dominate more notably in other conditions and is rarely outperformed.

2. Least-Squares and (Quasi) Maximum-Likelihood with Spatial Lags (inconsistent: simultaneity bias)

3. Spatial Method-of-Moments Estimators:

a) Spatial Instrumental-Variables / Two-Stage-Least-Squares (consistent but inefficient)

b) Spatial Generalized-Method-of-Moments (consistent and asymptotically efficient)

4. (Fully Specified) Spatial Maximum-Likelihood (consistent and asymptotically efficient and normal)

B. Estimating Lag Models

The spatial lag model has become a very popular specification in social science research. One might arrive to this model via batteries of diagnostic tests or directly from theory. The theory-driven approach starts by estimating the spatial model, and then uses Wald, LR, and related tests to refine the specification. We begin with OLS estimation of spatial lag models, which we label spatial OLS (S-OLS).

1. Spatial OLS

Spatial OLS is inconsistent. To see this, we start by rewriting the spatial lag model as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \text{ where } \mathbf{Z} = \begin{bmatrix} \mathbf{Wy} & \mathbf{X} \end{bmatrix} \text{ and } \boldsymbol{\delta} = \begin{bmatrix} \rho & \boldsymbol{\beta} \end{bmatrix}' \qquad (21).$$

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

$$\text{plim } \hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \text{plim}\left[\left(\frac{\mathbf{Z'Z}}{n}\right)^{-1}\frac{\mathbf{Z'}\boldsymbol{\varepsilon}}{n}\right]. \qquad (22).$$

In the case where $\mathbf{Z}$ is a single exogenous regressor, $\mathbf{x}$, $\left(k=1, \operatorname{cov}(\boldsymbol{\varepsilon},\mathbf{x})=0\right)$, we can rewrite (22) as

$$\operatorname{plim}\hat{\boldsymbol{\delta}}=\begin{bmatrix}\rho\\ \beta\end{bmatrix}+\begin{bmatrix}\dfrac{\operatorname{var}(\mathbf{x})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2} & \dfrac{-\operatorname{cov}(\mathbf{x},\mathbf{Wy})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2}\\ \dfrac{-\operatorname{cov}(\mathbf{x},\mathbf{Wy})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2} & \dfrac{\operatorname{var}(\mathbf{Wy})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2}\end{bmatrix}\begin{bmatrix}\operatorname{cov}(\boldsymbol{\varepsilon},\mathbf{Wy})\\ \operatorname{cov}(\boldsymbol{\varepsilon},\mathbf{x})\end{bmatrix} \quad (23).$$

If we define $\boldsymbol{\Psi}=\operatorname{plim}\left(\dfrac{\mathbf{Z'Z}}{n}\right)$ and $\boldsymbol{\Gamma}=\operatorname{plim}\left(\dfrac{\mathbf{Z'}\boldsymbol{\varepsilon}}{n}\right)$, and do the matrix multiplication, (23) simplifies to

$$\operatorname{plim}\hat{\boldsymbol{\delta}}=\begin{bmatrix}\rho\\ \beta\end{bmatrix}+\begin{bmatrix}\dfrac{\Psi_{22}\Gamma_{11}}{|\boldsymbol{\Psi}|}\\ -\dfrac{\Psi_{12}\Gamma_{11}}{|\boldsymbol{\Psi}|}\end{bmatrix} \quad (24).$$

Since $\boldsymbol{\Psi}$ is a variance-covariance matrix, its determinant is strictly positive. With positive (negative) spatial dependence in the data, the covariances Ψ_{12} and Γ_{11} are positive (negative), and S-OLS will overestimate (underestimate) ρ and underestimate (overestimate) β. This is one of the analytic results we stressed repeatedly earlier: the simultaneity biases of S-OLS tend to induce exaggerated estimates of interdependence strength and correspondingly deflated estimates of the importance of non-spatial factors.

The S-OLS estimates are provided in column 2 of Table 1. Consistent with the results from our diagnostic tests, the estimated coefficient on the spatial lag is large, positive and statistically significant. The OLS estimates most affected by the switch to a spatial-lag specification are the party-competition and tax-effort coefficients, which become statistically insignificant. Conversely to S-OLS's simultaneity biases, the OLS coefficient estimates on these two variables may, because they cluster spatially, have suffered from omitted variable bias that would have inflated those estimates.

Franzese and Hays (2004, 2006a, 2007b) conclude that spatial OLS, despite its simultaneity, can perform acceptably under low-to-moderate interdependence-strength and reasonable sample-dimensions. Given our results, S-OLS is clearly preferable to OLS. In this particular case, however, both the size of the spatial-lag coefficient and the fact that no other coefficients are statistically significant should raise concern about simultaneity bias. We have advised using some consistent estimator under conditions like these. We discuss three consistent estimators below, starting with spatial-2SLS and spatial-GMM.

2. Spatial 2SLS and Spatial GMM

Spatial-2SLS and spatial-GMM provide consistent estimates for the coefficient on the spatial lag and use spatially weighted values of the exogenous variables in other units as instruments. The latter extends the former to account the heteroscedasticity in the quadratic form of the sample orthogonality conditions. If this particular form of heteroscedasticity is present, the S-GMM estimator yields smaller asymptotic variance than the spatial-2SLS estimator. If it is absent, the two estimators are equivalent.[29] Note that a mixed spatial autoregressive model of the form in **(51)** would suffer from heteroscedasticity, making S-GMM more efficient for estimating $\boldsymbol{\delta}$. (In this particular case, a generalized S-2SLS estimator using a Cochrane-Orcutt like transformation of the data is also available; see Kelejian and Prucha 1998, 1999).

To see how we estimate the spatial lag model (21) using S-2SLS, define the linear prediction of $\mathbf{Wy}$:

$$\widehat{\mathbf{Wy}} = \boldsymbol{\Pi}[\boldsymbol{\Pi})'(\boldsymbol{\Pi})]^{-1}\boldsymbol{\Pi}'\mathbf{Wy} \tag{25),}$$

where $\boldsymbol{\Pi}$ is the full set of *exogenous* variables including, at least, $\mathbf{X}$ and $\mathbf{WX}$. $\mathbf{WX}$ provides *spatial*-instruments.[30] Thus, $\boldsymbol{\Pi}$ is an N x L matrix, where $L \geq 2k$. The orthogonality condition for the 2SLS estimator is formally written as $E\left[\boldsymbol{\Pi\varepsilon}\right]=0$. Next, define $\hat{\mathbf{Z}}$ as an N*x*(k+1) matrix of the predicted values of $\mathbf{Wy}$ and $\mathbf{X}$,

$$\hat{\mathbf{Z}} = \left[\widehat{\mathbf{Wy}} \quad \mathbf{X}\right]. \tag{26).}$$

Using this definition, the spatial-2SLS estimator is

$$\hat{\boldsymbol{\delta}}_{\text{S2SLS}} = (\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'\mathbf{y} \tag{27);}$$

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\delta}}_{\text{S2SLS}}\right) = s^2(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1} \tag{28).}$$

where s^2 is calculated from residuals in the original structural model, (21), with $\hat{\boldsymbol{\delta}}_{\text{S2SLS}}$ substituted for $\boldsymbol{\delta}$.

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

---

[29] When the number of excluded exogenous variables exactly equals the number of endogenous variables, the GMM, 2SLS, and ILS estimators are equivalent. Therefore, we could more accurately say that GMM improves on 2SLS when the coefficients in the system/equation are overidentified and heteroscedasticity exists. In this case, we have one endogenous regressor, the spatial lag, in one equation. Provided the number of exogenous variables in $\mathbf{X}$ exceeds one, the number of spatial instruments will exceed one, making the coefficient on the spatial lag overidentified.

[30] One can also include higher order spatial instruments in $\boldsymbol{\Pi}$ --that is, $\left\{\mathbf{W}^2\mathbf{X}, \mathbf{W}^3\mathbf{X}, \mathbf{W}^4\mathbf{X}, ...\right\}$.

$$q = E[\boldsymbol{\mu}(\boldsymbol{\delta})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}(\boldsymbol{\delta})'] \,, \tag{29},$$

with the corresponding moment conditions:

$$\boldsymbol{\mu}(\boldsymbol{\delta}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\pi}_i(y_i - \mathbf{z}_i'\boldsymbol{\delta}) \tag{30};$$

$$\boldsymbol{\Sigma} = E\left[\boldsymbol{\mu}(\delta)\boldsymbol{\mu}(\delta)'\right] = \frac{1}{N}E\left[\sum_{i=1}^{N}\boldsymbol{\pi}_i\boldsymbol{\pi}_i'(y_i - \mathbf{z}_i'\boldsymbol{\delta})^2\right] = \frac{1}{N}\sum_{i=1}^{N}\omega\boldsymbol{\pi}_i\boldsymbol{\pi}_i = \frac{1}{N}\left(\boldsymbol{\Pi}'\boldsymbol{\Omega}\boldsymbol{\Pi}\right) \tag{31}.$$

In these equations, $\boldsymbol{\pi}_i$ is a column vector (lx1) that is the transpose of the i^{th} row of $\boldsymbol{\Pi}$ (representing the i^{th} observation) and, similarly, $\mathbf{z}_i$ is a (k+1)x1 vector that is the transpose of the i^{th} row of $\mathbf{Z}$. The GMM weighting matrix is calculated by inverting a consistent estimate of the variance-covariance matrix of the moment conditions.[31] White's estimator provides a consistent non-parametric estimate of $\boldsymbol{\Sigma}$ provided we have a consistent estimator of $\boldsymbol{\delta}$ (Anselin 2006). Fortunately, spatial-2SLS can provide these. Thus, the estimate for $\boldsymbol{\Sigma}$ is

$$\mathbf{S}_0 = \sum_{i=1}^{N}\boldsymbol{\pi}_i\boldsymbol{\pi}_i'\left(y_i - \mathbf{z}_i'\hat{\boldsymbol{\delta}}_{\text{S2SLS}}\right)^2 \tag{32},$$

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

$$\hat{\boldsymbol{\delta}}_{\text{SGMM}} = \left[\mathbf{Z}'\boldsymbol{\Pi}(\mathbf{S}_0)^{-1}\boldsymbol{\Pi}'\mathbf{Z}\right]^{-1}\left[\mathbf{Z}'\boldsymbol{\Pi}(\mathbf{S}_0)^{-1}\boldsymbol{\Pi}'\mathbf{y}\right] \tag{33};$$

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\delta}}_{\text{SGMM}}\right) = \left[\mathbf{Z}'\boldsymbol{\Pi}\left(\hat{\mathbf{S}}_0^{-1}\right)\boldsymbol{\Pi}'\mathbf{Z}\right]^{-1} \tag{34}.$$

We present the S-2SLS and S-GMM estimates for the spatial-lag model of welfare policy generosity in columns 3 and 4 of Table 1. The S-2SLS estimates for this particular specification and dataset are troubling as the spatial-lag coefficient estimate exceeds one, giving a non-stationary spatial process. This is a bit surprising when compared with the smaller S-OLS result, given that the S-OLS estimator has likely-inflationary simultaneity biases and S-2SLS likely does not. Of course, this can happen with a single sample and/or if the exogeneity of the instruments is violated.[32] The S-GMM estimates are better. The spatial-lag coefficient estimate is well below one (though it is still large) and the standard errors are about 5% smaller than the S-2SLS standard errors on average, as expected given the likely efficiency of the GMM estimator. The coefficients on government

[31] The logic here is similar to that behind the WLS estimator. OLS is consistent in the presence of heteroscedasticity, but WLS is more efficient. Likewise, 2SLS is consistent under heteroscedasticty, but GMM is asymptotically more efficient.
[32] Franzese and Hays (2004) show that the exogeneity at issue here is that the y_i must not cause the x_j, a condition we call (no) cross-spatial endogeneity. Such reverse "diagonal" causality seems unlikely to arise in many substantive contexts, although we do note also that spatial correlation among the other regressors plus the typical endogeneity from y to x would create it.

ideology and on party competition are statistically significant. The results suggest that, *ceteris paribus*, welfare benefits are highest in states with non-competitive party systems and liberal governments.

3. Spatial Maximum Likelihood

Implementing S-ML is not complicated, although the spatial-lag model adds a slight wrinkle to the standard linear additive case, and the maximization can be computationally intense for large samples. To see the minor complication, start by isolating the stochastic component of the spatial-lag model:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \Rightarrow \boldsymbol{\varepsilon} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \equiv \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \qquad (35).$$

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

$$L(\boldsymbol{\varepsilon}) = \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right) \qquad (36),$$

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

$$L(\mathbf{y}) = |\mathbf{A}|\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \qquad (37),$$

and the log-likelihood takes the form

$$\ln L(\mathbf{y}) = \ln|\mathbf{A}| - \left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln\sigma^2 - \left(\frac{1}{2\sigma^2}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \qquad (38).$$

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 $\prod_i \lambda_i$ because the eigenvector $\boldsymbol{\lambda}$ in this approximation does not depend on ρ. Then $|\mathbf{I}- \rho\mathbf{W}| = \prod_i (1-\lambda_i)$, which requires the estimation routine only to recalculate a product, not a determinant, as it updates.[33] 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 same approximation serves there.

[33] Unfortunately, the approximation may be numerically unstable (Anselin 1988, 2001; Kelejian and Prucha 1998). On the other hand, S-ML may enjoy a practical advantage over S-2SLS in multiple-W models in that S-ML does not require differentiated instrumentation for each **W** to gain distinct leverage on its corresponding ρ. The instruments, **WX**, would differ by virtue of **W** differing for the alternative interdependence processes, so S-2SLS is estimable for multiple-**W** models even with identical **X** in the **WX** instruments, but we harbor doubts about the practical identification leverage obtainable thereby.

Typically, estimation proceeds by maximizing a concentrated-likelihood. Given an estimate of the spatial-lag coefficient, ρ, an analytic optimum estimate of the non-spatial coefficients can be found as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \rho(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} = \hat{\boldsymbol{\beta}}_{\mathbf{O}} - \rho.\hat{\boldsymbol{\beta}}_{\mathbf{L}} \qquad (39).$$

Note that the first term in the second two expressions of (39) 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 of these rely only on observables, (except for ρ), and so are readily calculable given some ρ (estimate). Next, define these terms:

$$\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{O}} \text{ and } \hat{\boldsymbol{\varepsilon}}_{\mathbf{L}} = \mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{L}} \qquad (40).$$

It then follows that

$$\hat{\sigma}^2 = (1/N)(\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\varepsilon}}_{\mathbf{L}})'(\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\varepsilon}}_{\mathbf{L}}) \qquad (41)$$

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

$$\ln L_C(\mathbf{y}) = -\left(\frac{N}{2}\right)\ln\pi + \ln|\mathbf{A}| - \frac{N}{2}\ln\left(\frac{1}{N}(\boldsymbol{\varepsilon}_{\mathbf{O}} - \rho\boldsymbol{\varepsilon}_{\mathbf{L}})'(\boldsymbol{\varepsilon}_{\mathbf{O}} - \rho\boldsymbol{\varepsilon}_{\mathbf{L}})\right) \qquad (42)$$

yields the S-ML estimate of ρ which is substituted into (39) to get $\hat{\boldsymbol{\beta}}$. 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.

The S-ML estimates for our spatial lag model of welfare policy generosity are provided in column 5 of Table 1. These estimates are mostly similar to the S-GMM estimates. The most noteable difference is in the estimate of ρ. The S-ML coefficient is approximately 36% smaller than the S-GMM coefficient, and it is estimated much more precisely, the standard error being about half the size of the S-GMM standard error. Three of the coefficients in this model are statistically significant including the tax effort coefficient. The S-ML estimates imply welfare benefits are systematically larger, all else equal, in states with high taxes, liberal governments, non-competitive party systems. Franzese and Hays (2004, 2006a, 2007b) find that S-ML generally outperforms S-2SLS on mean squared error grounds. S-GMM lessens the efficiency advantage for S-ML over the IV class of estimators.

C. Classes of Spatial-Econometric Models

1. Spatial-Error Models (spatial interdependence in the stochastic component)

a) Utility & Limits of Consistent Variance-Covariance Estimation (PCSE's)

b) A Simple, Data-Driven Efficiency-Enhancement of PCSE: Utility, Limits

1.1. Spatial Model – 3 (Spatial Error Model)

1.1.1. Proposed Empirical Model

[10] [11] Those who take the standard "nuisance" approach to spatial dependence (OLS-PCSE) assume, hopefully correctly, that the spatial dependence in the true data generating process is limited to the error term and completely uncorrelated with the systematic component (i.e., with anything else in the model). Thinking in terms of Persson and Tabellini, this would imply that only unexpected changes in other countries tax policies would cause governments to reform their own.[34] In this case, we have what is referred to in the spatial econometrics literature as a spatial error or spatial moving average model, which takes the form

$$\begin{aligned} &\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}, \text{ where} \\ &\boldsymbol{\varepsilon} = \boldsymbol{\lambda\varepsilon} + \mathbf{u} \\ &\boldsymbol{\lambda} = \rho\mathbf{W} \end{aligned} \quad . \tag{6}$$

In the time-series-cross-sectional (TSCS) version of this model, $\mathbf{y}$ is an $NTx1$ vector of observations on the dependent variable stacked by unit (i.e., unit 1, time 1 to T, then unit 2, time 1 to T, *etc.* through unit N), and $\mathbf{W}$ is an $NTxNT$ block-diagonal spatial-weighting matrix (with elements w_{ij}). Thus, $\boldsymbol{\lambda\varepsilon}$ is an $NTx1$ matrix of spatially correlated disturbances. The $NTx1$ matrix $\mathbf{u}$ contains i.i.d. disturbances. The diagonal elements of the off-diagonal TxT blocks in $\mathbf{W}$, which reflect the contemporaneous effect of the column unit on the row unit, are the w_{ij} that reflect the degree of connection from unit j to i—so, unlike a variance-covariance matrix, $\mathbf{W}$ need not be symmetric. The spatial autoregressive coefficient, ρ, reflects the impact of the outcomes in the other ($j{\neq}i$) spatial units, as weighted by w_{ij}, on the outcome in i. Thus, ρ gauges the overall strength of diffusion, whereas the w_{ij} describe the relative magnitudes of the diffusions paths between the sample units.

The coefficients in this model can be estimated consistently using OLS. However, the standard error estimates are biased downward. Moreover, OLS is inefficient. The feasible generalized least squares (FGLS) estimator, developed for the TSCS context by Parks (1967), is efficient, but very few political scientists use Parks FGLS since the publication of Beck and Katz

[34] Perhaps capital flows are more sensitive to unexpected tax policy changes.

(1995). Beck and Katz argued, and demonstrated with Monte Carlo simulations, that Parks' method underestimates the true variance of the FGLS estimator's sampling distribution when N and T are small. Moreover, they argued the efficiency gains are modest in sample sizes typical in political science research. They concluded that the small efficiency gains come at a high cost in terms of overconfidence (i.e., significantly underestimating the true standard errors). Their recommendation, which is widely followed, is to report OLS coefficient estimates with their robust panel corrected standard errors (PCSE's).

A problem with both Parks-FGLS and OLS-PCSE is that because the spatial dependence is relegated to the error term of the model, it is almost always ignored by analysts. We propose a simple spatial OLS estimator (S-OLS) that uses the spatial interdependence in the data to create a spatial lag that, for each unit in the dataset, is a weighted average of the contemporaneous OLS residuals from the other units.

More specifically, our approach to S-OLS, which is only appropriate when the conditions in (6) hold, follows four steps:

1) Estimate the regression of $\mathbf{y}$ on $\mathbf{X}$ using non-spatial OLS
2) Use the OLS residuals from step one to estimate $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\lambda}\hat{\boldsymbol{\varepsilon}} + \mathbf{u}$ again using OLS
3) Implement S-OLS by regressing $\mathbf{y}$ on $\mathbf{X}$ and $\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\varepsilon}}$
4) Estimate panel corrected standard errors (Beck and Katz 1995)

Using the experimental design from Beck and Katz (1995), Franzese and Hays (2005) show this estimator matches most of the efficiency gains of Parks-FGLS in small samples without the problem of overconfidence, in part because S-OLS can be combined with PCSE estimates. The main results from their study are provided in the appendix (Tables A1 and A2). [10] [11]

1.2. Simulation – 2 (Spatial Error Model)

1.2.1. Experimental Design

[10] [11] In our experiments, we replicate and extend the simulations in Beck and Katz (1995), focusing our attention on the overconfidence and efficiency issues addressed in their Tables 4 and 5 respectively. (See pages 638-9 for a detailed discussion of their experimental design.) Beck and Katz use a version of equations in (6) above to generate their experimental data, which feature disturbances that are both contemporaneously correlated across units and panel heteroscedastic. The contemporaneous correlation or spatial dependence is experimentally manipulated through $\boldsymbol{\lambda}$ while panel heteroscedasticity is created by allowing the variance of $\mathbf{u}$ to be unit specific. In these

experiments, N is fixed at 16 while T is manipulated. [10] [11]

1.2.2. Simulation Results

[10] Our results with respect to standard error estimates and overconfidence are reported in the Table 1. The entries in this table gauge the accuracy of the standard error estimates (in percentages) by comparing them with the true variability in the coefficient estimates across the 1,000 trials of our experiment. Percentages above 100 indicate that the standard error estimates understate the true variability. The columns in Table 1 labeled OLS and PCSE replicate the findings in Beck and Katz. OLS standard errors underestimate the true variability in the coefficient estimates while panel corrected standard errors report this variability accurately. While we do not report them, the Parks standard errors are very inaccurate in small samples tending to underestimate the true variability of the coefficient estimates.[35] The column labeled S-OLS (PSCE) presents a new set of results, which show that PCSE estimates suffer very little from the inclusion of a spatial lag in the model. For example, when both panel heteroscedasticity and contemporaneous correlation are present, which is likely in real data, the difference between PCSE with and without the spatial lag 3%, 5%, and 6% for T's of 20, 30 and 40 respectively. Note that, for the experiments reported in their Table 4, Beck and Katz generate X's and ε's that are contemporaneously correlated and panel heteroscedastic. Since this specification implies spatial dependence in both the systematic and stochastic components of y, we use $\hat{\boldsymbol{\lambda}}\mathbf{y}$ as our spatial lag instead of $\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\varepsilon}}$.

[35] See Beck and Katz (1995) Table 2. The Parks standard errors are highly inaccurate when the number of observations per unit (T) is close to or not much greater than the number of units in the sample (N). The degree of overconfidence when N and T = 20 is more than 600% meaning that the true variability in the coefficient estimates is more than six times larger than that suggested by the standard error estimates.

Table 1. Replication and Extension of Beck and Katz (1995, Table 4)

			Overconfidence (%)		
T	Heterosce-dasticity	Contempo-raneous Correlation	OLS	PCSE	S-OLS (PCSE)
10	0	0	101	102	NA
10	0	0.25	136	105	NA
10	0.3	0	114	100	NA
10	0.3	0.25	145	104	NA
20	0	0	102	102	102
20	0	0.5	220	107	115
20	0.3	0	123	106	106
20	0.3	0.5	224	106	109
30	0	0	96	97	97
30	0	0.5	214	102	115
30	0.3	0	115	99	99
30	0.3	0.5	232	108	113
40	0	0	99	99	99
40	0	0.5	214	101	118
40	0.3	0	114	98	98
40	0.3	0.5	227	104	110

Notes: The spatial weighting matrix (W) was estimated using the OLS residuals. This matrix cannot be estimated unless T > N. S-OLS includes a spatial lag of the dependent variable on the right-hand-side (W*Y). These results are based on 1,000 trials. The number of units was fixed at 16 as opposed to 15

In Table 2 we report the efficiency comparisons. Efficiency is measured in percentages relative to non-spatial OLS. We use the same experimental parameters (N, T, and the degree of contemporaneous correlation) in Beck and Katz (1995, Table 5). The columns labeled Parks replicate their findings for the relative efficiency of FGLS to OLS. The columns labeled S-OLS compare the efficiency of our spatial OLS estimator to OLS. The first thing to note about these results is that S-OLS more or less matches the efficiency of Parks FGLS. The relative efficiency of both estimators relative to non-spatial OLS is determined by the degree of contemporaneous correlation in the data and the ratio of N to T in the sample. Our experiments suggest that on efficiency grounds S-OLS outperforms non-spatial OLS when the degree of contemporaneous correlation is above 0.25 and T is at least twice as large as N.

Table 2. Replication and Extension of Beck and Katz (1995, Table 5)

		Relative Efficiency of Parks and S-OLS to OLS (Over 100% indicates superiority of OLS)							
		Contemporaneous Correlation of the Errors							
N	T	0.0		0.25		0.5		0.75	
		Parks	S-OLS	Parks	S-OLS	Parks	S-OLS	Parks	S-OLS
10	10	102	NA	102	NA	99	NA	97	NA
	20	112	110	104	106	91	95	73	82
	30	110	110	104	102	86	88	66	73
	40	111	110	99	96	82	84	63	68
15	15	102	NA	101	NA	99	NA	98	NA
	20	110	111	101	108	94	103	88	99
	30	107	109	104	105	91	95	74	80
	40	112	113	102	103	87	90	67	72
20	20	101	NA	100	NA	99	NA	98	NA
	25	105	109	101	107	96	105	88	100
	30	113	113	102	108	93	100	83	96
	40	114	112	101	101	89	92	74	82

Notes: The spatial weighting matrix (W) was estimated using the OLS residuals. This matrix cannot be estimated unless T > N. S-OLS includes a spatial lag of the residuals on the right-hand-side (W*E). These results are based on 1,000 trials. Number of units was fixed at 16 as opposed to 15

To sum, we believe the S-OLS approach outlined above allows one to capture the efficiency benefit of Parks FGLS without paying the cost of inaccurate standard error estimates, and its use as an alternative to OLS-PCSE can be justified on purely econometric grounds in many circumstances. We stress once again that the experimental conditions we created are the most favorable to the "nuisance" approach to spatial dependence, which makes our results all the more remarkable. We show later that OLS-PCSE performs significantly worse than S-OLS once we relax the assumption that all the spatial dependence is in the error term. Most importantly, S-OLS focuses one's attention on the degree of spatial interdependence in the data, which is too easily ignored with OLS-PCSE. We demonstrate this problem in our reanalysis of the tax regressions in Swank and Steinmo (2002).

c) True Spatial-Error Models

C. Estimating Error Models

If specification tests indicate that spatial dependence is of the form in **Error! Reference source not found.**, OLS coefficient estimates are consistent, but standard-error estimates will be biased. One could combine OLS coefficient estimates with robust standard errors (e.g., PCSE's: Beck and Katz 1995, 1996). Another option is to estimate a spatial-error model. In this section we consider the maximum-likelihood estimator for this model. Again, we start by isolating the stochastic component, which, in this case, is

$$\mathbf{y}=\mathbf{X\beta}+\boldsymbol{\varepsilon}\equiv\mathbf{X\beta}+\left(\mathbf{I}-\lambda\mathbf{W}\right)^{-1}\mathbf{u}\Rightarrow\mathbf{u}=\left(\mathbf{I}-\lambda\mathbf{W}\right)\left(\mathbf{y}-\mathbf{X\beta}\right)\equiv\mathbf{B}\left(\mathbf{y}-\mathbf{X\beta}\right) \qquad (43).$$

The likelihood for a spatial error process is

$$L(\mathbf{y})=|\,\mathbf{B}\,|\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{N}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\mathbf{B}'\mathbf{B}\left(\mathbf{y}-\mathbf{X\beta}\right)\right) \qquad (44),$$

where $|\mathbf{B}|=|\mathbf{I}-\lambda\mathbf{W}|$, and the log-likelihood takes the form

$$\ln L(\mathbf{y})=\ln|\,\mathbf{B}\,|-\left(\frac{N}{2}\right)\ln\left(2\pi\right)-\left(\frac{N}{2}\right)\ln\sigma^2-\left(\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\mathbf{B}'\mathbf{B}\left(\mathbf{y}-\mathbf{X\beta}\right)\right) \qquad (45).$$

We first calculate OLS residuals, and then estimate λ by maximizing the concentrated likelihood:

$$\ln L_C(\mathbf{y})=-\left(\frac{N}{2}\right)\ln\left(2\pi\right)+\ln|\,\mathbf{B}\,|-\frac{N}{2}\ln\left(\frac{1}{N}(\hat{\boldsymbol{\varepsilon}}'\mathbf{B}'\mathbf{B}\hat{\boldsymbol{\varepsilon}})\right) \qquad (46).$$

Given λ, the ML estimates for $\boldsymbol{\beta}$ are calculated using FGLS

$$\hat{\boldsymbol{\beta}}_{ML}=(\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{y} \qquad (47).$$

The asymptotic variance covariance matrix for $\hat{\boldsymbol{\beta}}_{ML}$ is

$$\operatorname{var}\left(\hat{\boldsymbol{\beta}}_{ML}\right)=\hat{\sigma}^2\left(\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X}\right)^{-1} \qquad (48).$$

where $\hat{\sigma}^2=(1/N)(\hat{\boldsymbol{\varepsilon}}'\mathbf{B}'\mathbf{B}\,\hat{\boldsymbol{\varepsilon}})$ and $\hat{\boldsymbol{\varepsilon}}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML}$. The asymptotic variance for λ is

$$\operatorname{var}\left(\hat{\lambda}\right)=2\operatorname{tr}\left(\mathbf{W}\mathbf{B}^{-1}\right)^2 \qquad (49).$$

The S-ML estimates for the spatial error model of welfare policy generosity are provided in the last column of Table 1. We note only that the log-likelihood value for the error model is less than the log-likelihood for the lag model, and this is consistent with the robust LM specification test results.

2. Spatial-Lag Models (spatial interdependence in systematic component or *y*)

a) Can address Galton; key is the relative & absolute accuracy and power of the spatial and non-spatial aspects of the empirical model.

b) As in time-series, modern view is increasingly that this is way to proceed.

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

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (50),$$

where $\mathbf{y}$ is a $NTx1$ 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, etc.);[36] ρ is the spatial-autoregressive coefficient; $\mathbf{W}$ is an $NTxNT$ 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 $NTxK$ matrix of observations on K independent variables; $\boldsymbol{\beta}$ is a $Kx1$ vector of coefficients on $\mathbf{X}$; $\boldsymbol{\varepsilon}$ is a $NTx1$ residual vector. Note that each TxT block along $\mathbf{W}$'s block-diagonal,[37] which is the block multiplying $\mathbf{y}_i$ itself in the spatial-lag weighted-sum, is all zeros;[38] each of the off-diagonal TxT blocks has zero off-diagonal elements,[39] but non-zero diagonal elements reflecting the contemporaneous spatial correlation in $\mathbf{y}$. Note also that, as the w_{ij} elements of $\mathbf{W}$ reflect the relative connectivity from unit j to i, $\mathbf{W}$ may not be symmetric.[40] 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} are determined by theoretical and substantive argument as to which units will have greatest effect on outcomes in which others. ρ is the coefficient on $\mathbf{W}$'s pre-specified spatial-weights, giving the strength of interdependence along the pre-specified paths.[41] 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

---

[36] 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 spatio-temporal dependence below, so we prefer to retain the i,t subscripts here.
[37] The methodological literature on spatial dependence mostly focuses on cross-section data (T=1). In this case, each block referenced here in the text and surrounding notes has just one element.
[38] If $\mathbf{y}$ also manifests temporal dynamics, only $\mathbf{W}$'s prime diagonal is zero; the off-diagonal elements of the TxT block-diagonal are non-zero and reflect these temporal correlations.
[39] I.e., unless $\mathbf{y}$ exhibits spatial, cross-temporal interdependence so y_{it} affects y_{js} for some $i\neq j$ and $t\neq s$.
[40] In fact, symmetric $\mathbf{W}$ is unlikely in most C&IPE contexts (at least), where it would imply, e.g., equal-strength effects US→Belgium and Belgium→US. 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 $\mathbf{W}$ may be more likely in some more-homogenous contexts.
[41] The accuracy of $\mathbf{W}$'s pre-specification, both absolutely and relative to the non-spatial components of the model is of crucial empirical, theoretical, and substantive importance. Strategies for parameterizing $\mathbf{W}$ and estimating such models are of great interest but as yet mostly remain for future work to develop.

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 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 *co-membership*; our simulations below will reflect a special case of co-membership interdependence where all sample units are co-members of the same group and affect each other equally, implying uniform weights of $(N-1)^{-1}$.

C. Combined Lag and Error Models

A third model combines the two. Different externalities in modeled and unmodeled effects (a.k.a. systematic and stochastic components, which relaxes the previously noted constraint in the spatial-lag model. The resulting mixed SAR model is:

$$\begin{aligned}\mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \mathbf{u}\end{aligned} \qquad (51).$$

Analogously, a mixed SARMA model would be:

$$\begin{aligned}\mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}_2 \mathbf{u} + \mathbf{u}\end{aligned} \qquad (52).$$

3. Spatio-Temporal-Lag Models

a) Time-lagged spatial interdependence: utility and limits. Spatial LS or ML consistent under some conditions. Give and discuss those conditions.

b) Contemporaneous spatial interdependence

C. Spatio-temporal Models for Panel Data

The spatio-temporal autoregressive (STAR) lag model can write in matrix notation as

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \phi \mathbf{V}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \qquad (53)$$

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).[42] 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. $\mathbf{Wy}$ is thus the spatial lag; i.e., for each observation y_{it}, $\mathbf{Wy}$ gives a weighted sum of the y_{jt}, with weights, w_{ij}, given by the relative connectivity from j to i. Notice how $\mathbf{Wy}$ thus directly and straightforwardly reflects the dependence of each unit i's policy dependence on unit j's policy, exactly as in the formal model and theoretical arguments reviewed above. The parameter ϕ is the temporal autoregressive coefficient, and $\mathbf{V}$ is an $NT{\times}NT$ matrix with ones on the minor diagonal, i.e., at coordinates $(N{+}1,1)$, $(N{+}2,2)$, $\ldots,(NT,NT\text{-}N)$, and zeros elsewhere, so $\mathbf{Vy}$ 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 equation (53), $\boldsymbol{\varepsilon}$, is an $NT{\times}1$ vector of disturbances, assumed to be independent and identically distributed.[43]

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:

$$\text{Log} f_{\mathbf{y}_t,\mathbf{y}_{t-1},\ldots,\mathbf{y}_2|\mathbf{y}_1} = -\frac{1}{2}N\left(T-1\right)\log\left(2\pi\sigma^2\right)+\left(T-1\right)\log\left|\mathbf{I}-\rho\mathbf{W}\right|-\frac{1}{2\sigma^2}\sum_{t=2}^{T}\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t \tag{54}$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \rho\mathbf{W}_N\mathbf{y}_t - \phi\mathbf{I}_N\mathbf{y}_{t-1} - \mathbf{X}_t\boldsymbol{\beta}$.

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 $\mathbf{A} = \phi\mathbf{I}$, $\mathbf{B} = \mathbf{I} - \rho\mathbf{W}$, and ω as a characteristic root of $\mathbf{W}$, the statio-temporal process generating the data is covariance stationary if

$$\left|\mathbf{AB}^{-1}\right| < 1$$

or, equivalently, if

[42] With some work, nonrectangular panels and/or missing data are manageable, but we assume rectangularity and completeness for simplicity of exposition.

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

$$\begin{cases} |\phi| < 1 - \rho\omega_{\max}, & \text{if } \rho \geq 0 \\ |\phi| < 1 - \rho\omega_{\min}, & \text{if } \rho < 0 \end{cases} \tag{55}$$

If W is row-standardized and both the temporal and spatial dependence are positive ($\rho > 0$ and $\phi > 0$), stationarity requires simply that $\phi + \rho < 1$.

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).

$$\begin{aligned} \operatorname{Log} f_{\mathbf{y}_t,\ldots,\mathbf{y}_1} = & -\frac{1}{2} NT\log\left(2\pi\sigma^2\right) + \frac{1}{2}\sum_{i=1}^{N}\log\left(\left(1-\rho\omega_i\right)^2 - \phi^2\right) + \left(T-1\right)\sum_{i=1}^{N}\log\left(1-\rho\omega_i\right) \\ & -\frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t'\varepsilon_t - \frac{1}{2\sigma^2}\varepsilon_1'\left(\left(\mathbf{B}\text{-}\mathbf{A}\right)'\right)^{-1}\left(\mathbf{B}'\mathbf{B}\text{-}\mathbf{B}'\mathbf{A}\mathbf{B}^{-1}\left(\mathbf{B}'\mathbf{A}\mathbf{B}^{-1}\right)'\right)^{-1}\left(\mathbf{B}\text{-}\mathbf{A}\right)^{-1}\varepsilon_1 \end{aligned} \tag{56}$$

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.

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

$$\mathbf{y}_t = \eta\mathbf{W}\mathbf{y}_{t-1} + \phi\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t. \tag{57}$$

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

$$\begin{aligned} \operatorname{Log} f_{\mathbf{y}_t,\ldots,\mathbf{y}_1} = & -\frac{1}{2} NT\log\left(2\pi\sigma^2\right) + \frac{1}{2}\sum_{i=1}^{N}\log\left(1-\left(\phi+\eta\omega_i\right)^2\right) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t'\varepsilon_t \\ & -\frac{1}{2\sigma^2}\varepsilon_1'\left(\left(I-A\right)'\right)^{-1}\left(I-AA'\right)^{-1}\left(I-A\right)^{-1}\varepsilon_1 \end{aligned} \tag{58}$$

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 $\mathbf{A} = \phi\mathbf{I} + \eta\mathbf{W}$. 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.

We present estimates for a panel model of welfare policy generosity in Table 3. The data are annual observations from 1981-1990 on the contiguous 48 states. The dependent variable is the maximum AFDC benefit, and the independent variables remain unchanged. All the regressions include fixed state effects. The first column contains a non-spatial model estimated with OLS.

Clearly, from Moran's I statistic and the two-directional LM statistics, there is spatial dependence in the dataset. The diagnostics do not provide clear evidence in favor of a spatial lag or error specification, however. We estimate both with contemporaneous spatial lags. The second column contains a spatio-temporal lag model, and the third column contains a combined temporal lag and spatial error model. Interestingly, the retail wage variable is statistically significant and positive in all three regressions. Once again, the tax effort coefficient becomes statistically insignificant with the change from a non-spatial to spatial specification.

Table 3. State Welfare Policy (Maximum AFDC Benefit, 1981-1990)

Independent Variables	OLS	Spatial AR Lag (MLE)	Spatial AR Error (MLE)
Poverty Rate	-.855	-.911	-.903
	(1.130)	(1.050)	(1.198)
Retail Wage	.217***	.204***	.197***
	(.036)	(.034)	(.037)
Government Ideology	.053	.059	.027
	(.087)	(.081)	(.083)
Inter-party Competition	18.960	25.540	18.633
	(24.046)	(22.442)	(22.382)
Tax Effort	.388*	.322	.349
	(.223)	(.208)	(.218)
Federal Share	.483	.859*	.750
	(.521)	(.491)	(.510)
Temporal AR	.663***	.628***	.666***
	(.030)	(.031)	(.030)
Spatial AR		.143***	.200***
		(.044)	(.058)
Moran I-statistic	3.296***		
$LM_{\rho\lambda}$	11.896***		
LM_{ρ}	9.976***		
LM_{ρ}^{*}	1.446		
LM_{λ}	10.450***		
LM_{λ}^{*}	1.921		
Log-likelihood		-1991.357	-1991.290
Adj.-R^2	.981	.981	.982
Obs.	480	480	480

Notes: 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. All the spatial weights matrices are row-standardized. ***Significant at the 1% Level; **Significant at the 5% Level; *Significant at the 10% Level.

4. Spatial- and Spatio-Temporal-Lag Qualitative-Dependent-Variable Models:

a) Spatial Probit

B. Spatial Models for Binary Outcomes

The methods for estimating and analyzing spatial latent variable models for categorical data have received significant attention in the literature recently. Much of the methodological research has focused on the spatial probit model (e.g., McMillen 1992, LeSage 2000). This is also one of the most frequently used models in the applied research (Beron et al. 2003, Simmons and Elkins 2004). In this section we consider spatial models for binary outcomes, starting with the spatial lag probit model.

1. Spatial Lag Probit Models

The structural model for the spatial probit takes the form

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{59}$$

which can be written in its reduced form as

$$\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{60}$$

where $\mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$ and $\mathbf{y}^*$ is a latent variable is linked to the observed variable $\mathbf{y}$ through the measurement equation

$$y_i = \begin{cases} 1 \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \leq 0 \end{cases}. \tag{61}$$

The marginal probabilities are calculated as follows

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \Pr\left((\mathbf{I} - \rho\mathbf{W})_i^{-1}\mathbf{x}_i'\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})_i^{-1}\varepsilon_i > 0\right)$$

or

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \Pr\left(u_i < \frac{(\mathbf{I} - \rho\mathbf{W})_i^{-1}\mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i}\right) \tag{62}$$

using the marginal distribution from a multivariate normal with variance-covariance matrix $[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}$. The denominator in equation (62), which is the square root of the variance for unit *i*, is attributable to the heteroscedasticity induced by the spatial dependence. This heteroscedasticity distinguishes the spatial probit from the conventional probit and makes the estimator for the latter inconsistent for the spatial case. The fact that the u_i are interdependent also makes the standard probit estimator inappropriate for the spatial model. One does not sum the log of *n* one-dimensional probabilities to estimate the model, but rather calculates the log of one *n*-dimensional normal probability.

Beron et al. (2003) proposed estimation by simulation using recursive importance sampling

(RIS), which is discussed extensively in Vijverberg (1997). LeSage (2000) has suggested using Bayesian Markov Chain Monte Carlo (MCMC) methods. The MCMC approach is mostly straightforward. The full conditional distributions are standard except one, and therefore the Gibbs sampler can be used. The parameter ρ has a non-standard conditional distribution. Metropolis-Hastings sampling is used to draw values from this posterior.

We estimate several spatial lag probits in Table 2 using both standard ML and MCMC methods. In keeping with our state welfare spending example, we switch the dependent variable from maximum AFDC benefits to whether or not a state's CHIP (Children's Health Insurance Program) includes a monthly premium payment (Volden 2006). We keep the same independent variables since this dependent variable also reflects the generosity of the welfare program.

Table 2. State Welfare Policy (Monthly CHIP Premium)

Independent Variables	Probit MLE	Probit MCMC	Spatial AR Lag Probit	Spatial AR Error Probit
Constant	-4.978	-5.163	-5.606	-5.531
	(6.260)	(6.292)	(10.159)	(7.337)
Poverty Rate	-.244	-.265**	-.374**	-.243*
	(.153)	(.156)	(.231)	(.157)
Retail Wage	.004	.004*	.006*	.004*
	(.003)	(.003)	(.004)	(.003)
Government Ideology	.011	.011	.014	.014
	(.013)	(.013)	(.020)	(.014)
Inter-party Competition	2.174	2.108	1.473	2.636
	(3.388)	(3.478)	(6.134)	(3.794)
Tax Effort	-.014	-.014	-.020	-.017
	(.019)	(.019)	(.034)	(.021)
Federal Share	.045	.048	.065	.043
	(.063)	(.064)	(.095)	(.066)
Spatial AR	.079	.102	.200***	.297***
	(.798)	(.815)	(.148)	(.196)
Pseudo-R^2	.222	.220	.607	.574
Obs.	48	48	48	48

Notes: In the first two columns, the models are estimated assuming the spatial lags are exogenous. The model in the first column is estimated using standard ML techniques. The parentheses in this column contain estimated standard errors and the hypothesis tests assume that the asymptotic t-statistics are normally distributed. The models in columns two through four are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient estimate is the mean of the posterior density based on 10,000 observations after a 1000 observation burn-in period. The number in parentheses is the standard deviation of the posterior density. The p-values are also calculated using the posterior density. The last two models are estimated with true spatial estimators described in the text. In third column, 30 of the 10,000 spatial AR coefficients sampled from the posterior distribution were negative. In the fourth column, none of the 10,000 sampled spatial AR coefficients were negative. ***p-value <.01, **p-value<.05, *p-value <.10.

In the first two columns, the models are estimated assuming the spatial lags are exogenous. The model in the first column is estimated using standard ML techniques. The parentheses in this column contain estimated standard errors and the hypothesis tests assume that the asymptotic t-statistics are normally distributed. The models in columns two and three are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient estimate is the mean of the posterior distribution based on 10,000 observations after a 1000 observation burn-in period. The number in parentheses is the standard deviation of the posterior distribution. The p-values are also calculated using the posterior. The results in columns two and three are very similar, as they should be given our diffuse priors. Because the estimator used in column two incorrectly treats the spatial lag as exogenous (i.e., like any other right-hand-side variable) the likelihood is misspecified and the sampler draws from the wrong posterior distribution for the spatial coefficient ρ. This specification error has serious consequences for drawing inferences about the importance of spatial interdependence.

The model in column three is estimated with the true spatial estimator described above. The draws for ρ are taken from the correct (non-standard) posterior distribution using Metropolis-Hastings. In this case only 30 of the 10,000 spatial AR coefficients sampled from the posterior distribution were negative. Thus, there is strong evidence of positive spatial interdependence in states' decisions to include a monthly premium in their CHIP. In addition, these probit results suggest that a state's poverty rate and average monthly retail wage are also important determinants.

2. Spatial Error Probit Models

The spatial error version of the probit model takes the form

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \,, \tag{63}$$

where $\mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$. In this case, the marginal probabilities are calculated as

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \Pr\left(u_i < \frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i} \right) \tag{64}$$

using the marginal distribution from a multivariate normal with variance-covariance matrix $[(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}$. The same estimation techniques used for the spatial lag model can be used for the spatial error model. We present estimates for the spatial error model in column four of Table 2. In this case, none of the 10,000 sampled spatial AR coefficients were negative. We do not discuss specification tests (*lag* vs. *error*) for the spatial probit, but not that they are covered in Anselin (2006).

b) Spatial Over- and Under-dispersion in Count Models

D. Challenges for Estimation and Inference

1. Simultaneity of the simple estimation strategies (LS/ML, S-OLS/Quasi-ML)

2. Specification & Measurement Error

a) In the non-spatial components of the model

b) In the spatial-connectivity matrix

c) In instruments' exogeneity or in likelihood

3. W: The Spatial-Connectivity Matrix of Relative Interdependence

a) W commonly pre-specified; crucial stage of spatial analysis

(1) Examples given, discussion

(2) Achen's Problem with lagged-dependent-variable (LDV) models

(a) Review: Time-LDV v. AR(p) residuals

(b) Analogous issue in spatial-LDV v. spatial-error models

b) Estimators' properties tend to vary, often in complicated fashion, with W

c) Multiple-W Models: raise issues beyond the usual ones associated with controls (e.g., interpretation, colinearity) due to simultaneity and "overlap"

d) Inroads into parameterization and estimation of W

IV. (CHAPTER 4) Estimator Properties

A. Analytic Results

1. In 2-unit, 2-regressor case: omitted-variable bias (OVB) in non-spatial OLS and simultaneity in Spatial OLS

2. General Derivation of Simultaneity Biases in Spatial and Spatio-Temporal Model

We now demonstrate analytically, in a case simplifying (2) to a single regressor, **x**, that both non-spatial and spatial OLS estimates are biased and inconsistent in the presence of interdependence, and we specify those biases precisely. S-OLS estimation of (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 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 (2) as:

$$\mathbf{y} = \mathbf{Q}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \text{ where } \mathbf{Q} = \begin{bmatrix}\mathbf{Wy} & \mathbf{x}\end{bmatrix} \text{ and } \boldsymbol{\delta} = \begin{bmatrix}\rho & \beta\end{bmatrix}' \qquad (65).$$

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

$$\text{plim } \hat{\boldsymbol{\delta}}_{\text{S-OLS}} = \boldsymbol{\delta} + \text{plim}\left[\left(\tfrac{\mathbf{Q}'\mathbf{Q}}{n}\right)^{-1}\left(\tfrac{\mathbf{Q}'\boldsymbol{\varepsilon}}{n}\right)\right] \qquad (66).$$

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

$$\text{plim } \hat{\boldsymbol{\delta}}_{\text{S-OLS}} = \begin{bmatrix}\rho \\ \beta\end{bmatrix} + \frac{1}{|\boldsymbol{\Psi}|}\begin{bmatrix}\text{cov}\left(\mathbf{Wy},\boldsymbol{\varepsilon}\right)\times\text{var}(\mathbf{x}) \\ -\text{cov}\left(\mathbf{Wy},\boldsymbol{\varepsilon}\right)\times\text{cov}(\mathbf{Wy},\mathbf{x})\end{bmatrix} \text{ where } \boldsymbol{\Psi} = \text{plim}\left(\frac{\mathbf{Q}'\mathbf{Q}}{n}\right) \qquad (67).$$

So, e.g., in the likely common case of positive interdependence and positive covariance of spatial-lag and exogenous regressors, *S-OLS* would generally over-estimate 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 non-spatial OLS—is simply omitted-variable bias, the formula for which is well-known to be $\mathbf{F}\boldsymbol{\beta}$ 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:

$$\text{plim } \hat{\beta}_{\text{OLS}} = \beta + \rho\times\frac{\text{cov}(\mathbf{Wy},\mathbf{x})}{\text{var}(\mathbf{x})} \qquad (68).$$

$\hat{\rho}_{\text{OLS}} \equiv 0$, 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):

$$\begin{aligned} y_1 &= \rho_{12}y_2 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \rho_{21}y_1 + \beta_2 x_2 + \varepsilon_2 \end{aligned} \qquad (69).$$

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

$$S-OLS\ bias\ of\ \hat{\rho}_{21}=\tfrac{1}{n}(\mathbf{Q'Q})^{-1}_{22}\times\left(\tfrac{\rho_{12}}{1-\rho_{12}\rho_{21}}\right)\times Var(\varepsilon_2) \quad (70).$$

The $\{2,2\}^{\text{th}}$ element of ***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}.[44] Note the subscript reversal; it is intuitive. If Japan affects the US negatively and the US affects Japan positively, e.g., then S-OLS's simultaneity-biased estimates of US→Japan interdependence would, by ignoring dampening feedback from the negative Japan→US effect, induce underestimation of the positive US→Japan dependence. Trying to estimate the negative Japan→US interdependence by S-OLS would conversely incur positive bias by ignoring the countervailing US→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 (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 (non-spatial) model factors. To restate this substantively, S-OLS will suffer sizable simultaneity biases only if and insofar as spatial interdependence is strong <u>and</u> unit-level (domestic) and contextual (exogenous-external) factors account for relatively little systematic variation.

Finally, the omitted-variable biases in non-spatial OLS are larger as domestic or exogenous-external factors correlate more spatially. With fully common exogenous-external factors (e.g., time-period fixed-effects, as in our simulations), $x_1=x_2$ and OLS 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 $\hat{\rho}_{12}=\rho_{12}+\frac{\beta_1\beta_2+\rho_{21}\beta_1^2}{(1-\rho_{12}\rho_{21})(\beta_2^2+\beta_1^2\rho_{21}^2+\rho_{21}^2\,\text{var}(\varepsilon_1)+\text{var}(\varepsilon_2))}$.

In sum, S-OLS estimates of (1) or (2) suffer simultaneity bias; OLS estimates of (1) or (2) excluding the spatial lag suffer omitted-variable bias. S-OLS estimates of interdependence from *j* to *i*

[44] Note that we assume throughout that |ρ|<1. Spatial unit-roots, |ρ|≥1, are highly problematic statistically but, thankfully, also highly improbable substantively. This implies that $|\rho_{12}\rho_{21}|<1$ also.

will have bias of the same sign as the interdependence from *i* to *j*. 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 non-spatial 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 non-spatial 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, for example, 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, instrumental-variables (IV) and maximum-likelihood (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 is difficult. We turn therefore to Monte Carlo simulation to compare these estimators in richer, more realistic scenarios.

OLS estimation of model **Error! Reference source not found.**, 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

$$\mathbf{y} = \mathbf{Q}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \qquad (1.71),$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Wy} & \mathbf{My} & \mathbf{X} \end{bmatrix} \text{ and } \boldsymbol{\delta} = \begin{bmatrix} \rho & \phi & \boldsymbol{\beta} \end{bmatrix}' \qquad (1.72).$$

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

$$\text{plim } \hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \text{plim} \left[\left(\frac{\mathbf{Q}'\mathbf{Q}}{n} \right)^{-1} \frac{\mathbf{Q}'\boldsymbol{\varepsilon}}{n} \right] \qquad (1.73).$$

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 (22) as

$$\text{plim } \hat{\boldsymbol{\delta}} = \begin{bmatrix} \rho \\ \phi \\ \beta \end{bmatrix} + \frac{1}{|\boldsymbol{\Psi}|} \begin{bmatrix} \text{cov}(\mathbf{Wy}, \boldsymbol{\varepsilon}) \times \text{var}(\mathbf{My}) \times \text{var}(\mathbf{x}) \\ -\text{cov}(\mathbf{Wy}, \boldsymbol{\varepsilon}) \times \text{cov}(\mathbf{Wy}, \mathbf{My}) \times \text{var}(\mathbf{x}) \\ -\text{cov}(\mathbf{Wy}, \boldsymbol{\varepsilon}) \times \text{cov}(\mathbf{Wy}, \mathbf{x}) \times \text{var}(\mathbf{My}) \end{bmatrix} \qquad (1.74),$$

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 (23) 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.

Table 1: 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.

3. Simultaneity Biases in the multiple-W case.

A. Spatial Autoregressive Models with Multiple Lags

One innovation in the booming literature on policy and institutional diffusion in recent years is the use of spatial autoregressive models with multiple lags to evaluate distinct diffusion mechanisms (Simmons and Elkins 2004; Elkins, Guzman and Simmons 2006; Lee and Strang 2006). This section briefly highlights some of the difficulties involved in estimating these models, focusing on the linear additive case. Again, there are two main versions of this model. Brandsma and Ketellapper (1979) and Dow (1984) estimate biparametric error models of the form

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda_1 \mathbf{W}_1 \boldsymbol{\varepsilon} + \lambda_2 \mathbf{W}_2 \boldsymbol{\varepsilon} + \mathbf{u} \end{aligned}, \tag{75.}$$

using the maximum likelihood technique described in IV.C. In this case, the concentrated likelihood contains $|\mathbf{B}|=|\mathbf{I}-\lambda_1\mathbf{W_1}-\lambda_2\mathbf{W_2}|$ (see (46)). Again, the OLS estimator for $\boldsymbol{\beta}$ is consistent but inefficient when the spatial dependence takes the form in (75). Lacombe (2004) estimates a biparametric lag model:

$$\mathbf{y} = \rho_1 \mathbf{W}_1 \mathbf{y} + \rho_2 \mathbf{W}_2 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{76.}$$

As with the single-spatial-lag model, S-OLS estimation of the biparametric model suffers simultaneity bias. However, the problem is potentially worse in the case of multiple spatial lags with its two or more endogenous variables rather than one. To see this, first rewrite the model (without exogenous regressors):

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\rho} + \boldsymbol{\varepsilon} \;\text{ where }\; \mathbf{Z} = \begin{bmatrix} \mathbf{W}_1\mathbf{y} & \mathbf{W_2}\mathbf{y} \end{bmatrix} \;\text{ and }\; \boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \rho_2 \end{bmatrix}' \tag{77.}$$

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

$$\text{plim } \hat{\boldsymbol{\rho}} = \boldsymbol{\rho} + \text{plim}\left[\left(\frac{\mathbf{Z'Z}}{n}\right)^{-1}\frac{\mathbf{Z'\varepsilon}}{n}\right] \quad (78),$$

which can be written as

$$\text{plim } \hat{\boldsymbol{\rho}} = \begin{bmatrix}\rho_1\\ \rho_2\end{bmatrix} + \begin{bmatrix} \dfrac{\text{var}(\mathbf{W}_2\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - \left[\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})\right]^2} & \dfrac{-\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - \left[\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})\right]^2} \\ \dfrac{-\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - \left[\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})\right]^2} & \dfrac{\text{var}(\mathbf{W}_1\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - \left[\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})\right]^2} \end{bmatrix} \begin{bmatrix}\text{cov}(\boldsymbol{\varepsilon},\mathbf{W}_1\mathbf{y})\\ \text{cov}(\boldsymbol{\varepsilon},\mathbf{W}_2\mathbf{y})\end{bmatrix} \quad (79).$$

If we define $\boldsymbol{\Psi} = \text{plim}\left(\frac{\mathbf{Z'Z}}{n}\right)$ and $\boldsymbol{\Gamma} = \text{plim}\left(\frac{\mathbf{Z'\varepsilon}}{n}\right)$, and carry out the matrix multiplication, equation (79) simplifies to

$$\text{plim } \hat{\boldsymbol{\rho}} = \begin{bmatrix}\rho_1\\ \rho_2\end{bmatrix} + \begin{bmatrix} \dfrac{\Psi_{22}\Gamma_{11} - \Psi_{12}\Gamma_{21}}{|\boldsymbol{\Psi}|} \\ \dfrac{\Psi_{11}\Gamma_{21} - \Psi_{12}\Gamma_{11}}{|\boldsymbol{\Psi}|} \end{bmatrix} \quad (80).$$

Since $\boldsymbol{\Psi}$ is a variance-covariance matrix, its determinant is strictly positive $\left(\left[\Psi_{11}\Psi_{22} - \Psi_{12}^2\right] > 0\right)$. Therefore, if we assume that 1) there is positive spatial dependence $\left(\boldsymbol{\rho}, \Psi_{12}, \Gamma_{11} > 0\right)$, 2) the spatial lags have the same degree of endogeneity $\left(\Gamma_{11} = \Gamma_{21}\right)$, and 3) the variance of $\mathbf{W}_2\mathbf{y}$ is less than the variance of $\mathbf{W}_1\mathbf{y}$ $\left(\Psi_{22} < \Psi_{11}\right)$, it follows that S-OLS will underestimate ρ_1 and overestimate ρ_2 *asymptotically*.

Fortunately, the maximum likelihood estimator can be implemented in almost the same manner described in IV.B.3. In the biparametric case, the error term is

$$\boldsymbol{\varepsilon} = (\mathbf{I}_N - \rho_1\mathbf{W}_1 - \rho_2\mathbf{W}_2)\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \equiv \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \quad (81).$$

With this change, the likelihood function in (42) can be used for estimation. The main practical difficulty in maximizing the concentrated likelihood is calculating the log-determinant of $\mathbf{A}$. Lacombe (2004) addresses this difficulty by calculating $\log|\mathbf{A}|$ over a grid of values for ρ_1 and ρ_2 prior to estimation. His routine calls values from this table during the optimization process.

B. Simulation Results: using generic empirical model of interdependence, and:

1. Interdependence Processes/Patterns of Interdependence:

a) Uniform/Homogenous Interdependence: The full set of experiments

described are conducted first with this specification; then repeated for all of the following, on which the reports will be briefer.

A. Design of the Simulation Exercises (Monte Carlo Experiments)

The true model generating the data for our experiments is the reduced-form solution of (2):[45]

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \qquad (82),$$

with matrix **X** comprising $\mathbf{d}=\{d_{it}\}$, $\mathbf{s}=\{s_t\}$, $\mathbf{ds}=\{\mathrm{d}_{it}s_t\}$, as in our generic context-conditional model (1). Residuals, **ε**, are an *NTx1* vector of *i.i.d.* draws from a standard-normal distribution. **d** is also an *NTx1* 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 *NTx1* stack of *T* vectors of size *Nx1*. 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.[46] 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 ρ and relative connectivities from units *j* to *i* of w_{ij}.

We then generate **y** by (9) using coefficients $(\beta_D,\beta_S,\beta_{DS},\rho)=(1,1,1,\rho)$, varying ρ from 0.1 to 0.9 and sample dimensions over subsets of $N=\{5,10,25,40,50\}$ *x* $T=\{20,35,40,50\}$ 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** while ρ gives the general strength of interdependence following this pattern. Thus, larger ρ (for a given $\overline{w}_{ij}$ [47]) implies stronger interdependence. We assume the spatial dependence is time-invariant—i.e., the w_{ij} connecting *j* to *i* persists all *T* periods without change—so each diagonal element of each *TxT* 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

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

[46] Note: **s** is one regressor, not a set of time-period dummies. This assumes 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.

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

alleviates estimation issues from endogeneity. However, many substantive contexts and dataset structures will combine to suggest interdependence across units occurs within the time span of one observation period, and, even if not, time-lagged spatial-lags only alleviate the endogeneity insofar as temporal *and spatio-temporal* dynamics are modeled adequately.[48]

Finally, all data in our experiments are temporally uncorrelated; effectively, we assume temporal dependence successfully modeled elsewise (e.g., by time-lagged dependent-variables) or absent (as likely, e.g., in pooled independent surveys).[49] We then set the interdependence pattern between spatial units by fixing all of these w_{ij} elements of **W** to $(N\text{-}1)^{-1}$. I.e., 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.[50]

To estimate spatial-lag models, researchers must pre-specify this spatial-weights matrix, **W**. This pre-specification is a crucial theoretical and empirical step in studying interdependence.[51] As already stressed, distinguishing between and evaluating the relative strength of interdependence and unit-level or contextual effects relies firstly upon the relative precision with which these alternative sources of spatial correlation are specified. Our simulations, on the other hand, simply set all non-zero elements (i.e., the off-block-diagonal diagonals) of the **W** specified for estimation to $(N\text{-}1)^{-1}$. I.e., 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, interdependence is truly homogenous, and the hypothetical researcher has correctly specified the estimation weighting-matrix to equal the true one exactly.[52]

We evaluate non-spatial OLS, and S-OLS, S-2SLS, and S-ML, the LS estimators with and without PCSE's (i.e., estimates of the variance-covariance matrix of the coefficient estimates that are

---

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

[49] 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 spatio-temporal interactions are important there.

[50] 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.

[51] Strategies for parameterizing and estimating **W** according to theoretical/substantive expectations would be of great interest but are, as yet, undeveloped. See also note 27.

[52] 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.

"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}, ρ={.1,..,.9}.[53] Each table reports 1000-trial experiments; the figures show parameter sweeps in 100-trial experiments. We report means of coefficient estimates and of LS, ML, standard errors and PCSE's; actual standard deviations of coefficient estimates; and root mean-squared errors (RMSE) of coefficient estimates. Comparing mean parameter estimates to their true values gives their (small-sample) biases. By comparing mean reported standard-errors to true standard deviations of coefficient estimates, we observe standard-error accuracy and how well PCSE's 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).

C. Simulation Results

We note first that noticeable estimation problems arise virtually exclusively in $\hat{\rho}$ and $\hat{\beta}_s$, the latter intuitively because omission or misestimation of $\hat{\rho}$ will induce biases primarily in the factor(s) 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 *ρ={.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, non-spatial 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 non-spatial 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

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

percentage (although not in absolute) terms.

TABLE 1: Average Coefficient Estimates Across 1000 Trials (BIAS)							
			Coeff.	**OLS**	**S-OLS**	**S-2SLS**	**S-ML**
ρ=.1	**N=5**	**T=20**	$\hat{\beta}_s$	1.112	1.027	***1.003***	1.048
			$\hat{\rho}$	*	.078	***.097***	.063
		T=40	$\hat{\beta}_s$	1.112	.991	***1.001***	1.021
			$\hat{\rho}$	*	.108	***.099***	.082
	N=40	**T=20**	$\hat{\beta}_s$	1.112	1.049	***.994***	1.050
			$\hat{\rho}$	*	.055	***.105***	.054
		T=40	$\hat{\beta}_s$	1.112	***.999***	1.003	1.026
			$\hat{\rho}$	*	***.101***	.098	.077
ρ=.5	**N=5**	**T=20**	$\hat{\beta}_s$	1.999	.837	***.998***	1.050
			$\hat{\rho}$	*	.579	***.499***	.475
		T=40	$\hat{\beta}_s$	2.001	.826	***1.000***	1.029
			$\hat{\rho}$	*	.587	***.500***	.487
	N=40	**T=20**	$\hat{\beta}_s$	2.004	.861	***1.008***	1.050
			$\hat{\rho}$	*	.570	***.497***	.474
		T=40	$\hat{\beta}_s$	2.000	.844	***1.002***	1.025
			$\hat{\rho}$	*	.578	***.499***	.487
Number of Unbiased Wins/Ties				**0**	**2**	**14**	**0**
#Times Noticeable (≥5%) Bias				**16**	**10**	**0**	**10**
NOTES: Unbiasedness Winner in ***Bold-Italics***; Notable or Appreciable Biases in Shaded							

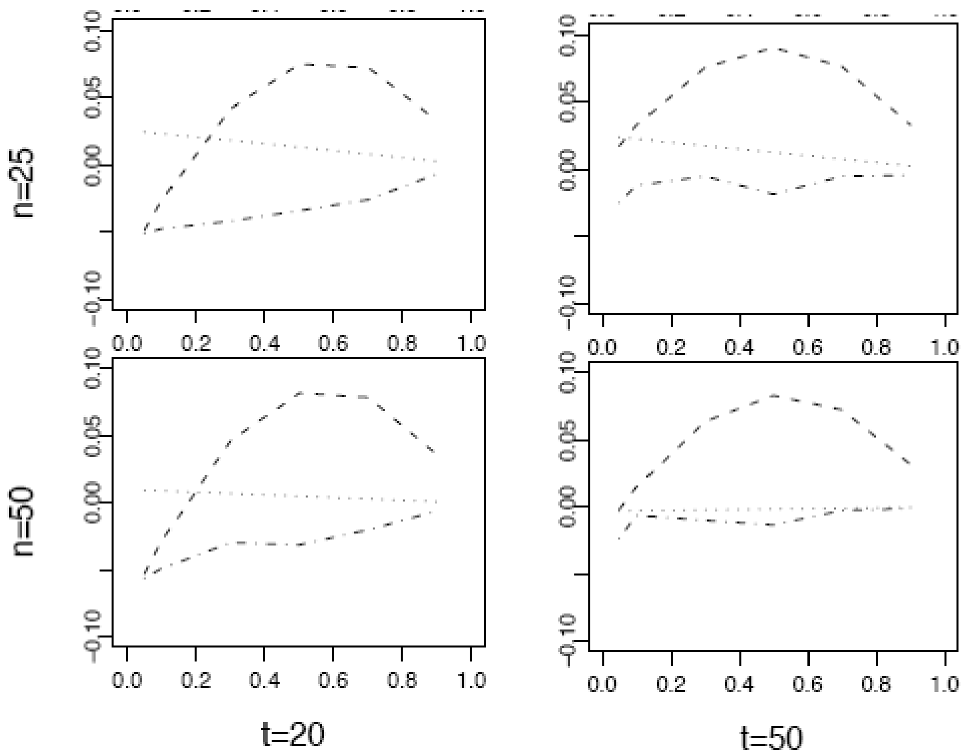
We also plot the percentage biases for each of the estimators across a range of interdependence-strengths, ρ={.1,..,.9}, in grids of graphs arrayed by sample dimensions—we show *N={25,50}*, *T={20,50}*—such as in Figures 1 for $\hat{\rho}$ and Figure 2 for $\hat{\beta}_s$. The *x*-axes in both figures are the true ρ. Figure 1's *y*-axis is the estimated bias in $\hat{\rho}$ (i.e., the average $\hat{\rho}$ minus the true ρ); Figure 2's is 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 Figure 1 and to one in Figure 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 (perceptibly larger) negative one; i.e., their convergence is from above and from below, respectively.[54] 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 ρ=0.1 value reported in Table 1 and ρ=0.2 or so. Over most of the range, $\hat{\rho}_{\text{S-OLS}}$ suffers positive (i.e., inflation) bias, as we would generally expect. Interestingly, the bias peaks in absolute terms at around the ρ=.5 value reported in of Table 1, but the magnitudes of these biases do not seem to depend in any intuitive way on sample-dimensions.[55] As seen in Figure 2, once again, the induced bias in $\hat{\beta}_s$ maintains opposite sign 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 non-spatial OLS "estimate" of ρ is, of course, -100%. (Plotted, it would be a $-45°$ line from 0 to -1.) The induced biases in the non-spatial 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 ρ=0, of course, but bias grows dramatically with ρ. In fact, the bias magnitude crosses 100% at ρ=.5, echoing the bottom of Table 1. Bias continues to rocket from there, formulaically by $\left(\frac{1}{1-\rho}-1\right)x100\%$, 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 incurs even moderately should now be abundantly clear.

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

[55] The proportionate bias, not shown, peaks at around ρ=.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).

<u>Figure 1</u>: Estimated Bias in $\hat{\rho}$ Plotted Against True ρ

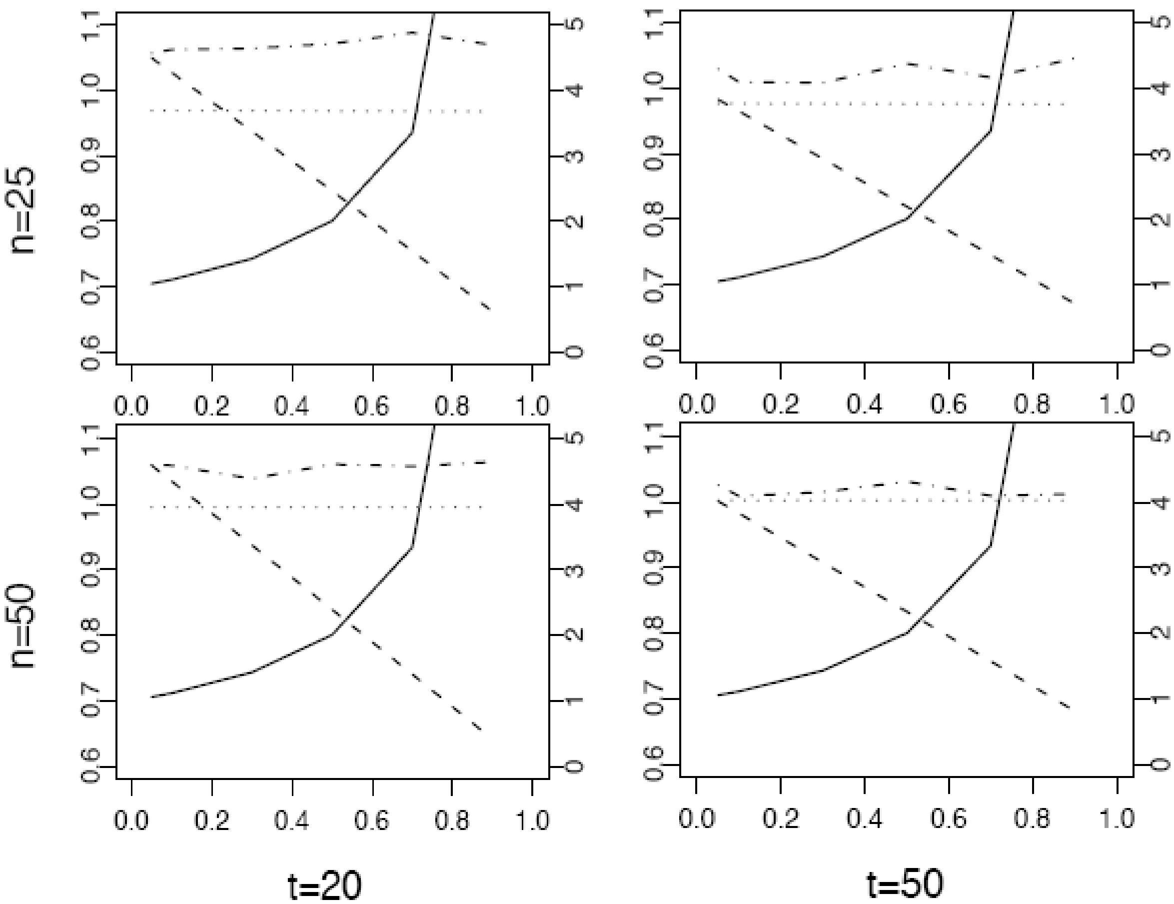
Across Representative N*x*T Sample-Dimensions



Notes: Dashed Line: S-OLS. Dotted Line: S-2SLS-IV. Dashed-Dotted Line: S-ML

<u>Figure 2</u>: Estimated $\hat{\beta}_S$ Plotted Against True ρ

Across Representative N*x*T Sample-Dimensions



Notes: Solid Line: Non-spatial OLS. Dashed Line: S-OLS. Dotted Line: S-2SLS-IV. Dashed-Dotted Line: S-ML. Non-Spatial OLS results plotted against the larger-scale 2nd y-axis on the right.

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. Root mean-squared-error (RMSE) adds the square of bias plus variance (then takes the square root), thus weighing those terms equally, to summarize these concerns. Table 2 reports the RMSE of each parameter estimate under the same 2*x*2*x*2 subset combinations of ρ, N, and T values as Table 1.

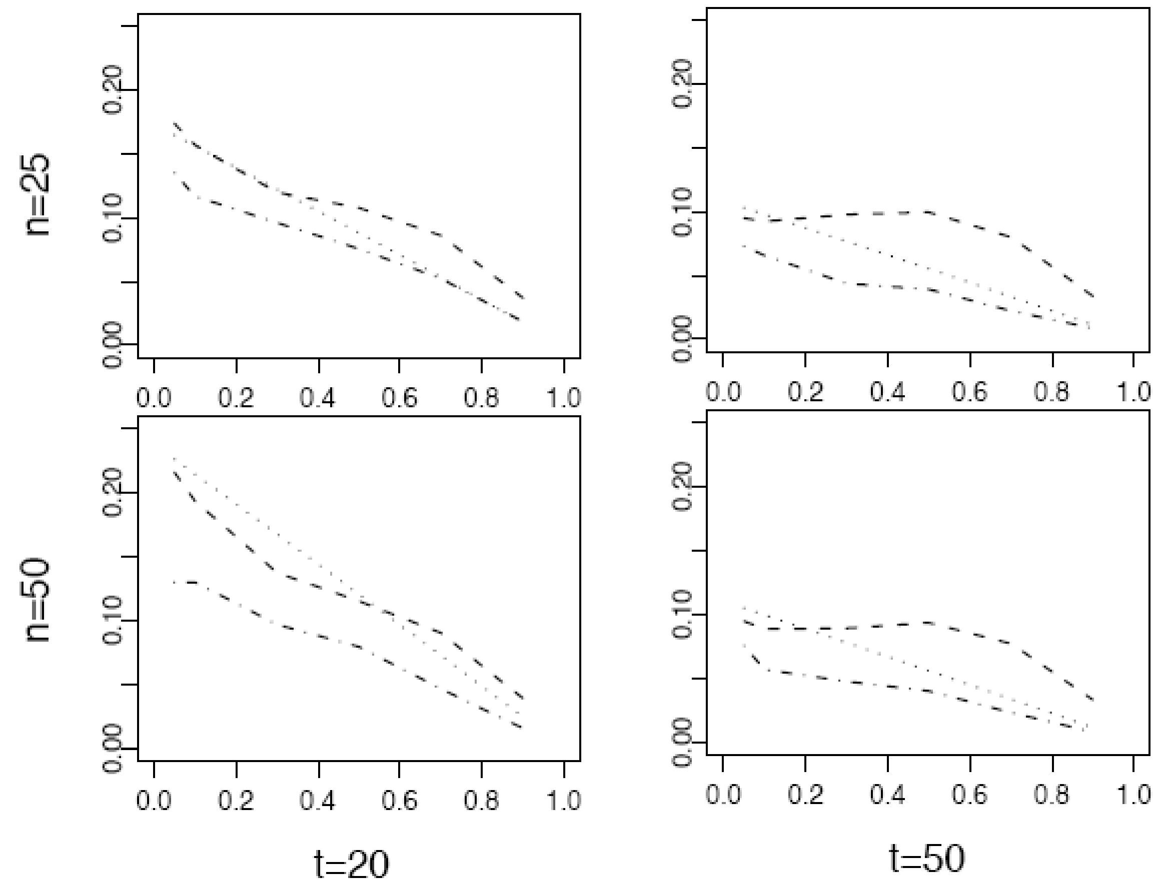
TABLE 2: Root Mean-Squared-Error (RMSE) Across 1000 Trials (EFFICIENCY & BIAS)

			Coeff.	**OLS**	**S-OLS**	**S-2SLS**	**S-ML**
ρ=.1	**N=5**	**T=20**	$\hat{\beta}_s$	.167	.216	.235	***.162***
			$\hat{\rho}$	*	.115	.177	***.106***
		T=40	$\hat{\beta}_s$	.139	.125	.139	***.107***
			$\hat{\rho}$	*	.093	.103	***.070***
	N=40	**T=20**	$\hat{\beta}_s$	.119	.215	.211	***.104***
			$\hat{\rho}$	*	.191	.188	***.096***
		T=40	$\hat{\beta}_s$	.116	.119	.136	***.062***
			$\hat{\rho}$	*	.105	.120	***.057***
ρ=.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**
# Clearly (50+%>) Dominated				**13**	**8**	**7**	**0**

NOTES: RMSE Winner in ***Bold-Italics***; Clearly Dominated RMSE's Shaded.

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. Non-spatial 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 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 Figure 3 we can easily see that (*i*) all three spatial estimators yield smaller RMSE of $\hat{\rho}$ as *N*, *T*, and/or true ρ increase, (*ii*) 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 (*iii*) 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 non-spatial OLS by any spatial estimator is underscored dramatically by the necessity of the larger-scale 2nd *y*-axis.

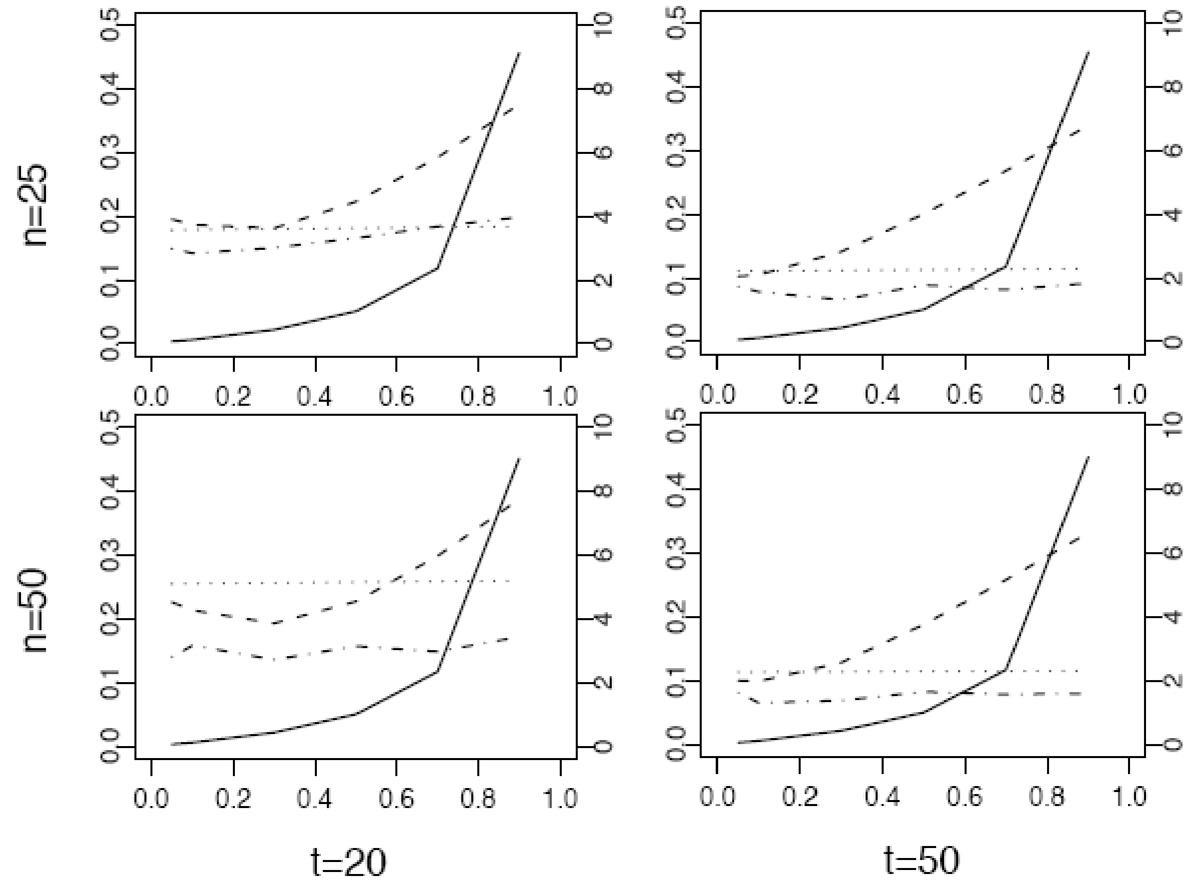
Figure 3: RMSE $\hat{\rho}$ Plotted Against True ρ Across Representative N*x*T Sample-Dimensions (BIAS+EFFICIENCY)



Notes: Dashed Line: S-OLS. Dotted Line: S-2SLS-IV. Dashed-Dotted Line: S-ML.

<u>Figure 4</u>: RMSE $\hat{\beta}_S$ Plotted Against True ρ Across Representative

N*x*T Sample-Dimensions (BIAS+EFFICIENCY)



Notes: Solid line: Non-spatial OLS. Dashed Line: S-OLS. Dotted Line: S-2SLS-IV. Dashed-Dotted Line: S-ML. Non-spatial OLS results plotted against the larger-scale 2nd y-axis on the right.

A third property of interest to empirical researchers is the accuracy of the estimators' reported standard errors 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 standard-error (or PCSE) across the 1000 trials to the actual standard deviation of the estimated parameters across those trials. Standard-error-accuracy ratios equal to (less than, greater than) one imply honest reporting (understatement/overconfidence, overstatement) on average. For the most part, non-spatial OLS reports inaccurate standard errors for $\hat{\beta}_s$ except at small ρ. That is, at stronger interdependence, not only do omitted-variable biases favor overstating non-spatial factors' importance, but inaccurate standard-errors foster egregious overconfidence (64±% in one case) in that erroneous conclusion! PCSE only partially ameliorates this aspect of non-spatial 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 non-spatial 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 $\hat{\beta}_s$ standard errors. Here, PCSE near-uniformly *worsens* matters, yielding 12-32% overconfidence. Intuitively, PCSE improves OLS standard-error estimates because the pattern of unmodeled spatial correlation there will correlate with **s**'s parts of the **X'X** matrix, which is precisely that to which PCSE is consistent. It worsens S-OLS standard-error estimation because simultaneity bias in $\hat{\rho}$ is all that induces the spatial-correlation that remains in the **ee'** matrix. This pattern will also correlate with **Wy**'s and **s**'s parts of the **X'X** matrix, thereby leading PCSE to "correct" unnecessarily. S-2SLS produces standard-error estimates within 11% of true estimation variation except in the strong-interdependence, N=40*x*T=20 sample-dimension experiment, where reported standard errors are 22-23% overconfident. S-ML consistently produces standard-error estimates within 11% of true estimation variation without exception, and so emerges as nearly (weakly) dominant on this desiderata also.

TABLE 3: Mean Reported Standard Error Divided by Standard Deviation Across 1000 Trials (STANDARD-ERROR ACCURACY)

			Coeff.	OLS		S-OLS		S-2SLS	S-ML
				s.e.	PCSE	s.e.	PCSE		
ρ=.1	N=5	T=20	$\hat{\beta}_S$	0.871	0.895	0.814	0.758	0.901	***0.938***
			$\hat{\rho}$	*	*	0.773	**0.727**	0.893	***0.943***
		T=40	$\hat{\beta}_S$	0.914	***0.975***	0.928	0.880	0.971	0.965
			$\hat{\rho}$	*	*	0.859	0.826	***0.971***	0.955
	N=40	T=20	$\hat{\beta}_S$	0.927	***0.951***	**0.727**	**0.689**	0.943	0.911
			$\hat{\rho}$	*	*	**0.715**	**0.677**	***0.931***	0.890
		T=40	$\hat{\beta}_S$	0.867	0.933	0.849	0.807	0.926	***0.956***
			$\hat{\rho}$	*	*	0.829	0.790	0.925	**0.942**
ρ=.5	N=5	T=20	$\hat{\beta}_S$	**0.491**	0.799	***1.034***	0.831	0.901	0.936
			$\hat{\rho}$	*	*	***0.961***	0.776	0.907	0.916
		T=40	$\hat{\beta}_S$	**0.485**	0.822	1.017	0.857	***0.993***	.979
			$\hat{\rho}$	*	*	0.980	0.837	1.016	***1.010***
	N=40	T=20	$\hat{\beta}_S$	**0.364**	0.782	0.933	**0.745**	0.770	***0.938***
			$\hat{\rho}$	*	*	0.914	**0.728**	0.779	***0.936***
		T=40	$\hat{\beta}_S$	**0.370**	0.836	***1.010***	0.840	0.941	0.924
			$\hat{\rho}$	*	*	***1.000***	0.837	0.954	0.914
# 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**

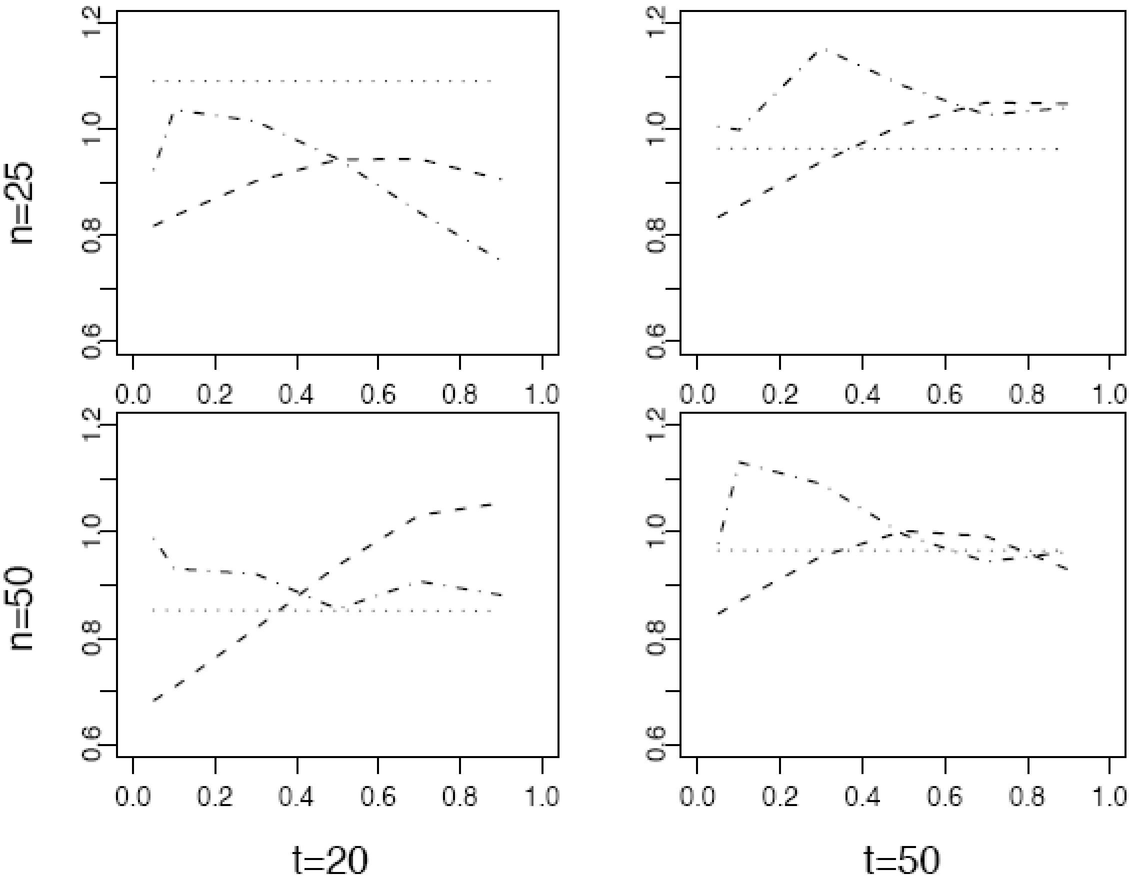
NOTES: SE-Accuracy Winner in ***Bold-Italics***; Notable (>12.5%) Inaccuracies Shaded. Massive (>25%) Inaccuracies **Shaded & Bold.**

Figures 5 and 6 plot standard-error-accuracy ratios in the now-familiar grid of broader experimental conditions, generally confirming the conclusions drawn from the (different) tabulated subset. Non-spatial OLS is not only greatly and increasingly biased at moderate through to strong interdependence but wildly overconfident regarding $\hat{\beta}_s$ over that range also. S-OLS offers acceptably honest standard-error 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 standard-errors 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.[56] Indeed, reminiscent of Beck and Katz's (1995, 1996) critique of Parks-Kmenta, 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 standard-error accuracy. S-2SLS and S-ML thus emerge as the clear winners on standard-error accuracy.

[56] LeSage's S-ML MatLab code uses **ee'** in the variance-covariance formula whereas the maintained assumptions, most directly the sphericity of V(**ε**), would allow the tighter **e'e**/n. Using **ee'** is roughly equivalent to calculating PCSE's for the ML estimates. Our tables impose **e'e**/n; our figures use LeSage's original **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 standard-errors performance there.

Figure 5: Standard-Error $\hat{\rho}$ Accuracy Plotted Against True ρ

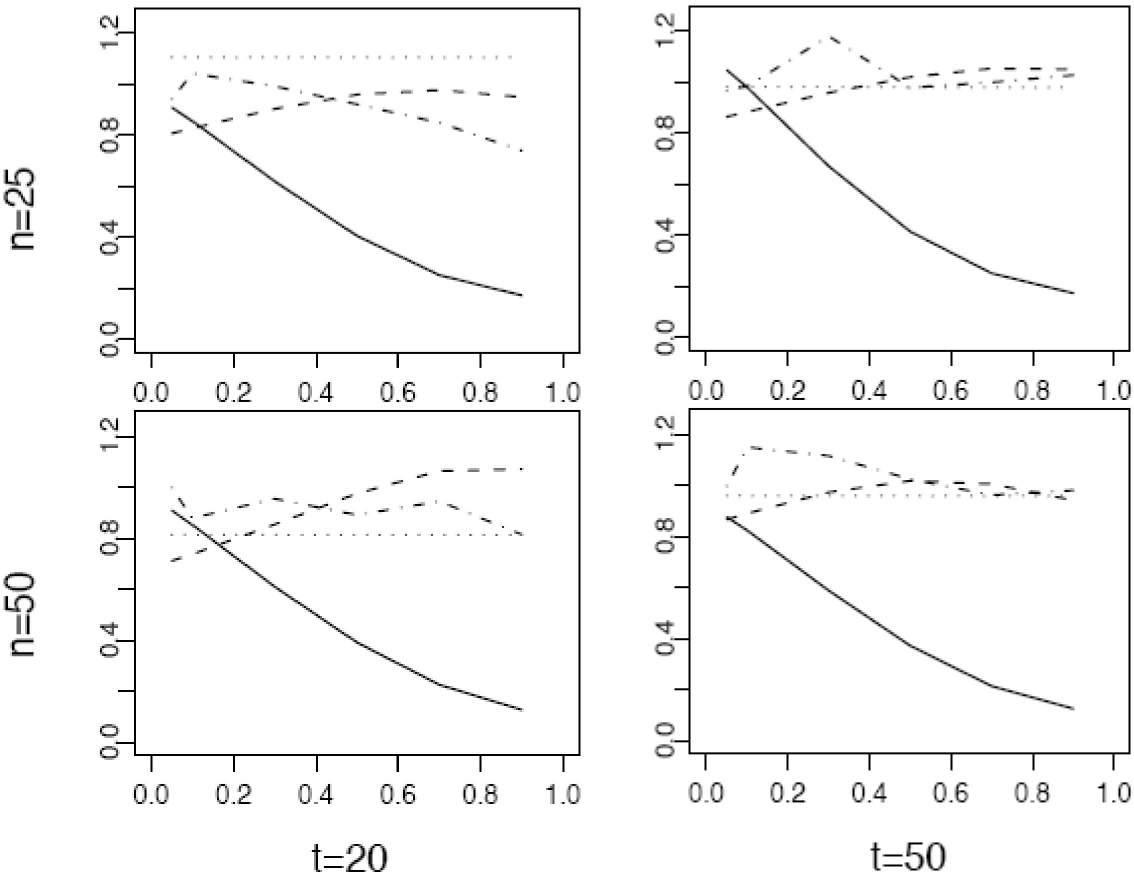
Across Representative N*x*T Sample-Dimensions



Notes: Dashed Line: S-OLS. Dotted line: S-2SLS-IV. Dashed-Dotted Line: S-ML. Standard error accuracy is gauged by the ratio of the average estimated standard error to the true standard deviation of the sampling distribution. Values less than one indicate overconfidence.

Figure 6: Standard-Error $\hat{\beta}_S$ Accuracy Plotted Against True ρ

Across Representative N*x*T Sample-Dimensions



Notes: Solid Line: Non-spatial OLS. Dashed Line: S-OLS. Dotted line: S-2SLS-IV. Dashed-Dotted Line: S-ML. Standard error accuracy is gauged by the ratio of the average estimated standard error to the sampling-distribution standard deviation. Values less than one indicate overconfidence.

D. Conclusions from Analytic and Simulation Results

In summary, we have shown analytically that non-spatial OLS ignores spatial interdependence and suffers omitted-variable biases fostering overestimation of non-spatial 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 (ρ>.1±) and gargantuan at greater ρ. The biases concentrate in common shocks for homogenous diffusion processes. Standard errors for these overestimated effects are also dramatically underestimated in these ranges, and PCSE offers only limited amelioration. In short, given any noticeable interdependence, non-spatial 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 non-spatial factors' roles. However, our simulations demonstrated that these simultaneity biases generally remain mild through a small-to-moderate interdependence-range (ρ<.3±). S-OLS is also rather efficient, so only its problems with standard-error 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, are available to handle other contexts.[57] Our simulations show S-2SLS to be admirably unbiased in both parameter and standard-error estimation,[58] especially as sample sizes increase, but efficiency can be an issue.[59] S-ML, for its part, weakly dominates or nearly 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 standard-error 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[60]—common conditions, we suspect—favor adequacy of simpler LS over IV or ML spatial estimators. Conversely, when interdependence is stronger, a consistent estimators should be chosen over S-OLS, whose simultaneity bias grows. S-ML has efficiency advantages over S-2SLS, although these come at some non-negligible computational costs. The downward small-sample bias of S-ML vs. S-2SLS's upward one, plus their generally accurately reported standard errors, 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, non-spatial OLS is absolutely disastrous and to be avoided unless interdependence is known to be very weak or non-existent.

b) Interdependence proportional to some economic/strategic

---

[57] 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 `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 standard errors of the spatial-lag coefficient*. One might also want to tighten the standard-error formula as suggested in note 56.

[58] Whether PCSE's might further improve or worsen standard-error estimation awaits further research.

[59] GMM extensions of S-2SLS (see, e.g., Kelejian 1993) might improve efficiency.

[60] Franzese and Hays (2004) show imperfect instruments add inconsistency to inefficiency concerns.

connection; w_{ij} drawn from actual datasets in lit: trade/capital flows, trade/capital mobility,

c) Strategic complements/substitutes (cooperation/competition) relations between units; w_{ij} drawn from uniform-distribution {-1…1}.

d) Unit interdependence due to "co-membership" (borders, organizations, culturo-linguistic-religious groups, dyads, etc.); w_{ij} drawn from actual datasets in literature. Considerations in exploring this parameter space:

(1) Overlapping v. non-overlapping group-memberships.

(2) Complete group membership (no isolates)

(3) Size & number of these groups relative to totals

e) Unit interdependence proportional to geographic or Euclidean distance b/w units (drawn from actual datasets in the literature)

6. Simulation Results: Trade, Borders, and Complements/Substitutes Connectivities

As Tables 11-22 reveal, our findings evolve and begin to clarify notably when we consider three other spatial-connectivity matrices one might see in substantive applications: trade interconnections (Tables 11-14), borders or similar group comemberships (Tables 15-18), and complement/substitute relations (Tables 19-22). In Tables 11-14, we evaluate the performance of our four estimators (OLS, S-OLS, S-2SLS-IV, and S-ML) in the same model **(5)**, and over the same sample dimensions (Table 11, 13: N=5; Tables 12, 14: N=40) and overall interdependence-strengths (Table 11-12: ρ=.1; Tables 13-14: ρ=.5), but for a connectivity matrix, **W**, with w_{ij} given by $(X_{i\to j}+M_{j\to i})/GDP_i$. I.e., relative $j\to i$ connectivity is assumed proportional to the sum of i's exports to j and i's imports from j as shares of i's GDP. This would be one common way of specifying in **W** interdependence of i on j theoretically/substantively expected to derive from economic ties between the two units. To further enhance the match of our experimental **W** to those actually used in such substantive studies, we draw our w_{ij} from actual data compiled for this purpose, and generously shared publicly, by Kristian Skrede Gleditsch (see note **Error! Bookmark not defined.**). The results are striking.

Table 11. Comparing Estimators (N=5, ρ = 0.1, w_{ij} = Export$_{ij}$ + Import$_{ij}$, 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	0.448	0.118	0.562	0.723	0.124	0.294	0.941	0.184	0.166	1.000	0.123	0.103
	$s.e.(\beta_1)$	0.105	0.015		0.100	0.011		0.155	0.057		0.117	0.015	
	$pcse(\beta_1)$												
	β_2	0.436	0.109	0.574	0.719	0.121	0.298	0.937	0.180	0.167	1.003	0.123	0.109
	$s.e.(\beta_2)$	0.106	0.016		0.101	0.016		0.155	0.050		0.120	0.022	
	$pcse(\beta_2)$												
	β_3	0.450	0.164	0.561	0.725	0.142	0.294	0.943	0.183	0.167	1.004	0.126	0.110
	$s.e.(\beta_3)$	0.110	0.019		0.104	0.019		0.157	0.057		0.123	0.026	
	$pcse(\beta_3)$												
	ρ				0.051	0.011	0.050	0.090	0.027	0.024	0.101	0.011	0.046
	$s.e.(\rho)$				0.008	0.001		0.022	0.010		0.010	0.002	
	$pcse(\rho)$												
T=40	β_1	0.458	0.086	0.547	0.736	0.088	0.273	0.980	0.133	0.110	1.003	0.083	0.072
	$s.e.(\beta_1)$	0.074	0.007		0.069	0.005		0.109	0.029		0.081	0.007	
	$pcse(\beta_1)$												
	β_2	0.432	0.079	0.573	0.723	0.088	0.286	0.973	0.133	0.114	1.002	0.087	0.074
	$s.e.(\beta_2)$	0.074	0.008		0.070	0.007		0.110	0.026		0.083	0.010	
	$pcse(\beta_2)$												
	β_3	0.457	0.119	0.548	0.734	0.100	0.275	0.978	0.129	0.111	1.002	0.087	0.074
	$s.e.(\beta_3)$	0.075	0.009		0.071	0.009		0.109	0.030		0.083	0.012	
	$pcse(\beta_3)$												
	ρ				0.051	0.008	0.049	0.096	0.020	0.017	0.101	0.008	0.032
	$s.e.(\rho)$				0.006	0.000		0.016	0.005		0.007	0.001	
	$pcse(\rho)$												

Table 12. Comparing Estimators (N=40, $\rho = 0.1$, w_{ij} = Export$_{ij}$ + Import$_{ij}$, 1,000 trials)

		OLS			**S-OLS**			**S-2SLS**			**ML**		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	0.432	0.194	0.598	0.987	0.036	0.039	0.992	0.082	0.075	0.998	0.035	0.036
	$s.e.(\beta_1)$	0.186	0.035		0.036	0.002		0.075	0.047		0.036	0.002	
	$pcse(\beta_1)$												
	β_2	0.951	0.248	0.193	0.996	0.039	0.037	0.995	0.068	0.050	0.998	0.039	0.038
	$s.e.(\beta_2)$	0.187	0.034		0.037	0.006		0.050	0.065		0.038	0.007	
T=20	$pcse(\beta_2)$												
	β_3	0.415	0.284	0.616	0.985	0.041	0.041	0.991	0.097	0.081	0.998	0.039	0.038
	$s.e.(\beta_3)$	0.192	0.036		0.038	0.007		0.081	0.060		0.038	0.007	
	$pcse(\beta_3)$												
	ρ				0.097	0.001	0.003	0.099	0.013	0.011	0.099	0.001	0.005
	$s.e.(\rho)$				0.001	0.000		0.011	0.008		0.001	0.000	
	$pcse(\rho)$												
	β_1	0.425	0.141	0.590	0.986	0.026	0.029	0.995	0.051	0.049	0.995	0.025	0.025
	$s.e.(\beta_1)$	0.132	0.018		0.025	0.001		0.049	0.013		0.025	0.001	
	$pcse(\beta_1)$												
	β_2	0.951	0.179	0.141	0.998	0.027	0.025	0.998	0.030	0.029	1.000	0.027	0.026
	$s.e.(\beta_2)$	0.132	0.017		0.025	0.003		0.029	0.008		0.026	0.003	
T=40	$pcse(\beta_2)$												
	β_3	0.428	0.203	0.588	0.986	0.027	0.029	0.996	0.054	0.050	0.996	0.026	0.026
	$s.e.(\beta_3)$	0.134	0.017		0.026	0.003		0.050	0.016		0.026	0.003	
	$pcse(\beta_3)$												
	ρ				0.098	0.001	0.002	0.099	0.007	0.007	0.099	0.001	0.003
	$s.e.(\rho)$				0.000	0.000		0.007	0.002		0.000	0.000	
	$pcse(\rho)$												

Table 13. Comparing Estimators (N=5, ρ = 0.5, w_{ij} = Export$_{ij}$ + Import$_{ij}$, 1,000 trials)

		OLS			**S-OLS**			**S-2SLS**			**ML**		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	0.406	0.120	0.604	0.711	0.125	0.306	0.924	0.234	0.190	0.997	0.123	0.104
	$s.e.(\beta_1)$	0.107	0.015		0.099	0.011		0.174	0.096		0.117	0.015	
	$pcse(\beta_1)$												
	β_2	0.399	0.112	0.611	0.710	0.123	0.307	0.920	0.223	0.191	1.001	0.122	0.109
	$s.e.(\beta_2)$	0.108	0.016		0.100	0.015		0.173	0.084		0.121	0.022	
T=20	$pcse(\beta_2)$												
	β_3	0.409	0.169	0.602	0.713	0.145	0.305	0.927	0.228	0.191	1.001	0.125	0.111
	$s.e.(\beta_3)$	0.112	0.020		0.104	0.019		0.176	0.092		0.123	0.026	
	$pcse(\beta_3)$												
	ρ				0.261	0.054	0.243	0.441	0.164	0.136	0.504	0.051	0.149
	$s.e.(\rho)$				0.039	0.005		0.123	0.072		0.048	0.007	
	$pcse(\rho)$												
	β_1	0.417	0.089	0.588	0.724	0.090	0.284	0.973	0.153	0.125	1.001	0.081	0.072
	$s.e.(\beta_1)$	0.075	0.007		0.069	0.005		0.122	0.043		0.082	0.007	
	$pcse(\beta_1)$												
	β_2	0.394	0.080	0.611	0.714	0.089	0.295	0.966	0.153	0.129	1.000	0.086	0.075
	$s.e.(\beta_2)$	0.075	0.008		0.070	0.007		0.124	0.040		0.083	0.010	
T=40	$pcse(\beta_2)$												
	β_3	0.415	0.123	0.590	0.722	0.102	0.287	0.970	0.147	0.126	1.000	0.087	0.074
	$s.e.(\beta_3)$	0.077	0.009		0.071	0.009		0.123	0.042		0.084	0.012	
	$pcse(\beta_3)$												
	ρ				0.264	0.039	0.237	0.475	0.110	0.092	0.501	0.035	0.029
	$s.e.(\rho)$				0.027	0.002		0.089	0.035		0.033	0.004	
	$pcse(\rho)$												

Table 14. Comparing Estimators (N=40, $\rho = 0.5$, w_{ij} = Export$_{ij}$ + Import$_{ij}$, 1,000 trials)

		OLS			**S-OLS**			**S-2SLS**			**ML**		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	-0.544	1.470	2.080	1.000	0.035	0.036	0.980	0.178	0.118	1.001	0.035	0.037
	$s.e.(\beta_1)$	1.393	0.304		0.036	0.002		0.117	0.112		0.036	0.002	
	$pcse(\beta_1)$												
	β_2	1.941	0.981	1.690	0.999	0.038	0.037	0.995	0.085	0.070	0.999	0.039	0.038
	$s.e.(\beta_2)$	1.404	0.303		0.037	0.006		0.070	0.051		0.038	0.007	
	$pcse(\beta_2)$												
	β_3	-0.534	1.987	2.106	1.000	0.039	0.038	0.979	0.170	0.125	1.000	0.039	0.038
	$s.e.(\beta_3)$	1.443	0.313		0.038	0.007		0.123	0.115		0.038	0.007	
	$pcse(\beta_3)$												
	ρ				0.499	0.001	0.001	0.495	0.045	0.034	0.500	0.001	0.002
	$s.e.(\rho)$				0.000	0.000		0.034	0.033		0.001	0.000	
	$pcse(\rho)$												
T=40	β_1	-0.592	0.964	1.872	0.997	0.025	0.026	0.993	0.086		0.998	0.025	0.025
	$s.e.(\beta_1)$	0.984	0.149		0.025	0.001		0.074	0.062		0.025	0.001	
	$pcse(\beta_1)$												
	β_2	1.933	0.642	1.356	1.001	0.026	0.026	1.000	0.048	0.044	1.001	0.027	0.026
	$s.e.(\beta_2)$	0.984	0.149		0.026	0.003		0.044	0.034		0.026	0.003	
	$pcse(\beta_2)$												
	β_3	-0.549	1.435	1.843	0.998	0.026	0.026	0.993	0.090	0.076	0.999	0.026	0.026
	$s.e.(\beta_3)$	1.000	0.153		0.026	0.003		0.076	0.064		0.026	0.003	
	$pcse(\beta_3)$												
	ρ				0.499	0.000	0.001	0.499	0.023	0.021	0.500	0.000	0.001
	$s.e.(\rho)$				0.000	0.000		0.021	0.020		0.000	0.000	
	$pcse(\rho)$												

First, the biases in non-spatial OLS are dramatic, with only two of the non-spatial coefficient-estimates being within 5% of their true values (β_S, remarkably, for the N=40, ρ=.1 case). In the rest of the cases, coefficient-estimates are biased by more than 50%; in one case (N=40, ρ=.5), $\hat{\beta}_S$ is almost 100% too large and the other two coefficient-estimates over 150% too small (reversing sign). Furthermore, these huge biases no longer concentrate in the exogenous-shock parameter-estimates, β_S, but distribute more equally across all three estimates, reflecting the fact that the common-shocks, **S**, no longer represent the decidedly closer available correlate for OLS to the erroneously omitted spatial lag. The unit-time uniquely variant domestic factors and context-conditional aspects, despite being randomly drawn, correlate in limited samples at least as well with that omitted variable as the only time-variant shocks. The reported standard errors are also frequently grossly overconfident, and every non-spatial parameter estimate under every experimental condition has greater RMSE than that of any of the spatial estimators, at minimum by a factor of about two. If non-spatial estimators were a bad idea given homogenous interdependence processes, they are a horrible idea given interdependence patterns that describe economic interconnections.

S-OLS improves matters appreciably, reducing biases in the non-spatial coefficient-estimates by about 50% in the smaller-N samples, and likewise estimating interdependence-strength at least 50% of its true magnitude. RMSE's are likewise cut by about half or more, but standard errors are regularly over-confident by about 20% or more. However, the most remarkable feature of S-OLS performance under these conditions is how well it does as the number of units, N, increases. At N=40, regardless of T or ρ, simple S-OLS estimates every parameter within 1% of its true value, reports standard errors accurate to within 2-3%, with RMSE's nearly as low, and as-often lower, as any of the alternative estimators! Substantive C&IPE scholars must take strong caution against employing non-spatial models from the preceding paragraph, but can take almost as strong solace from this one: as sample-size increases in the cross-sectional dimension, simple S-OLS seems to become a dominant strategy, at least under these experimental conditions (n.b., no specification-error in **W**, *inter alia*). S-OLS is not such a great strategy, however, if N remains small.

Meanwhile, S-2SLS-IV offers nearly unbiased parameter estimates and reasonably accurate standard errors under either the smaller-N or the larger-N case, regardless of interdependence strength, although non-negligible overconfidence does arise in the smaller-N cases. However, S-ML dominates the IV estimator in this case on all fronts: bias, standard-error-accuracy, and RMSE, for all coefficients. Under these conditions, in fact, S-ML basically "nails it" and, essentially, at least weakly dominates any of the alternatives, under any sample or model conditions, for all parameter estimates. The conclusion, then, for this kind of trade-connectivity matrix, and with perfect specification of that single **W**, is: S-ML weakly dominates, although the much simpler S-OLS does as well in larger-N samples (in fact, S-OLS just beats S-ML on 1 or 2 bases for 1 or 2 parameters in such samples) and so might be preferred there. S-2SLS-IV also generally performs acceptably in any of these N- or ρ-size conditions, but its inefficiency relative to S-ML and, in larger-N samples, S-OLS alternatives reduce its appeal unless S-ML is unavailable (e.g., >2 interdependence-processes operating). The crucial point, though, remains that non-spatial OLS does horribly.

Table 15. Comparing Estimators (N=5, ρ = 0.1, w_{ij} = 1 if MinD<=950km, otherwise w_{ij}=0, 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	1.014	0.109	0.110	0.994	0.107	0.106	0.996	0.107	0.107	0.998	0.106	0.104
	$s.e.(\beta_1)$	0.109	0.012		0.106	0.012		0.107	0.012		0.104	0.012	
	$pcse(\beta_1)$												
	β_2	1.140	0.126	0.178	0.959	0.139	0.137	1.003	0.146	0.142	1.008	0.131	0.125
	$s.e.(\beta_2)$	0.110	0.020		0.130	0.021		0.142	0.027		0.125	0.020	
	$pcse(\beta_2)$												
	β_3	1.013	0.121	0.117	0.996	0.117	0.113	0.998	0.116	0.114	1.000	0.116	0.111
	$s.e.(\beta_3)$	0.116	0.025		0.113	0.025		0.114	0.025		0.111	0.024	
	$pcse(\beta_3)$												
	ρ				0.130	0.064	0.061	0.099	0.068	0.067	0.094	0.047	0.045
	$s.e.(\rho)$				0.053	0.007		0.067	0.015		0.045	0.004	
	$pcse(\rho)$												
T=40	β_1	1.012	0.079	0.076	0.996	0.077	0.072	0.999	0.077	0.073	1.000	0.077	0.903
	$s.e.(\beta_1)$	0.075	0.006		0.072	0.006		0.073	0.006		0.072	0.006	
	$pcse(\beta_1)$												
	β_2	1.138	0.084	0.157	0.957	0.091	0.098	0.999	0.096	0.096	1.003	0.085	0.907
	$s.e.(\beta_2)$	0.075	0.009		0.088	0.010		0.096	0.012		0.085	0.009	
	$pcse(\beta_2)$												
	β_3	1.013	0.078	0.078	0.996	0.077	0.075	0.999	0.076	0.075	1.000	0.076	0.903
	$s.e.(\beta_3)$	0.077	0.011		0.075	0.011		0.075	0.011		0.074	0.011	
	$pcse(\beta_3)$												
	ρ				0.131	0.042	0.048	0.099	0.046	0.045	0.097	0.031	0.031
	$s.e.(\rho)$				0.036	0.004		0.045	0.007		0.031	0.002	
	$pcse(\rho)$												

Table 16. Comparing Estimators (N=40, ρ = 0.1, w_{ij} = 1 if MinD<=950km, otherwise w_{ij}=0, 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	1.092	0.091	0.121	0.999	0.035	0.036	1.002	0.036	0.037	1.001	0.035	0.036
	$s.e.(\beta_1)$	0.078	0.010		0.036	0.002		0.037	0.002		0.036	0.002	
	$pcse(\beta_1)$												
	β_2	-0.491	0.131	1.493	1.038	0.051	0.061	1.000	0.080	0.080	1.019	0.046	0.047
	$s.e.(\beta_2)$	0.079	0.013		0.047	0.007		0.080	0.016		0.044	0.006	
T=20	$pcse(\beta_2)$												
	β_3	1.092	0.115	0.123	0.999	0.039	0.038	1.001	0.040	0.039	1.000	0.039	0.038
	$s.e.(\beta_3)$	0.081	0.014		0.038	0.007		0.039	0.007		0.038	0.007	
	$pcse(\beta_3)$												
	ρ				0.103	0.002	0.003	0.100	0.005	0.005	0.101	0.001	0.002
	$s.e.(\rho)$				0.002	0.000		0.005	0.001		0.001	0.000	
	$pcse(\rho)$												
	β_1	1.087	0.066	0.103	0.997	0.025	0.026	0.999	0.026	0.026	0.998	0.025	0.025
	$s.e.(\beta_1)$	0.055	0.005		0.025	0.001		0.026	0.001		0.025	0.001	
	$pcse(\beta_1)$												
	β_2	-0.489	0.091	1.490	1.038	0.036	0.050	0.997	0.057	0.057	1.020	0.031	0.036
	$s.e.(\beta_2)$	0.055	0.006		0.033	0.003		0.057	0.009		0.030	0.003	
T=40	$pcse(\beta_2)$												
	β_3	1.091	0.088	0.107	0.998	0.026	0.026	1.000	0.026	0.026	0.999	0.026	0.026
	$s.e.(\beta_3)$	0.056	0.007		0.026	0.003		0.026	0.003		0.026	0.003	
	$pcse(\beta_3)$												
	ρ				0.103	0.001	0.003	0.100	0.003	0.003	0.101	0.001	0.002
	$s.e.(\rho)$				0.001	0.000		0.003	0.001		0.001	0.000	
	$pcse(\rho)$												

Table 17. Comparing Estimators (N=5, ρ = 0.5, w_{ij} = 1 if MinD<=950km, otherwise w_{ij}=0, 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	1.530	0.293	0.590	0.933	0.110	0.125	0.996	0.111	0.109	1.004	0.109	0.107
	$s.e.(\beta_1)$	0.259	0.037		0.105	0.012		0.109	0.013		0.107	0.012	
	$pcse(\beta_1)$												
	β_2	2.806	0.439	1.824	0.766	0.134	0.271	0.994	0.173	0.164	1.019	0.144	0.138
	$s.e.(\beta_2)$	0.260	0.046		0.137	0.022		0.164	0.034		0.137	0.022	
T=20	$pcse(\beta_2)$												
	β_3	1.520	0.368	0.588	0.936	0.121	0.129	0.997	0.121	0.115	1.005	0.119	0.114
	$s.e.(\beta_3)$	0.273	0.056		0.112	0.024		0.115	0.025		0.114	0.025	
	$pcse(\beta_3)$												
	ρ				0.566	0.028	0.070	0.501	0.038	0.035	0.495	0.025	0.024
	$s.e.(\rho)$				0.025	0.004		0.035	0.008		0.023	0.003	
	$pcse(\rho)$												
	β_1	1.529	0.211	0.559	0.934	0.080	0.098	0.998	0.081	0.074	1.002	0.079	0.074
	$s.e.(\beta_1)$	0.180	0.018		0.072	0.006		0.074	0.006		0.074	0.006	
	$pcse(\beta_1)$												
	β_2	2.802	0.289	1.811	0.769	0.091	0.249	0.995	0.115	0.110	1.007	0.095	0.093
	$s.e.(\beta_2)$	0.180	0.021		0.093	0.010		0.110	0.015		0.093	0.010	
T=40	$pcse(\beta_2)$												
	β_3	1.519	0.249	0.551	0.936	0.081	0.098	0.999	0.080	0.076	1.002	0.078	0.076
	$s.e.(\beta_3)$	0.185	0.025		0.074	0.011		0.076	0.011		0.076	0.011	
	$pcse(\beta_3)$												
	ρ				0.564	0.019	0.066	0.501	0.025	0.024	0.498	0.017	0.016
	$s.e.(\rho)$				0.017	0.002		0.024	0.004		0.016	0.002	
	$pcse(\rho)$												

Table 18. Comparing Estimators (N=40, ρ = 0.5, w_{ij} = 1 if MinD<=950km, otherwise w_{ij}=0, 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	-29.665	16.438	33.345	0.995	0.036	0.037	1.001	0.043	0.043	0.995	0.036	0.005
	$s.e.(\beta_1)$	13.098	2.858		0.036	0.002		0.043	0.008		0.002	0.002	
	$pcse(\beta_1)$												
	β_2	-23.754	13.218	28.051	0.993	0.038	0.038	0.997	0.042	0.040	0.993	0.039	0.039
	$s.e.(\beta_2)$	13.194	2.789		0.037	0.006		0.040	0.007		0.038	0.007	
T=20	$pcse(\beta_2)$												
	β_3	-29.599	21.934	33.472	0.995	0.039	0.039	1.001	0.046	0.044	0.995	0.039	0.038
	$s.e.(\beta_3)$	13.567	2.909		0.038	0.007		0.044	0.011		0.038	0.007	
	$pcse(\beta_3)$												
	ρ				0.500	0.000	0.000	0.500	0.000	0.000	0.500	0.000	0.000
	$s.e.(\rho)$				0.000	0.000		0.000	0.000		0.000	0.000	
	$pcse(\rho)$												
	β_1	-29.499	11.758	31.875	0.992	0.025	0.027	0.999	0.030	0.030	0.992	0.025	0.026
	$s.e.(\beta_1)$	9.266	1.417		0.025	0.001		0.030	0.005		0.025	0.001	
	$pcse(\beta_1)$												
	β_2	-23.825	9.813	26.501	0.995	0.026	0.026	1.000	0.028	0.028	0.995	0.027	0.026
	$s.e.(\beta_2)$	9.276	1.473		0.025	0.003		0.028	0.003		0.026	0.003	
T=40	$pcse(\beta_2)$												
	β_3	-30.342	16.299	32.726	0.993	0.026	0.008	1.000	0.030	0.030	0.993	0.026	0.027
	$s.e.(\beta_3)$	9.417	1.499		0.003	0.003		0.030	0.006		0.026	0.003	
	$pcse(\beta_3)$												
	ρ				0.500	0.000	0.000	0.500	0.000	0.000	0.500	0.000	0.000
	$s.e.(\rho)$				0.000	0.000		0.000	0.000		0.000	0.000	
	$pcse(\rho)$												

The situation changes in some striking respects when we turn to a geographic-proximity kind of connectivity matrix, drawn from an actual **W** used in substantive research (once again, provided by Gleditsch; see note **Error! Bookmark not defined.**), although the overall message remains similar. Non-spatial OLS seems to suffer only a moderate (10%) bias in its common-shock parameter-estimate in the small-N, small-ρ sample (Table 15), with those problems once again concentrating in the common-shock estimate, that having become the closest available proxy to the omitted spatial-lag again. OLS performance quickly becomes simply atrocious, though, on all bases, as either N or ρ grows (Tables 16-18), with biases, standard-error inaccuracies, and RMSE's orders-of-magnitude worse than any of the spatial alternatives (e.g., biases exceed -2,000%, i.e., -20*x*, in the most extreme case: Table 18). And once again, S-OLS does fairly well in redressing those biases in the smaller-N samples, but it leaves 12-30% overestimation of ρ, and a corresponding underestimation of β_S of 4-24%. S-OLS standard errors are reasonably accurate too, and its efficiency on par with the consistent alternatives in these smaller-N cases. Even more remarkable, though, is its again-stellar performance in the larger-N samples, with biases only for the weaker-ρ case (somewhat surprisingly) and in just the 3-4% range there ($\hat{\rho}$ inflated, $\hat{\beta}$ attenuated), with accurate standard-errors and RMSE as good or nearly so as the best alternative. Here again, finally, S-ML "nails it" or nearly so, but now S-2SLS-IV does slightly better in the larger-N samples—about equal efficiency and standard-error-accuracy, slightly lesser bias (small by either estimator)—and comparably in the smaller-N samples too—slightly lesser bias (small by either estimator), and little efficiency loss, only in a few instances noticeable. Thus, the conclusion for this pattern of connectivity is similar: (1) non-spatial OLS is a very bad idea; (2) S-OLS redresses most of the problem in smaller-N samples, although not as completely or as well as the consistent alternatives, and all of the problem as well as either alternative in larger-N samples; and (3) either of the consistent estimators works very well in larger- or smaller-N conditions, with perhaps a slight edge to S-2SLS-IV.

Finally, we consider in Tables 19-22 a connectivity matrix uniformly [-1,1] distributed w_{ij}, intended to reflect, e.g., a set of units that are evenly distributed substitutes or complements for each other (perhaps similar to what a researcher might specify in a model of FDI flows). Notice that the mean interdependence, $\rho\mathbf{W}$, is zero in this case. From that perspective, and recalling our analytical results regarding the omitted-variable and simultaneity biases afoot here, it is not surprising that any of the estimates, non-spatial OLS included, performs reasonably well in bias terms under any N- or ρ-size conditions. The weakness of non-spatial OLS appears here only at higher ρ where the relative inefficiency of OLS failing to model the interdependence can make itself felt. Indeed, with higher ρ

and higher N, the latter perhaps reflecting more chances to draw a high w_{ij}, we rediscover huge biases, standard-error inaccuracies, and inefficiencies for non-spatial OLS here too. S-OLS again seems to redress most of the issues well in most cases, and, again, either S-2SLS-IV or S-ML seems to offer viable further improvements thereupon, the latter more efficiently in smaller sample. In these examples, though, across all the experimental conditions, S-OLS redresses the bias fully or almost fully (because $\rho\mathbf{W}$ has mean zero). Its efficiency, too, is comparable with the more-efficient S-ML alternative. Only in the accuracy of its standard errors for interdependence-strength (i.e., ρ) estimates, which are highly inaccurate do we see any appreciable advantage to the more complex estimators, and perhaps PCSE's (to be considered in future work) may redress even that issue in these cases.

Table 19. Comparing Estimators (N=5, ρ = 0.1, w_{ij} = U(-1,1), 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	0.996	0.104	0.107	0.995	0.102	0.105	0.996	0.103	0.105	1.006	0.105	0.103
	$s.e.(\beta_1)$	0.107	0.012		0.105	0.012		0.105	0.012		0.103	0.012	
	$pcse(\beta_1)$												
	β_2	0.999	0.123	0.108	1.003	0.110	0.109	1.002	0.111	0.111	1.002	0.114	0.109
	$s.e.(\beta_2)$	0.108	0.019		0.109	0.019		0.111	0.021		0.109	0.021	
	$pcse(\beta_2)$												
	β_3	0.997	0.115	0.114	0.998	0.112	0.112	0.998	0.113	0.113	1.003	0.114	0.110
	$s.e.(\beta_3)$	0.114	0.025		0.112	0.025		0.113	0.025		0.110	0.025	
	$pcse(\beta_3)$												
	ρ				0.098	0.049	0.047	0.098	0.060	0.057	0.098	0.048	0.046
	$s.e.(\rho)$				0.047	0.009		0.057	0.015		0.046	0.009	
	$pcse(\rho)$												
T=40	β_1	1.001	0.075	0.074	1.000	0.073	0.073	1.001	0.073	0.073	1.000	0.073	0.072
	$s.e.(\beta_1)$	0.074	0.006		0.073	0.005		0.073	0.005		0.072	0.005	
	$pcse(\beta_1)$												
	β_2	1.001	0.091	0.075	0.999	0.074	0.075	1.000	0.076	0.076	0.999	0.074	0.074
	$s.e.(\beta_2)$	0.075	0.010		0.075	0.010		0.076	0.010		0.074	0.009	
	$pcse(\beta_2)$												
	β_3	1.001	0.082	0.077	1.000	0.079	0.075	1.000	0.079	0.075	0.998	0.077	0.074
	$s.e.(\beta_3)$	0.077	0.012		0.075	0.012		0.075	0.012		0.074	0.011	
	$pcse(\beta_3)$												
	ρ				0.100	0.034	0.033	0.102	0.039	0.039	0.100	0.034	0.032
	$s.e.(\rho)$				0.033	0.005		0.039	0.008		0.032	0.005	
	$pcse(\rho)$												

Table 20. Comparing Estimators (N=40, ρ = 0.1, w_{ij} = U(-1,1), 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	0.999	0.047	0.046	0.999	0.038	0.036	0.999	0.038	0.036	1.002	0.037	0.036
	$s.e.(\beta_1)$	0.046	0.003		0.036	0.002		0.036	0.002		0.036	0.002	
	$pcse(\beta_1)$												
	β_2	1.002	0.080	0.046	1.000	0.038	0.037	1.000	0.038	0.037	1.001	0.037	0.038
	$s.e.(\beta_2)$	0.046	0.007		0.037	0.006		0.037	0.006		0.038	0.007	
	$pcse(\beta_2)$												
	β_3	1.002	0.055	0.048	1.000	0.038	0.038	1.000	0.038	0.038	0.999	0.040	0.038
	$s.e.(\beta_3)$	0.048	0.008		0.038	0.007		0.038	0.007		0.038	0.007	
	$pcse(\beta_3)$												
	ρ				0.100	0.005	0.005	0.100	0.006	0.006	0.100	0.005	0.005
	$s.e.(\rho)$				0.005	0.001		0.006	0.001		0.005	0.001	
	$pcse(\rho)$												
T=40	β_1	1.001	0.032	0.032	1.000	0.025	0.025	1.000	0.025	0.025	0.999	0.025	0.025
	$s.e.(\beta_1)$	0.032	0.001		0.025	0.001		0.025	0.001		0.025	0.001	
	$pcse(\beta_1)$												
	β_2	1.000	0.070	0.032	1.000	0.026	0.025	1.000	0.026	0.026	1.000	0.027	0.026
	$s.e.(\beta_2)$	0.032	0.003		0.025	0.003		0.026	0.003		0.026	0.003	
	$pcse(\beta_2)$												
	β_3	1.003	0.037	0.033	1.001	0.025	0.026	1.001	0.025	0.026	1.000	0.026	0.026
	$s.e.(\beta_3)$	0.032	0.004		0.026	0.003		0.026	0.003		0.026	0.003	
	$pcse(\beta_3)$												
	ρ				0.100	0.003	0.003	0.100	0.004	0.004	0.100	0.003	0.003
	$s.e.(\rho)$				0.003	0.000		0.004	0.000		0.003	0.000	
	$pcse(\rho)$												

Table 21. Comparing Estimators (N=5, ρ = 0.5, w_{ij} = U(-1,1), 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	1.044	0.366	0.192	0.987	0.103	0.106	0.996	0.104	0.107	1.006	0.106	0.104
	$s.e.(\beta_1)$	0.186	0.118		0.105	0.012		0.107	0.012		0.104	0.013	
	$pcse(\beta_1)$												
	β_2	1.034	0.522	0.191	0.993	0.111	0.109	1.003	0.113	0.113	1.002	0.114	0.109
	$s.e.(\beta_2)$	0.188	0.124		0.109	0.020		0.113	0.022		0.109	0.022	
	$pcse(\beta_2)$												
	β_3	1.034	0.354	0.199	0.991	0.112	0.112	0.998	0.113	0.114	1.003	0.118	0.111
	$s.e.(\beta_3)$	0.196	0.125		0.112	0.025		0.114	0.026		0.111	0.026	
	$pcse(\beta_3)$												
	ρ				0.492	0.058	0.042	0.498	0.062	0.057	0.498	0.044	0.057226
	$s.e.(\rho)$				0.042	0.013		0.057	0.021		0.041	0.014	
	$pcse(\rho)$												
T=40	β_1	1.037	0.296	0.132	0.992	0.074	0.073	1.001	0.073	0.073	0.999	0.073	0.072
	$s.e.(\beta_1)$	0.127	0.081		0.073	0.005		0.073	0.006		0.072	0.006	
	$pcse(\beta_1)$												
	β_2	1.059	0.574	0.141	0.991	0.078	0.075	1.000	0.077	0.077	0.999	0.075	0.075
	$s.e.(\beta_2)$	0.128	0.083		0.075	0.010		0.077	0.012		0.075	0.010	
	$pcse(\beta_2)$												
	β_3	1.032	0.264	0.135	0.992	0.080	0.075	1.000	0.080	0.076	0.998	0.077	0.074
	$s.e.(\beta_3)$	0.131	0.089		0.075	0.012		0.076	0.012		0.074	0.011	
	$pcse(\beta_3)$												
	ρ				0.493	0.048	0.030	0.501	0.040	0.040	0.500	0.031	0.029
	$s.e.(\rho)$				0.029	0.008		0.039	0.012		0.029	0.009	
	$pcse(\rho)$												

Table 22 Comparing Estimators (N=40, ρ = 0.5, w_{ij} = U(-1,1), 1,000 trials)

		OLS			S-OLS			S-2SLS			ML		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	0.570	20.798	1.855	0.992	0.039	0.037	0.997	0.058	0.060	1.001	0.037	0.037
	$s.e.(\beta_1)$	1.804	10.057		0.036	0.002		0.060	0.042		0.037	0.002	
	$pcse(\beta_1)$												
	β_2	0.076	12.653	2.008	0.993	0.040	0.037	0.997	0.066	0.067	1.000	0.037	0.038
	$s.e.(\beta_2)$	1.783	10.060		0.037	0.006		0.067	0.060		0.038	0.007	
T=20	$pcse(\beta_2)$												
	β_3	0.892	29.509	1.866	0.993	0.039	0.038	0.997	0.059	0.064	0.998	0.040	0.038
	$s.e.(\beta_3)$	1.863	10.686		0.038	0.007		0.064	0.050		0.038	0.007	
	$pcse(\beta_3)$												
	ρ				0.495	0.006	0.005	0.499	0.027	0.025	0.500	0.002	0.002
	$s.e.(\rho)$				0.002	0.001		0.025	0.030		0.002	0.001	
	$pcse(\rho)$												
	β_1	0.043	10.697	1.516	0.994	0.026	0.026	1.000	0.044		0.999	0.025	0.025
	$s.e.(\beta_1)$	1.176	8.198		0.025	0.001		0.046	0.043		0.025	0.001	
	$pcse(\beta_1)$												
	β_2	0.630	17.452	1.257	0.993	0.028	0.026	0.999	0.056	0.054	1.000	0.027	0.026
	$s.e.(\beta_2)$	1.202	8.901		0.025	0.003		0.054	0.105		0.026	0.003	
T=40	$pcse(\beta_2)$												
	β_3	0.136	5.588	1.487	0.994	0.026	0.026	1.000	0.046	0.047	0.999	0.026	0.026
	$s.e.(\beta_3)$	1.211	8.937		0.026	0.003		0.047	0.044		0.026	0.003	
	$pcse(\beta_3)$												
	ρ				0.495	0.006	0.005	0.500	0.023	0.021	0.500	0.002	0.001
	$s.e.(\rho)$				0.001	0.001		0.021	0.027		0.001	0.001	
	$pcse(\rho)$												

2. **Specification and Measurement Error:**

    a) Fully accurate specification and measurement: The full set of experiments are conducted first without specification or measurement error; then repeated for all of the following.

    b) Measurement/specification error in exogenous-external factors

4. Simulation Results: Exogenous-External or Interdependence Specification Errors

We consider next the implications of researcher misspecification of the pattern of interdependence, **W**, which is crucial because all of the spatial-lag estimators considered here require pre-specification of **W**. Table 5 reports the results for ρ=0.5 and w_{ij}=1/(N-1)+U[-0.1,+0.1]. With this experiment, we are exploring the consequences of *random* specification error in **W**; i.e., importantly, the specification error is orthogonal to all other components of the model. Misspecifications correlated with any part of **X**, **W**, or **ε** would likely have worse consequences than those found here. Note that the true proportionate variation in the relative strength of cross-unit connections is quite sizable in this example. With N=5, 1/(N-1)=.25, so ±.1 is ±40%. With N=40, 1/(N-1)≈.025, so ±.1 is ±400% roughly. Still, given that the true spatial weights are randomly distributed about those used by the analyst in the estimation, we might expect little change in the bias properties of either OLS or S-OLS while the sampling variability for both estimators should increase. Furthermore, since the S-OLS estimator uses an imperfect (although unbiased) spatial weighting matrix, we might expect these estimates to offer a lesser improvement over non-spatial OLS than it did in the example of Table 4 where the estimator used exactly the right weighting matrix.

Some of these expectations obtain; others do not. The standard errors for all three estimators[61] do increase with the introduction of random noise to the spatial weighting matrix, but, surprisingly, the biases in S-OLS estimates of β_2 and ρ actually decrease with the introduction of random noise to **W**. We suspect this occurs because the random draws in true **W** effectively add measurement error to the analyst's spatial lag, which (given the orthogonality of these errors) induces (only) attenuation bias that works in the opposite direction as the simultaneity bias. Meanwhile, the improvement of the S-2SLS-IV estimator over S-OLS declines (for related reasons). In Table 4, e.g., we saw that, when

[61] We have abandoned for now exploration of S-ML given its computational demands, which would delay, perhaps unwarrantedly given the uninspiring performance found above, accumulation of results for the other, simpler estimators.

analysts use the true spatial weighting matrix for estimation and T=40, S-2SLS-IV has better RMSE than S-OLS (.065 *vs.* .092) under these sample and interdependence-strength conditions. In Table 5, however, the advantage is eliminated. The problem for S-2SLS-IV is that the random noise in **W** matrix reduces the predictive power of the spatial instruments. Weak instruments are well known to worsen IV estimates (see, e.g., Bartels 1991). In other words, as the researcher provides better specification for the pattern of interdependence, her ability to obtain better estimates of interdependence strength improves absolutely and relative to domestic, exogenous-external, and context-conditional alternatives not only by the enhanced distinction between these mechanisms but also by the enhanced ability of spatial-lag instruments to redress the simultaneity biases in estimating ρ.

Table 5. Comparing Estimators (N=40, ρ = 0.5, w_{ij} = (1/N-1)+U[-.1,+.1], 100 trials)

		OLS			S-OLS			S-2SLS		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	1.002	0.044	0.044	0.991	0.041	0.042	0.992	0.041	0.042
	$s.e.(\beta_1)$	0.042	0.002		0.04	0.002		0.04	0.002	
	$pcse(\beta_1)$	0.041	0.003		0.04	0.002				
	β_2	1.999	0.168	1.013	0.911	0.185	0.205	1.034	0.255	0.256
	$s.e.(\beta_2)$	0.043	0.007		0.171	0.04		0.229	0.097	
T=20	$pcse(\beta_2)$	0.085	0.021		0.132	0.041				
	β_3	1.003	0.053	0.052	0.995	0.049	0.049	0.995	0.049	0.049
	$s.e.(\beta_3)$	0.043	0.007		0.042	0.007		0.042	0.007	
	$pcse(\beta_3)$	0.042	0.008		0.041	0.007				
	P				0.541	0.102	0.109	0.48	0.134	0.135
	$s.e.(\rho)$				0.083	0.02		0.112	0.049	
	$Pcse(\rho)$				0.064	0.021				
	β_1	1.016	0.026	0.03	1.002	0.026	0.026	1.004	0.026	0.026
	$s.e.(\beta_1)$	0.029	0.001		0.028	0.001		0.028	0.001	
	$pcse(\beta_1)$	0.029	0.001		0.028	0.001				
	β_2	1.977	0.144	0.987	0.874	0.119	0.173	1.006	0.157	0.156
	$s.e.(\beta_2)$	0.029	0.003		0.111	0.019		0.141	0.041	
T=40	$pcse(\beta 2)$	0.061	0.012		0.088	0.018				
	β_3	1.011	0.032	0.033	0.999	0.031	0.031	1	0.031	0.031
	$s.e.(\beta_3)$	0.029	0.003		0.028	0.003		0.028	0.003	
	$pcse(\beta_3)$	0.029	0.003		0.028	0.003				
	P				0.556	0.068	0.088	0.489	0.087	0.088
	$s.e.(\rho)$				0.054	0.01		0.07	0.022	
	$Pcse(\rho)$				0.043	0.01				

c) Measurement/specification error in W

In Table 6, we consider the analogous researcher-error with respect to the exogenous-external variable, **S**. We generate the true exogenous-external shock, recall, by summing two independent random draws from normal distributions. Here, however, we estimate our models using only one of these two equal components. Thus, the hypothetical researcher specifies the exogenous-external factor, **S**, only half correctly; of course, this misspecification will affect the context-conditional component, **DS**, also. As in Table 5, this misspecification error is independently random and so, crucially, orthogonal to all the other components of the model. Thus, we expect and do find an attenuation bias in S-OLS estimates. This time the attenuation biases in β estimates reinforce the simultaneity bias in ρ estimates, so the overall bias in the S-OLS estimator increases. Thus, as expected, inadequate modeling of exogenous external-stimuli mechanisms attenuates estimates of exogenous-external effects[62] and inflates estimates of interdependence effects. This motivates our most-central advice regarding the primacy of effectively modeling both the spatial-interdependence and the alternative domestic, exogenous-external, and context-conditional causal mechanisms to distinguishing between them and evaluating their relative weight empirically.

[62] Somewhat surprisingly to us, these biases concentrated almost exclusively in β_S, and not noticeably in β_{DS}, estimates. We expect that the independent normality of the specification errors and the domestic conditions that enter the context-conditional component leaves β_{DS} estimates unaffected *on average*, but notice the very large increase in its true standard error: once can see this as their error-variance doubling but remaining mean-zero.

Table 6. Comparing Estimators (N=40, ρ = 0.5, w_{ij} = (1/N-1), η = η₁ + η₂, 100 trials)

		OLS			S-OLS			S-2SLS		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	1.004	0.175	0.174	0.991	0.166	0.165	0.996	0.167	0.166
	$s.e.(\beta_1)$	0.065	0.009		0.044	0.003		0.047	0.006	
	$pcse(\beta_1)$	0.07	0.012		0.055	0.006				
	β_2	2.001	0.468	1.104	0.113	0.102	0.892	0.488	0.623	0.804
	$s.e.(\beta_2)$	0.099	0.023		0.092	0.014		0.265	0.229	
	$pcse(\beta_2)$	0.472	0.127		0.079	0.024				
	β_3	1.023	0.249	0.249	1.017	0.244	0.243	1.021	0.243	0.242
	$s.e.(\beta_3)$	0.1	0.023		0.067	0.013		0.071	0.015	
	$pcse(\beta_3)$	0.103	0.027		0.081	0.017				
	ρ				0.95	0.034	0.451	0.759	0.295	0.391
	$s.e.(\rho)$				0.032	0.005		0.129	0.114	
	$pcse(\rho)$				0.027	0.009				
T=40	β_1	1.007	0.136	0.136	0.982	0.126	0.127	0.987	0.129	0.129
	$s.e.(\beta_1)$	0.046	0.004		0.031	0.001		0.037	0.03	
	$pcse(\beta_1)$	0.049	0.005		0.035	0.002				
	β_2	2.028	0.358	1.088	0.114	0.082	0.89	0.844	1.457	1.458
	$s.e.(\beta_2)$	0.067	0.01		0.063	0.008		0.412	1.759	
	$pcse(\beta2)$	0.321	0.055		0.057	0.014				
	β_3	1.031	0.194	0.196	1.012	0.19	0.189	1.017	0.193	0.193
	$s.e.(\beta_3)$	0.067	0.01		0.045	0.006		0.054	0.04	
	$pcse(\beta_3)$	0.069	0.012		0.05	0.007				
	ρ				0.946	0.023	0.446	0.575	0.861	0.86
	$s.e.(\rho)$				0.021	0.002		0.222	1.088	
	$pcse(\rho)$				0.02	0.004				

d) Exogenous unit-specific and exogenous-external explanators vs. varying magnitudes of endogeneity from left- to right-hand-side as well as varying magnitudes spatial interdependence among left- and right-hand-side.

5. Simulation Results: Cross-Spatial Endogeneity and Imperfect Instruments

In Tables 7-10, finally, we wish to consider the consequences of *cross-spatial endogeneity*, our term for the condition which would render spatial-lag **X** instruments imperfectly exogenous, for OLS, S-OLS, and S-2SLS-IV. By *cross-spatial endogeneity* we mean direct or indirect causal correlation from y_i to x_j, i.e. from outcomes in unit i to the non-spatial domestic (and context-conditional[63]) regressors in unit j, for example, from tax policies in, say, France, to government partisanship in Germany. Such "diagonal" causal arrows may often seem, at least at first, implausible substantatively, which is why we suggested previously that the spatially interdependent structure of the data themselves suggest plausible instruments. However, cross-spatial endogeneity, i.e., diagonal arrows, may also arise from the combination of horizontal and vertical arrows, i.e., causal correlation of some **X** (e.g., government partisanship) across units *and* endogeneity of **X** to **y** (e.g., government partisanship to tax). Thus, exploring the consequences of cross-spatial endogeneity for S-2SLS-IV, which may be emerging as a preferred alternative to S-OLS when interdependence is strong, in particular and relative to other estimators, becomes important. To generate such cross-spatial endogeneity, we simultaneously draw **X** and $\boldsymbol{\varepsilon}$ from a multivariate normal distribution with correlation matrix **C**.

We set the strength of the cross-spatial endogeneity equal to the overall strength of interdependence by giving the appropriate elements of **C** (c_{ij}) a value of ρw_{ij} (i.e., $\rho/(N-1)$). The weak interdependence and weak cross-spatial endogeneity results for our small and large N experiments are presented in Tables 7 and 8 respectively. These results show relatively little difference from the experiments without cross-spatial endogeneity (compare with Tables 1 and 2). As one might expect, the RMSE increase for each of the estimators, but the same basic conclusions hold. S-OLS remains the preferred estimator when interdependence is weak. However, when

[63] The common exogenous-external factors, even while exogenous, do not provide identification leverage precisely because they are common to all units. The context-conditional components of those effects may provide leverage.

interdependence is stronger, we know from above that the potential gains on all criteria from using S-2SLS-IV grow. With perfect instruments and a large *T*, S-2SLS-IV outperforms S-OLS in RMSE terms (Tables 3 and 4; T=40), but this is no longer true when cross-spatial endogeneity is also strong. The performance of both S-OLS and S-2SLS-IV deteriorate under these circumstances, but the decline is so much greater for the latter estimator that the advantages of using it—including its relative unbiasedness and the accuracy of standard error estimates—disappear. When N=5 and T=40 the RMSE for the S-2SLS-IV estimate of β_2 more than doubles and for ρ it quadruples. At this sample size, the RMSE for the S-OLS estimate of β_2 and ρ increase by 79% and 111% respectively. With N=40 and T=40, cross-spatial endogeneity has similar consequences. Even if interdependence is strong, therefore, S-2SLS-IV shows no real advantage over S-OLS when its conditions are equally heavily violated. Thus, the relative advantage of S-2SLS-IV over simple S-OLS, like all IV *vs.* OLS comparisons, depends on the combination of how strong are the instruments, how perfectly they purge the endogeneity, and how strong is the endogeneity to begin. We need at this point to provide more simulations across the ranges of those conditions (and likely helpful to plot rather than tabulate them) to understand better when S-OLS and when S-2SLS-IV may be the wiser estimation strategy.

Table 7. Comparing Estimators (N=5, ρ = 0.1, w_{ij} = (1/N-1), c_{ij} = .025, 100 trials)

		OLS			S-OLS			S-2SLS		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	1.028	0.124	0.126	1.022	0.123	0.124	1.022	0.122	0.123
	$s.e.(\beta_1)$	0.106	0.012		0.106	0.012		0.107	0.012	
	$pcse(\beta_1)$	0.104	0.012		0.104	0.012				
	β_2	1.125	0.135	0.183	1.017	0.220	0.219	0.962	0.219	0.221
	$s.e.(\beta_2)$	0.108	0.021		0.172	0.039		0.205	0.069	
T=20	$pcse(\beta_2)$	0.111	0.027		0.156	0.041				
	β_3	0.994	0.110	0.109	0.994	0.111	0.111	0.994	0.110	0.109
	$s.e.(\beta_3)$	0.110	0.024		0.111	0.025		0.113	0.026	
	$pcse(\beta_3)$	0.108	0.025		0.109	0.025				
	P				0.100	0.133	0.132	0.145	0.162	0.167
	$s.e.(\rho)$				0.118	0.024		0.154	0.055	
	$pcse(\rho)$				0.108	0.028				
	β_1	1.040	0.074	0.084	1.034	0.073	0.081	1.033	0.074	0.081
	$s.e.(\beta_1)$	0.074	0.006		0.073	0.006		0.073	0.006	
	$pcse(\beta_1)$	0.072	0.006		0.072	0.006				
	β_2	1.109	0.089	0.14	0.957	0.119	0.126	0.938	0.123	0.137
	$s.e.(\beta_2)$	0.075	0.009		0.114	0.015		0.130	0.024	
T=40	$pcse(\beta_2)$	0.081	0.012		0.105	0.016				
	β_3	1.001	0.081	0.081	1.001	0.080	0.08	1.001	0.080	0.08
	$s.e.(\beta_3)$	0.076	0.011		0.076	0.010		0.076	0.011	
	$pcse(\beta_3)$	0.075	0.010		0.074	0.011				
	P				0.138	0.077	0.086	0.155	0.083	0.099
	$s.e.(\rho)$				0.078	0.010		0.096	0.019	
	$pcse(\rho)$				0.072	0.011				

Table 8. Comparing Estimators (N=40, ρ = 0.1, w_{ij} = (1/N-1), c_{ij} = .003, 100 trials)

		OLS			S-OLS			S-2SLS		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
T=20	β_1	1.011	0.037	0.032	1.010	0.037	0.038	1.010	0.037	0.038
	$s.e.(\beta_1)$	0.037	0.002		0.037	0.002		0.037	0.002	
	$pcse(\beta_1)$	0.036	0.002		0.037	0.002				
	β_2	1.112	0.041	0.118	0.979	0.176	0.176	0.921	0.150	0.169
	$s.e.(\beta_2)$	0.038	0.007		0.142	0.030		0.177	0.059	
	$pcse(\beta_2)$	0.041	0.010		0.129	0.038				
	β_3	1.001	0.047	0.045	1.001	0.047	0.046	1.001	0.046	0.046
	$s.e.(\beta_3)$	0.038	0.007		0.038	0.007		0.038	0.007	
	$pcse(\beta_3)$	0.038	0.007		0.038	0.007				
	P				0.118	0.156	0.156	0.170	0.130	0.147
	$s.e.(\rho)$				0.123	0.026		0.155	0.053	
	$pcse(\rho)$				0.112	0.033				
T=40	β_1	1.003	0.027	0.027	1.002	0.027	0.027	1.002	0.027	0.027
	$s.e.(\beta_1)$	0.025	0.001		0.025	0.001		0.025	0.001	
	$pcse(\beta_1)$	0.025	0.001		0.025	0.001				
	β_2	1.112	0.033	0.117	0.967	0.118	0.122	0.947	0.117	0.128
	$s.e.(\beta_2)$	0.026	0.003		0.097	0.014		0.118	0.025	
	$pcse(\beta_2)$	0.028	0.004		0.089	0.018				
	β_3	1.000	0.026	0.026	1.000	0.025	0.025	1.000	0.025	0.025
	$s.e.(\beta_3)$	0.026	0.003		0.026	0.003		0.026	0.003	
	$pcse(\beta_3)$	0.026	0.003		0.026	0.003				
	P				0.131	0.100	0.104	0.149	0.099	0.111
	$s.e.(\rho)$				0.084	0.012		0.103	0.022	
	$pcse(\rho)$				0.077	0.015				

Table 9. Comparing Estimators (N=5, ρ = 0.5, w_{ij} = (1/N-1), c_{ij} = .125, 100 trials)

		OLS			S-OLS			S-2SLS		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	1.314	0.167	0.355	1.045	0.113	0.121	1.038	0.115	0.121
	$s.e.(\beta_1)$	0.151	0.017		0.105	0.012		0.106	0.012	
	$pcse(\beta_1)$	0.143	0.023		0.107	0.013				
	β_2	1.971	0.312	1.019	0.724	0.136	0.307	0.690	0.138	0.339
	$s.e.(\beta_2)$	0.156	0.033		0.161	0.036		0.173	0.047	
T=20	$pcse(\beta_2)$	0.260	0.060		0.105	0.025				
	β_3	1.061	0.160	0.17	0.965	0.128	0.132	0.963	0.129	0.133
	$s.e.(\beta_3)$	0.164	0.040		0.111	0.025		0.111	0.025	
	$pcse(\beta_3)$	0.152	0.041		0.111	0.024				
	P				0.634	0.047	0.142	0.652	0.054	0.161
	$s.e.(\rho)$				0.061	0.013		0.068	0.018	
	$pcse(\rho)$				0.042	0.009				
	β_1	1.331	0.111	0.349	1.063	0.075	0.098	1.060	0.076	0.097
	$s.e.(\beta_1)$	0.105	0.007		0.072	0.005		0.072	0.005	
	$pcse(\beta_1)$	0.102	0.009		0.074	0.006				
	β_2	2.061	0.199	1.079	0.756	0.092	0.261	0.737	0.090	0.277
	$s.e.(\beta_2)$	0.108	0.011		0.111	0.016		0.117	0.019	
T=40	$pcse(\beta_2)$	0.184	0.021		0.076	0.011				
	β_3	1.103	0.146	0.179	0.982	0.081	0.082	0.981	0.081	0.083
	$s.e.(\beta_3)$	0.109	0.013		0.073	0.010		0.073	0.010	
	$pcse(\beta_3)$	0.106	0.016		0.073	0.010				
	P				0.624	0.035	0.129	0.633	0.036	0.138
	$s.e.(\rho)$				0.040	0.006		0.044	0.008	
	$pcse(\rho)$				0.029	0.004				

Table 10. Comparing Estimators (N=40, ρ = 0.5, w_{ij} = (1/N-1), c_{ij} = .013, 100 trials)

		OLS			**S-OLS**			**S-2SLS**		
		Mean	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*	*Mean*	*Std. Dev.*	*RMSE*
	β_1	1.030	0.045	0.054	1.004	0.038	0.038	1.004	0.038	0.038
	$s.e.(\beta_1)$	0.039	0.002		0.037	0.002		0.037	0.002	
	$pcse(\beta_1)$	0.038	0.002		0.036	0.002				
	β_2	2.000	0.116	1.007	0.773	0.114	0.254	0.748	0.117	0.278
	$s.e.(\beta_2)$	0.040	0.007		0.140	0.033		0.160	0.050	
T=20	$pcse(\beta_2)$	0.096	0.024		0.091	0.026				
	β_3	1.002	0.043	0.042	0.994	0.039	0.039	0.994	0.039	0.039
	$s.e.(\beta_3)$	0.041	0.007		0.038	0.006		0.038	0.006	
	$pcse(\beta_3)$	0.040	0.007		0.038	0.007				
	P				0.614	0.055	0.127	0.627	0.057	0.138
	$s.e.(\rho)$				0.067	0.015		0.077	0.023	
	$pcse(\rho)$				0.044	0.012				
	β_1	1.036	0.025	0.044	1.007	0.022	0.023	1.007	0.022	0.023
	$s.e.(\beta_1)$	0.027	0.001		0.025	0.001		0.026	0.001	
	$pcse(\beta_1)$	0.027	0.001		0.025	0.001				
	β_2	2.011	0.075	1.014	0.757	0.078	0.255	0.766	0.080	0.247
	$s.e.(\beta_2)$	0.028	0.003		0.088	0.013		0.098	0.019	
T=40	$pcse(\beta_2)$	0.071	0.011		0.060	0.010				
	β_3	1.003	0.031	0.031	0.990	0.026	0.028	0.990	0.026	0.028
	$s.e.(\beta_3)$	0.028	0.003		0.026	0.003		0.026	0.003	
	$pcse(\beta_3)$	0.028	0.003		0.026	0.003				
	P				0.623	0.038	0.129	0.619	0.038	0.125
	$s.e.(\rho)$				0.042	0.006		0.047	0.009	
	$pcse(\rho)$				0.029	0.005				

3. Spatio-Temporal-Lag Models

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

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \phi\mathbf{M}\mathbf{y} + \beta\mathbf{x} + \boldsymbol{\varepsilon} \qquad (1.83),$$

where

$$\mathbf{x} = \left(\mathbf{z} + \theta\mathbf{W}\mathbf{z}\right) \qquad (1.84)$$

and

$$\boldsymbol{\varepsilon} = g \times \left(\mathbf{u} + \lambda\mathbf{M}\mathbf{u}\right) \qquad (1.85).$$

The reduced form is thus

$$\mathbf{y} = \left(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{M}\right)^{-1}\left[\beta\left(\mathbf{z} + \theta\mathbf{W}\mathbf{z}\right) + \left(\mathbf{u} + \lambda\mathbf{M}\mathbf{u}\right)\right] \qquad (1.86)$$

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.

4. Multiple-W models; Parameterization & Estimation of W

5. Spatial- and Spatio-Temporal-Lag Models of Qualitative Dependent-Variables

V. (CHAPTER 5) Interpretation & Presentation of Spatial & Spatio-Temporal Effects & Dynamics:

A. Calculating Spatial and Spatio-Temporal Multipliers, Effects, and Dynamics

1. The Spatial Multiplier and Contemporaneous Spatial Effects

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. In empirical models containing *spatial dynamics*, as in those with only temporal dynamics, coefficients on explanatory variables give only the *pre-dynamic* impetuses to the outcome variable from increases in those variables. This represents the pre-interdependence impetus, which, incidentally, is unobservable if spatial dynamics are instantaneous (i.e., incur within observation period). This section discusses calculation of spatial multipliers, which allow expression of the *effects* of counterfactual shocks across units, and it applies the delta-method to compute standard errors for these *effects*.[64]

$$
\begin{aligned}
\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
&= (\mathbf{I}_{\mathrm{N}} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \begin{bmatrix} 1 & -\rho w_{1,2} & \cdots & \cdots & -\rho w_{1,\mathrm{N}} \\ -\rho w_{2,1} & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & -\rho w_{(\mathrm{N}-1),\mathrm{N}} \\ -\rho w_{\mathrm{N},1} & \cdots & \cdots & -\rho w_{\mathrm{N},(\mathrm{N}-1)} & 1 \end{bmatrix}^{-1} \left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \\
&= \mathbf{M}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)
\end{aligned}
\tag{87}
$$

Denote the i^{th} column of $\mathbf{M}$ as $\mathbf{m}_i$ and its estimate as $\hat{\mathbf{m}}_i$. If we are interested in the spatial effects of a one-unit increase in explanatory variable *k* in country *i*, we calculate $\frac{dx_{i,k}\beta_{i,k}}{dx_{i,k}}$. The effect of this change on country *i*'s neighbors is $\mathbf{M}\frac{dx_{i,k}\beta_{i,k}}{dx_{i,k}}$ or simply, $\mathbf{m}_i\beta_k$.

The standard errors calculation, using the delta method, is

$$
\mathrm{var}\left(\hat{\mathbf{m}}_i\hat{\beta}_k\right) = \left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\mathrm{var}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]', \text{ where } \hat{\boldsymbol{\theta}} = \begin{bmatrix}\hat{\rho} \\ \hat{\beta}_k\end{bmatrix} \text{ and } \left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right] = \begin{bmatrix}\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}} & \mathbf{m}_i\end{bmatrix} \tag{88}
$$

The vector $\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}}$ is the i^{th} column of $\beta_k\frac{\partial\hat{\mathbf{M}}}{\partial\hat{\rho}}$. Since $\mathbf{M}$ is an inverse matrix, the derivative in equation (88) is calculated as $\frac{\partial\hat{\mathbf{M}}}{\partial\hat{\rho}} = -\hat{\mathbf{M}}\frac{\partial\hat{\mathbf{M}}^{-1}}{\partial\hat{\rho}}\hat{\mathbf{M}} = -\hat{\mathbf{M}}\left(\mathbf{I} - \rho\mathbf{W}\right)\hat{\mathbf{M}} = -\hat{\mathbf{M}}(-\mathbf{W})\hat{\mathbf{M}} = \hat{\mathbf{M}}\mathbf{W}\hat{\mathbf{M}}$. We do not calculate and present the spatial effects implied by the models in Table 1. Instead, we

[64] For an excellent discussion of spatial multipliers, see Anselin (2003). For an application (without standard errors), see Kim, Phipps, and Anselin (2003).

concentrate on calculating spatio-temporal effects using one of the panel models in the next section. These spatio-temporal calculations are slightly more complicated than the purely spatial ones.

2. The Spatio-Temporal Multiplier and Spatio-Temporal Response-Paths and Long-Run Spatio-Temporal Steady-States

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[65]), 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 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.[66]

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

$$\mathbf{y}_t = \rho\mathbf{W}\mathbf{y}_t + \phi\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (1.89).$$

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 (1.89) and solve. This gives the steady-state effect, assuming stationarity and that the exogenous

---

[65] 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.

[66] For an excellent discussion of spatial multipliers, see Anselin (2003).

RHS terms, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, remain permanently fixed to their hypothetical/counterfactual levels:[67]

$$\begin{aligned}
\mathbf{y}_t &= \rho\mathbf{W}\mathbf{y}_t + \phi\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\
&= \left(\rho\mathbf{W} + \phi\mathbf{I}\right)\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\
&= \left[\mathbf{I}_{\mathrm{N}} - \rho\mathbf{W} - \phi\mathbf{I}_{\mathrm{N}}\right]^{-1}\left(\mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right) \\
&= \begin{bmatrix}
1-\phi & -\rho w_{1,2} & \cdots & \cdots & -\rho w_{1,\mathrm{N}} \\
-\rho w_{2,1} & 1-\phi & & & \vdots \\
\vdots & & \ddots & & \vdots \\
\vdots & & & 1-\phi & -\rho w_{(\mathrm{N}-1),\mathrm{N}} \\
-\rho w_{\mathrm{N},1} & \cdots & \cdots & -\rho w_{\mathrm{N},(\mathrm{N}-1)} & 1-\phi
\end{bmatrix}^{-1}\left(\mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right) \\
&\equiv \mathbf{S}\left(\mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right)
\end{aligned} \tag{1.90}$$

To offer standard-error estimates for these steady-state estimates, one could use the delta method. I.e., give a first-order[68] Taylor-series linear-approximation to nonlinear (1.90) around the estimated 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

$$\widehat{\mathbf{V}\left(\hat{\mathbf{s}}_i\hat{\beta}_k\right)} = \left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\widehat{\mathbf{V}\left(\hat{\boldsymbol{\theta}}\right)}\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]' \tag{1.91}$$

where $\hat{\boldsymbol{\theta}} \equiv \left[\hat{\rho} \quad \hat{\phi} \quad \hat{\beta}_k\right]'$, $\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right] \equiv \left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\rho}} \quad \frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\phi}} \quad \hat{\mathbf{s}}_i\right]$, 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 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 $N\times1$ vector of unit outcomes, $\mathbf{y}_t$, to the exogenous RHS terms, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, could also emerge by rearranging (1.89) to isolate $\mathbf{y}_t$ on the LHS:

---

[67] 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.

[68] 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.

$$\mathbf{y}_t = \left[\mathbf{I}_N - \rho\mathbf{W}_N\right]^{-1}\left\{\phi\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right\} \quad (1.92).$$

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 (1.90) as $\mathbf{S} \equiv \left[\mathbf{I}_{\mathrm{NT}} - \rho\mathbf{W} - \phi\mathbf{M}\right]^{-1}$ and follow the steps outlined above. We calculate these effects for the presentation of our empirical reanalysis below, for example.

B. Calculating Standard Errors for S and ST Multipliers, Effects, and Dynamics

1. Delta Method

[See preceding two sections of material.]

2. Simulation Methods

C. Presenting Estimates of Spatial and Spatio-Temporal Effects and Dynamics

1. Spatial and Spatio-Temporal Effect-Grids

2. Plotting Spatio-Temporal Response-Paths

3. Maps (and Dynamic Maps) of Spatial and Spatio-Temporal Effects

[See illustrative empirical applications below]

VI. (<u>CHAPTER 6</u>) Empirical Applications: Replications & extensions (of existing work using spatial lags or failing to use them when theory & substance imply them, ordered by mechanism of interdependence)

A. Competition: Globalization, Capital-Tax Competition for Investment, & Domestic-Policy Autonomy (Swank&Steinmo APSR '02; Hays WP '03; Basinger&Hallerberg APSR '04)

Replication and Reanalysis (Swank and Steinmo)

Swank and Steinmo stress the importance of domestic factors—particularly budgetary dynamics, the level of public sector debt, and macroeconomic performance—in their empirical study of tax policy reform. They also find some external factors, specifically a country's capital account and trade openness, are important determinants of tax reform. A few of their findings are counterintuitive—for example, that increased capital mobility and trade put downward pressure on marginal statutory corporate tax rates but not on effective capital tax rates and that increased capital mobility leads to lower effective tax rates on labor. They argue this is because statutory rate reductions are combined with the elimination of specific investment incentives leaving effective tax burdens unaffected. The finding that increased capital mobility leads to lower effective tax rates on labor income is explained by arguing that labor taxes raise the non-wage costs of employment, cutting into profits. Swank and Steinmo recognize their data are spatially interdependent—they report panel corrected standard errors—but they treat this dependence as a "nuisance" rather than as additional evidence of the importance of external factors in determining tax policy.

Swank and Steinmo implicitly assume that any spatial dependence is limited to the error term in their model. For the purposes of our reanalysis we accept their assumption about the nature of the spatial dependence as true and therefore estimate a model that includes a weighted average of the OLS residuals from the other units (countries in this case) as a spatial lag. We focus on the results for effective tax rates on capital income, labor income, and consumption that are reported in Table 2 from their appendix (pp. 653-4). Their sample covers 13 countries over the period 1981-1995 (N=13 and T=15).

Table 1. Replication and Reanalysis of Swank and Steinmo (Appendix, Table 2)

	Effective Tax Rate on Capital		Effective Tax Rate on Labor		Effective Rate on Consumption	
	Repl.	Reanal.	Repl.	Reanal.	Repl.	Reanal.
Liberalization	1.146	1.165	-0.261**	-0.256**	-0.033	0.057
	(0.991)	(0.945)	(0.126)	(0.107)	(0.211)	(0.184)
Trade	-0.018	-0.022	-0.009	-0.009	0.000	-0.001
	(0.073)	(0.062)	(0.031)	(0.027)	(0.016)	(0.013)
Structural	-1.147***	-0.943***	-0.359	-0.244	0.147	0.303***
Unemployment	(0.292)	(0.273)	(0.26)	(0.211)	(0.117)	(0.088)
Public Sector	0.089**	0.059	0.053***	0.051***	-0.007	-0.012*
Debt	(0.044)	(0.037)	(0.017)	(0.015)	(0.01)	(0.007)
Elderly Population	1.264**	1.297**	-0.018	-0.026	0.05	-0.005
	(0.594)	(0.534)	(0.227)	(0.219)	(0.133)	(0.117)
Temporal Lag	0.809***	0.788***	0.671***	0.658***	0.73***	0.677***
	(0.072)	(0.069)	(0.082)	(0.068)	(0.075)	(0.058)
Growth	0.23	0.259	-0.008	-0.016	0.038	0.032
	(0.176)	(0.175)	(0.054)	(0.046)	(0.036)	(0.025)
% Change Profits	0.127*	0.168**				
	(0.073)	(0.075)				
Domestic	0.066	0.062				
Investment	(0.066)	(0.063)				
Inflation			0.115**	0.104**	-0.041	-0.026
			(0.05)	(0.048)	(0.033)	(0.029)
Unemployment			0.28**	0.257***	-0.114	-0.161***
			(0.114)	(0.090)	(0.07)	(0.042)
Left Government	0.018*	0.017	0.008**	0.009**	-0.003	-0.005*
	(0.011)	(0.013)	(0.004)	(0.003)	(0.003)	(0.003)
Christian	0.041**	0.038**	0.001	0.003	-0.004	-0.011**
Democratic Gov	(0.016)	(0.018)	(0.009)	(0.010)	(0.007)	(0.006)
Spatial		0.595***		0.569***		0.985***
Lag		(0.160)		(0.136)		(0.134)

Nostes: The country and year dummy variable coefficients are omitted to save space.

Table 1 presents the results of our replication and reanalysis. We follow the steps for S-OLS outlined above with one notable exception. We assume that the pattern of spatial dependence in countries' capital, labor, and consumption taxes are similar and, therefore, pool the OLS residiuals from all three of Swank and Steinmo's tax regressions to estimate the spatial weighting matrix $(\boldsymbol{\lambda})$. This gives us KxT observations (3*15 = 45) instead of T to estimate the N-1 independent non-zero elements in each country's *TxNT* block row of $\boldsymbol{\lambda}$.[69] Comparing the original OLS-PCSE results with our S-OLS (PCSE) results suggest there are efficiency gains with the spatial lag model. There are thirty-three common coefficients estimates. For twenty-eight of these coefficients, the standard errors are smaller with S-OLS (PCSE) than with OLS-PCSE.[70] Most importantly, we come to

[69] In our estimation of $\boldsymbol{\lambda}$ we allowed the OLS residuals from each tax regression to have separate means and variances.
[70] Part of this change is attributable to the overconfidence of S-OLS (PCSE). Distinguishing overconfidence from increased efficiency, in these results, is not an easy task.

different conclusions about the importance of international factors for capital taxes.[71] In each model, the coefficient estimate on the spatial lag is statistically significant. In discussing their results, Swank and Steinmo (2002, 650) write, "[they] are consistent with the argument that while internationalization has influenced the shift in the content of tax policy, the combined effect of statutory tax rate cuts and base-broadening reductions in investment incentives has left the effective tax burden on capital largely unchanged." They come to this conclusion because they largely ignore the spatial dependence in their data except to make standard error corrections for their coefficient estimates. The spatial dependence is "out of sight, and out of mind." When a spatial lag is included on the right-hand-side of their regression model we see this conclusion about the effects of international, external factors is likely incorrect. Unexpected changes in effective capital tax rates in one country have statistically significant consequences for effective capital tax rates in other countries.

Here is an example feedback grid, containing both short-term (within one observational period) and long-term-steady-state feedback estimates, with standard errors and significance of the former indicated.

[71] Interestingly, the S-OLS (PCSE) estimates lead to a very different set of conclusions about the determinants of effective tax rates on consumption in that we find statistically significant unemployment and partisan effects on these tax rates. This could be due to either omitted variable or simultaneity biases.

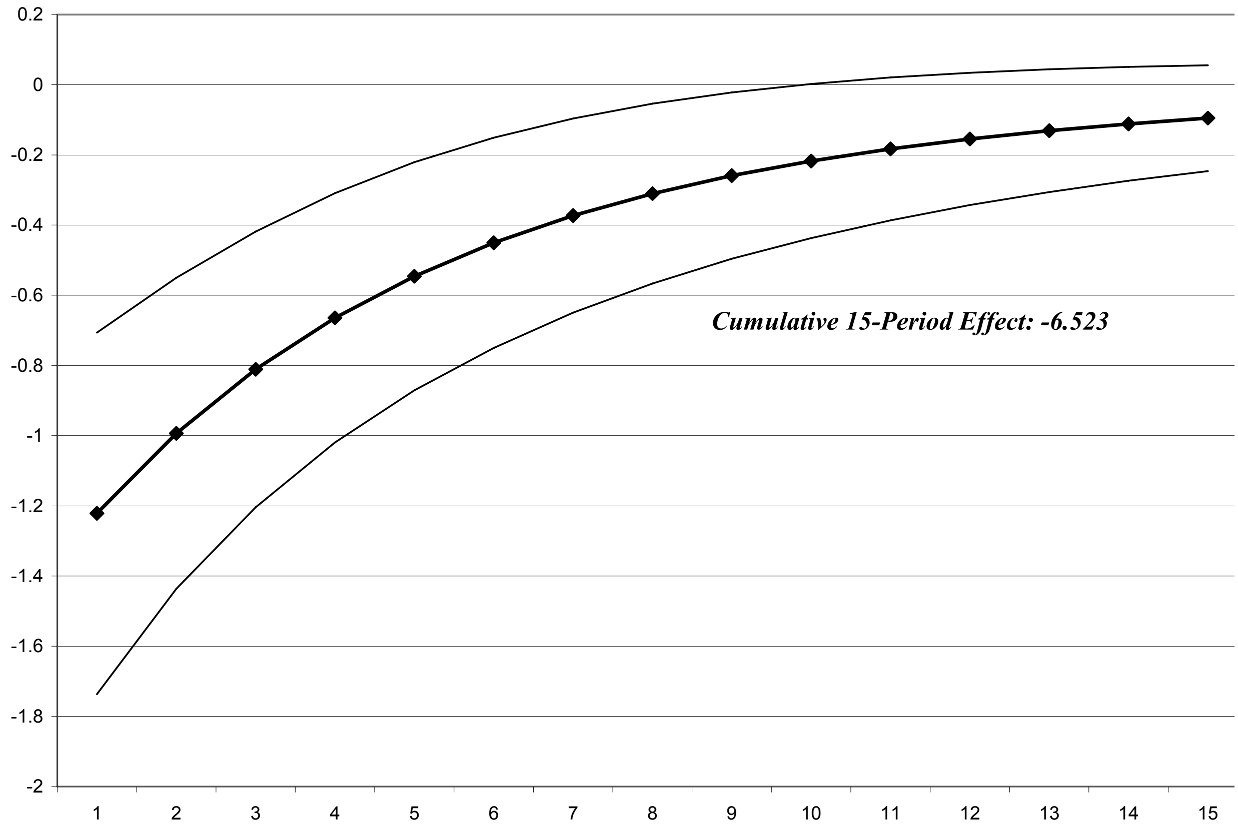
Table 3. Short-Run and Steady-State Spatial Effects from a Shock to Structural Unemployment

	BEL	CAN	FIN	FRA	GER	ITA	NTH	NOR	SWE	GBR	USA
	-1.222**	0	0	-0.034*	-0.034*	-0.001	-0.034*	0	0	-0.034*	0
BEL	0.307	0	0	0.021	0.02	0.001	0.021	0	0	0.02	0
	-7.403	0	0	-1.672	-1.51	-0.227	-1.549	0	0	-1.51	0
	0	-1.231**	0	0	0	0	0	0	0	0	-0.128*
CAN	0	0.309	0	0	0	0	0	0	0	0	0.074
	0	-8.994	0	0	0	0	0	0	0	0	-4.876
	0	0	-1.225**	0	0	0	0	-0.067*	-0.067*	0	0
FIN	0	0	0.307	0	0	0	0	0.04	0.04	0	0
	0	0	-7.954	0	0	0	0	-2.958	-2.958	0	0
	-0.034*	0	0	-1.224**	-0.033*	-0.032*	-0.003	0	0	-0.033*	0
FRA	0.021	0	0	0.307	0.019	0.018	0.003	0	0	0.019	0
	-1.672	0	0	-7.643	-1.395	-1.036	-0.731	0	0	-1.395	0
	-0.045*	0	0	-0.044*	-1.222**	-0.001	-0.044*	0	0	-0.004	0
GER	0.027	0	0	0.026	0.306	0.001	0.026	0	0	0.004	0
	-2.013	0	0	-1.859	-7.187	-0.252	-1.723	0	0	-0.836	0
	-0.004	0	0	-0.127*	-0.003	-1.221**	0	0	0	-0.003	0
ITA	0.004	0	0	0.073	0.004	0.306	0.001	0	0	0.004	0
	-0.907	0	0	-4.144	-0.756	-6.912	-0.396	0	0	-0.756	0
	-0.045*	0	0	-0.004	-0.044*	0	-1.222**	0	0	-0.044*	0
NTH	0.028	0	0	0.005	0.026	0	0.307	0	0	0.026	0
	-2.066	0	0	-0.974	-1.723	-0.132	-7.253	0	0	-1.723	0
	0	0	-0.067*	0	0	0	0	-1.225**	-0.067*	0	0
NOR	0	0	0.04	0	0	0	0	0.307	0.04	0	0
	0	0	-2.958	0	0	0	0	-7.954	-2.958	0	0
	0	0	-0.067*	0	0	0	0	-0.067*	-1.225**	0	0
SWE	0	0	0.04	0	0	0	0	0.04	0.307	0	0
	0	0	-2.958	0	0	0	0	-2.958	-7.954	0	0
	-0.045*	0	0	-0.044*	-0.004	-0.001	-0.044*	0	0	-1.222**	0
GBR	0.027	0	0	0.026	0.004	0.001	0.026	0	0	0.306	0
	-2.013	0	0	-1.859	-0.836	-0.252	-1.723	0	0	-7.187	0
	0	-0.128*	0	0	0	0	0	0	0	0	-1.231**
USA	0	0.074	0	0	0	0	0	0	0	0	0.309
	0	-4.876	0	0	0	0	0	0	0	0	-8.994

Notes: The elements of the table report the effect of a one-unit increase in the column country's level of structural unemployment on the row country's capital tax rate. The first number reported in each cell is the estimated short-run effect (direct effect plus spatial feedback). The second number is the standard error of this estimate. The final number is the estimated long-run steady-steady effect. Australia and Japan are omitted from the table because they have no "neighbors" in the sample. **Significant at the5% Level; *Significant at the 10% Level.

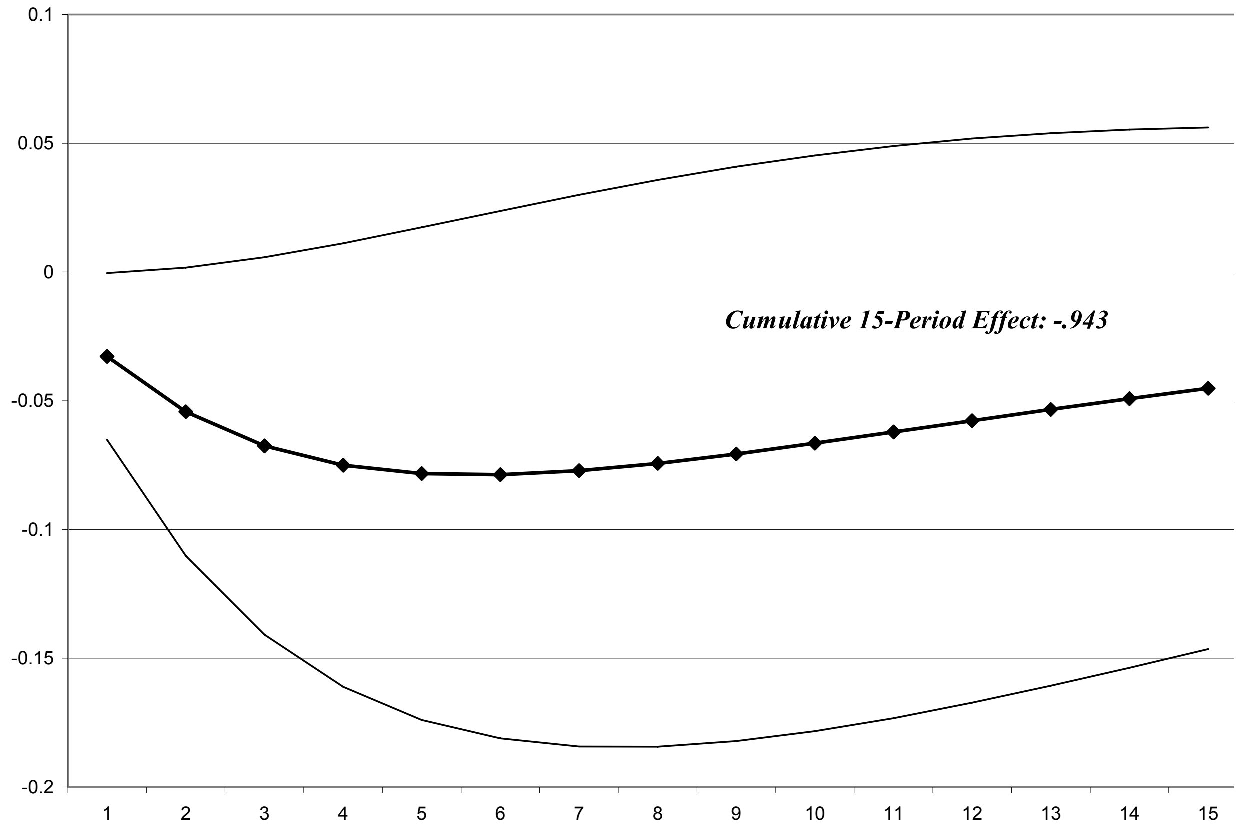
Here is an example of a first-order spatio-temporal-response-path plot (with confidence interval) which is the time-path of a unit's response to its own shocks, but including the spatial feedback to itself of that own shock:

Figure 2. Spatio-Temporal Effects on the German Capital Tax Rate from a Positive One-Unit Counterfactual Shock to Structural Unemployment in Germany (with a 90% C.I.)



Here is an example of a second-order spatio-temporal-response-path plot (with confidence interval), which is the time-path of one unit's response to shocks to another unit, including all spatial feedback to both parties (and all others):

Figure 3. Spatio-Temporal Effects on the French Capital Tax Rate from a Positive One-Unit Counterfactual Shock to Structural Unemployment in Germany (with a 90% C.I.)



Replication and Reanalysis (Hays and Basinger and Hallerberg)

Although all theoretical models of and arguments regarding tax competition, and, indeed, the very substance of its supposed process, clearly imply the spatial interdependence of capital taxes, few scholars have empirically modeled such interdependence directly. Two recent exceptions, Hays (2003) and Basinger and Hallerberg (2004), however, do estimate spatial-lag models of international capital-tax competition, using *OLS*. In the next section, we discuss the empirical work in these two papers and then conduct a reanalysis of their respective regression models. Our results suggest that both studies underestimate the degree of international interdependence in capital tax policymaking, though for very different reasons: Hays uses arbitrary weights to generate his spatial lag and Basinger and Hallerberg ignore that their spatial lag is likely endogenous.[72]

Hays (2003) argues that the effect of globalization—specifically increased international capital

[72] Overall, Hays probably overestimates the coefficient on his spatial lag variable. This is because the simultaneity bias in his S-OLS estimates, which he also ignores, is inflating.

mobility—on a country's capital tax rate depends on its capital endowment and political institutions. An exogenous increase in international capital mobility affects the capital tax rate in two ways. First, it shifts the revenue maximizing tax rate downward. And second, by making the supply of capital less (more) elastic, it increases the marginal gain from increasing (decreasing) the capital tax rate when it is below (above) the revenue maximizing level. How far globalization shifts the revenue maximizing tax rate downward is a function of a country's capital endowment: the drop is large for capital rich countries and relatively small for capital poor ones. The impact of changes to the elasticity of the supply of capital on tax rates depends on a country's political institutions. The capital-supply elasticity determines the marginal revenue-gain from changing tax rates while political institutions determine the marginal cost. Hays argues that increased international capital mobility will have the greatest negative impact on capital tax rates in relatively closed and capital-rich countries with majoritarian political institutions (e.g., the U.K.).

To test his hypothesis, Hays estimates a spatial-lag model with a temporal lag and fixed country effects. He uses the Mendoza et al. (1994, 1997) capital tax rates as the dependent variable. The key independent variables are the degree of capital mobility—measured by Quinn's (1997) indices of capital and financial openness—and capital mobility interacted with a measure of each country's capital endowment and its consensus democracy score (Lijphart 1999).[73] For each country, Hays uses the average tax rate, i.e., the average of the dependent variable, $\mathbf{y}$, in the N-1 other countries as the spatial lag. In other words, all the off-diagonal elements of the spatial weighting matrix from (7) are set to 1/(N-1). For Hays' purposes, the spatial lag controls for the possibility that the observed changes in capital taxation are being driven by tax competition between countries.[74] Hays estimates the model using OLS and reports panel corrected standard errors (S-OLS with PCSE's).

Basinger and Hallerberg (2004) estimate spatial-lag models to test the following hypotheses derived from their theoretical model of tax competition: 1) countries will undergo tax reform more frequently if the political costs of such reforms are low and/or the decisiveness of reforms in determining the patterns of investment flows is high, 2) countries will engage in tax reform when the political costs of reform in competitor countries is low, 3) the domestic political costs of reform and the decisiveness of reform will determine the sensitivity of countries' tax-policies to tax changes in

[73] The capital endowment data are provided in the Penn World Table. Hays uses the capital stock per worker in 1965 as a measure of each country's initial capital endowment.

[74] While Hays' regression models allow for tax competition his theoretical model does not. He makes a small country assumption so the global net-of-tax return to capital is exogenous. Tax competition is not inconsistent with the theory, but Hays' focus is on strategic interaction (among producer groups) within countries rather than on tax competition between countries.

their competitors. Basinger and Hallerberg (2004) include both spatially weighted **X**'s and spatially weighted **Y**'s (i.e., spatial lags) on the right-hand-side of their regression models. Hypothesis 1 is operationalized with a set of domestic **X**'s; they test Hypothesis 2 using a set of spatially weighted **X**'s and Hypothesis 3 with spatial lags interacted with domestic **X**'s.

The dependent variable in their empirical analysis is the change in the *capital tax rate*. In addition to the Mendoza et al. capital-tax rates, the same variable used by Swank and Steinmo (2002) and Hays (2003), Basinger and Hallerberg include the top marginal capital-tax rates (both central government and overall). They identify two kinds of domestic political costs as independent variables: transaction and constituency costs. *Ideological distances* between veto players are used to measure transaction costs. The greater the ideological distance between political actors that can block policy change the harder it is to alter the status quo (in this case, adopt capital-tax reform). *Partisanship* is used to measure constituency costs; the constituency costs associated with capital tax reform will be higher when left governments are in power. A third independent variable of interest, the degree of capital mobility, is measured using *capital controls* on outflows. The degree of capital mobility determines the decisiveness of capital taxes in determining the location of international investments.

Basinger and Hallerberg use four different spatial weighting matrices: a symmetric 1/(N-1) weighting matrix (the same one used by Hays), which makes the spatial lag for each unit equal to the simple average of the **Y**'s in the other units, and three weighted averages using GDP, FDI, and Fixed Capital Formation (FCF) as weights. For each row, the last three spatial weighting matrices have cell entries that differ across columns, but the rows themselves are identical. In other words, for every country in the sample, the US—because of its large GDP, capital stock, and flows of FDI—is weighted more heavily than Finland, but the effect of American tax rates on other tax rates is the same for all countries. American tax rates have the same effect on Canada as they do on Austria, for example. The spatial weights are time varying. Basinger and Hallerberg (2004) include fixed country effects in their models, but, unlike Hays, do not lag the dependent variable directly. The do include the lagged level of the tax rate, though, which makes their model with changes as the dependent variable essentially the same as a partial-adjustment (lagged-dependent-variable) model in levels. The models are estimated using OLS; panel corrected standard errors are reported. In our reanalysis, we focus on Hypothesis 3 (Tables 3-5 from Basinger and Hallerberg). In these models, the spatial lag is differenced and lagged temporally. (See discussion of equations (15) and (16) above.)

Both Hays (2003) and Basinger and Hallerberg (2004) find the coefficient on the spatial lag is positive and statistically significant. The problem with both sets of analyses is they do not account for the fact that the spatial lag is endogenous making the *S-OLS* estimator biased and inconsistent. Both may have underspecified *common-conditions* sorts of arguments also.

We conduct a reanalysis of Hays' (2003:99)[75] regressions using a new spatial-weights matrix and two consistent estimators—spatial two-stage least squares and spatial maximum likelihood (Tables 2 and 3). As with our earlier reanalysis of Swank and Steinmo, we assume that Hays' spatial lag specification is the correct one. Overall, the results of our reanalysis with respect to the spatial lag are mixed. Given Hays' setup, the endogeneity and measurement-induced biases are pushing his coefficient estimates in different directions. This is clear from Table 2, which presents our renalaysis of the capital-account openness models. The original estimates are reported in the second column labeled "Spatial OLS" and "Symmetric Diffusion." By symmetric diffusion we mean that Hays used a spatial weighting matrix with off-diagonal elements that all take a value of 1/(N-1). In our reanalysis, we also include an asymmetric weighting matrix based on observed cross-national correlations in capital tax rates. For each country's row in the spatial weights matrix we enter ones for the countries with which its capital tax rates have a statistically significant positive correlation. We then row-standardize the spatial weighting matrix.[76] The weighting matrix is asymmetric because country 1's importance in determining country 2's capital tax rate may not be the same as Country 2's importance in determining Country 1's tax rate.[77]

We report non-spatial *OLS* estimates in the first column of Table 2 to demonstrate that there is omitted variable bias when the spatial lag is left out of the model. Most importantly, the non-spatial *OLS* estimate for the consensus democracy interaction term is about 35% smaller than the original *S-OLS* estimate and statistically insignificant. Two things worry us about Hays' original estimates in the second column. First, he uses *S-OLS*, which, because the spatial lag is endogenous, is likely to give an inflated estimate of the coefficient. This simultaneity bias can induce bias in the other coefficient estimates (Franzese and Hays 2004). Second, Hays uses an arbitrary spatial weighting matrix. Each country's capital tax rate in the sample is assumed to be important (and equally so) in determining every other country's tax rate. This convenient assumption gives a simple unweighted

---

[75] The original estimates are reported in Hays' Table 2.

[76] Standardization is done by replacing the ones in each country's row in the weighting matrix with 1/N, where N is the number of countries with which its tax rate is correlated. In other words, if a country's capital tax rate is positively correlated with five other countries, the appropriate cells in the weighting matrix take a value of 0.2.

[77] If Country 1's tax rate is correlated with five other countries and Country 2's tax rate is only correlated with Country 1, the importance of Country 1's tax rate (i.e., its weight in the spatial weighting matrix) in determining Country 2's tax rate will be greater than the reverse.

average of the capital tax rates in the other countries as the spatial lag. If this assumption is wrong, which it almost certainly is, the spatial lag contains measurement error, which may cause attenuation bias in the spatial-lag coefficient estimate.[78] Note that these biases work in opposite directions.

[78] There is no strong reason to think this measurement error would be systematic. See Franzese and Hays (2004) for a discussion of this attenuation bias.

Table 2. Capital Tax Rates and International Capital Mobility (Capital Account Openness)

Independent Variables	Capital Account Openness	Capital Account Openness	Capital Account Openness	Capital Account Openness	Capital Account Openness	Capital Account Openness
Capital Mobility	1.918**	2.223**	2.159**	1.620*	1.695*	1.729*
	(.919)	(.930)	(1.045)	(.859)	(.996)	(1.013)
Capital Mobility Interacted with:						
Capital Endowment	-.070*	-.069*	-.069**	-.033	-.0425	-.048
	(.040)	(.040)	(.033)	(.039)	(.030)	(.040)
Consensus Democracy	.484	.746*	.691	1.245***	1.053**	1.121**
	(.431)	(.434)	(.472)	(.428)	(.485)	(.534)
Corporatism	-1.186	-2.229	-2.008	-3.047**	-2.578*	-2.453
	(1.339)	(1.359)	(1.399)	(1.318)	(1.357)	(1.641)
Left Government	.370*	.286	.304	.304	.321	.331
	(.196)	(.195)	(.209)	(.186)	(.196)	(.215)
Population	-1.79e-07	-9.77e-06**	-7.74e-06*	5.79e-07	3.88e-07	.001
	(3.49e-06)	(3.98e-06)	(4.03e-06)	(3.30e-06)	(3.60e-06)	(.004)
European Union	-.204	-.465***	-.410**	-.520***	-.440**	-.442***
	(.161)	(.170)	(.185)	(.161)	(.176)	(.168)
Temporal Lag	.834***	.754***	.771***	.686***	.723***	.706***
	(.034)	(.039)	(.028)	(.043)	(.031)	(.038)
Spatial Lag		.280***	.221***	.316***	.237***	.267***
		(.066)	(.048)	(.049)	(.035)	(.044)
Obs.	465	465	465	465	465	465
Estimation	Non-spatial OLS	Spatial OLS	Spatial 2SLS	Spatial OLS	Spatial 2SLS	Spatial ML
Diffusion		Symmetric	Symmetric	Asymmetric	Asymmetric	Asymmetric

Notes: The regressions were estimated with fixed country effects. (Coefficients for country dummies not shown.)
Parentheses for the OLS estimates contain panel corrected standard errors.
Parentheses for the 2SLS estimates contain robust standard errors clustered by year.
Parentheses for the ML estimates contain robust standard errors.
** * Significant at 1%, ** Significant at 5%, * Significant at 10%

The estimates in the third and fourth columns are consistent with our expectations. When we estimate using *S-2SLS* the estimate on the spatial lag drops from .280 to .221 (a 21% reduction) and when we use the asymmetric spatial weighting matrix the estimate increases to .316 (+13%). Both "corrections" (asymmetric spatial weighting matrix and consistent estimator) are made in columns 5 (S-2SLS) and 6 (S-ML). Overall, these results, which are very similar across the two estimators, suggest that Hays overestimated the coefficient on the spatial lag (simultaneity bias) and underestimated the coefficients on the capital-mobility variable and the capital-mobility*consensus-democracy interaction variable (induced biases). In other words, because of the endogeneity of the spatial lag, Hays likely overestimated the importance of international factors (tax competition) at the expense of domestic (consensus democracy) and common external factors (capital mobility), which is what we would expect. Our reanalysis of the financial openness model in Table 3 tells a similar story.

Table 3. Capital Tax Rates and International Capital Mobility (Financial Openness)

Independent Variables	Financial Openness	Financial Openness	Financial Openness	Financial Openness	Financial Openness	Financial Openness
Capital Mobility	.858***	.988***	.958**	.725**	.758**	.741**
	(.338)	(.342)	(.359)	(.313)	(.345)	(.322)
Capital Mobility Interacted with:						
Capital Endowment	-.029*	-.034**	-.033**	-.024*	-.025**	-.028*
	(.015)	(.016)	(.013)	(.014)	(.011)	(.014)
Consensus Democracy	.209	.306*	.283*	.422***	.369**	.369**
	(.154)	(.157)	(.165)	(.151)	(.161)	(.168)
Corporatism	-.534	-.817*	-.751	-.888**	-.799	-.656
	(.471)	(.490)	(.603)	(.451)	(.567)	(.612)
Left Government	.099*	.085	.088	.089*	.0916	.095
	(.054)	(.054)	(.057)	(.051)	(.055)	(.059)
Population	2.45e-07	-2.10e-06*	-1.55e-06	2.13e-07	2.21e-07	.000
	(1.04e-06)	(1.23e-06)	(1.11e-06)	(9.83e-07)	(9.79e-07)	(.001)
European Union	-.075*	-.148***	-.131**	-.156***	-.136***	-.131***
	(.046)	(.050)	(.050)	(.045)	(.045)	(.044)
Temporal Lag	.825***	.751***	.768***	.682***	.718***	.702***
	(.036)	(.040)	(.030)	(.044)	(.031)	(.038)
Spatial Lag		.261***	.200***	.309***	.231***	.261***
		(.066)	(.053)	(.047)	(.036)	(.043)
Obs.	465	465	465	465	465	465
Estimation	Non-spatial OLS	Spatial OLS	Spatial 2SLS	Spatial OLS	Spatial 2SLS	Spatial ML
Diffusion		Symmetric	Symmetric	Asymmetric	Asymmetric	Asymmetric

Notes: The regressions were estimated with fixed country effects. (Coefficients for country dummies not shown.)
Parentheses for the OLS estimates contain panel corrected standard errors.
Parentheses for the 2SLS estimates contain robust standard errors clustered by year.
Parentheses for the ML estimates contain robust standard errors.
** * Significant at 1%, ** Significant at 5%, * Significant at 10%

First, non-spatial *OLS* produces serious omitted variable bias (Column 1, Table 3). And second, Hays (2003) probably overestimates the coefficient on the spatial lag and underestimates the coefficients on the capital mobility and consensus democracy interaction variables (Columns 5 and 6 vs. Column 2).

Next, we do a reanalysis of several regressions from Basinger and Hallerberg's Tables 3, 4, and 5. We focus on the regressions that use the Mendoza et al. capital tax rates and levels of fixed capital formation (FCF) for the dependent variable and spatial weights matrix respectively.[79] In these regressions, the dependent variable is the one period change and the spatial lag is differenced and lagged one period. This particular specification raises the concern that unmodeled common shocks are causing Basinger and Hallerberg to underestimate the degree of spatial interdependence. We address this potential problem in two ways—we estimate the model using S-2SLS, which purges the spatial lag of it correlation with the omitted common shocks, and we estimate a model with fixed period effects. The results are reported in these tables:

[79] Our results are similar across dependent variables and weighting schemes.

Table 3. Conditional Coefficients for the Effects of a Change in Tax Rates on Capital in Competitor Countries Given Domestic Partisanship Level (Mendoza et al. Tax Rates, GDP Weights)

Partisanship	*APSR*			
Regression Coefficients				
Coefficient, change in competitor	-.24	1.39*	-.83	-.11
countries	(.40)	(.71)	(.89)	(.21)
Coefficient, change in competitor	.67	-2.13*	1.47	.40
countries*distance, party, cap controls	(.63)	(1.21)	(1.52)	(.37)
0 partisanship	-.24	1.39*	-.83	-.11
(no country)	(.40)	(.71)	(.89)	(.21)
0.2 partisanship	-.10	.96*	-.54	-.03
(Norway, 1989)	(.27)	(.48)	(.60)	(.14)
0.4 partisanship	.03	.53*	-.24	.05
(Netherlands, 1982-88)	(.16)	(.28)	(.32)	(.07)
0.6 partisanship	.16*	.10	.05	.13*
(Austria, 1987-97)	(.07)	(.21)	(.19)	(.06)
0.8 partisanship	.30*	-.32	.35	.21*
(Ireland, 1990-92)	(.13)	(.36)	(.39)	(.11)
Estimator	S-OLS	S-OLS	S-ML	S-ML
	Incorrect	Correct	Correct	Binary
Weights Matrix	GDP	GDP	GDP	Contiguity
	Weights	Weights	Weights	Weights

Notes: See Basinger and Hallerberg (2004), Table 4.

Table 4. Conditional Coefficients for the Effects of a Change in Tax Rates on Capital in Competitor Countries Given Domestic Use of Capital Controls (Mendoza et al. Tax Rates, GDP Weights)

Capital Controls	*APSR*			
Regression Coefficients				
Coefficient, change in competitor	.26*	.11	.16	.08
countries	(.10)	(.28)	(.21)	(.06)
Coefficient, change in competitor	-.96	.60	-1.88	.20
countries*distance, party, cap controls	(.76)	(1.86)	(1.51)	(.33)
0 capital controls	.26*	.11	.16	.08
(United States, 1980-97)	(.10)	(.28)	(.21)	(.06)
0.25 capital controls	.02	.26	-.31	.13*
(France, 1980-89)	(.14)	(.35)	(.33)	(.07)
0.5 capital controls	-.22	.41	-.78	.18
(Portugal, 1980-85)	(.32)	(.77)	(.68)	(.14)
0.75 capital controls	-.46	.56	-1.25	.23
(Greece, 1981)	(.50)	(1.23)	(1.05)	(.22)
Estimator	S-OLS	S-OLS	S-ML	S-ML
Weights Matrix	Incorrect GDP Weights	Correct GDP Weights	Correct GDP Weights	Binary Contiguity Weights

Notes: See Basinger and Hallerberg (2004), Table 5.

As expected, we find that Basinger and Hallerberg underestimate the sensitivity of countries' capital tax policies to changes in competitor countries when the cost of capital tax reform domestically is low. Specifically, when the ideological distance between veto players is 0, the coefficient estimate for the spatial lag increases to .33 and .81 (from .21) with the S-2SLS and fixed period estimators respectively. When the government leans right (partisanship score of 0.8), these same estimate increases from .30 to .57 and .84. For countries with no capital controls, the estimate rises from .23 to .37 and .79.

B. Learning & Emulation: Volden (AJPS 2006) on US State Adoption of Policy Innovations

To help illustrate how these tests can be used in empirical research, we present OLS estimates for a non-spatial model of welfare policy generosity in column 1 of Table 1. All variables in our illustrative analysis are states' averages over the five years 1986-1990. The dependent variable is the maximum monthly AFDC benefit, and the independent variables are the state's poverty rate, average monthly wage in the retail sector, government ideology (ranging from 0=conservative to 100=liberal), degree of interparty competition (ranging from .5=competitive to 1.0=non-competitive, tax effort (revenues as a percentage of tax "capacity"), and the portion of AFDC benefits paid by the federal government. We use a standardized binary contiguity-weights matrix, which begins by coding wij = 1 for states i and j that share a border and wij = 0 for states that do not border. Then, we row-standardize (as commonly done in spatial-econometrics) the resulting matrix by dividing each cell in a row by that row's sum. This gives the unweighted average of the dependent variable in "neighboring" (so-defined) states.

Table 1. State Welfare Policy (Maximum AFDC Benefit)

Independent Variables	OLS	Spatial AR Lag (S-OLS)	Spatial AR Lag (S-2SLS)	Spatial AR Lag (S-GMM)	Spatial AR Lag (S-MLE)	Spatial AR Error (S-MLE)
Constant	54.519	-246.76	-422.09	-500.05	-156.282	676.120
	(531.830)	(450.75)	(437.74)	(413.02)	(429.130)	(471.965)
Poverty Rate	-6.560	8.04	13.205	7.29	3.657	3.239
	(11.262)	(10.022)	(9.977)	(8.452)	(8.917)	(10.062)
Retail Wage	-.121	.016	.089	-.008	-.025	-.344
	(.226)	(.193)	(.187)	(.201)	(.181)	(.243)
Government	1.513	1.397	1.359*	1.655**	1.432*	1.696**
Ideology	(1.030)	(.863)	(.825)	(.761)	(.806)	(.822)
Inter-party	621.799**	368.65	286.98	438.9**	444.677*	263.887
Competition	(290.871)	(250.55)	(243.72)	(197.47)	(226.911)	(238.419)
Tax Effort	3.357**	2.022	1.553	2.397	2.423*	2.936**
	(1.587)	(1.364)	(1.328)	(1.493)	(1.262)	(1.213)
Federal Share	-4.405	-5.818	-6.012	-3.654	-5.393	-6.882*
	(5.001)	(4.20)	(4.014)	(3.415)	(3.901)	(4.099)
Spatial AR		.767***	1.069***	.840***	.537***	.565***
		(.178)	(.232)	(.237)	(.122)	(.131)
Moran I-statistic	3.312***					
$LM_{\rho\lambda}$	12.322***					
LM_{ρ}	11.606***					
LM_{ρ}^{*}	6.477**					
LM_{λ}	5.845***					
LM_{λ}^{*}	.716					
Log-likelihood					-270.763	-272.728
Adj.-R^2	.461	.622	.595	.606	.510	.588
Obs.	48	48	48	48	48	48

Notes: The spatial lags are generated with a binary contiguity weighting matrix. All the spatial weights matrices are row-standardized. ***Significant at the 1% Level; **Significant at the 5% Level; *Significant at the 10% Level.

The results for our non-spatial model suggest that high tax effort and low party competition are associated with more generous AFDC benefit payments. This seems reasonable. However, if the data exhibit spatial dependence, we need to worry about validity of these inferences. To check this possibility, we implement the diagnostic tests outlined above starting with Moran's I. The value of the standardized Moran-I test statistic is 3.312, which is statistically significant. We can reject the null hypothesis of no spatial dependence. We also include the LM tests. The result of the two-directional test leads to the same conclusion. Both the standard one-directional tests seem, predictably, statistically significant, which, unfortunately, gives us little guidance for specification. As expected, the robust one-directional tests are more helpful in this regard. The robust test against the spatial lag alternative is statistically significant while the robust test against the spatial error alternative is not. This suggests a spatial lag specification. We conclude with a warning. Ignoring evidence of spatial dependence can be extremely problematic, especially if the data suggest the true source of dependence is a spatial-lag process. In this case, simple OLS is likely to provide inaccurate coefficient estimates, particularly for variables that happen to cluster spatially (e.g., Franzese and Hays 2004, 2006a, 2007b).

The S-OLS estimates are provided in column 2 of Table 1. Consistent with the results from our diagnostic tests, the estimated coefficient on the spatial lag is large, positive and statistically significant. The OLS estimates most affected by the switch to a spatial-lag specification are the party-competition and tax-effort coefficients, which become statistically insignificant. Conversely to S-OLS's simultaneity biases, the OLS coefficient estimates on these two variables may, because they cluster spatially, have suffered from omitted variable bias that would have inflated those estimates.

We present the S-2SLS and S-GMM estimates for the spatial-lag model of welfare policy generosity in columns 3 and 4 of Table 1. The S-2SLS estimates for this particular specification and dataset are troubling as the spatial-lag coefficient estimate exceeds one, giving a non-stationary spatial process. This is a bit surprising when compared with the smaller S-OLS result, given that the S-OLS estimator has likely-inflationary simultaneity biases and S-2SLS likely does not. Of course, this can happen with a single sample and/or if the exogeneity of the instruments is violated.[80] The S-GMM estimates are better. The spatial-lag coefficient estimate is well below one (though it is still large) and the standard errors are about 5% smaller than the S-2SLS standard errors on average, as expected given the likely efficiency of the GMM estimator. The coefficients on government ideology and on party competition are statistically significant. The results suggest that, *ceteris paribus*, welfare benefits are highest in states with non-competitive party systems and liberal governments.

The S-ML estimates for our spatial lag model of welfare policy generosity are provided in column 5 of Table 1. These estimates are mostly similar to the S-GMM estimates. The most noteable difference is in the estimate of ρ. The S-ML coefficient is approximately 36% smaller than the S-GMM coefficient, and it is estimated much more precisely, the standard error being about half the size of the S-GMM standard error. Three of the coefficients in this model are statistically significant including the tax effort coefficient. The S-ML estimates imply welfare benefits are systematically larger, all else equal, in states with high taxes, liberal governments, non-competitive party systems. Franzese and Hays (2004, 2006a, 2007b) find that S-ML generally outperforms S-2SLS on mean squared error grounds. S-GMM lessens the efficiency advantage for S-ML over the IV class of

[80] Franzese and Hays (2004) show that the exogeneity at issue here is that the y_i must not cause the x_j, a condition we call (no) cross-spatial endogeneity. Such reverse "diagonal" causality seems unlikely to arise in many substantive contexts, although we do note also that spatial correlation among the other regressors plus the typical endogeneity from y to x would create it.

estimators.

The S-ML estimates for the spatial error model of welfare policy generosity are provided in the last column of Table 1. We note only that the log-likelihood value for the error model is less than the log-likelihood for the lag model, and this is consistent with the robust LM specification test results.

We estimate several spatial lag probits in Table 2 using both standard ML and MCMC methods. In keeping with our state welfare spending example, we switch the dependent variable from maximum AFDC benefits to whether or not a state's CHIP (Children's Health Insurance Program) includes a monthly premium payment (Volden 2006). We keep the same independent variables since this dependent variable also reflects the generosity of the welfare program.

Table 2. State Welfare Policy (Monthly CHIP Premium)

Independent Variables	Probit MLE	Probit MCMC	Spatial AR Lag Probit	Spatial AR Error Probit
Constant	-4.978	-5.163	-5.606	-5.531
	(6.260)	(6.292)	(10.159)	(7.337)
Poverty Rate	-.244	-.265**	-.374**	-.243*
	(.153)	(.156)	(.231)	(.157)
Retail Wage	.004	.004*	.006*	.004*
	(.003)	(.003)	(.004)	(.003)
Government Ideology	.011	.011	.014	.014
	(.013)	(.013)	(.020)	(.014)
Inter-party Competition	2.174	2.108	1.473	2.636
	(3.388)	(3.478)	(6.134)	(3.794)
Tax Effort	-.014	-.014	-.020	-.017
	(.019)	(.019)	(.034)	(.021)
Federal Share	.045	.048	.065	.043
	(.063)	(.064)	(.095)	(.066)
Spatial AR	.079	.102	.200***	.297***
	(.798)	(.815)	(.148)	(.196)
Pseudo-R^2	.222	.220	.607	.574
Obs.	48	48	48	48

Notes: In the first two columns, the models are estimated assuming the spatial lags are exogenous. The model in the first column is estimated using standard ML techniques. The parentheses in this column contain estimated standard errors and the hypothesis tests assume that the asymptotic t-statistics are normally distributed. The models in columns two through four are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient estimate is the mean of the posterior density based on 10,000 observations after a 1000 observation burn-in period. The number in parentheses is the standard deviation of the posterior density. The p-values are also calculated using the posterior density. The last two models are estimated with true spatial estimators described in the text. In third column, 30 of the 10,000 spatial AR coefficients sampled from the posterior distribution were negative. In the fourth column, none of the 10,000 sampled spatial AR coefficients were negative. ***p-value <.01, **p-value<.05, *p-value <.10.

In the first two columns, the models are estimated assuming the spatial lags are exogenous. The model in the first column is estimated using standard ML techniques. The parentheses in this column contain estimated standard errors and the hypothesis tests assume that the asymptotic t-statistics are normally distributed. The models in columns two and three are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient estimate is the mean of the posterior distribution based on 10,000 observations after a 1000 observation burn-in period. The number in parentheses is the standard deviation of the posterior distribution. The p-values are also calculated using the posterior. The results in columns two and three are very similar, as they should be given our diffuse priors. Because the estimator used in column two incorrectly treats the spatial lag as exogenous (i.e., like any other right-hand-side variable) the likelihood is misspecified and the sampler draws from the wrong posterior distribution for the spatial coefficient ρ. This specification error has serious consequences for drawing inferences about the importance of spatial interdependence.

The model in column three is estimated with the true spatial estimator described above. The draws for ρ are taken from the correct (non-standard) posterior distribution using Metropolis-Hastings. In this case only 30 of the 10,000 spatial AR coefficients sampled from the posterior distribution were negative. Thus, there is strong evidence of positive spatial interdependence in states' decisions to include a monthly premium in their CHIP. In addition, these probit results suggest that a state's poverty rate and average monthly retail wage are also important determinants.

We present estimates for the spatial error model in column four of Table 2. In this case, none of the 10,000 sampled spatial AR coefficients were negative. We do not discuss specification tests (*lag* vs. *error*) for the spatial probit, but not that they are covered in Anselin (2006).

We present estimates for a panel model of welfare policy generosity in Table 3. The data are annual observations from 1981-1990 on the contiguous 48 states. The dependent variable is the maximum AFDC benefit, and the independent variables remain unchanged. All the regressions include fixed state effects. The first column contains a non-spatial model estimated with OLS. Clearly, from Moran's I statistic and the two-directional LM statistics, there is spatial dependence in the dataset. The diagnostics do not provide clear evidence in favor of a spatial lag or error specification, however. We estimate both with contemporaneous spatial lags. The second column contains a spatio-temporal lag model, and the third column contains a combined temporal lag and spatial error model. Interestingly, the retail wage variable is statistically significant and positive in all three regressions. Once again, the tax effort coefficient becomes statistically insignificant with the change from a non-spatial to spatial specification.

Table 3. State Welfare Policy (Maximum AFDC Benefit, 1981-1990)

Independent Variables	OLS	Spatial AR Lag (MLE)	Spatial AR Error (MLE)
Poverty Rate	-.855	-.911	-.903
	(1.130)	(1.050)	(1.198)
Retail Wage	.217***	.204***	.197***
	(.036)	(.034)	(.037)
Government Ideology	.053	.059	.027
	(.087)	(.081)	(.083)
Inter-party Competition	18.960	25.540	18.633
	(24.046)	(22.442)	(22.382)
Tax Effort	.388*	.322	.349
	(.223)	(.208)	(.218)
Federal Share	.483	.859*	.750
	(.521)	(.491)	(.510)
Temporal AR	.663***	.628***	.666***
	(.030)	(.031)	(.030)
Spatial AR		.143***	.200***
		(.044)	(.058)
Moran I-statistic	3.296***		
$LM_{\rho\lambda}$	11.896***		
LM_{ρ}	9.976***		
LM_{ρ}^{*}	1.446		
LM_{λ}	10.450***		
LM_{λ}^{*}	1.921		
Log-likelihood		-1991.357	-1991.290
Adj.-R^2	.981	.981	.982
Obs.	480	480	480

Notes: 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. All the spatial weights matrices are row-standardized. ***Significant at the 1% Level; **Significant at the 5% Level; *Significant at the 10% Level.

To calculate marginal spatio-temporal effects (non-cumulative) or plot the over-time path of the effect of a permanent one-unit change in an explanatory variable (cumulative), and their standard errors, simply solve for $\mathbf{y}$ in (53):

$$\begin{aligned}\mathbf{y} &= \rho\mathbf{W}\mathbf{y}+\phi\mathbf{V}\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\\ &=\left(\rho\mathbf{W}+\phi\mathbf{V}\right)\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\\ &=\left[\mathbf{I}_{\mathrm{NT}}-\rho\mathbf{W}-\phi\mathbf{V}\right]^{-1}\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\\ &\equiv\mathbf{M}\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\end{aligned} \tag{93}$$

Denote the i^{th} column of $\mathbf{M}$ as $\mathbf{m}_i$ and its estimate as $\hat{\mathbf{m}}_i$. The spatial effects of a one-unit increase in explanatory variable *k* in country *i* are $\mathbf{m}_i\beta_k$ with delta method standard errors calculated as

$$\text{var}\left(\hat{\mathbf{m}}_i\hat{\beta}_k\right)=\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\text{var}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]' \quad , \tag{94}$$

where $\hat{\boldsymbol{\theta}}=\begin{bmatrix}\hat{\rho}\\ \hat{\phi}\\ \hat{\beta}_k\end{bmatrix}$, $\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]=\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}} \quad \frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\phi}} \quad \hat{\mathbf{m}}_i\right]$, and the vectors $\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}}$ and $\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\phi}}$ are the i^{th} columns of $\hat{\beta}_k\hat{\mathbf{M}}\mathbf{W}\hat{\mathbf{M}}$ and $\hat{\beta}_k\hat{\mathbf{M}}\hat{\mathbf{M}}$ respectively. In Table 4, we present the immediate and long-run (steady-state) spatial effects on regional AFDC benefits from a permanent \$100 increase to monthly retail wages in Missouri using the calculations in equations (93) and (94). The immediate (steady-state) effects range from a low of \$0.44 (\$3.68) in Kentucky to a high of \$0.77 (\$6.38) in Kansas.

Table 4. Spatial Effects on AFDC Benefits from a \$100 Counterfactual Shock to Monthly Retail Wages in Missouri

Neighbor	Immediate Spatial Effect	Long-Run Steady State Effect
Arkansas	.51 [.16,.87]	4.26 [1.01,7.52]
Illinois	.62 [.19,1.04]	5.11 [1.25,8.97]
Iowa	0.52 [.15,.88]	4.37 [.99, 7.75]
Kansas	0.77 [.23,1.31]	6.38 [1.60,11.17]
Kentucky	0.44 [.13,.75]	3.68 [.87,6.50]
Nebraska	0.52 [.15,.89]	4.44 [.99,7.90]
Oklahoma	0.52 [.15,.89]	4.47 [.96,7.98]
Tennessee	0.38 [.12,.65]	3.21 [.75,5.67]

Notes: Effects calculated using estimates from the spatial AR lag model in Table 3. Brackets contain a 95% confidence interval.

In Figure 1, we present the spatio-temporal effects on AFDC benefits in Missouri from a permanent \$100 increase to monthly retail wages in Missouri (with 95% C.I.). The marginal effects decay rapidly with most of the total effect experienced within the first 2 years after the shock. The cumulative 10-year effect is approximately \$55.75. In Figure 2, we present the spatio-temporal effects on AFDC benefits in Nebraska from a \$100 counterfactual increase to monthly retail wages in Missouri (with 95% C.I.). The cumulative 10-year effect is about \$4.11. Interestingly, the maximum effects in Nebraska are not experienced until one or two years after the initial shock. This serves to highlight an important point, namely the contemporaneous spatial lag specification does not imply all (or even most) of the spatial effects are instantaneous.

Figure 1. Spatio-Temporal Effects on AFDC Benefits in Missouri from a $100 Counterfactual Shock to Monthly Retail Wages in Missouri (with 95% C.I.)

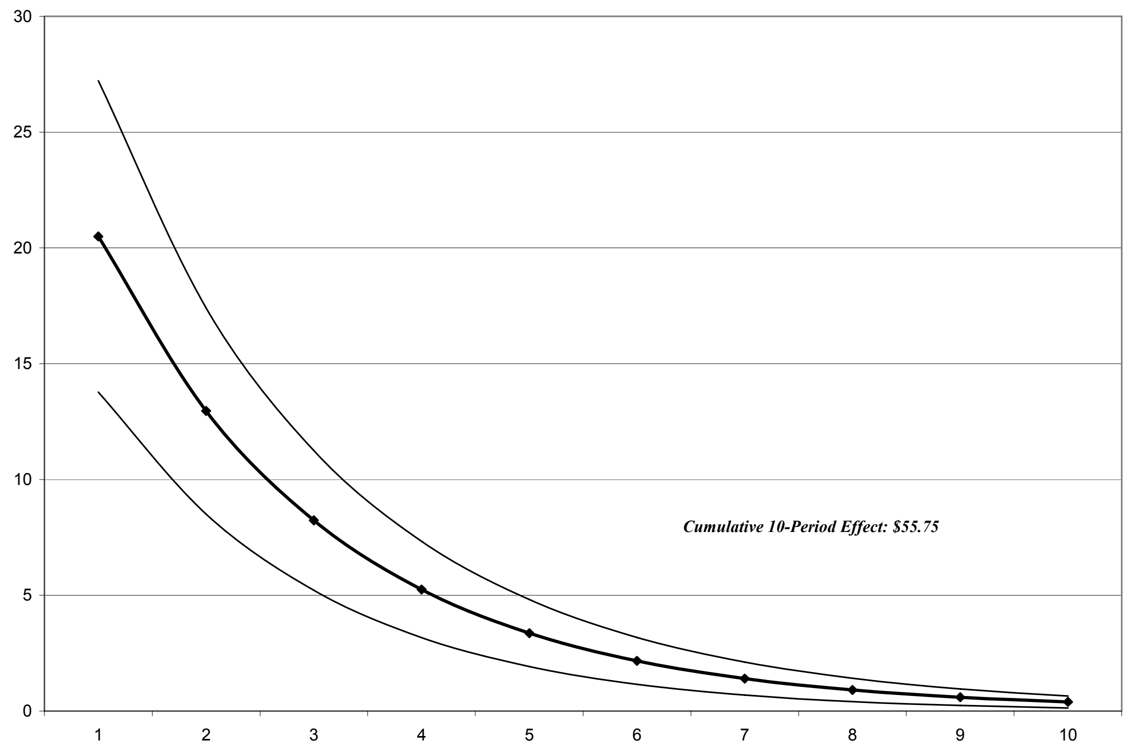
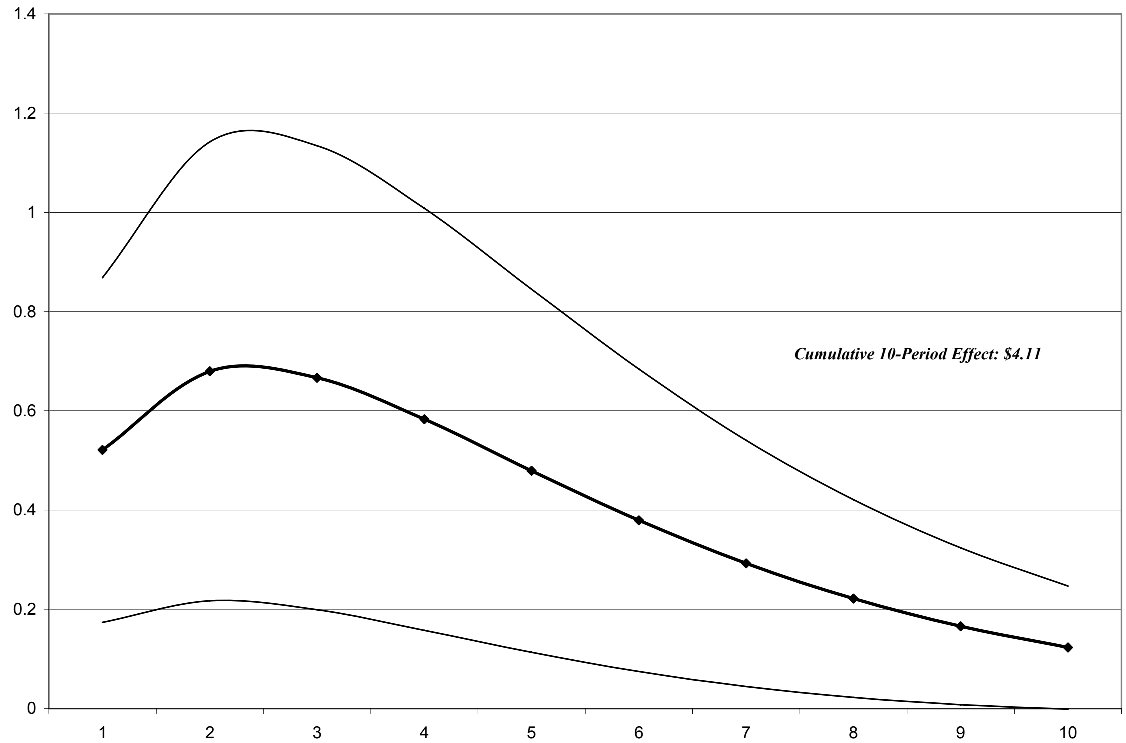


Figure 2. Spatio-Temporal Effects on AFDC Benefits in Nebraska from a $100 Counterfactual Shock to Monthly Retail Wages in Missouri (with 95% C.I.)



C. Competition and Negative Externalities: Beck et al. (ISQ 2006) on International Conflict and Trade

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 theretoforeapparent lack of an unambiguously superior estimator in such data,[81] 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[82]). 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 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.[83] 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

[81] 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.

[82] 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.

[83] 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.

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.[84] Our coefficient estimate for the temporal lag is also 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.

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

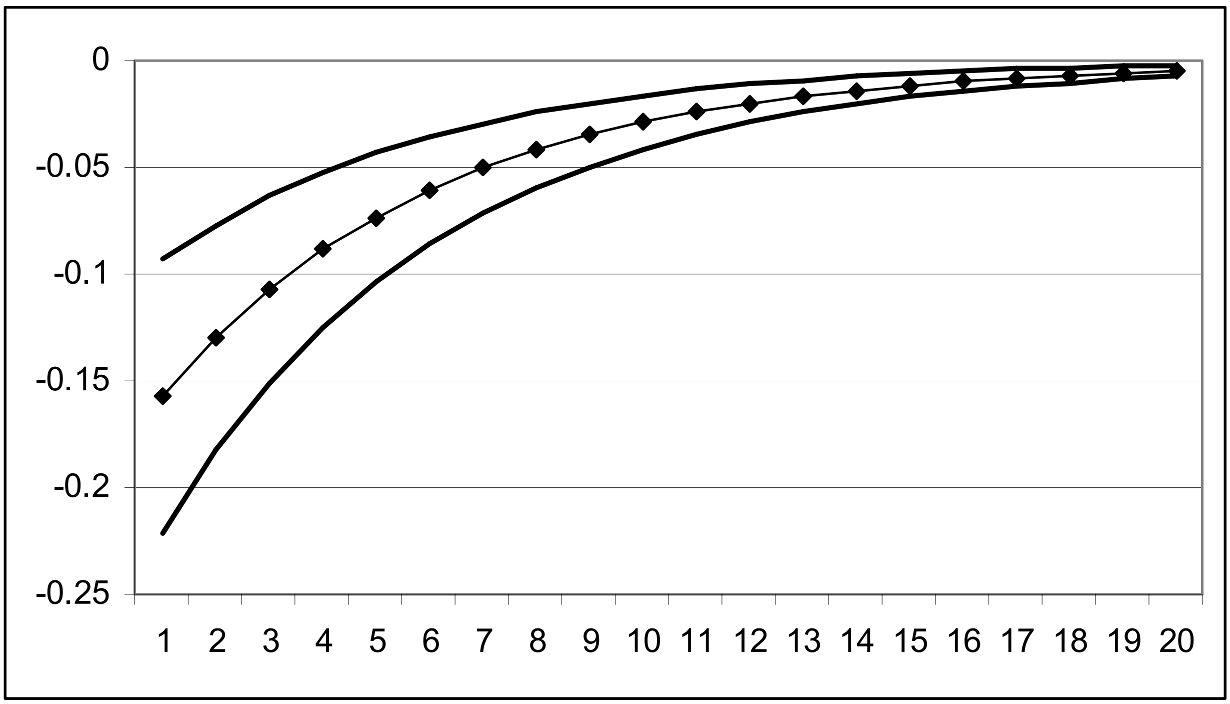
Spatio-Temporal Model Estimates					
$R^2 = 0.9702$				$\bar{R}^2 = 0.9687$	
$\hat{\sigma}^2 = 0.0939$				N, k = 2565, 120	
LL = 282.677				# of iterations = 12	
Variable	**BG&B Coeff**	**Our Coeff**	**BG&B s.e.**	**Our s.e.**	**p-level**
constant	0.17	-0.5888	(0.11)	0.6130	0.3367
Ln GNP A	0.02	*0.0276*	(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

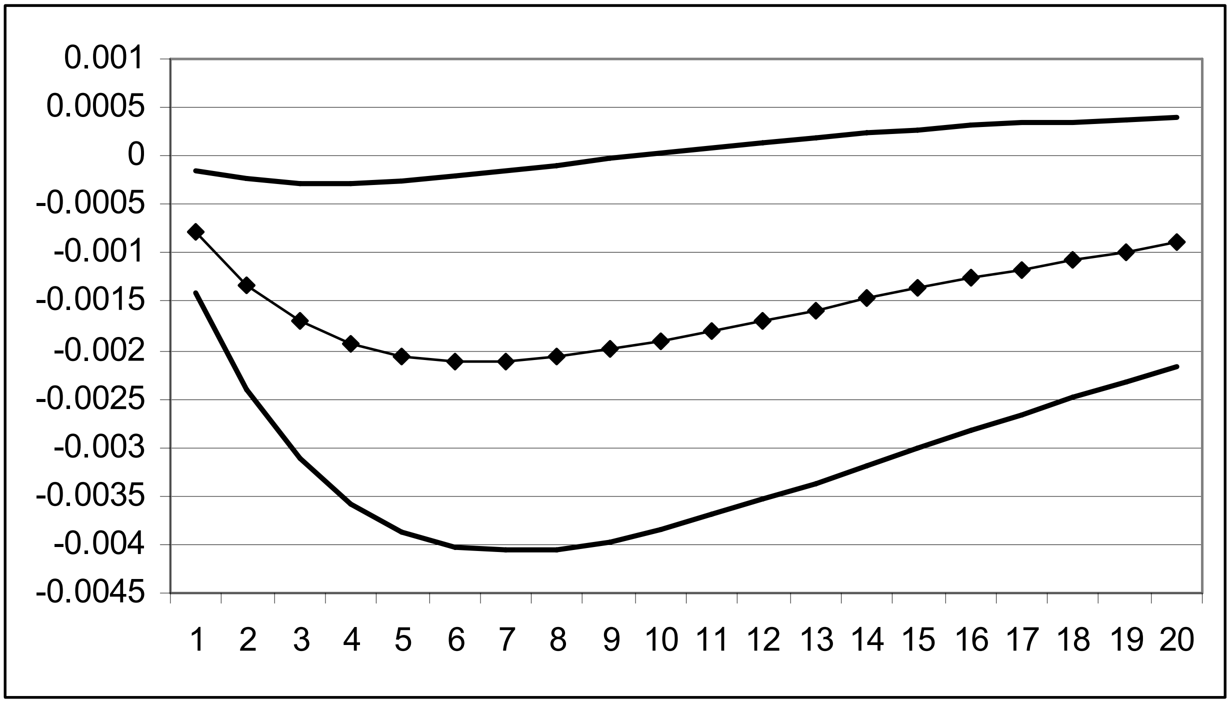
[84] Due to political division and reunification, distance, a seemingly time invariant measure, does change for the dyads including Germany.

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 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.

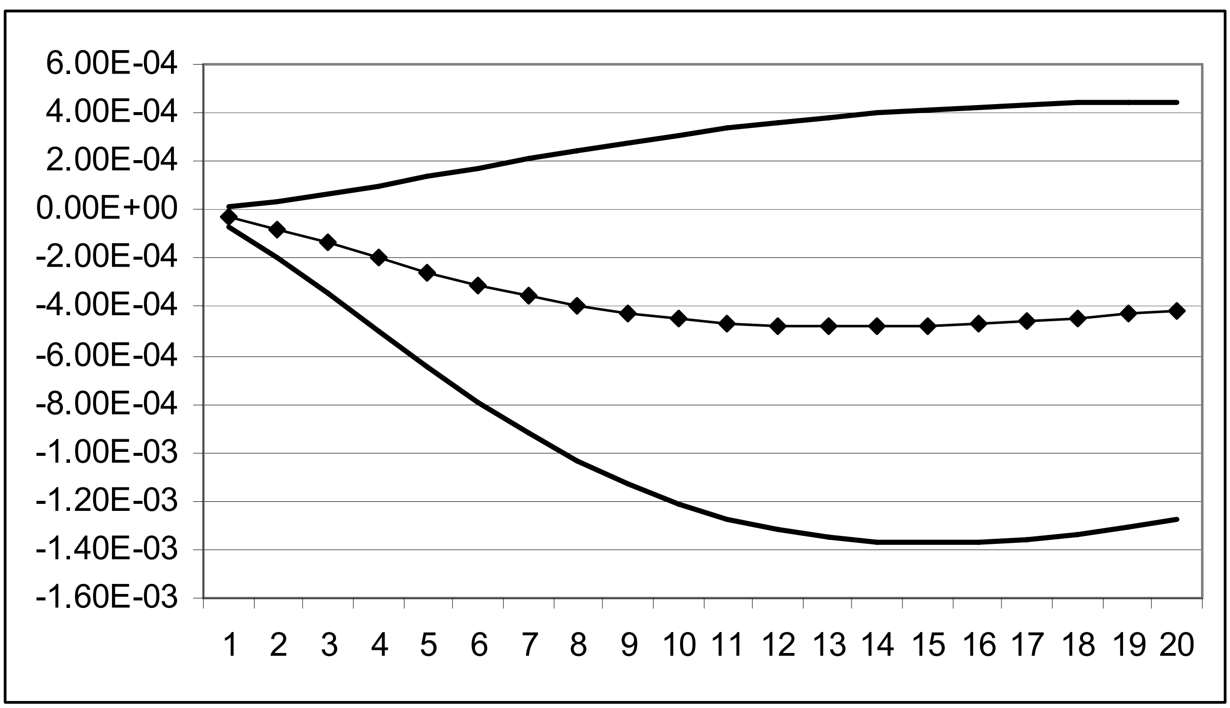
Figure 1: Temporal Effects with Spatial Feedback (E.g., US Exports to Russia response to US-Russia MID)



<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)



D. Cooperation & Positive Externalities: Franzese & Hays (EUP 2006) on ALMP in the EU

Five years ago at the Lisbon Summit the EU committed to becoming "the most competitive and dynamic knowledge-based economy in the world by 2010," (European Council 2000: 1). Active labor market (ALM) policies are a critical part of the plan designed to achieve this objective, the European Employment Strategy (EES). ALM programs are supposed to improve job seekers' prospects of finding employment and increase the productivity and earning potential of workers. They include spending on public employment, labor market training, and other policies intended to promote employment among the unemployed. While ALM policies—particularly training and education programs—seem almost inherently necessary to create the kind of workforce and economy EU leaders envisage, coordinating these policies through an EES system that relies heavily on the principle of subsidiarity, may be problematic. Subsidiarity in the EES implies that member states create their own programs and implement them on a mostly voluntary basis, yet individualistic voluntarism leaves policy susceptible to positive-externality induced underinvestment. Has this theoretically possible negative interdependence of European ALM policies actually arisen empirically? If so, are these spillovers and the detrimental interdependence they induce sufficiently sizable to warrant concern and redress?

We argue and present evidence that ALM policies do indeed entail significant externalities that spill across national boundaries and that, apparently, these spillovers are sufficiently sizable to generate appreciable political and economic incentives for European governments to free ride off the efforts of their neighbors. That is, we provide empirical evidence that the national best-response functions for ALM spending, worker training programs in particular, are statistically significantly and substantively appreciably downward-sloping: an increase in expenditures in one country decreases equilibrium expenditures in its neighbors. This leads us to conclude that current levels of ALM expenditures may indeed be too low and that, apparently, the limited (although increasing) coordination of the EES framework is insufficient to internalize positive ALM policy externalities noticeably. Stronger enforcement procedures would seem to be necessary if the European Union is to achieve its EES objectives.

The paper structures these explorations as follows. In the first section, we briefly review the history of the EES starting with the Luxembourg Jobs Summit. We cover the generic theory of strategic policy complementarity and substitutability (positive and negative externalities, respectively) in section two. Section three contains our empirical analysis, and sections four and five discuss the results and offer our conclusions, respectively.

Historical Overview of the European Employment Strategy

The key elements of the European Employment Strategy (EES), adopted by EU governments in November 1997 at the Luxembourg Jobs Summit, are contained within the Amsterdam Treaty's Title on Employment (see Goetschy, 1999). This section of the treaty, among other things, makes unemployment a common European concern, places job creation alongside macroeconomic stability as one of the EU's primary objectives, and creates an EU-level institutional mechanism for the oversight and evaluation of member states' employment policies. Since the Luxembourg Summit, the objectives and coordination procedures of the EES have been refined at several European Councils.

For our interests in subsidiarity and policy coordination, some of the most important changes came at the Berlin Council in 1999 where EU member states decided to use Structural Funds to finance EES programs aimed at developing human resources.[85] Regulation (EC) No 1260/1999, in addition to incorporating this change and setting new objectives, created a system of ex-ante, mid-term, and ex-post program evaluations combined with performance rewards and punishments (i.e. performance reserve-allocations). As a result of this new regulation, the EU budget and employment promotion processes became entwined under "Objective 3" (i.e. human resource development) funding. It also established a limited system of centralized enforcement with respect to employment policy coordination.

The (annual) coordination cycle of the EES follows four steps. Each year the European Council adopts a set of guidelines—developed by the European Commission—for EU member states' employment policies. These guidelines are intended to be instrumental in achieving full employment, improved quality and productivity at work, and labor market inclusiveness. National governments respond by writing action plans to describe how these guidelines are being (or will be) implemented domestically. The Commission and Council then review these plans and publish a joint economic report. If necessary, the Council makes country-specific recommendations.

In 2004, the most recently completed cycle, the Council promulgated the following

---

[85] The management of EU Structural Funds is guided by the principle of 'concentration' which holds that assistance should be focused on achieving a few core objectives. At Berlin the number of objectives was cut from seven to three: 1) support for the poorest regions, 2) conversion of regions facing structural difficulties, and 3) human resource development (European Commission, 1999).

guidelines for national employment policies: they should increase the adaptability of workers to changing conditions, attract more people into the labor market, and promote investment in human capital. Community-wide and country-specific recommendations that stressed the importance of labor-market training-programs were issued (European Commission, 2004). In fact, the Council encouraged each of the original fifteen members to increase participation in such programs.

Since 1988 the Council has set a multi-annual EU budget in its 'Fiscal Perspectives' report. The current budget agreement (the 'Agenda 2000' plan), which covers spending from 2000 to 2006, is the first governed by Regulation (EC) No 1260/1999. Midterm reviews, including reviews of programs financed under "Objective 3", were conducted by member states in 2003 and submitted to the Commission by 31 December. On the basis of these reviews, the Commission allocated the performance reserves on 23 March 2004.

Despite the ostensible strengthening of the EES framework over time, employment policies (ALM programs in particular) remain primarily the prerogative of national governments. The Council sets guidelines, but member countries choose their own response and, with the minor exceptions of country performance reports and performance reserve allocations, no enforcement mechanism is in place should they fail to follow through. Thus, the situation post-Luxembourg is not fundamentally different from the one existing prior to the 1997 Council meeting.

How have EU member governments fared in the provision of ALM policies? Outside of Scandinavia perhaps, the consensus seems to be that EU member governments are behind in designing and implementing policies to upgrade the skills of their workers.[86] Strategic interdependence among European countries in the making of active labor market policies could explain this. Two kinds of interactions in particular, *race-to-the-bottom* dynamics and *policy free-riding*, would induce suboptimal expenditures on employment policies.

Race-to-the-Bottom Dynamics and Policy Free Riding

In theory, race-to-the-bottom (RTB) dynamics occur when policies are strategic complements across jurisdictions—that is, when policy changes adopted in one jurisdiction create incentives for other jurisdictions to adopt similar changes. The RTB argument has been applied *inter alia* to capital taxation, environment regulations, and labor standards. Cuts in taxes and the elimination of

[86] The 2004 Joint Economic Report asked six of the original fifteen members to strengthen their ALM policies. Five of the six later received a C-grade for their response (partial and limited). One received a B (in progress). The Council asked every member country to improve its investment in human capital in one or more ways. The modal response of member governments to these recommendations was "partial and limited" (European Commission, 2005). See Murray and Wanlin (2005) for another disappointing report card.

regulations and standards in one jurisdiction increase the costs to others of maintaining high taxes, regulations, and standards causing the effected jurisdictions to follow suit in their own policies. By contrast, free riding occurs when policies are strategic substitutes—that is, when policy changes in one jurisdiction create incentives for governments in others to adopt change in the opposite direction. For example, an increase in defense expenditures in one country might lower the marginal security benefit from defense spending in its military allies, creating an incentive for them to free ride (see Redoano, 2003).

More formally, consider a two-country world (i,j), each with homogenous populations and domestic welfare that, due to externalities, are a function of government policy in both countries:

$$W^i \equiv W^i(p_i, p_j) \quad ; \quad W^j \equiv W^j(p_j, p_i) \qquad (1).$$

When the government in country *i* chooses its policy, p_i, to maximize its own social welfare, this affects the optimal policy-choice in country *j*, and *vice versa*. We can express such *strategic interdependence* between countries *i* and *j* with a pair of *best-response functions*, giving optimal policies of *i*, p_i^*, as a function of the policy chosen in *j*, and *vice versa*:

$$p_i^* \equiv \mathrm{Argmax}_{\mathrm{p_i}} W^i(p_i, p_j) \equiv R(p_j) \quad ; \quad p_j^* \equiv \mathrm{Argmax}_{\mathrm{p_j}} W^j(p_j, p_i) \equiv R(p_j) \qquad (2).$$

Explicitly, country *i*'s optimum policy is obtained by maximizing $W^i(p_i, p_j)$ with respect to p_i, taking p_j as given (fixed); i.e. setting the first derivative of the welfare function with respect to p_i equal to zero and solving for p_i^* as a function of p_j (and then verifying that the second derivative is negative). Equation (2) expresses the result as the best-response function $p_i^* = R(p_j)$. The slopes of these best-response functions, the signs of which determine whether RTB or free-riding dynamics will occur, depend on the following ratios of second cross-partial derivatives:

$$\frac{\partial p_i^*}{\partial p_j} = -W^i{}_{p_i p_j} / W^i{}_{p_i p_i} \quad ; \quad \frac{\partial p_j^*}{\partial p_i} = -W^j{}_{p_j p_i} / W^j{}_{p_j p_j} \qquad (3).$$

If the government is welfare maximizing, the second order condition guarantees the denominator in (3) is negative. Therefore, the slopes will depend directly on the signs of the second cross-partial derivatives (i.e. the numerator). If $W^{i,j}_{p_i p_j} > 0$, i.e. if policies are strategic complements, we see from (3) that policy reaction-functions will slope upward. If $W^{i,j}_{p_i p_j} < 0$, policies are strategic substitutes, and the reaction functions slope downward. If the second cross-partial derivative is zero, strategic interdependence does not materialize and the best-response functions are flat (Brueckner, 2003).

Notice that *positive* externalities induce strategic-*substitute* policy-interdependence and *negative* externalities induce strategic-*complement* policy-interdependence (and lack of externalities yields policy-independence). In the national-defense example discussed above, spending in one country induces free riding in others due to the positive security externalities (among allies) and diminishing returns of military expenditures. If ALM expenditures create positive employment externalities and exhibit diminishing returns, the same problem could arise in this context. In other words, if reducing unemployment requires increasing amounts of spending—€1000 per worker to reduce unemployment from 6% to 5%, €2000 to reduce from 5% to 4%, €4000 to reduce from 4% to 3%, etc.—and ALM spending in one country, *i*, helps reduce unemployment in another, *j*, an increase in expenditures in country *i* will reduce the marginal benefit to *j*'s of its (marginal increment of) spending, inducing lower equilibrium spending in *j*. Figure 1 illustrates this situation graphically. This strategic context also creates a first-mover disadvantage—the country that spends first will bear a larger portion of the cost of reducing unemployment—and the potential for war-of-attrition dynamics that would push the equilibrium ALM spending of both countries is even lower.

<Figure 1 About Here>

Do cross-border positive employment externalities of ALM policies exist among European countries; and if so are they sufficient to induce this kind of fiscal free-riding in ALM policy? Evidence from a number of recent studies, at both the micro and macro levels, shows that ALM policies have increased employment in Europe and other OECD countries (see, e.g., Martin, 2000; Estevao, 2003; European Commission, 2005). That ALM spending would exhibit diminishing returns also seems reasonable. For instance, if labor-market training-programs increase employment by raising workers' marginal productivity, then, in any given macroeconomic conditions, some workers will just miss being employed because their marginal productivity was just below a threshold beyond which firms find hiring them profitable and some whose productivity was far below this threshold. In this case, a little spending might get the first group of workers hired, but much more spending per worker would be required to get members of the less-productive second group employed. If unemployed low- and high-productivity workers are spatially concentrated in regions that span national boundaries, this could create incentives for fiscal free riding. Below, we describe several other mechanisms by which such cross-border spillovers may arise.

A large literature examines the regional patterns of unemployment in Europe (see, e.g., Elhorst, 2003; Puga, 2002; and Overman and Puga, 2002). This research shows that, in many cases, differences in employment between bordering regions are much smaller, even if the regions lie in different countries, than the differences between more distant regions within countries. In other words, geographic proximity is more important than nationality in understanding spatial patterns of unemployment in Europe. Labor-market performance in Languedoc-Roussillion in southern France on the Mediterranean, for example, is likely to resemble much more closely that in Cataluna in northeastern Spain than that in Paris. Similarly, employment outcomes in Nord-Pas-de-Calais on the French border with Belgium correlate highly with those in Region Wallonne across the border than with employment patterns in the center of France.[87]

Consider the implications of (effective) French ALM spending for Belgium, for example. Effective French ALM policies enhance Belgian workers' abilities to obtain training in France, and return, more employable, to work in Belgium. Effective French ALM policies also enhance Belgian workers' abilities to find work in France. Effective French ALM policies also enhance the pool of workers (quantity, quality, and diversity) available to employers along the Belgian border, thereby luring employers to both sides of the border. These and other agglomeration effects all yield positive externalities of (effective) French ALM policies to Belgian workers (and citizens more generally). Of course, Belgians cannot provide political support to French policymakers in response to these spillover effects, so French policymakers ignore these spillover benefits in determining French ALM spending. Accordingly, ALM spending by national policymakers will exhibit negative interdependence, reflecting the positive externalities.

Given what we know about spatial patterns of unemployment in Europe and the employment effects of ALM policies, fiscal free-riding seems plausible. In the next section, we examine the empirical record to gauge the evidence of its existence and magnitude.

Empirical Analysis

To evaluate strategic interaction in national ALM expenditures empirically, we estimate a dynamic spatial-lag model. This approach has been used by Case et al. (1993), Brueckner and Saavedra (2001), Fredriksson and Millimet (2002), Redoano (2003), and Allers and Elhorst (2005) among others. Our sample includes annual data from 1987-98 for 15 European countries: Austria, Belgium, Denmark, Germany, Greece, Finland, France, Ireland, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK.[88,89] ALM programs include spending on public

[87] Overman and Puga (2002) attribute the growing importance of spatial proximity to changes in the demand for labor. They identify, test, and find empirical support for three sources of demand change over the period 1986-1996: the regional concentration of skilled and unskilled labor, the spatial clustering of industries, and what they call agglomeration effects (all of which are illustrated in the examples given next in the text).

[88] Our paper is motivated by *EU* employment policy coordination, but limiting our sample to EU *countries* makes little sense. Empirically, we are evaluating the implications of strategic interaction among European neighbors in EU-member national policymaking. If such strategic interdependence exists, it should be evidenced in all neighboring country pairs

employment, labor-market training, and other directly active policies intended to promote employment among the unemployed. In this analysis we focus exclusively on labor-market training (LMT) programs. Our dependent variables are LMT expenditures per unemployed worker (2000 €, PPP) and the ratio LMT expenditures to GDP. [90] The key independent variable, which allows us to evaluate the hypothesis of strategic interdependence, is the *spatial lag* of LMT spending. We also include a number of control variables in the analysis. These are discussed below.

More precisely, the model we estimate is

$$\mathbf{Y_t} = \phi\mathbf{Y_{t-1}} + \rho\mathbf{W}\mathbf{Y_t} + \mathbf{X_t}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{t} \qquad (4).$$

$\mathbf{Y_t}$ is an *Nx1* vector of LMT observations in the N=15 countries for each year, *t*. ρ is the spatial-autoregressive coefficient (explained below). **W** is an *NxN* (*15x15*) *spatial-weighting matrix*, with elements w_{ij} reflecting the relative degree of connection from unit *j* to *i* (elaborated below). $\mathbf{WY_t}$ is thus the *spatial lag*; i.e. for each LMT observation in each country *i*, y_{it}, $\mathbf{WY_t}$ gives the weighted sum of the other countries' LMT y_{jt}, with the weights given by the relative connectivity from *j* to *i*. Note that each element of **W**'s diagonal, which would be the w_{ii} multiplying y_{it} itself in the weighted-sum spatial-lag, is zero. Thus, ρ gives the impact of the LMT outcomes in the other ($j \neq i$) countries, as weighted by w_{ij}, on LMT in county *i*; i.e. ρ gauges the overall strength of interdependence, whereas the w_{ij} describe the relative magnitudes of the specific interdependence relations from one country to another. $\mathbf{X_t}$ is an *NxK* matrix of observations on *K* independent variables, including fixed country and period effects (i.e. country and year dummies). $\boldsymbol{\beta}$ is a *Kx1* vector of coefficients on those **X**, and $\boldsymbol{\varepsilon}$ is a *Nx1* vector of residuals.

We calculated $\mathbf{WY_t}$ using two alternative spatial weights matrices. The first is a standardized *binary contiguity-weights matrix* which begins by coding w_{ij}=1 for countries *i* and *j* that share a border and w_{ij}=0 for countries that do not border. As exceptions, we code France, Belgium, and the Netherlands as contiguous with Britain, and Denmark as contiguous with Sweden. Then, we *row-standardize* (as commonly done in spatial-econometrics research) the resulting matrix by dividing

regardless of whether one, both, or neither are members of the EU. Therefore, we include Norway and Switzerland, as well as Austria, Sweden, and Finland which were not EU members for most of the sample period.

[89] Data availability limits our sample to the period before the Amsterdam Treaty entered into force. This should not affect our theoretical conclusions qualitatively since the lack of EES enforcement leaves the pre-Amsterdam strategic incentives largely unchanged. Empirically, the post-Amsterdam behavior of EU member states with respect to employment policy seems to have changed little.

[90] The labor-market-training expenditures data are from the OECD Labor Market Statistics Database. We chose the time period to maximize the number of countries (because the spatial-interdependence structure renders appropriate treatment of non-rectangular time-series cross-sections complicated). This database covers the period 1985-2002, but the data start for Denmark in 1986 and end for Greece in 1998, leaving 15 countries 1986-1998 preserving rectangularity. Qualitatively, our results are robust to using total ALM spending instead of LMT spending.

each cell in a row by that row's sum. So, e.g., standardized $w_{1,6}$=½=.5 and $w_{15,2}$=¼=.25.

$$\mathbf{W}=\begin{array}{c|ccccccccccccccc} & AUT & BEL & DEN & FIN & FRA & DEU & GRE & IRE & NTH & NOR & PRT & ESP & SWE & CHE & GBR \\ \hline AUT & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ BEL & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ DEN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ FIN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ FRA & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ DEU & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ GRE & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ IRE & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ NTH & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ NOR & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ PRT & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ ESP & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ SWE & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ CHE & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ GBR & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \qquad (5)$$

The second **W** we consider uses shared border-length for the spatial weights; that is, we substituted actual border-length for each 1 in (5) and then standardized the rows. Austria, for example, shares a 784km border with Germany and a 164km border with Switzerland. After standardization, the non-zero cells in Austria's row of **W** are .83 and .17, respectively.

We estimated the models by Maximum Likelihood using MATLAB code written by James P. LeSage and J. Paul Elhorst. The likelihood function for the spatial-lag model involves only one complicating modification of the likelihood for the standard linear additive model. To see this, start by expressing the simple spatial-lag model with the stochastic component on the left-hand side:

$$\mathbf{y}=\rho\mathbf{WY}+\mathbf{XB}+\boldsymbol{\varepsilon}\Rightarrow\boldsymbol{\varepsilon}=(\mathbf{I}-\rho\mathbf{W})\mathbf{Y}-\mathbf{XB}\equiv\mathbf{AY}-\mathbf{XB} \qquad \textbf{(6).}$$

The likelihood function for the stochastic component, **ε**, is then the usual linear-normal likelihood:

$$L(\boldsymbol{\varepsilon})=\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{NT}{2}}\exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right) \qquad \textbf{(7)},$$

which, in this case, will produce a likelihood in terms of **y** as follows:

$$L(\mathbf{y})=|\mathbf{A}|\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{NT}{2}}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{AY}-\mathbf{XB})'(\mathbf{AY}-\mathbf{XB})\right) \qquad \textbf{(8).}$$

This resembles the typical linear-normal likelihood, although the transformation from **ε** to **y** is not by the usual factor, 1, but by |**A**|=|**I**- ρ**W**|. The maximum likelihood is found numerically.[91]

Table 1 presents our regression results. The first column of results provides estimates for our base model, which includes a time-lag of the dependent variable plus country and year dummies to

---

[91] Some (surmountable) technical issues do arise in that search (see Franzese and Hays, 2004), but those issues do not affect the present substance and so need not detain us here.

account for temporal dependence and unit- and period-heterogeneity. The period dummies provide a flexible way to model common trends and/or common (random) shocks in LMT expenditures. As we show elsewhere (Franzese and Hays, 2005), by far the most important issue methodologically in obtaining good estimates of the strength of interdependence, i.e. of ρ, is to model well any alternative mechanisms by which the outcomes might correlate spatially, such as common exogenous shocks (e.g., global economic conditions) or correlated domestic factors. From that perspective, the country and year dummies serve as a powerfully conservative way to account for almost any sort of outside shock or spatially correlated domestic factor. Failure to account for such alternative mechanisms will bias spatial-lag coefficient estimates, usually in a positive direction (Franzese and Hays, 2004). The first column reports results using the standardized binary-contiguity matrix for our spatial weights. Again, the estimated coefficient on the spatial lag gives an estimate of the strength of strategic interdependence in ALM policymaking, assuming border-contiguity as the source of the employment spillovers.

Our estimate of the spatial-lag coefficient in the base model is statistically significantly negative. This coefficient gives some indication of short-run effect on country *i*'s LMT expenditures of a one-unit positive shock to all of its neighbors' spending. Insofar as we can credit that indication at this point, these results imply that a one-unit increase in neighbor spending leads to an immediate .258 decrease in domestic spending, suggesting some degree of fiscal free-riding among European countries in LMT spending. Because of spatial feedback—this effect then itself affects LMT in those neighbors, which feeds back into *i*'s LMT, and so on, recursively—this indication will be reasonably accurate regarding the magnitude of these effects only when ρ, the strength of the feedback, is fairly small, and, intuitively, it will become increasingly misleading about the ultimate magnitude of feedback effects as ρ increases. The sign of ρ remains directly informative about the sign of the feedback and interdependence, however. We will provide the exact calculations below when we discuss the substantive magnitudes of our preferred estimates and their implications.

Table 1. Labor Market Training Expenditures in Europe (1987-1998)

	LMT/Unemp.	LMT/Unemp.	LMT/Unemp.	LMT/Unemp.	LMT/GDP	LMT/Unemp.
Temporal Lag	0.657***	0.553***	0.514***	0.490***	0.691***	0.547***
	(.055)	(.064)	(.068)	(.068)	(.054)	(.069)
Spatial Lag	-0.258***	-0.277***	-0.286***	-0.284***	-0.109*	-0.130**
	(.068)	(.066)	(.067)	(.068)	(.065)	(.064)
Real GDP Per		-0.964	-1.203*	-0.863	-0.477**	-588
Capita		(.645)	(.655)	(.798)	199	(.833)
Unemployment		-054***	-0.092***	-0.092***	-0.015**	-094***
Rate		(.019)	(.026)	(.028)	006	(.029)
Union Density			0.000	0.000	0.003	0.000
			(.001)	(.001)	(.002)	(.001)
Deindustrialization			0.008	0.011	0.004	0.019**
			(.008)	(.009)	(.007)	(.009)
Trade Openness			0.051*	0.048*	0.001	0.056*
			(.028)	(.029)	(.002)	(.030)
Foreign Direct			0.071	0.027	0.001	0.012
Investment			(.043)	(.050)	(.005)	(.052)
Old Age				0.000	0.008	-0.000
				(.007)	(.012)	(.008)
Left Cabinet Seats				-0.030	0.000	-0.025
				(.021)	(.000)	(.022)
Christian Dem.				-0.002	0.000	-0.012
Cabinet Seats				(.013)	(.001)	(.013)
Left Libertarian				-0.003	-0.008**	-0.002
Vote				(.003)	(.003)	(.003)
Government				0.026	0.024***	0.035
Consumption				(.032)	(.008)	(.033)
Spatial Weights	Binary	Binary	Binary	Binary	Binary	Border
Matrix	Contiguity	Contiguity	Contiguity	Contiguity	Contiguity	Length
Log-Likelihood	-27.256	-22.818	-19.626	-17.600	234.573	-23.292
σ^2	0.077	0.074	0.071	0.069	0.004	0.075

Note: All regressions include fixed period and unit effects; those coefficient-estimates suppressed to conserve space. The first five 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. The last spatial lag is generated using shared border length for the spatial weights (see text for details). All the spatial weights matrices are row-standardized.

In the next model (column 2), we control for a country's macroeconomic performance by adding real GDP per capita and the unemployment rate to the regression. As their economies grow, governments might provide more public goods and services (Wagner's Law of Increasing State Activity). If so, we would expect a positive coefficient estimate for the GDP per capita variable. Some alternative theories, such as Baumol's Disease, which refers to an argued decreasing relative productivity in service sectors rendering financing of public services increasingly difficult as economies develop and shift toward heavier service-sector employment, would suggest a negative relationship between GDP per capita and LMT expenditures. We would certainly expect a negative coefficient on the unemployment rate though. As the unemployment rate rises, financing *per-worker* LMT expenditures at existing levels clearly becomes more costly and might therefore suggest declining LMT. However, here too, one could offer an alternative that rising unemployment might increase political demand for LMT expenditures. In any event, the spatial-lag coefficient-estimate remains negative and statistically significant in this model. The unemployment-rate coefficient, meanwhile, is negatively signed and statistically significant, supporting the cost-constraint more than the demand-spurring argument. The GDP coefficient is negative but statistically insignificant, favoring neither Baumol nor Wagner particularly strongly.

Next we control for a number of structural features of a country's economy related to its labor markets and exposure to external shocks. The labor-market variables are union density and Iversen and Cusak's (2000) measure of deindustrialization. Because higher union density increases the influence of organized labor, we would expect it to go hand in hand with greater LMT spending. With respect to deindustrialization, Iversen and Cusak (2000) argued that workers cross significant skill barriers when they move out of manufacturing and agricultural and into services. Thus, we would expect deindustrialization to induce higher levels of LMT expenditures also. Many scholars have argued that exposure to the international economy leads to increased government spending, especially on programs that help workers adjust to external shocks (e.g. Ruggie, 1982; Cameron, 1978; Katzenstein, 1985; Rodrik, 1997; Hays et al., 2005). We use trade openness and foreign-direct-investment flows as our measures of exposure. In this case, all of the coefficient estimates on the structural variables are positively signed, but only the trade openness coefficient achieves statistical significance (Table 1, column 3). The GDP coefficient is still negatively signed and now marginally significant. Most centrally, the spatial-lag coefficient remains negative and statistically significant.

We also consider several political variables (column 4): the retired population as a percentage of the total population, the percentage of cabinet seats held by left and Christian Democratic parties, the percentage of general-election votes won by left-libertarian parties, and government consumption as a percentage of GDP. Working-age voters are the natural constituency for LMT programs. For retired voters, the benefits are indirect at best. Therefore, the political pressure for LMT programs should be lower when the percentage of retired individuals is high. Social Democratic, Christian Democratic, and Left-Libertarian parties have all been identified as key supporters of active social-policies, albeit of different precise natures (see, e.g., Garrett, 1998; Swank, 2002; and Kitschelt, 1994). Government consumption is a direct measure of government size and intervention into the economy. We would expect, *ceteris paribus*, governments with high consumption to GDP ratios to spend more on LMT programs.[92] Interestingly, none of the coefficient estimates on the political variables is statistically significant. The coefficients on the spatial lag, the unemployment rate, and trade openness, however, remain correctly signed and statistically significant.

Finally, we explore alternative plausible operationalizations of the dependent variable and spatial lag. First, we substitute LMT spending as a percentage of GDP for spending per unemployed worker as the dependent variable. We would argue that LMT spending per unemployed worker is more directly connected to governments' actual policy decision than is LMT spending as a percent of GDP; i.e. governments set the amount of labor-market-training spending per recipient more than they set such LMT spending proportionally to GDP. Even using the percentage-of-GDP measure, however, the estimated strategic interdependence (column 5) remains significantly negative ($p<.10$), albeit not as strongly so. Next, we consider an alternative spatial-weight matrix based on the length of shared borders. The combination of binary indicators for borders and row standardization in our contiguity-weights matrix implies that the effect of country j on country i is equal for all j, and declines for each j proportionally to many neighbors country i has. If a country has five neighbors, the individual weights are .2 for each bordering country, whereas the w_{ij} for a country i with two neighbors are both .5. Substantively, one might argue that the employment spillovers between two countries that create their ALM-policy interdependence depend on how much those countries border each other more than on how many neighbors each country has. Austria, e.g., shares borders with both Switzerland and Germany, but the lengths of those borders differ radically (784 km vs. 164 km). However, the political-economic significance of borders may not correlate any more strongly with their geographic length than with their number. For example, Britain shares no land border with France, Belgium, and the Netherlands, and Denmark shares none with Sweden (and sea-border lengths are difficult to gauge comparably). Furthermore, some countries share long borders with lesser population and/or economic activity on either side whereas other borders are shorter with more; e.g., the northern Finlandian- Norwegian border is 727km whereas the Finlandian-Swedish border is only 614km). Ideally, therefore, we would weight border length (or, even better, the ratio of border length to land area of country i) by some measure of the population and economic significance of the border regions. Unfortunately, such finely tuned weighting measures proved impossible to obtain and construct. Using the length of countries' shared territorial borders

[92] We recognize that LMT spending is one component of total government consumption. Our concerns here, however, are to be as thorough and cautious as possible in controlling for domestic factors that might explain LMT spending and especially those explanators that might themselves be spatially correlated, as such as these latter are easily confused empirically with interdependence (Galton's Problem; see Franzese and Hays, 2005).

(standardized) as an alternative spatial-weight matrix, despite its possible faults, the estimated negative LMT spending-interdependence nonetheless remains significant (column 6).

We are satisfied, then, that EU members' ALM policymaking exhibits statistically significant negative interdependence, but what do these statistically significant results tell us of the estimated substantive magnitude of this interdependence, i.e. of the effects individual EU countries' ALM policies have on their neighbors' policymakers? To answer this question we need to calculate the so-called spatial multiplier, which is given in the following reduced form of equation (4):

$$\begin{aligned}\mathbf{Y_t} &= \rho\mathbf{W}\mathbf{Y_t} + \phi\mathbf{Y_{t-1}} + \mathbf{X_t}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{t} \Rightarrow \\ \mathbf{Y_t}(\mathbf{I} - \rho\mathbf{W}) &= \phi\mathbf{Y_{t-1}} + \mathbf{X_t}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{t} \Rightarrow \\ \mathbf{Y_t} &= (\mathbf{I} - \rho\mathbf{W})^{-1}(\phi\mathbf{Y_{t-1}} + \mathbf{X_t}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{t})\end{aligned} \tag{9}$$

The spatial multiplier, $(\mathbf{I} - \rho\mathbf{W})^{-1}$, captures the feedback from, say, Belgium on France and other countries, and back from France and those others on Belgium, and so on recursively. The immediate time-*t* effect on the vector of policy-outcomes in the 15 countries, $\mathbf{Y_t}$, including that recursive feedback, can now be calculated with this spatial multiplier by considering certain counter-factual shocks to the rest of the right-hand side of (9). Specifically, multiplying $(\mathbf{I} - \rho\mathbf{W})^{-1}$ by an *Nx1* column vector with 1 in row *i* and 0 elsewhere gives the immediate effect of a unit-shock to country *i* on policies in the other *(N-1)* countries *j*. For example, multiplying $(\mathbf{I} - \rho\mathbf{W})^{-1}$ by a *15x1* column vector with 0 in all rows except that corresponding to Austria, which gets a 1, will give a 15x1 column-vector containing the estimated effects of a unit-shock in Austria on the other 14 countries in their respective rows.[93] Another informative counter-factual to consider is the "post-feedback" effect of unit-shocks to all of the other 14 countries on Austria. One obtains this estimate by multiplying $(\mathbf{I} - \rho\mathbf{W})^{-1}$ and a *15x1* column vector with 1 in all rows except Austria's row, which gets a 0. Table 2 reports estimates of these two different counterfactuals: the effect of each country on each of the others and the effect of all of the others on each country. Specifically, the off-diagonal elements of Table 2 report the effect of a one-unit positive shock in the column country's LMT expenditures on the other countries in the sample. So, for example, a €1 positive shock to British LMT spending reduces spending in Ireland by €0.29. The diagonal elements of Table 2 report the effect of a one-

[93] In Austria's row will be the estimated effect after feedback of a unit-shock to the rest of Austria's right-hand side on Austria itself, which, in this case, will be somewhat more than the original unit because an increase in Austria's LMT spending induces other EU members to cut theirs which induces Austria to further raise its spending and so on, recursively.

unit positive spending shock in all the column country's counterparts on its own LMT spending.[94]

Table 2. Short-Run Spatial Effects of Labor Market Training Expenditures (Binary Contiguity Weights Matrix)

	AUT	BEL	DEN	FIN	FRA	DEU	IRE	NTH	NOR	PRT	ESP	SWE	CHE	GBR
AUT	-.241	.005	.006	.000	.020	-.135	.000	.006	.000	.000	-.001	-.001	-.140	-.002
BEL	.002	-.239	.003	.000	-.066	-.064	.005	-.065	.000	-.001	.004	.000	.006	-.064
DEN	.006	.006	-.242	.012	.006	-.148	.000	.007	.012	.000	.000	-.148	.006	-.001
FIN	.000	.000	.012	-.252	.000	-.002	.000	.000	-.134	.000	.000	-.129	.000	.000
FRA	.008	-.053	.002	.000	-.245	-.051	.004	.010	.000	.009	-.061	.000	-.057	-.057
DEU	-.045	-.043	-.049	-.001	-.043	-.248	-.001	-.046	-.001	.000	.003	.007	-.040	.010
IRE	.000	.018	.000	.000	.020	-.004	-.242	.020	.000	.000	-.001	.000	-.001	-.294
NTH	.004	-.086	.004	.000	.017	-.093	.007	-.238	.000	.000	-.001	-.001	.003	-.093
NOR	.000	.000	.012	-.134	.000	-.002	.000	.000	-.252	.000	.000	-.129	.000	.000
PRT	.000	-.002	.000	.000	.043	-.002	.000	.000	.000	-.264	-.298	.000	-.002	-.002
ESP	-.001	.008	.000	.000	-.152	.008	-.001	-.002	.000	-.149	-.272	.000	.008	.008
SWE	-.001	-.001	-.099	-.086	-.001	.014	.000	-.001	-.086	.000	.000	-.260	-.001	.000
CHE	-.093	.009	.004	.000	-.095	-.080	.000	.003	.000	-.001	.006	-.001	-.244	.005
GBR	-.001	-.064	-.001	.000	-.071	.015	-.074	-.070	.000	-.001	.004	.000	.003	-.257

Note: The off-diagonal elements of the table report the effect of a one-unit increase in the column country's labor-market-training expenditures on its European counterparts. The diagonal elements report the effect of a one-unit increase in all of the other countries' labor-market-training expenditures on the column country's own spending. These numbers are calculated using the spatial multiplier matrix $(\mathbf{I}-\boldsymbol{\rho}\mathbf{W})^{-1}$ and thus reflect all feedback effects.

In addition to these spatial dynamics, our model of LMT spending includes a time-lag of the dependent variable and corresponding temporal dynamics. We could, therefore, plot the evolution of the one-period effects from Table 2 over time to illustrate the spatio-temporal dynamics of responses to various counterfactuals. More compactly and perhaps more comprehensibly, we can calculate the long-run steady-state effect, including the feedback effects, of *permanent* hypothetical shocks to one country or to the set of other countries such as those considered above. To find these long-run steady state effects, we set $\mathbf{Y_{t-1}}$ in (9) equal to $\mathbf{Y_t}$, which gives:

$$\mathbf{Y_t} = (\mathbf{I}-(\mathbf{I}-\rho\mathbf{W})^{-1}\phi)^{-1}(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{X_t}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_\mathbf{t}) \qquad (10).$$

Table 3 reports these calculations, the column-entries corresponding to the same hypotheticals as those in Table 2. Not surprisingly, long-run steady-state effects are much larger. In the long run, a €1 positive shock to British LMT spending reduces spending in Ireland by €1.26. Again, this effect assumes a *permanent* increase in British spending and would take many years to materialize. In this sense, the calculation likely represents an upper bound for our spatial effects. One would probably not expect Britain to maintain the spending increase permanently, perhaps especially noting the

[94] These latter calculations are (a) close but (b) not equal to the coefficient estimate for ρ because (b) they include spatial feedback effects but (a) ρ is reasonably small.

cumulative Irish response over the equally long run, so we would probably never directly observe this full long-run-equilibrium degree of fiscal free-riding empirically.

Table 3. Steady-State Spatial Effects of Labor Market Training Expenditures (Binary Contiguity Weights Matrix)

	AUT	BEL	DEN	FIN	FRA	DEU	IRE	NTH	NOR	PRT	ESP	SWE	CHE	GBR
AUT	-.854	.027	.052	.002	.159	-.530	.005	.050	.002	.006	-.021	-.016	-.557	-.033
BEL	.013	-.828	.023	.001	-.254	-.238	.033	-.236	.001	-.009	.033	-.007	.047	-.237
DEN	.052	.047	-.881	.094	.051	-.640	.003	.056	.094	.002	-.007	-.648	.039	-.025
FIN	.002	.002	.094	-.938	.002	-.028	.000	.002	-.520	.000	.000	-.493	.002	-.001
FRA	.064	-.203	.020	.001	-.907	-.207	.034	.082	.001	.080	-.286	-.006	-.240	-.247
DEU	-.177	-.158	-.213	-.009	-.173	-.911	-.012	-.191	-.009	-.006	.023	.065	-.133	.083
IRE	.009	.132	.007	.000	.172	-.069	-.876	.163	.000	.006	-.023	-.002	-.015	-1.257
NTH	.033	-.314	.038	.002	.137	-.382	.054	-.835	.002	.005	-.018	-.011	.011	-.390
NOR	.002	.002	.094	-.520	.002	-.028	.000	.002	-.938	.000	.000	-.493	.002	-.001
PRT	.012	-.037	.004	.000	.398	-.038	.006	.015	.000	-1.076	-1.345	-.001	-.044	-.045
ESP	-.021	.067	-.007	.000	-.714	.068	-.011	-.027	.000	-.672	-1.155	.002	.079	.082
SWE	-.011	-.009	-.432	-.329	-.010	.129	-.001	-.011	-.329	.000	.001	-1.005	-.008	.005
CHE	-.371	.062	.026	.001	-.400	-.266	-.005	.011	.001	-.015	.053	-.008	-.874	.037
GBR	-.017	-.237	-.012	-.001	-.309	.125	-.314	-.293	-.001	-.011	.041	.004	.028	-.998

Note: The off-diagonal elements of the table report the effect of a one-unit increase in the column country's labor-market-training expenditures on its European counterparts. The diagonal elements report the effect of a one-unit increase in all of the other countries' labor-market-training expenditures on the column country's own spending. These numbers are calculated using the long-run spatio-temporal-multiplier matrix $\left[\mathbf{I}-(\mathbf{I}-\rho\mathbf{W})^{-1}\phi\right]^{-1}\left[\mathbf{I}-\rho\mathbf{W}\right]^{-1}$.

Discussion

Much of the research into policy interdependence explores contexts in which either or both complement and substitute relations might be expected. In environmental policy, e.g., the positive externality of cleaner domestic skies when others tighten environmental regulations may induce policy free-riding by relaxing domestic regulation (strategic substitutes); on the other hand, if others tighten environmental regulations and this entails some economic-distortionary costs, the domestic costs of following suit to tighten environmental regulations are reduced (strategic complements). In other frequently studied contexts, like tax-competition, strategic-complement relations are widely expected to dominate although, theoretically, either are possible. Previous empirical studies in these and other policy contexts have found evidence of RTB dynamics or little- or mixed-evidence more often than free-riding. In other words, most studies have reported positive spatial lag coefficient estimates, including those in the recent literature on policy and institutional diffusion as well (see, e.g., Simmons and Elkins, 2004). Exceptionally, several of the results in Case et al. (1993) and those given in Redoano (2003) do find evidence for fiscal free-riding (i.e. negative spatial lag coefficient estimates). The former study examines spatial interdependence in public expenditures among American states. Though the authors stress a different set of results, they find a negative and statistically significant estimate for the coefficient on the spatially lagged dependent-variable when they operationalize the mechanism of interdependence by geography, specifically in a row-standardized binary contiguity matrix like ours. Redoano explores fiscal interdependence among European countries in a number of different policy areas: tax rates, expenditures on education, public health, social security and defense. She uses geographical distance between capitals for her spatial weights which result in negative coefficient estimates for her spatial lags of social security and defense expenditures.

Why do these studies and ours here find evidence of free riding while others do not? One methodological similarity between both of these studies and ours is that the empirical models include time-period dummies as a way to account for common trends and shocks. The failure to account for common shocks will bias spatial-lag coefficient estimates, usually in a positive direction (Franzese and Hays, 2004). So, in the case of ALM policies, for example, two countries may experience a common shock that reduces expenditures in both countries (e.g., something like the convergence requirements of the Maastricht Treaty), and, if this common external stimuli is not or is insufficiently recognized in the empirical model, their policy behavior will be confounded with the RTB dynamics induced by the strategic interdependence that it resembles.

Another similarity is that all three of these studies use geographically defined spatial weights as opposed to other notions of space or distance—e.g., economic space/distance (Case et al., 1993). As we suggested above, many policy areas might have different aspects that induce strategic-complement and strategic-substitute relations. The latter, we conjecture, might be more likely to materialize along lines of geographic proximity. To give an example, for a country to be a competitor to another for mobile capital, does not require that the countries be geographically proximate but rather that they be economic competitors. Similarly, *learning* mechanisms for interdependence— i.e. that countries learn about policies and their effects—do not necessarily require geographic proximity and might more readily transmit along cultural or demographic or other non-geographic proximity bases. Conversely, for a country to experience the spillovers that tend to induce strategic-substitute relations and policy free-riding, geographic proximity seems more critical, in many aspects such as environmental quality or, in our case, labor-market connectivity.

Our empirical results suggest that the EU may have to play a stronger role enforcing the coordination of its member states' employment policies if it is to achieve its EES objectives. Offering precise and definitive recommendations about what should be done is beyond this paper's scope, but the results here do demand one option receive serious consideration: strengthening the performance-review system. What most clearly distinguishes EU efforts to coordinate on policies affecting macroeconomic stability (The Stability and Growth Pact) from those affecting employment (The European Employment Strategy) is that the latter efforts lack an effective enforcement mechanism to encourage governments to implement the policies they commit to in their National Action Plans. We suspect this contributes to the under-provision of ALM policies.

Under the Performance Reserve System adopted in Regulation (EC) No 1260/1999, 4% of the structural funds allocated to member states are kept in reserve. The Commission monitors program implementation and performance and then disperses the reserves after a successful midterm review. While the enforcement-mechanism design seems reasonable, its magnitude, the proportion of funds, is far too small to alter significantly the behavior of member states. One obvious reform to strengthen the system would be to increase both the percentages held in reserve and the number of installments paid to recipient countries, say, annual reviews and payments. This would make the mechanism more similar to the IMF system of credit tranches based on conditionality and phasing.

Additionally, the allocation of performance reserves, which are currently project specific, could be linked to general compliance with EES commitments. In other words, the Commission could withhold reserve funds from successful EES/Structural Fund projects for the failure of other EES programs or commitments that are not financed with EU Structural Funds. These changes might be politically unpopular and would undoubtedly make Commission evaluations and decisions much more controversial than they are today, but they would increase the incentives countries have to follow through on their action plans.

Conclusion

A large and growing literature on regional patterns of unemployment in Europe suggests labor-market outcomes are spatially clustered across the continent without tremendous regard for national borders. The unemployment rate in one region affects unemployment rates in its (regional) neighbors, and whether these regions are separated by national boundaries does not matter much for the degree to which these labor-market experiences are shared. Moreover, the available evidence shows that ALM programs are effective at reducing unemployment. In theory, these conditions create strategic incentives for national governments to free ride on the ALM policies of neighboring states. We provide evidence via estimation of dynamic spatial models that such strategic interaction among European governments in ALM policymaking does in fact exist, that these national best-response functions are indeed generally downward sloping, and that these effects are of substantively appreciable magnitude. Thus, we conclude that the EU should play a more active role in enforcing the policy commitments governments make throughout the EES process.

Further presentational options for the Franzese & Hays (2006) results:

A "short-run" response grid:

Franzese & Hays *EUP* '06

Table 2. Short-Run Spatial Effects of Labor Market Training Expenditures in Europe (Binary Contiguity Weights Matrix)

	AUT	BEL	DEN	FIN	FRA	DEU	IRE	NTH	NOR	PRT	ESP	SWE	CHE	GBR
AUT		0.005* (.0019)	0.006 (.0031)	0.000 (.0001)	0.020* (.0095)	-0.135* (.0317)	0.000 (.0001)	0.006* (.0029)	0.000 (.0001)	0.000 (.0002)	-0.001 (.0009)	-0.001 (.0007)	-0.140* (.0338)	-0.002 (.0015)
BEL	0.002* (.0010)		0.003* (.0014)	0.000 (.0000)	-0.066* (.0152)	-0.064* (.0142)	0.005* (.0021)	-0.065* (.0144)	0.000 (.0000)	-0.001* (.0004)	0.004* (.0019)	0.000 (.0003)	0.006* (.0029)	-0.064* (.0142)
DEN	0.006 (.0031)	0.006* (.0029)		0.012* (.0058)	0.006* (.0030)	-0.148* (.0383)	0.000 (.0001)	0.007* (.0033)	0.012* (.0058)	0.000 (.0001)	0.000 (.0003)	-0.148* (.0387)	0.006* (.0026)	-0.001 (.0011)
FIN	0.000 (.0001)	0.000 (.0083)	0.012* (.0058)		0.000 (.0001)	-0.002 (.0013)	0.000 (.0000)	0.000 (.0001)	-0.134* (.0314)	0.000 (.0000)	0.000 (.0000)	-0.129* (.0294)	0.000 (.0001)	0.000 (.0000)
FRA	0.008* (.0038)	-0.053* (.0122)	0.002 (.0012)	0.000 (.0000)		-0.051* (.0120)	0.004* (.0020)	0.010* (.0049)	0.000 (.0000)	0.009* (.0044)	-0.061* (.0164)	0.000 (.0003)	-0.057* (.0143)	-0.057* (.0144)
DEU	-0.045* (.0106)	-0.043* (.0095)	-0.049* (.0128)	-0.001* (.0004)	-0.043* (.0100)		-0.001 (.0005)	-0.046* (.0114)	-0.001* (.0004)	0.000 (.0003)	0.003* (.0012)	0.007 (.0036)	-0.040* (.0081)	0.010* (.0049)
IRE	0.000 (.0003)	0.018* (.0084)	0.000 (.0002)	0.000 (.0000)	0.020* (.0099)	-0.004 (.0031)		0.020* (.0097)	0.000 (.0000)	0.000 (.0002)	-0.001 (.0009)	0.000 (.0000)	-0.001 (.0007)	-0.294* (.0755)
NTH	0.004* (.0020)	-0.086* (.0192)	0.004 (.0022)	0.000 (.0001)	0.017* (.0082)	-0.093* (.0227)	0.007* (.0032)		0.000 (.0001)	0.000 (.0001)	-0.001 (.0007)	-0.001* (.0005)	0.003* (.0010)	-0.093* (.0231)
NOR	0.000 (.0001)	0.000 (.0001)	0.012* (.0058)	-0.134* (.0314)	0.000 (.0001)	-0.002 (.0013)	0.000 (.0000)	0.000 (.0001)		0.000 (.0000)	0.000 (.0000)	-0.129* (.0294)	0.000 (.0001)	0.000 (.0000)
PRT	0.000 (.0003)	-0.002 (.0016)	0.000 (.0001)	0.000 (.0000)	0.043 (.0220)	-0.002 (.0016)	0.000 (.0002)	0.000 (.0004)	0.000 (.0000)		-0.298* (.0788)	0.000 (.0000)	-0.002 (.0018)	-0.002 (.0018)
ESP	-0.001 (.0009)	0.008* (.0038)	0.000 (.0003)	0.000 (.0000)	-0.152* (.0411)	0.008* (.0037)	-0.001* (.0004)	-0.002 (.0011)	0.000 (.0000)	-0.149* (.0394)		0.000 (.0001)	0.008 (.0043)	0.008 (.0043)
SWE	-0.001* (.0005)	-0.001* (.0004)	-0.099* (.0258)	-0.086* (.0196)	-0.001* (.0004)	0.014 (.0073)	0.000 (.0000)	-0.001* (.0005)	-0.086* (.0196)	0.000 (.0000)	0.000 (.0000)		-0.001* (.0004)	0.000 (.0001)
CHE	-0.093* (.0225)	0.009* (.0039)	0.004* (.0017)	0.000 (.0000)	-0.095* (.0238)	-0.080* (.0161)	0.000 (.0002)	0.003 (.0010)	0.000 (.0000)	-0.001 (.0006)	0.006* (.0029)	-0.001* (.0004)		0.005* (.0021)
GBR	-0.001 (.0007)	-0.064* (.0142)	-0.001 (.0005)	0.000 (.0000)	-0.071* (.0180)	0.015 (.0073)	-0.074* (.0189)	-0.070* (.0173)	0.000 (.0000)	-0.001* (.0004)	0.004 (.0021)	0.000 (.0001)	0.003 (.0016)	

Notes: The off-diagonal elements of the table report the effect of a one-unit increase in the column country's labor-market-training expenditures on its European counterparts. These numbers are calculated using the spatial multiplier matrix $(\mathbf{I}-\rho\mathbf{W})^{-1}$ and thus reflect all feedback effects. Parentheses contain standard errors calculated by the delta method.

A "long-run steady-state" response grid:

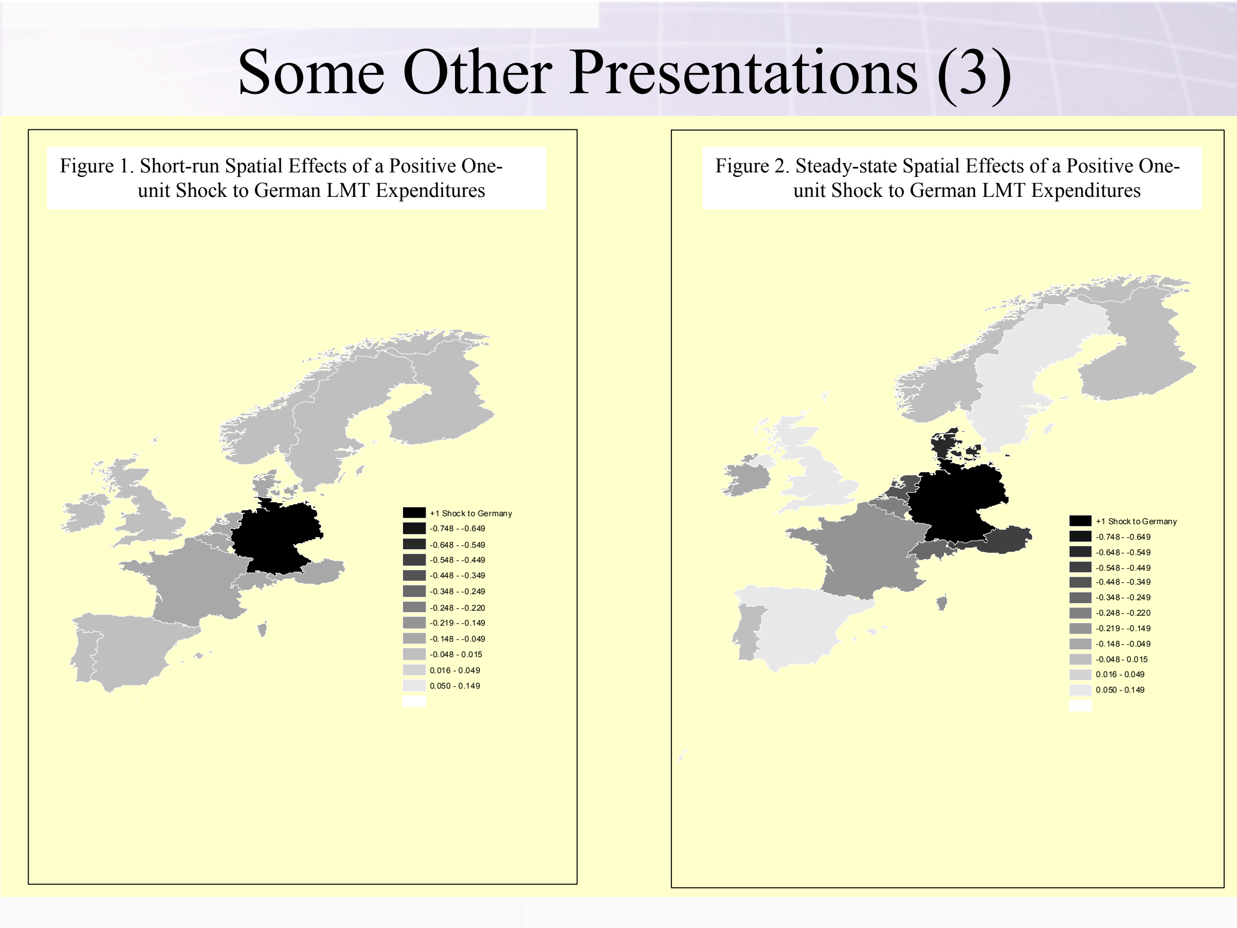
Franzese & Hays *EUP* '06

Table 3. Steady-State Spatial Effects of labor Market Training Expenditures in Europe (Binary Contiguity Weights Matrix)

	AUT	BEL	DEN	FIN	FRA	DEU	IRE	NTH	NOR	PRT	ESP	SWE	CHE	GBR
		0.027	0.052	0.002	0.159	-0.530*	0.005	0.050	0.002	0.006	-0.021	-0.016	-0.557*	-0.033
AUT		(.0136)	(.0296)	(.0024)	(.0914)	(.1588)	(.0050)	(.0292)	(.0024)	(.0067)	(.0189)	(.0134)	(.1663)	(.0277)
	0.013		0.023	0.001	-0.254*	-0.238*	0.033	-0.236*	0.001	-0.009	0.033	-0.007	0.047	-0.237*
BEL	(.0068)		(.0125)	(.0010)	(.0723)	(.0636)	(.0167)	(.0581)	(.0010)	(.0079)	(.0204)	(.0058)	(.0237)	(.0636)
	0.052	0.047		0.094	0.051	-0.640*	0.003	0.056	0.094	0.002	-0.007	-0.648*	0.039*	-0.025
DEN	(.0407)	(.0346)		(.0505)	(.0305)	(.2218)	(.0038)	(.0335)	(.0505)	(.0022)	(.0062)	(.2271)	(.0183)	(.0211)
	0.002	0.002	0.094		0.002	-0.028	0.000	0.002	-0.520*	0.000	0.000	-0.493*	0.002	-0.001
FIN	(.0785)	(.0640)	(.0739)		(.0570)	(.2043)	(.0096)	(.0621)	(.0076)	(.0057)	(.0149)	(.0403)	(.0266)	(.0489)
	0.064*	-0.203*	0.020	0.001		-0.207*	0.034	0.082*	0.001	0.080*	-0.286*	-0.006	-0.240*	-0.247*
FRA	(.0031)	(.0027)	(.0505)	(.2121)		(.0228)	(.0002)	(.0026)	(.1477)	(.0001)	(.0004)	(.1402)	(.0016)	(.0014)
	-0.177*	-0.158	-0.213*	-0.009	-0.173		-0.012	-0.191*	-0.009	-0.006	0.023	0.065	-0.133	0.083
DEU	(.0502)	(.0864)	(.0122)	(.0010)	(.2112)		(.0209)	(.0466)	(.0010)	(.0514)	(.1164)	(.0054)	(.0808)	(.0910)
	0.009	0.132	0.007	0.000	0.172	-0.069		0.163	0.000	0.006	-0.023	-0.002	-0.015	-1.257*
IRE	(.0131)	(.0928)	(.0075)	(.0005)	(.1045)	(.0576)		(.0940)	(.0005)	(.0074)	(.0212)	(.0029)	(.0127)	(.4242)
	0.033	-0.314*	0.038	0.002	0.137	-0.382*	0.054		0.002	0.005	-0.018	-0.011	0.011	-0.390*
NTH	(.0267)	(.1186)	(.0223)	(.0018)	(.0777)	(.1241)	(.0313)		(.0018)	(.0057)	(.0161)	(.0099)	(.0007)	(.1319)
	0.002	0.002	0.094	-0.520*	0.002	-0.028	0.000	0.002		0.000	0.000	-0.493*	0.002	-0.001
NOR	(.0031)	(.0027)	(.0505)	(.1477)	(.0024)	(.0228)	(.0002)	(.0026)		(.0001)	(.0004)	(.1402)	(.0016)	(.0014)
	0.012	-0.037	0.004	0.000	0.398	-0.038	0.006	0.015	0.000		-1.345*	-0.001	-0.044	-0.045
PRT	(.0175)	(.0424)	(.0044)	(.0003)	(.2572)	(.0341)	(.0074)	(.0171)	(.0003)		(.5073)	(.0016)	(.0398)	(.0425)
	-0.021	0.067	-0.007	0.000	-0.714*	0.068	-0.011	-0.027	0.000	-0.672*		0.002	0.079	0.082
ESP	(.0252)	(.0559)	(.0062)	(.0004)	(.2910)	(.0448)	(.0106)	(.0242)	(.0004)	(.2537)		(.0025)	(.0524)	(.0566)
	-0.011	-0.009	-0.432*	-0.329*	-0.010	0.129	-0.001	-0.011	-0.329*	0.000	0.001		-0.008	0.005
SWE	(.0119)	(.0103)	(.1514)	(.0935)	(.0090)	(.0805)	(.0010)	(.0099)	(.0935)	(.0005)	(.0016)		(.0059)	(.0056)
	-0.371*	0.062	0.026	0.001	-0.400*	-0.266*	-0.005	0.011	0.001	-0.015	0.053	-0.008		0.037
CHE	(.1644)	(.0439)	(.0122)	(.0011)	(.1346)	(.0532)	(.0042)	(.0007)	(.0011)	(.0133)	(.0350)	(.0059)		(.0217)
	-0.017	-0.237*	-0.012	-0.001	-0.309*	0.125	-0.314*	-0.293*	-0.001	-0.011	0.041	0.004	0.028	
GBR	(.0185)	(.0959)	(.0105)	(.0007)	(.1137)	(.0734)	(.1061)	(.0989)	(.0007)	(.0106)	(.0283)	(.0042)	(.0163)	

Notes: The off-diagonal elements of the table report the effect of a one-unit increase in the column country's labor-market-training expenditures on its European counterparts. These numbers are calculated using the long-run spatio-temporal-multiplier matrix $\left[\mathbf{I}-(\mathbf{I}-\rho\mathbf{W})^{-1}\phi\right]^{-1}\left[\mathbf{I}-\rho\mathbf{W}\right]^{-1}$. Parentheses contain standard errors calculated by the delta method.

A cartographic representation of the same two types of effects, for one specific counterfactual:



VII. <u>(Web) Appendices</u>:

A. Stata™ code to implement all of the procedures described in the text.

B. Data and code to replicate all of the estimation results given in the text.