

Sebastian Juhl ^{id}

University of Mannheim, Collaborative Research Center 884, B6, 30–32, Room 324, 68159 Mannheim, Germany.
Email: sebastian.juhl@gess.uni-mannheim.de

Abstract

Distinguishing substantively meaningful spillover effects from correlated residuals is of great importance in cross-sectional studies. Both forms of spatial dependence not only hold different implications for the choice of an unbiased estimator but also for the validity of inferences. To guide model specification, different empirical strategies involve the estimation of an unrestricted spatial Durbin model and subsequently use the Wald test to scrutinize the nonlinear restriction of common factors implied by pure error dependence. However, the Wald test's sensitivity to algebraically equivalent formulations of the null hypothesis receives scant attention in the context of cross-sectional analyses. This article shows analytically that the noninvariance of the Wald test to such reparameterizations stems from the application of a Taylor series expansion to approximate the restriction's sampling distribution. While asymptotically valid, Monte Carlo simulations reveal that alternative formulations of the common factor restriction frequently produce conflicting conclusions in finite samples. An empirical example illustrates the substantive implications of this problem. Consequently, researchers should either base inferences on bootstrap critical values for the Wald statistic or use the likelihood ratio test which is invariant to such reparameterizations when deciding on the model specification that adequately reflects the spatial process generating the data.

Keywords: spillover effects, common factors, Wald test, spatial econometrics

1 Motivation

The correct specification of the inherently unknown spatial process generating observable patterns of interrelatedness among the units of analysis constitutes a considerable challenge in cross-sectional studies. In particular, distinguishing substantively meaningful indirect spillover effects from spatially correlated random shocks is imperative as there is a serious risk of making incorrect inferences when estimating a misspecified model (e.g., LeSage and Pace 2009; Darmofal 2015). Unfortunately, while unfocused tests for spatial autocorrelation commonly applied in empirical research, like Moran's I (e.g., Cliff and Keith Ord 1981), help to detect spatial clustering in model residuals, they do not provide guidance on the exact process generating these dependencies. Since spatial regression models differ with respect to the implied pathways of dependence, these simple diagnostic tools do not allow researchers to identify the adequate model specification.

To address this specification problem, many empirical model selection strategies proposed in the literature feature the estimation of the spatial Durbin model (SDM) as an unrestricted nesting model and utilize the Wald test to scrutinize nonlinear common factor restrictions implied by pure error dependence (e.g., Burrige 1981).¹ Since the Wald test is asymptotically equivalent to the likelihood ratio (LR) and the Lagrange multiplier (LM) tests, which are the two alternative likelihood-based specification tests, the choice of a test statistic is oftentimes motivated by convenience or familiarity (e.g., LeSage and Pace 2009, 55). However, while previous studies in the field of spatial

¹ There is an ongoing debate in the spatial econometrics literature whether the *specific-to-general* or the *general-to-specific* approach should be used in order to identify the true data-generating model (e.g., Florax *et al.* 2003; Florax *et al.* 2006; Hendry 2006; Elhorst 2014a). However, both approaches have their disadvantages and there is no conclusive evidence for the superiority of any of these search strategies (Mur and Angulo 2009; Rüttenauer 2019). Consequently, many search procedures rely on a combination of both approaches (e.g., Mur and Angulo 2006; Elhorst 2010; Elhorst 2014a).

Political Analysis (2020)

DOI: 10.1017/pan.2020.23

Corresponding author
Sebastian Juhl

Edited by
Jeff Gill

© The Author(s), 2020. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

econometrics report notable differences with respect to their finite sample properties (e.g., Mur and Angulo 2006; Mur and Angulo 2009), the Wald test's sensitivity to algebraically equivalent alternative formulations of the null hypothesis is somewhat overlooked. Given that this result is well-established in a time-series context (e.g., Gregory and Veall 1985; Lafontaine and White 1986; Breusch and Schmidt 1988; Dagenais and Dufour 1991; de Paula and Cribari-Neto 1993; Goh and King 1996) and regarding the importance of distinguishing spillover effects from residual correlation for substantive inferences, this negligence is startling.

By remedying this omission, the present study evaluates the Wald test's appropriateness for differentiating between alternative mechanisms that cause spatial clustering in cross-sectional data structures. It discusses the substantive and econometric implications of alternative spatial processes and shows analytically that the Wald test's lack of invariance to reparameterizations of nonlinear common factor restrictions stems from the necessity to approximate the restrictions' sampling distributions. While asymptotically valid, Monte Carlo experiments demonstrate that this approximation frequently leads to misleading inferences concerning the presence of spillover effects across a wide range of parameter settings in finite samples. An empirical example further illustrates the severity of this problem for applied research aiming to assess the support for distinct theoretical mechanisms against possible alternative explanations. Given that a misspecification of the process generating cross-sectional dependencies can bias substantive inferences, the results suggest that, irrespective of the specification search strategy employed, researchers should not base inferences on the Wald statistic's asymptotic χ^2 distribution. Instead, simulation techniques such as bootstrap methods allow researchers to use estimated critical values as an alternative to their asymptotic counterparts. The LR test also offers a valuable alternative procedure that is invariant to reparameterizations of the null hypothesis.

2 Substantive and Residual Dependence in Cross-Sectional Models

In regression analyses utilizing cross-sectional data, three different types of interaction effects can be distinguished that generate spatial autocorrelation in the dependent variable. First, endogenous interaction effects occur whenever the units' outcomes are intertwined. In these situations, the actions, decisions, or behaviors of the units are simultaneously determined and responsive to the other units' outcomes. Second, exogenous interaction effects cause spatial clustering by linking the response of each unit to the covariates of other units. Finally, cross-sectional dependencies can be a product of spatially correlated model residuals (e.g., Elhorst 2014b; Halleck Vega and Elhorst 2015). While endogenous and exogenous interaction effects are part of the regression's systematic component, correlation among the error terms is confined to the model residuals and does not affect the expectation of the outcome conditional on the regressors.

Despite their close correspondence (see e.g., Gibbons and Overman 2012), the distinction between the different mechanisms causing spatial dependencies in the data has far-reaching implications for the estimation and interpretation of the regression coefficients (LeSage and Pace 2009; Rüttenauer 2019). Importantly, substantively meaningful indirect spillover effects, loosely defined as the impact of changes in one unit's covariates on the other units' outcomes, only exist if cross-unit interactions are part of the regression's systematic component (Elhorst 2010; Darmofal 2015; Halleck Vega and Elhorst 2015). In these instances, the cross-partial derivative of unit i 's outcome y_i with respect to unit j 's covariate x_j is nonzero, signifying a systematic relationship between the units (e.g., LeSage and Pace 2009, 38). Otherwise, the regression model imposes the restriction of no spillovers by assumption.² To detect the existence of these spillover effects, different model specification search strategies suggest to utilize the unrestricted SDM model as a

2 By virtue of the Gauss–Markov assumptions, nonspatial regression models typically estimated by ordinary least squares (OLS) do not incorporate any spatial effects. In a regression model with a single regressor \mathbf{x} , the direct effect of a change in x_i for unit i on the unit's outcome y_i is $\partial E(y_i | x_i) / \partial x_i = \hat{\beta}_{OLS} = (\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' \mathbf{y}$ while this change is $\partial E(y_j | x_i) / \partial x_i = 0$ for all

general model featuring substantive as well as residual dependence and subsequently test several parameter restrictions (e.g., Mur and Angulo 2006; LeSage and Pace 2009; Elhorst 2010).

2.1 An Illustrative Example of the Different Spatial Processes

Before outlining the alternative spatial model specifications, it is useful to contrast the different spatial processes with respect to their substantive implications for empirical political science research.

Spillover effects occur whenever the behavior (endogenous interactions) or certain characteristics (exogenous interactions) of one unit—may this be a country, a (coalition) government, a political party, or any other entity of interest—affect adjacent units (Darmofal 2015, 5).³ For example, the municipalities' income tax revenues are directly related to their local economic performance. At the same time, the economic prosperity of its neighbors also exerts a positive impact on a municipality's income tax revenues as its residents might commute to work. These dependencies in income tax revenues produce spillovers between the municipalities: the effect of a change in one unit's characteristics propagates to its neighbors (e.g., LeSage and Pace 2009). Hence, adequately understanding the phenomenon of interest—cross-sectional variation in income tax revenues—necessitates the consideration of these exogenous interaction effects among the municipalities that generate substantively meaningful and theoretically relevant indirect spillover effects.

In contrast, cross-sectional dependencies in the disturbances constitute another spatial process that has different substantive implications and requires an alternative model specification. Several circumstances can cause correlation between the units' residuals, for example spatial clustering in measurement errors. Alternatively, omitting a spatially dependent explanatory variable that is part of the true data-generating process (DGP) from the regression equation leads to correlated errors (Elhorst 2014b; Darmofal 2015). With regards to the example given above, it is reasonable to expect that several unobservable but spatially dependent characteristics, like a region's general appeal as a place of residence, affect a municipality's revenue from income taxation as well. In contrast to the theorized exogenous interaction effects, these omitted characteristics merely affect the model residuals and there are no relevant indirect effects present in the process that generates the data. Since exactly modeling the true dependence structure is almost impossible (e.g., Juhl 2020b), omitting relevant variables that are spatially correlated can create linkages among the units' disturbances.

2.2 Common Factors in the Spatial Durbin Model

While the previous example illustrates the crucial differences in the substantive implications of the underlying process that links the observations to one another, model misspecification may cause severe econometric problems as well. In fact, neglecting cross-sectional interdependencies can induce correlation between the regressors and the residuals, resulting in the canonical endogeneity problem (e.g., Gibbons and Overman 2012; Betz, Cook, and Hollenbach 2019).

In order to illustrate the problem of omitted spillover effects, consider the stylized DGP in which a dependent variable y is entirely determined by two uncorrelated regressors, denoted x and z , such that $y = x\beta + z$.⁴ Assume that z is unobservable and follows a spatial autoregressive process such that $z = \rho Wz + u_1$, where ρ is a scalar parameter, W is an exogenously defined connectivity matrix, and u_1 is a vector of independent and identically distributed normal disturbances with zero

units j where $j \neq i$. By the same token, the spatial error model (SEM) specification does not feature spillover effects since residual dependence does not affect $E(y_i | x_i)$.

- 3 Different theoretical mechanisms can produce substantively meaningful spillover effects. Shipan and Volden (2008), for example, distinguish between four different mechanisms of policy diffusion: learning, economic competition, imitation, and coercion. Acknowledging this, I leave aside a thorough discussion of alternative mechanisms and restrict the focus to the empirical modeling of cross-sectional dependencies.
- 4 For the sake of simplicity, I set the coefficient associated with z to 1 and omit it from the equation.

mean and a fixed variance.⁵ This scenario leads to the spatial error model (SEM) specification:

$$y = x\beta + \underbrace{(I_n - \rho W)^{-1} u_1}_z. \tag{1}$$

Due to the uncorrelatedness of x and z , omitting the spatially autocorrelated variable does not lead to endogeneity concerns. Furthermore, the relationship depicted in Equation (1) implies no indirect spillover effects since the cross-unit interactions are confined to the residuals. Consequently, even a nonspatial OLS model specification would provide unbiased but inefficient parameter estimates (Lacombe and LeSage 2015).

Now consider a slightly modified scenario in which the included and the omitted regressors are no longer independent from one another but correlated. To induce correlation between the variables, suppose that the random variable u_1 in Equation (1) is replaced by u_2 which is an additive linear function of x and a stochastic disturbance term v such that $u_2 = x\gamma + v$. In addition to the spatial autocorrelation, the unobserved covariate z is now also correlated with x and the scalar $\gamma \in (0, 1]$ as well as the dispersion of v (σ_v^2) jointly determine the strength of the correlation. In this slightly modified scenario, the true DGP becomes

$$y = x\beta + \underbrace{(I_n - \rho W)^{-1} (x\gamma + v)}_z. \tag{2}$$

In contrast to the DGP shown in Equation (1), the correlation between the included regressor x and the spatially clustered variable z causes an endogeneity problem if z is omitted from the regression's systematic part. Importantly, the effect of a change in regressor x on the outcome y in Equation (2) is more complex and both a nonspatial OLS model as well as a SEM model do not yield unbiased estimates since they ignore indirect effects produced by the spatial patterning of the correlated omitted variable. Consequently, the effect of a change in x_i is not confined to unit i 's outcome y_i but rather pertains to all nonisolated units in the entire system through indirect spillover and instantaneous feedback effects (e.g., LeSage and Pace 2009; Betz *et al.* 2019).

In order to address the endogeneity problem and to identify meaningful spillover effects, the unrestricted SDM model plays an important role. In fact, it serves as a general nesting model in many specification search strategies since it comprises several simpler spatial regression models frequently employed in empirical studies (e.g., Mur and Angulo 2009; Elhorst 2010; Angulo and Mur 2011).⁶ By allowing researchers to test different parameter restrictions, the SDM model facilitates the specification of an econometric model that reflects the spatial process generating the data most appropriately. For the hypothetical scenario with one regressor x and the (possibly correlated) unobserved variable z discussed here, the SDM model to be estimated takes the following form:

$$y = \rho W y + x\beta + W x\theta + v. \tag{3}$$

While it is easy to verify that the SDM model reduces to the popular spatial autoregressive (SAR) model, also known as spatial lag model (e.g., Elhorst 2014b, 5), that features global spillover effects

5 Also assume that ρ is contained in the compact open interval $(\omega_{min}^{-1}; \omega_{max}^{-1})$, where ω_{min} and ω_{max} are the smallest and largest eigenvalues of W . This stability constraint ensures that the matrix $(I_n - \rho W)$ is positive definite and its inverse exists (e.g., LeSage and Pace 2009; Elhorst 2014b).
 6 While the general nesting spatial (GNS) model incorporates all possible types of cross-sectional interaction effects, it tends to be overparameterized. Hence, it provides no additional information and is rarely used in applied studies (e.g., Elhorst 2014b; Rüttenauer 2019).

if $\theta = 0$ and to the spatial lag of X (SLX) model featuring local spillovers when $\rho = 0$, it also subsumes the SEM model that rules out any substantive indirect effects by assumption.⁷ To illuminate the relationship between these models, it is useful to restate the SEM DGP displayed in Equation (1) by multiplying both sides of the equation by $(\mathbf{I}_n - \rho\mathbf{W})$ and rearranging terms which results in the following structural form (Burridge 1981; Anselin 2003):

$$\begin{aligned} (\mathbf{I}_n - \rho\mathbf{W})\mathbf{y} &= (\mathbf{I}_n - \rho\mathbf{W})\mathbf{x}\beta + \mathbf{u}_1 \\ \mathbf{y} - \rho\mathbf{W}\mathbf{y} &= (\beta\mathbf{I}_n - \rho\beta\mathbf{W})\mathbf{x} + \mathbf{u}_1 \\ \mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{x}\beta + \mathbf{W}\mathbf{x}(-\rho\beta) + \mathbf{u}_1. \end{aligned} \tag{4}$$

Equation (4) elucidates that the SEM process imposes a nonlinear common factor restriction on the parameter associated with $\mathbf{W}\mathbf{x}$ which can be assessed by using estimates from the unrestricted SDM model depicted in Equation (3).⁸ More precisely, if the estimates from the unrestricted SDM model satisfy the common factor restriction $\theta = -\rho\beta$, the model can be simplified to a SEM model specification because there are no discernible substantive spillover effects in the data. Given that $E(\hat{\beta}) - \beta = 0$ in Equation (4), the common factor restriction holds for SEM processes.

However, if \mathbf{x} and \mathbf{z} are correlated in the true DGP, an endogeneity problem occurs since the variable \mathbf{z} is unobserved and the common factor restriction no longer holds. Restating the DGP in Equation (2) that features this correlation in a similar fashion yields:

$$\begin{aligned} (\mathbf{I}_n - \rho\mathbf{W})\mathbf{y} &= (\mathbf{I}_n - \rho\mathbf{W})\mathbf{x}\beta + \mathbf{x}\gamma + \mathbf{v} \\ \mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{x}(\beta + \gamma) + \mathbf{W}\mathbf{x}(-\rho\beta) + \mathbf{v}. \end{aligned} \tag{5}$$

Given the discussion above, Equation (5) illustrates that, in the presence of an omitted variable that is (i) spatially clustered and (ii) correlated with an included regressor, the true DGP resembles the SDM specification. Although the SDM model shown in Equation (3) provides consistent estimates for $E(\hat{\rho}) = \rho$ and $E(\hat{\theta}) = -\rho\beta$, the estimate of β is asymptotically biased since $E(\hat{\beta}) = \beta + \gamma$. Based on these model estimates, the common factor restriction does not hold because of the endogeneity bias present in $E(\hat{\beta})$. Hence, a violation of the common factor restriction is indicative of the existence of indirect spillover effects that need to be included in the systematic part of the regression model.

As this discussion suggests, the spatially lagged exogenous variables in the SDM model specification can also be understood as instruments for omitted variables that are correlated with included regressors (e.g., Elhorst 2014b, 18). At the same time, as Gibbons and Overman (2012) emphasize, this strategy provides only weak identification of the model parameters and crucially depends on the exogeneity and the assumed perfect knowledge of \mathbf{W} . Consequently, the SDM model does not provide a general solution to the omitted variables problem that would allow researchers to identify causal effects. Just like in any other observational study, doing so requires the application of appropriate research designs (see also Betz *et al.* 2019; Rüttenauer 2019). Moreover, while the SDM specification features global spillover effects, local spillovers would require a different model specification like the spatial Durbin error model (e.g., Halleck Vega and Elhorst 2015).

7 Since this study is primarily concerned with the Wald test's ability to assess the common factor hypothesis, the reader may be referred to Halleck Vega and Elhorst (2015), Elhorst (2014b), Elhorst 2010, or LeSage and Pace (2009) who provide excellent treatments of alternative spatial regression models.
 8 The number of these common factor restrictions equals the number of regressors included in the model. To ease the exposition, I assume a single regressor throughout this study.

3 The Wald Test of Nonlinear Restrictions

Given the estimates from an unrestricted model, the Wald test is flexible enough to scrutinize several different and possibly nonlinear parameter constraints within the same model. Intuitively, the Wald test assesses the distance between the observed estimates and the restrictions imposed. As the distance grows, the restrictions become less likely. In contrast to the LM and the LR tests which constitute prominent and asymptotically equivalent alternative specification tests, the Wald test does not require the estimation of restricted alternative models (e.g., Burrige 1981).

Despite its advantages, a major drawback of the Wald test of nonlinear restrictions is its lack of invariance to algebraically equivalent expressions of the null hypothesis. Since the asymptotic distribution of a nonlinear restriction needs to be approximated by a Taylor series expansion, seemingly identical functional representations may produce different test statistics. Although this undesirable property is a well-known feature of the Wald test (e.g., Gregory and Veall 1985; Lafontaine and White 1986; Breusch and Schmidt 1988; Dagenais and Dufour 1991; de Paula and Cribari-Neto 1993; Goh and King 1996), its consequences for empirical spatial model search strategies have been neglected so far.

3.1 Analytical Derivation and Asymptotic Distribution of the Wald Statistic

Consider a situation in which a test needs to be constructed in order to evaluate a single nonlinear restriction $H_0 : g(\lambda) = 0$, where λ is a parameter vector and $g(\cdot)$ is some function that is continuously differentiable in a neighborhood of λ . For this general case, the Wald statistic is defined by

$$w = g(\hat{\lambda}) [\widehat{V}(g(\hat{\lambda}))]^{-1} g(\hat{\lambda}), \tag{6}$$

where $\widehat{V}(g(\hat{\lambda}))$ is the estimated variance of $g(\hat{\lambda})$. Under the null hypothesis, w asymptotically follows a χ^2 distribution with the number of degrees of freedom equal to the number of restrictions imposed.

The only complication involved here is that obtaining w necessitates knowledge about the sampling distribution of a nonlinear function. While it is straightforward to compute the value of $g(\hat{\lambda})$ at the parameter estimates, deriving the Wald statistic in Equation (6) additionally requires information about its variability which depends on the estimator $\hat{\lambda}$ and the restriction $g(\cdot)$. However, given the restriction's nonlinearity, exact distributional results become inapplicable (Greene 2012). Instead, the delta method provides an approximation of the restriction's asymptotic distribution.⁹ It is based on a first-order Taylor series expansion of $g(\hat{\lambda})$ around the true parameter vector λ . Assuming that $\hat{\lambda}$ is a consistent estimator with a limiting distribution defined by $\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ and that the standard regularity conditions hold (see e.g., Newey and McFadden 1994), the delta method implies that

$$\sqrt{n}(g(\hat{\lambda}) - g(\lambda)) \xrightarrow{d} \mathcal{N}(0, \mathbf{G}(\lambda)\Sigma\mathbf{G}(\lambda)'), \tag{7}$$

where $\mathbf{G}(\lambda) = \partial g(\lambda) / \partial \lambda'$ is a row vector of partial derivatives. It follows that the asymptotic distribution of the restriction under H_0 is $g(\hat{\lambda}) \stackrel{a}{\sim} \mathcal{N}(g(\lambda), \mathbf{G}(\lambda)n^{-1}\Sigma\mathbf{G}(\lambda)')$. By using the consistent estimates obtained from an unrestricted model, the restriction's sampling variability derived from the delta method is given by

$$\widehat{V}(g(\hat{\lambda})) = \mathbf{G}(\hat{\lambda})\widehat{\Sigma}\mathbf{G}(\hat{\lambda})', \tag{8}$$

9 Supplementary Material A contains further details on the delta method.

Table 1. Algebraically identical formulations of the common factor hypothesis.

	$g(\hat{\lambda}) = 0$	$G(\hat{\lambda})$		
		$\partial g(\hat{\lambda})/\partial \hat{\rho}$	$\partial g(\hat{\lambda})/\partial \hat{\beta}$	$\partial g(\hat{\lambda})/\partial \hat{\theta}$
$H_0(I)$	$\hat{\rho}\hat{\beta} + \hat{\theta}$	$\hat{\beta}$	$\hat{\rho}$	1
$H_0(II)$	$\hat{\beta} + \hat{\theta}/\hat{\rho}$	$-\hat{\theta}/\hat{\rho}^2$	1	$\hat{\rho}^{-1}$
$H_0(III)$	$\hat{\rho} + \hat{\theta}/\hat{\beta}$	1	$-\hat{\theta}/\hat{\beta}^2$	$\hat{\beta}^{-1}$
$H_0(IV)$	$\hat{\rho}\hat{\beta}/\hat{\theta} + 1$	$\hat{\beta}/\hat{\theta}$	$\hat{\rho}/\hat{\theta}$	$-\hat{\rho}\hat{\beta}/\hat{\theta}^2$

with $G(\hat{\lambda})$ evaluated at $\hat{\lambda}$ and $\hat{\Sigma} = n^{-1}\Sigma$ being a consistent estimator of the symmetric, positive definite asymptotic variance–covariance matrix. By substituting Equation (8) into Equation (6), the Wald test statistic can be calculated. Asymptotically, since $\text{plim}_{n \rightarrow \infty} \hat{\lambda} = \lambda$, the function $g(\hat{\lambda})$ converges in distribution to $g(\lambda)$ with a mean given by $\text{plim}_{n \rightarrow \infty} g(\hat{\lambda}) = g(\lambda)$. At the same time, the necessity to estimate the nonlinear restriction’s sampling variability can cause a mismatch between the Wald statistic’s asymptotic χ^2 distribution and its finite sample distribution which has considerable consequences for hypothesis testing (e.g., Lafontaine and White 1986; Phillips and Park 1988).

3.2 The Wald Test of Common Factors in Spatial Models

The analytical results derived above are directly applicable to the empirical assessment of the common factor restriction in spatial regression models because the null hypothesis can be expressed as a nonlinear function of the estimates obtained from an unrestricted SDM model. In order to calculate the test statistic, it is necessary to determine the functional representation of the null hypothesis of common factors. Yet, there are numerous algebraically equivalent alternative parameterizations that satisfy $g(\hat{\lambda}) = 0$, where $\hat{\lambda} = [\hat{\rho}, \hat{\beta}, \hat{\theta}]$ are the estimates obtained from the SDM model in Equation (3).

Table 1, for example, lists the four alternative expressions of the null hypothesis considered by Gregory and Veall (1986) in a time-series context. While all of the alternative statements declare the same restriction, they produce distinct test statistics and p values in finite samples because the approximation of the restriction’s sampling variability is based on the partial derivatives of the parameter estimates. Depending on the exact representation of the null hypothesis, the right part of Table 1 shows that the vector of partial derivatives obtained from these nonlinear functions differ.

As a result, alternative expressions of the common factor hypothesis use different estimators for the nonlinear restriction’s sampling variability which yields distinct test statistics and causes them to converge to the asymptotic χ^2 distribution at individual rates. In large samples, this circumstance is unproblematic as the accuracy of the Taylor series approximation increases in sample size while the contribution of the restriction’s estimated variability to the test statistic becomes negligible. In finite samples, however, the differences between the alternatives can be substantial (e.g., Gregory and Veall 1985). At worst, alternative and algebraically identical functional representations of the parameter restriction can indicate opposing conclusions regarding its validity despite the fact that the same model estimates are used to calculate the test statistic.

Importantly, while many statistical software packages used to estimate spatial regression models, like Stata or R packages, report results from a Wald test by default, they do not test for common factors. Instead, the null hypothesis these packages evaluate is cross-sectional independence, that is, $\hat{\rho} = 0$. Since this is a linear restriction, the Wald statistic’s noninvariance problem does not arise and the tests these packages perform are only indicative of the presence of nonrandom spatial clustering. They do not permit any inferences regarding the spatial process at work. Hence,

scrutinizing the common factor restriction requires researchers to amend the Wald test's null hypothesis.

3.3 Modifications of the Wald Test

Since the application of the Taylor series expansion results in distinct Wald statistics for algebraically equivalent formulations of the null hypothesis, the asymptotic χ^2 distribution might constitute an inappropriate approximation of the statistic's finite sample distribution for some parameterizations. Therefore, modifications proposed in the literature that attempt to address the Wald test's noninvariance problem primarily focus on adjusting the statistic's reference distribution.

Phillips and Park (1988), for example, show that an Edgeworth expansion of the Wald statistic provides additional information on the statistic's distribution which can be used to obtain corrected critical values and modified test statistics for each functional representation of the null hypothesis (e.g., de Paula and Cribari-Neto 1993; King and Goh 2002). Besides these corrections, simulation techniques allow researchers to generate the empirical distribution under the null hypothesis for each specification of the common factor restriction and base inferences on these reference distributions (e.g., Lafontaine and White 1986; Goh and King 1996). In particular, bootstrap methods provide a way to estimate critical values and use them as an alternative to the (corrected) asymptotic critical values which can be unreliable in finite samples (Godfrey and Veall 1998). Using the following procedure, it is straightforward to derive bootstrap critical values to test for common factors in spatial regression models:

1. Use the estimates from an unrestricted SDM model and calculate the observed Wald statistic w according to Equation (6) for a given restriction.
2. Estimate the restricted SEM model. With these estimates and the DGP shown in Equation (1), generate 100 bootstrap samples by resampling with replacement from the residual vector to obtain bootstrap disturbances.
3. Repeat step 1 for each bootstrap sample and store the Wald statistic in vector \tilde{w} .
4. Sort \tilde{w} in ascending order. The value with rank $(1 - \alpha) \times 100 + 1$ is the estimated bootstrap critical value, χ_{boot}^2 , corresponding to a predefined α -level (e.g., $\alpha = 0.05$).

By comparing w calculated in step 1 to the corresponding bootstrap critical value χ_{boot}^2 from step 4, its statistical significance can be assessed. This procedure can be repeated for any functional representation of the common factor restriction in order to obtain individual bootstrap critical values for each restriction and a given region of the parameter space. Thereby, researchers can base inferences on the empirical distribution under the null hypothesis instead of relying on the asymptotic χ^2 distribution. This is especially important since the performance of the Wald test not only depends on the specific expression of the common factor hypothesis but also on the particular region in the parameter space. In fact, previous research shows that there is no single formulation of the restriction that consistently outperforms all alternatives (e.g., Gregory and Veall 1986; Lafontaine and White 1986; Phillips and Park 1988).

While Goh and King (1996) demonstrate that both asymptotic modifications—the corrected critical values and the improved test statistics—might even deteriorate the Wald test's power and size properties, they conclude that the bootstrap approach constitutes a useful improvement for applied research (see also Lafontaine and White 1986; Godfrey and Veall 1998; King and Goh 2002). Of course, neither the Edgeworth corrections nor simulation techniques completely resolve the noninvariance problem inherent to the Wald test. Doing so requires the application of alternative tests such as the asymptotically equivalent LR test that is invariant to such reparameterizations (e.g., Mur and Angulo 2006). Yet, by providing corrections for the Wald test's empirical size, these

modifications reduce the possibility of intentionally manipulating the result by amending the functional expression of the null hypothesis (King and Goh 2002, 260).

4 Monte Carlo Analysis

4.1 Experimental Setup

In order to investigate the finite sample performance of the Wald test, I conduct Monte Carlo experiments in which I vary the sample size, the strength of the interdependence, and the severity of the omitted variables bias through the degree of correlation between the included and the omitted regressor. Using the spatial process depicted in Equation (2), I generate 1,000 samples of the outcome vector \mathbf{y} for each of the parameter configurations. In the simulations, $\beta = 2$ and $\sigma_v^2 = 1$ are held constant and \mathbf{x} is drawn from a standard normal distribution. The parameter space of γ ranges from 0 to 1 in steps of 0.2 while ρ takes on values between 0 and 0.8 in steps of 0.2.¹⁰ This setup includes a nonspatial DGP without omitted variables bias ($\rho = 0$ and $\gamma = 0$), nonspatial DGPs with omitted variables bias ($\rho = 0$ and $\gamma > 0$), SEM DGPs ($\rho > 0$ and $\gamma = 0$), and SDM DGPs ($\rho > 0$ and $\gamma > 0$). \mathbf{W} is a row-stochastic contiguity matrix based on the queen criterion of adjacency. In contrast to the rook connectivity scheme which links spatial units ordered on a lattice to their direct horizontal and vertical neighbors, the queen criterion additionally connects the units to their diagonal neighbors (e.g., Cliff and Keith Ord 1981).¹¹ The sample sizes specified here contain 49, 100, 225, and 400 observations distributed on regular grids (7×7 , 10×10 , 15×15 , and 20×20) to realistically reflect small to medium sized samples frequently encountered in political science.¹²

Since the consequences of model misspecification for the estimation of unbiased effect estimates have been studied elsewhere (e.g., LeSage and Pace 2009; Pace and LeSage 2010; Lacombe and LeSage 2015; Rüttenauer 2019), this Monte Carlo analysis focuses on the ability of the Wald test to identify the true spatial model and differentiate between substantive and residual dependence across a range of alternative DGPs.¹³ To this end, I investigate the performance of the Wald test using the four alternative null hypotheses of common factors summarized in Table 1.¹⁴

4.2 Performance of the Original Wald Test

Table 2 reports the rejection rates of the four expressions of the null hypotheses of common factors at an α -level of 0.05 across the simulations when the true DGP is that of the SEM model ($\gamma = 0$). Since there are no omitted spillovers in this scenario, the common factor restriction holds and the four variants of the Wald test are expected to reject the true null hypothesis in about 5% of the simulation trials with a 95% confidence interval of [3.65%; 6.35%].

Although Section 3 analytically shows that the alternative Wald tests are asymptotically equivalent, their type I error rates differ notably in finite samples. Especially $H_0(II)$ but also $H_0(IV)$ deviate considerably from the expected error rate. Across all sample sizes, $H_0(II)$ is too conservative when ρ is small. Since $\hat{\rho}$ appears in the restriction's denominator (see Table 1), the restriction has no derivative at zero which violates the assumed continuity of derivatives. However, the Wald test based on $H_0(II)$ remains valid as its asymptotic distribution is obtained under the null hypothesis which precludes the problematic value (Gregory and Veall 1985).¹⁵

- 10 Supplementary Material B contains information on the correlation between the regressor \mathbf{x} and the omitted variable \mathbf{z} for the different values of γ .
- 11 Since \mathbf{W} is row-stochastic, any value of ρ in the interior of ω_{min}^{-1} and 1 ensures matrix invertibility (LeSage and Pace 2009; Elhorst 2014b).
- 12 To perform the simulations, I rely on resources from the High Performance Computing Cluster bwHPC and use the R package `spdep` (Bivand and Piras 2015) to estimate all spatial regression models. Replication materials are available on the Political Analysis Dataverse (Juhl 2020a).
- 13 For an investigation of the substantive effects of spatial misspecification bias in nonspatial OLS and SEM models, see Supplementary Material C.1.
- 14 Supplementary Material C contains additional robustness tests including a scenario with negative spatial autocorrelation (C.2), an alternative specification of the connectivity scheme and an investigation of possible edge effects (C.3).
- 15 This situation also occurs at $\hat{\beta} = 0$ for $H_0(III)$ and at $\hat{\theta} = 0$ for $H_0(IV)$.

Table 2. Share of false positives (type I error rates) using asymptotic critical values.

<i>n</i> = 49					<i>n</i> = 100				
ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$	ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$
0	6.9	0.1	7.1	2.6	0	6.4	0.2	6.0	3.5
0.2	7.5	0.1	7.9	4.0	0.2	5.5	0.0	5.5	2.8
0.4	7.5	0.0	7.2	6.0	0.4	5.8	0.1	5.6	4.3
0.6	9.4	0.7	8.9	7.3	0.6	7.7	1.7	7.6	7.5
0.8	11.0	4.9	10.7	11.5	0.8	6.0	3.5	6.0	5.1
<i>n</i> = 225					<i>n</i> = 400				
ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$	ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$
0	5.2	0.0	5.0	4.9	0	5.9	0.0	5.8	2.0
0.2	5.4	0.0	5.4	5.8	0.2	4.3	0.3	4.3	4.6
0.4	6.3	1.2	6.7	5.3	0.4	4.7	2.4	4.8	5.9
0.6	6.2	4.3	6.5	6.1	0.6	3.7	3.0	3.8	4.2
0.8	5.6	4.8	5.5	4.2	0.8	4.9	4.6	5.0	4.9

Based on the χ^2 distribution with $df = 1$ and $\alpha = 0.05$, the asymptotic critical value used for all variants of the Wald test and across the different levels of spatial autocorrelation is $\chi^2_{asym} = 3.841$. The theoretically expected rejection rate across the 1,000 simulation iterations is 5% with a 95% (binomial proportion) confidence interval of [3.65%;6.35%].

At the same time, incorrectly rejecting a true null hypothesis might be less problematic in this case since the SDM model derives unbiased impact estimates even if only the residuals are spatially clustered and no substantive spillovers exist (e.g., Elhorst 2010). The only drawback is that the appropriate SEM specification would be more efficient which might affect inferences regarding the statistical significance of a regressor’s impact. Consequently, it is crucial for any test of the common factor hypothesis to have satisfactory power properties in order to reduce concerns about biased effect estimates.

Against this background, Figure 1 compares the performance of the alternative variants of the Wald test for different levels of correlation between the spatially dependent omitted variable and the included regressor by reporting their power. As the correlation increases, the tests should be more likely to reject the null hypothesis. In order to account for the effects of the sample size and the strength of the interdependence on the performance of the tests, Figure 1 is comprised of 16 panels. In each panel, the horizontal axis depicts the different values of γ and the vertical axis shows the observed share of rejections across the simulation trials.

A brief inspection of Figure 1 already confirms that alternative parameterizations of the null hypothesis—although algebraically equivalent—yield strikingly different results in finite samples. Even with a decently sized sample, there are pronounced differences in the rejection rates of the four Wald tests. While $H_0(I)$, which is considered to be the common way to express the restriction (Gregory and Veall 1986, 204), and $H_0(III)$ perform comparatively well in these simulations, the specifications based on $H_0(II)$ and $H_0(IV)$ have inferior power properties. The remarkably low rejection rates of these specifications of the Wald test increase the likelihood that researchers incorrectly infer the absence of meaningful spillover effects.

Moreover, the behavior of $H_0(IV)$ differs greatly from the expectation as its rejection rate initially decreases in almost all parameter settings as γ increases. This phenomenon—known as nonmonotonicity in the power function—makes the rejection of the null hypothesis even less likely as the difference between the true DGP and the restriction increases (King and Goh 2002,

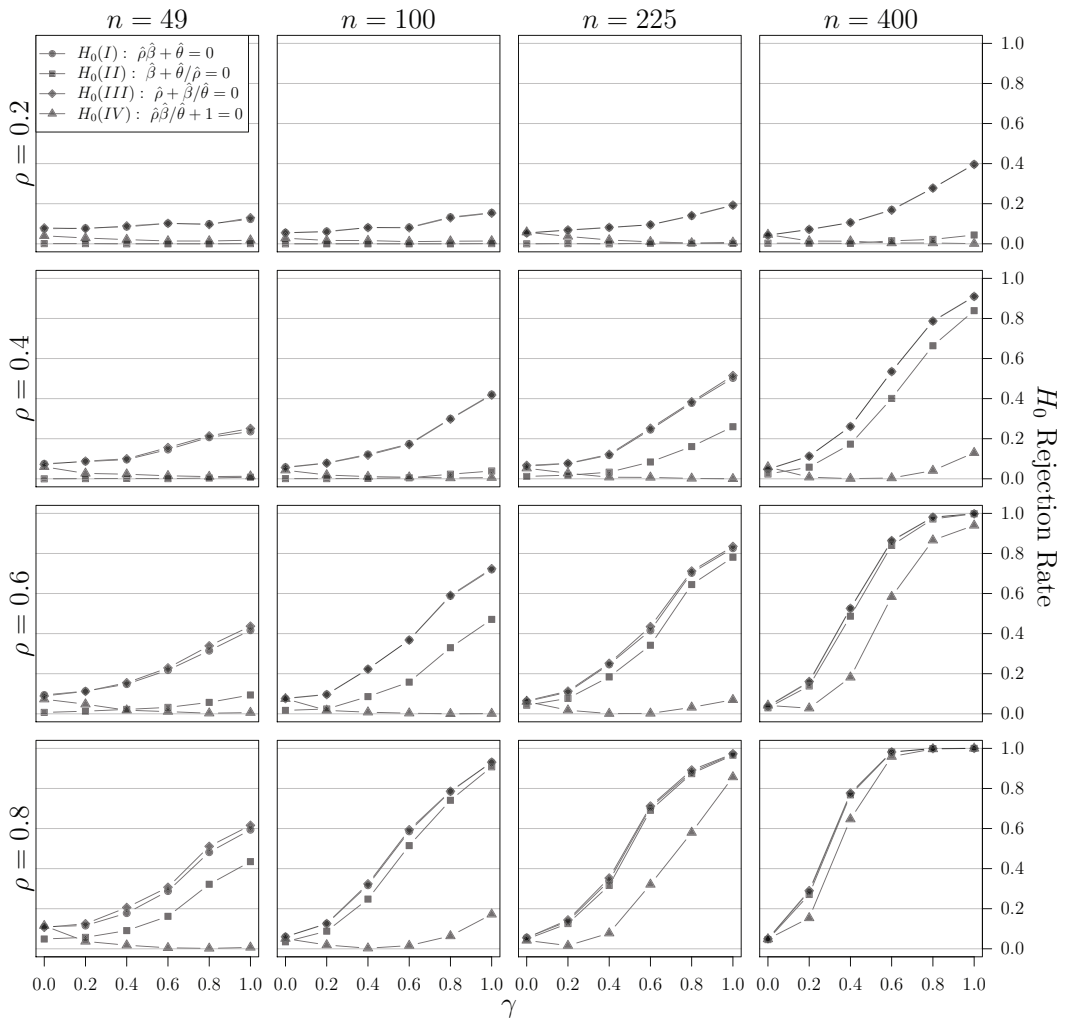


Figure 1. Share of null hypothesis rejections at a nominal significance level of 5%.

256–58).¹⁶ In practice, these tests would suggest that the data was generated by a DGP with spatially correlated errors even if there are sizable spillover effects. Researchers would incorrectly conclude that a SEM model or even a nonspatial OLS model appropriately represents the unobservable DGP. Given that these model specifications produce biased impact estimates if a SDM process generated the data, the low rejection rates are highly problematic for substantive inferences.

Although the different variants of the Wald test use the same data and identical parameter estimates, this simulation study shows that, depending on the functional representation of the null hypothesis, they can come to contradictory conclusions regarding the validity of the common factor restriction.¹⁷ In fact, Breusch and Schmidt (1988) analytically show that it is possible to obtain any desired Wald statistic by appropriately specifying the restriction which opens up the possibility to intentionally manipulate the test result (see also King and Goh 2002). Therefore, any search strategy utilizing the Wald test, like the basic *general-to-specific* approach or the multistep procedure suggested by Elhorst (2014a), is subject to this malfunctioning. Since there is no theoretically justified functional representation of the common factor hypothesis and given the

16 The Wald test based on $H_0(IV)$ is also strongly affected by β in the DGP. As Supplementary Material C.4 shows, the test's performance is even worse for smaller values of β .
 17 Supplementary Material C.5 provides a more detailed discussion on these inconsistencies and identifies regions of the parameter space where the alternative Wald tests diverge most frequently.

Table 3. Share of false positives (type I error rates) using bootstrap critical values.

<i>n</i> = 49					<i>n</i> = 100				
ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$	ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$
0	5.5 (4.696)	1.4 (1.533)	5.8 (4.648)	4.1 (3.330)	0	5.1 (4.308)	1.3 (1.449)	4.8 (4.293)	4.1 (3.218)
0.2	5.2 (4.839)	1.3 (1.364)	5.2 (4.734)	4.7 (3.491)	0.2	4.8 (4.409)	1.1 (1.359)	5.0 (4.379)	4.7 (3.244)
0.4	5.1 (4.963)	2.5 (1.462)	5.4 (4.923)	5.7 (3.853)	0.4	4.8 (4.476)	4.2 (1.788)	4.8 (4.428)	4.3 (3.675)
0.6	5.2 (5.192)	4.3 (2.032)	5.4 (5.132)	6.1 (4.757)	0.6	5.6 (4.600)	6.7 (2.611)	5.6 (4.542)	7.0 (4.387)
0.8	5.8 (5.380)	8.5 (3.083)	5.8 (5.365)	6.8 (5.894)	0.8	4.3 (4.690)	5.2 (3.637)	4.2 (4.646)	4.0 (4.862)

<i>n</i> = 225					<i>n</i> = 400				
ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$	ρ	$H_0(I)$	$H_0(II)$	$H_0(III)$	$H_0(IV)$
0	5.5 (4.125)	0.6 (1.305)	5.6 (4.097)	5.0 (3.629)	0	5.4 (4.038)	0.8 (1.283)	5.5 (4.012)	3.7 (2.917)
0.2	5.0 (4.147)	2.2 (1.593)	5.0 (4.131)	5.4 (4.002)	0.2	4.2 (4.082)	4.0 (1.863)	4.2 (4.074)	6.1 (3.374)
0.4	5.6 (4.168)	6.1 (2.583)	5.6 (4.173)	4.3 (4.515)	0.4	5.1 (4.128)	5.7 (3.037)	5.1 (4.111)	5.6 (3.817)
0.6	5.9 (4.249)	6.0 (3.509)	5.9 (4.193)	5.5 (4.628)	0.6	3.2 (4.137)	3.3 (3.707)	3.3 (4.130)	3.7 (3.850)
0.8	4.9 (4.307)	5.3 (3.999)	4.6 (4.264)	4.5 (4.313)	0.8	4.9 (4.179)	4.9 (4.010)	4.9 (4.150)	5.4 (3.955)

For the different levels of spatial autocorrelation, the median bootstrap critical values χ^2_{boot} for each variant of the Wald test at the nominal significance level of 5% are displayed in parentheses. Again, the theoretically expected rejection rate across the simulation trials is 5% [3.65%; 6.35%].

strikingly large share of inconsistent inferences across a range of parameter settings, the evidence presented here strongly caution against the use of the standard Wald test based on an asymptotic reference distribution.

4.3 Performance of the Modified Wald Test Based on Bootstrap Critical Values

While the simulations performed here illustrate that the standard Wald test based on asymptotic critical values is unreliable for the identification of the unobservable spatial process, this section investigates whether the application of simulated reference distributions improves the test's performance. To this end, I use the bootstrap procedure outlined in Section 3.3 and compare the observed Wald statistics based on the different formulations of the common factor restriction to their estimated critical values.

The results reported in Table 3 show that the estimated critical values from the bootstrap approach, χ^2_{boot} , displayed in parentheses not only differ from their asymptotic counterpart $\chi^2_{asym} = 3.841$ on which the original Wald test is based. They also reveal sizable discrepancies between the alternative parameterizations of the null hypothesis. While the estimated critical values for $H_0(I)$ and $H_0(III)$ are always higher than χ^2_{asym} , the simulated null distributions of $H_0(II)$ suggest much smaller critical values for this expression of the null hypothesis in most scenarios.

Since the functional expression of the nonlinear common factor restriction determines the Wald statistic's rate of convergence to the asymptotic χ^2 distribution, estimating critical values for each alternative parameterization improves the empirical size of the Wald test in finite samples. Compared to the original tests based on the asymptotically derived critical value, the observed rejection rates of each of the four variants of the Wald test is closer to the nominal significance level of 5% across all sample sizes. Whereas the observed rejection rate of $H_0(I)$ ranges from 6.9% to 11% across the different values of ρ for a sample size of $n = 49$ when relying on the asymptotic χ^2 distribution (see Table 2), its corresponding range is narrowed to 5.1%–5.8% when using bootstrap critical values. Similarly, the bootstrap critical values even improve the size of $H_0(IV)$ which performed poorly under the asymptotic reference distribution. For $n = 49$, basing inferences on the simulated null distribution narrows the range of rejection rates from 2.6%–11.5% to 4.1%–6.8% across the different levels of spatial autocorrelation.

In conclusion, the Monte Carlo evidence presented here demonstrate that using the simulated null distribution as a reference distribution and basing inferences on estimated rather than asymptotically derived critical values ameliorates the problems posed by the Wald test's lack of invariance to alternative parameterizations of the common factor hypothesis.¹⁸ Since the bootstrap critical values account for differences in the convergence rates of the Wald statistics, this modification constitutes a superior alternative that facilitates the empirical assessment of the common factor restriction in spatial regression models. Alternatively, the LR test constitutes another option that is invariant to such reparameterizations (Godfrey and Veall 1998).¹⁹ Hence, irrespective of the empirical model search strategy employed, researchers should utilize the modified Wald test based on the simulated null distribution or the LR test in order to empirically evaluate the appropriateness of the spatial model employed.

5 Empirical Example: Spatial Contagion Effects in Economic Voting

An empirical example helps to demonstrate the consequences of the problem for applied research aiming to evaluate the empirical evidence for a theorized mechanism while ruling out alternative mechanisms. To this end, I reanalyze a study conducted by Williams and Whitten (2015) that investigates spatial contagion effects, understood as the process by which “[...] a policy success or failure of one political party in the eyes of voters similarly affects those parties that are ideologically proximate” (Williams and Whitten 2015, 312). The utilization of different sample sizes and the availability of a plausible alternative mechanism make this study an ideal case to investigate the consequences of the Wald test's lack of invariance to reformulations of the common factor hypothesis.

Williams and Whitten (2015) argue that the electorate not only rewards or punishes the parties forming the current government for the country's economic performance at the ballot box as predicted by the economic voting hypothesis. Since voters group parties based on their ideological stances, the effect of economic prosperity also spills over to ideologically proximate parties irrespective of whether or not these parties also belong to the government. These indirect effects conjectured by the authors link the economic wellbeing of a country to the electoral performance of opposition parties. Therefore, the study contributes to the literature on electoral competition by combining insights from the hitherto separated literatures on economic voting and spatial party competition.

To assess the empirical support for the proposed mechanism, Williams and Whitten (2015, 315–16) analyze data on electoral contests in 23 parliamentary democracies from 1951 to 2005, where the parties constitute the unit of analysis. The change in a party's vote share between

¹⁸ The application of bootstrap critical values can also improve the Wald test's power as Supplementary Material C.6 shows.

¹⁹ Supplementary Material C.7 verifies the good performance of the LR test in the Monte Carlo simulations conducted here.

two consecutive elections is the dependent variable and the country's economic performance, measured by the real GDP per capita growth, is the main regressor of interest.²⁰ In their study, the authors emphasize the importance of spatial regression models which facilitate the estimation of the theoretically expected contagion effects in the form of spatial spillovers. They choose the SAR model specification in order to quantify (global) spillover effects (Williams and Whitten 2015, 313–14). In line with the economic voting literature suggesting that the voters' ability to clearly attribute the responsibility for the economic (mis)fortune is a necessary precondition for economic voting to occur, they estimate separate SAR models for elections with high ($n = 398$) and low levels of clarity ($n = 1,030$). While economic voting itself should be less pronounced in elections characterized by a low clarity of responsibility because it is harder for voters to hold a party accountable for the country's economic performance, the authors expect to find larger spatial contagion effects in this context. The argument proposed in the study is that in low clarity settings, the electorate is more experienced in switching their support from one party to an ideologically similar party. Therefore, the voters' sophistication in terms of reallocating their support creates stronger interdependencies between parties in low clarity elections as compared to high clarity settings where voters can easily identify the party who is responsible for the national state of the economy (Williams and Whitten 2015, 312–13).

Given these theoretical expectations, the SAR model constitutes an appropriate choice as it links the electoral fortune of a party to the performance of the other parties and allows to distinguish between direct and indirect effects of economic prosperity on the parties' vote shares. Yet, while unfocused diagnostics, like Moran's I , indicate the existence of spatial interdependencies, it is possible that an alternative spatial process caused the clustering detected in the data.²¹ Unmodeled election specific particularities, for example, that are unrelated to a country's economic performance—like the general appeal of a candidate or political scandals—might affect the election outcome of ideologically proximate parties as well. Since these factors are not part of the regression's systematic part, they potentially cause spatial clustering in the residuals. Instead of substantively meaningful contagion effects, this plausible alternative process implies no indirect effect of economic performance but a mere diffusion of shocks which would be adequately captured by the SEM model. Consequently, there is a risk that the SAR models specified by the authors lead to incorrect inferences regarding the existence of contagion effects.

To demonstrate the substantive differences between the two alternative spatial processes, Figure 2 displays the estimated direct and indirect effects of the main regressor of interest—economic performance—on the vote share of opposition parties derived from the SAR, SEM, and SDM model.²² As the theory suggests, spatial contagion effects should mitigate the negative effect of a strong economy for opposition parties. This is because the beneficial effect of positive economic conditions for governing parties spills over to ideologically neighboring opposition parties. In contrast, if merely the errors are spatially correlated, no contagion takes place and only a direct negative effect of a country's economic performance exists for opposition parties.

Despite a significant spatial parameter estimate, Figure 2 illustrates that the alternative spatial models suggest no indirect spillover effect of economic growth in low clarity elections.²³ In high clarity elections, only the SAR model identifies significant spillover effects. While the SEM model assumes no spillovers, the average indirect impact of economic growth on the change in vote share

- 20 Williams and Whitten (2015) present a more detailed discussion on the dataset as well as a comprehensive derivation of their theoretical expectations regarding spatial contagion effects.
- 21 Table 1 in the study by Williams and Whitten (2015), 316 reports the Moran's I and Geary's C tests of spatial interdependencies which both indicate spatial autocorrelation. Note that, while this table also reports results of a Wald test, this is not the Wald test of common factors but rather a test of the null hypothesis of no spatial dependence (see Section 3.2).
- 22 While Table 2 in the original work only contains the prespatial marginal effects, the estimates presented here explicitly disentangle direct and indirect (or contagion) effects.
- 23 This highlights the necessity to base substantive inferences on impact rather than coefficient estimates. Elhorst (2010) and LeSage and Pace (2009) provide a more detailed discussion on this important issue.

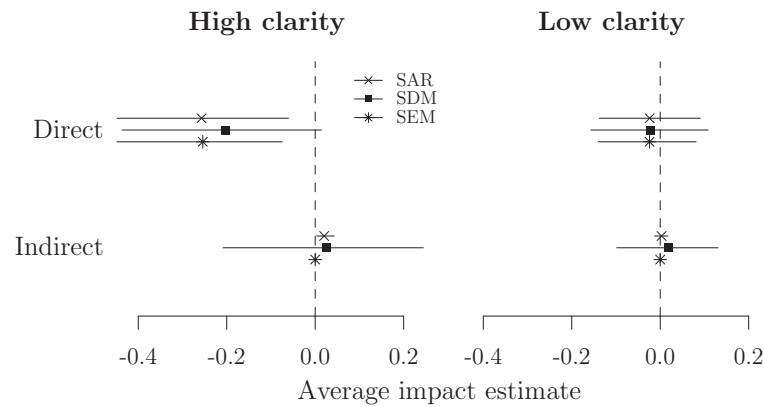


Figure 2. Average direct and indirect impact estimates of economic performance in low- and high-clarity elections.

for opposition parties as estimated by the SAR model is 0.020 with a simulated 95% confidence interval within [0.004;0.045].²⁴ In contrast, the estimate derived from the SDM specification is 0.025 [−0.208;0.244], suggesting no significant spillovers. Besides the SDM model’s remarkable efficiency loss, this example illustrates that the identification of the theorized contagion effects is contingent on the specification of the underlying spatial process. Although the SAR and SEM models produce similar and statistically indistinguishable total impact estimates of economic performance in high clarity elections, the results have very different theoretical implications.²⁵ Notably, the overall impact of economic growth on an opposition party’s vote share in the SEM model solely consists of the direct impact of x_i on y_i . In contrast, the SAR model also identifies significant indirect impacts. Therefore, while the SAR model supports the theory of spatial contagion effects, there are no substantive spillover effects in the SEM model which highlights the importance of adequately distinguishing between these alternative processes for substantive inferences.

In order to address the problem of model misspecification and to empirically distinguish the two plausible spatial processes, I implement the Wald test of common factors by using the SDM model estimates and the four specifications of the common factor restriction outlined in Table 1. If the data supports the theory of indirect contagion effects, the tests should reject the null hypothesis. Yet, Table 4 illustrates that in both high and low clarity contexts, the alternative Wald tests not only differ in their test statistics. Based on the asymptotic critical values, they also come to substantively different conclusions regarding the existence of the spillover effects. While $H_0(IV)$ supports the theory proposed by Williams and Whitten (2015), the other three alternative versions of the Wald test fail to reject the common factor hypothesis. Instead of substantively meaningful spillovers, these tests only indicate residual dependence which implies that no spatial contagion takes place among the parties. Given the rather large number of observations in the low clarity scenario, these differences become even more alarming. Alternatively, when inferences regarding the underlying spatial process are based on the simulated null distribution of each parameterization of the common factor restriction, all four variants of the Wald test fail to reject the null hypothesis at conventional significance levels.

Taken together, this empirical case study confirms the Monte Carlo evidence by demonstrating that relying on bootstrap critical values in order to identify statistically significant deviations from

- 24 In order to appropriately account for sampling uncertainty, I use the point estimates and the variance-covariance matrices obtained from the different spatial models to set up multivariate normal distributions from which I sample 1,000 sets of coefficients.
- 25 The SAR model’s ATI estimate in high clarity elections is −0.237 [−0.410;−0.056] while the ATI derived from the SEM model is −0.254 [−0.448;−0.075].

Table 4. Wald tests of common factors for the analysis of spatial contagion effects.

	High clarity				Low clarity			
	<i>w</i>	<i>p_{asym}</i>	χ^2_{boot}	<i>p_{boot}</i>	<i>w</i>	<i>p_{asym}</i>	χ^2_{boot}	<i>p_{boot}</i>
$H_0(I)$	18.02	0.39	38.15	0.41	19.81	0.41	29.43	0.42
$H_0(II)$	8.20	0.96	12.72	0.27	15.60	0.68	22.92	0.36
$H_0(III)$	9.19	0.93	20.40	0.51	12.73	0.85	19.30	0.35
$H_0(IV)$	1,596.76	0.00	11,276.70	0.36	465.51	0.00	1,323.40	0.58
<i>n</i>		398				1,030		
<i>df</i>		17				19		
χ^2_{asym}		27.59				30.14		

w is the observed Wald statistic. *p_{asym}* and *p_{boot}* denote *p* values based on asymptotic critical values χ^2_{asym} and on bootstrap critical values χ^2_{boot} respectively. While each restriction has an individual simulated critical value, they share a single asymptotic critical value which depends on the α -level and the number of degrees of freedom.

the Wald statistic’s null distribution improves its finite sample performance and alleviates the conflict between alternative parameterizations of the nonlinear common factor hypothesis. While the tests based on the asymptotic χ^2 distribution come to contradictory conclusions even with a sample size of more than 1,000 observations, the bootstrap procedure is able to correct for this undesirable circumstance. Regarding the theorized mechanism, this analysis finds insufficient evidence to convincingly dispel doubts that, instead of the theorized spatial contagion effects, correlation in the residuals caused the spatial clustering found in the data.

6 Conclusion

Distinguishing substantively meaningful indirect spillover effects from a mere diffusion of random shocks is essential as there is a serious risk of making incorrect inferences when estimating a misspecified model. Yet, the task of appropriately modeling the process underlying observable patterns of interrelatedness between the units poses notable difficulties for political scientists. Although many empirical specification search procedures rely on the Wald test to assess the nonlinear common factor restriction, the test’s lack of invariance to algebraically equivalent formulations of the null hypothesis poses a serious problem for the accuracy of inferences.

This study investigates the consequences of the Wald test’s sensitivity to alternative and algebraically equivalent expressions of the common factor hypothesis for its ability to guide the empirical model specification search. By presenting analytical evidence and using Monte Carlo simulations as well as an empirical example, it shows that the necessity to approximate the sampling variability of a nonlinear function by a Taylor series expansion causes the Wald test’s sensitivity to algebraically equivalent reparameterizations of the null hypothesis. While asymptotically valid, this approximation produces considerable differences in finite samples, depending on the restriction’s functional representation. In many instances, alternative null hypotheses even suggest contradictory conclusions regarding the underlying spatial process since they converge to the Wald statistic’s asymptotic χ^2 distribution at different rates. Given that there is no theoretical justification for any particular expression and since their performance is contingent on the relevant region of the parameter space, the results caution against relying on the Wald test’s asymptotic results in any specification search strategy. Instead, practitioners should either base inferences on a simulated null distribution by estimating bootstrap critical values or turn to the LR test which is invariant to such reparameterizations in order to avoid spurious inferences.

Subsequent research might continue this line of research by developing more reliable strategies that help practitioners to differentiate between substantive and residual dependence. As Mur and Angulo (2009) show, the evidence in favor of any search strategy proposed in the literature is mixed which explains the debate about the most appropriate strategy and prevents the development of general guidelines for the empirical identification of the correct model specification (Rüttenauer 2019, 16). In this regard, Lacombe and LeSage (2015), for example, demonstrate that Bayesian methods constitute a promising alternative to the frequentist null hypothesis significance testing approach. Additionally, multimodel inference might help overcoming the current fixation on model selection and instead allows researchers to focus on the identification of substantively meaningful spillover effects in the data (see also Juhl 2020b). Especially regarding the considerable difficulties researchers face when attempting to empirically distinguish between different spatial processes (e.g., Gibbons and Overman 2012), following this line of investigation will enhance model building and contribute to our understanding of different interaction effects among the units of analysis.

Spatial autocorrelation poses notable challenges for the correct specification and interpretation of statistical models as model misspecification can bias the substantive inferences. Notwithstanding these difficulties, interdependencies are paramount in social science theories which obliges researchers to carefully consider the process generating these dependencies when building empirical models in order to make valid inferences with respect to the theories. Consequently, especially in the absence of design-based identification strategies as proposed by Gibbons and Overman (2012), methodological research facilitating the appropriate specification of spatial models constitutes an important contribution for a thorough assessment of theoretical expectations.

Funding

This research was funded by the German Research Foundation (DFG)—Project-ID 139943784—SFB 884. I also gratefully acknowledge support by the state of Baden-Württemberg through the High Performance Computing Cluster bwHPC (INST 35/1134-1 FUGG) and the University of Mannheim's Graduate School of Economic and Social Sciences.

Acknowledgments

This project has been presented at the EPSA conference 2019 in Belfast, Northern Ireland. I would like to thank Lion Behrens, Thomas Bräuninger, Thomas Plümper, Akisato Suzuki, Garrett Vande Kamp, and Laron K. Williams, the participants of the 2019 CDSS Political Science Colloquium at the University of Mannheim, and four anonymous reviewers as well as the journal's editor Jeff Gill for helpful comments.

Data Availability Statement

Replication code for this article has been published in Code Ocean, a computational reproducibility platform that enables users to run the code, and can be viewed interactively at <https://doi.org/10.24433/CO.1459046.v1>. A preservation copy of the same code and data can also be accessed via Dataverse at Juhl (2020a).

Supplementary Material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2020.23>.

Bibliography

Angulo, A. M., and J. Mur. 2011. "The Likelihood Ratio Test of Common Factors under Non-Ideal Conditions." *Investigaciones Regionales* 21:37–52.

- Anselin, L. 2003. "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics." *International Regional Science Review* 26(2):153–166.
- Betz, T., S. J. Cook, and F. M. Hollenbach. 2019. "Spatial Interdependence and Instrumental Variable Models." *Political Science Research and Methods*. doi:10.1017/psrm.2018.61
- Bivand, R., and G. Piras. 2015. "Comparing Implementations of Estimation Methods for Spatial Econometrics." *Journal of Statistical Software* 63(18):1–36.
- Breusch, T. S., and P. Schmidt. 1988. "Alternative Forms of the Wald Test: How Long is a Piece of String?" *Communications in Statistics - Theory and Methods* 17(8):2789–2795.
- Burrige, P. 1981. "Testing for a Common Factor in a Spatial Autoregression Model." *Environment and Planning A: Economy and Space* 13(7):795–800.
- Cliff, A. D., and J. Keith Ord. 1981. *Spatial Processes: Models & Applications*. London: Pion.
- Dagenais, M. G., and J.-M. Dufour. 1991. "Invariance, Nonlinear Models, and Asymptotic Tests." *Econometrica* 59(6):1601–1615.
- Darmofal, D. 2015. *Spatial Analysis for the Social Sciences. Analytical Methods for Social Research*. New York: Cambridge University Press.
- de Paula Ferrari, S. L., and F. Cribari-Neto. 1993. "On the Corrections to the Wald Test of Non-Linear Restrictions." *Economics Letters* 42(4):321–326.
- Elhorst, J. P. 2010. "Applied Spatial Econometrics: Raising the Bar." *Spatial Economic Analysis* 5(1):9–28.
- Elhorst, J. P. 2014a. "Matlab Software for Spatial Panels." *International Regional Science Review* 37(3):389–405.
- Elhorst, J. P. 2014b. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Heidelberg: Springer.
- Florax, R. J.G.M., H. Folmer, and S. J. Rey. 2003. "Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology." *Regional Science and Urban Economics* 33(5):557–579.
- Florax, R. J.G.M., H. Folmer, and S. J. Rey. 2006. "A Comment on Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology: A Reply." *Regional Science and Urban Economics* 36(2):300–308.
- Gibbons, S., and H. G. Overman. 2012. "Mostly Pointless Spatial Econometrics?" *Journal of Regional Science* 52(2):172–191.
- Godfrey, L. G., and M. R. Veall. 1998. "Bootstrap-Based Critical Values for Tests of Common Factor Restrictions." *Economics Letters* 59(1):1–5.
- Goh, K.-L., and M. L. King. 1996. "Modified Wald Tests for Non-Linear Restrictions: A Cautionary Tale." *Economics Letters* 53(2):133–138.
- Greene, W. H. 2012. *Econometric Analysis*. 7th edn. Boston, MA: Pearson Education, Limited.
- Gregory, A. W., and M. R. Veall. 1985. "Formulating Wald Tests of Nonlinear Restrictions." *Econometrica* 53(6):1465–1468.
- Gregory, A. W., and M. R. Veall. 1986. "Wald Tests of Common Factor Restrictions." *Economics Letters* 22(2-3):203–208.
- Halleck Vega, S., and J. P. Elhorst. 2015. "The SLX Model." *Journal of Regional Science* 55(3):339–363.
- Hendry, D. F. 2006. "A Comment on "Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology"." *Regional Science and Urban Economics* 36(2):309–312.
- Juhl, S. 2020a. "Replication Data for: The Wald Test of Common Factors in Spatial Model Specification Search Strategies." <https://doi.org/10.7910/DVN/CY7YWE>, Harvard Dataverse, V1.
- Juhl, S. 2020b. "The Sensitivity of Spatial Regression Models to Network Misspecification." *Political Analysis* 28(1):1–19.
- King, M. L., and K.-L. Goh. 2002. "Improvements to the Wald Test." In *Handbook of Applied Econometrics and Statistical Inference*, edited by A. Ullah, A. T. K. Wan, and A. Chaturvedi, 251–275. New York: Dekker.
- Lacombe, D. J., and J. P. LeSage. 2015. "Using Bayesian Posterior Model Probabilities to Identify Omitted Variables in Spatial Regression Models." *Papers in Regional Science* 94(2):365–383.
- Lafontaine, F., and K. J. White. 1986. "Obtaining any Wald Statistic You Want." *Economics Letters* 21(1):35–40.
- LeSage, J. P., and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.
- Mur, J., and A. Angulo. 2006. "The Spatial Durbin Model and the Common Factor Tests." *Spatial Economic Analysis* 1(2):207–226.
- Mur, J., and A. Angulo. 2009. "Model Selection Strategies in a Spatial Setting: Some Additional Results." *Regional Science and Urban Economics* 39(2):200–213.
- Newey, W. K., and D. L. McFadden. 1994. Large Sample Estimation and Hypothesis Testing. In *Handbook of Econometrics*, Vol IV, edited by R. F. Engle, and D. L. McFadden, 2111–2245. Amsterdam, Netherlands: North-Holland.
- Pace, R. K., and J. P. LeSage. 2010. Omitted Variables Biases of OLS and Spatial Lag Models. In *Progress In Spatial Analysis: Theory and Computation, and Thematic Applications*, edited by A. Páez, J. Le Gallo, R. N. Buliung, and S. Dall'erba, 17–28. Berlin, Germany: Springer.
- Phillips, P. C. B., and J. Y. Park. 1988. "On the Formulation of Wald Tests of Nonlinear Restrictions." *Econometrica* 56(5):1065–1083.

- Rüttenauer, T. 2019. "Spatial Regression Models: A Systematic Comparison of Different Model Specifications Using Monte Carlo Experiments." *Sociological Methods & Research*. doi:10.1177/0049124119882467
- Shipan, C. R., and C. Volden. 2008. "The Mechanisms of Policy Diffusion." *American Journal of Political Science* 52(4):840–857.
- Williams, L. K., and G. D. Whitten. 2015. "Don't Stand So Close to Me: Spatial Contagion Effects and Party Competition." *American Journal of Political Science* 59(2):309–325.