PA

Measurement Error and the Specification of the Weights Matrix in Spatial Regression Models

Garrett N. Vande Kamp

Texas A&M University, Bush School of Government and Public Service, 4348 TAMU, College Station, TX 77843, USA.
Email: garrettvandekamp@tamu.edu

Abstract
While the spatial weights matrix W is at the core of spatial regression models, there is a scarcity of techniques for validating a given specification of W. I approach this problem from a measurement error perspective. When W is inflated by a constant, a predictable form of endogeneity occurs that is not problematic in other regression contexts. I use this insight to construct a theoretically appealing test and control for the validity of W that is tractable in panel data, which I call the K test. I demonstrate the utility of the test using Monte Carlo simulations.

Keywords: spatial regression, spatial autocorrelation, measurement error, time series cross-section data

1 Introduction

Spatial regression models are popular in the social sciences for modeling spatial dependencies within data. At the core of these models is the spatial weights matrix W, an $n \times n$ matrix with zeroes along the diagonal that, in the cross-sectional context, contains the set of spatial relationships in the data in its off-diagonal elements: each element w_{ij} represents the relative influence unit j has on unit i. More extreme values of w_{ij} indicate a stronger influence of unit j on unit i. Likewise, unit j has no influence on unit i when w_{ij} equals zero.

There are many approaches to specifying W, both theoretical and empirical (Bhattacharjee and Jensen-Butler 2013; Neumayer and Plumper 2016). There are also a number of challenges to specifying W, including using the correct conception of space, correct means of expressing the same conception of space, and measuring w_{ij} without error. Potential misspecification creates a problem for scholars, as the estimates from spatial regression models are conditional upon W. A misspecification of W, even by a constant, will cause inconsistent estimates (Neumayer and Plumper 2016). This is not the case in other regression contexts, as inflation of a variable by a constant will be absorbed into the intercept. Scholars largely agree that the misspecification of W creates problems for inference, though this view is not universal (Corrado and Fingleton 2012; Gibbons and Overman 2012, but see LeSage and Pace 2014).

Despite the risks of misspecifying W, few techniques exist for evaluating a particular specification. Some developed techniques adjudicate between competing W matrices (Leenders 2002; Stakhovych and Bijmolt 2009). But these approaches are limited in that they can only help judge the best W among a set of choices, leading to the possibility of picking the best out of a set of badly misspecified matrices (Harris, Moffat and Kravtsova 2011). Neumayer and Plumper offer a test to determine if the nondiagonal elements of W that are set to zero are valid (2016, p. 123).

The test, however, is only theoretically described and gives no indication about the potential misspecification of nonzero elements of $\boldsymbol{W}$.

This letter highlights the problematic bias resulting from the misspecification of $\boldsymbol{W}$ and provides a test and control for some misspecifications. Specifically, each $w_{ij}, i \neq j$, within $\boldsymbol{W}$ is thought to be measured with some amount of error. From this approach, I develop a theoretically appealing regression-based test of the specification of $\boldsymbol{W}$ that is tractable when using panel data. I call this the K test, which is simply another spatial lag where all nondiagonal elements of $\boldsymbol{W}$ equal one. The test is able to detect when $\boldsymbol{W}$ is inflated by a constant in expectation. Inclusion of the K test in the regression model controls for part of the endogeneity resulting from this type of misspecification, at least in SLX models, much like how including an intercept in a regression model controls for the inflation of a variable by a constant.

2   The Specification of W as a Measurement Error Problem

To begin, I examine classical measurement error in a simple regression context. Consider the following data-generating process:

$$ \boldsymbol{y} = \boldsymbol{\iota}\alpha + \boldsymbol{x}^*\beta + \boldsymbol{\varepsilon} \tag{1} $$

in which $\boldsymbol{\varepsilon}$ is a well-behaved error term and $\boldsymbol{\iota}$ is a vector of ones associated with the intercept parameter α. Suppose a scholar would like to model the relationship between $\boldsymbol{x}^*$ and $\boldsymbol{y}$ but does not have access to $\boldsymbol{x}^*$. Instead, the scholar has access to $\boldsymbol{x}$, a function of $\boldsymbol{x}^*$:

$$ \boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{u}^* \tag{2} $$

in which $\boldsymbol{u}^*$ is an error variable with mean c that is uncorrelated to $\boldsymbol{x}^*$. It will be useful for our purposes to decompose $\boldsymbol{u}^*$ into mean-zero measurement error and a constant:

$$ \boldsymbol{x} = \boldsymbol{x}^* + (\boldsymbol{u}^* - \boldsymbol{c}) + \boldsymbol{c} \equiv \boldsymbol{x}^* + \boldsymbol{u} + \boldsymbol{c} \tag{3} $$

in which $\boldsymbol{u}$ has a mean of zero but is otherwise similarly distributed to $\boldsymbol{u}^*$ and $\boldsymbol{c}$ is a vector of the constant c. By substituting equation (3) into equation (1), we observe the problems associated with measurement error in the independent variable:

$$ \boldsymbol{y} = \boldsymbol{\iota}\alpha + (\boldsymbol{x} - \boldsymbol{u} - \boldsymbol{c})\beta + \boldsymbol{\varepsilon} \tag{4} $$

$$ \Leftrightarrow \boldsymbol{y} = \boldsymbol{\iota}\alpha + \boldsymbol{x}\beta + (\boldsymbol{\varepsilon} - \boldsymbol{u}\beta - \boldsymbol{c}\beta) \tag{5} $$

$$ \Leftrightarrow \boldsymbol{y} = (\boldsymbol{\iota}\alpha - \boldsymbol{c}\beta) + \boldsymbol{x}\beta + (\boldsymbol{\varepsilon} - \boldsymbol{u}\beta). \tag{6} $$

In equation (6), there are two sources of endogeneity in the model. The first results because $\boldsymbol{u}\beta$ is contained within the error term. The terms $\boldsymbol{u}$ and $\boldsymbol{x}$ are necessarily correlated given the data-generating process in equation (3). In the bivariate regression case, the estimate of β will be biased toward zero; this bias is commonly known as attenuation bias. In the multiple-regression context, bias will permeate all independent variables in unknown directions in addition to the attenuation bias in the mismeasured variable. Scholars have developed a number of approaches to circumventing attenuation bias (see Buonaccorsi 2010). In general, however, the onerous requirements for these solutions means that they are rarely implemented in the applied literature (Greene 2017).

The second source of endogeneity comes from $\boldsymbol{c}\beta$, which results in the expectation of the error term being nonzero. This endogeneity is transferred from the error term to the intercept parameter α; as a consequence, only α is biased. This is rarely problematic for scholars, as the quantity

of interest is usually not α but rather β. Thus, while inferences from regression models are not immune to measurement error, part of the endogeneity from nonzero measurement error is solved by including an intercept.

2.1 The Problem of Measurement Error Within $\mathbf{W}$

As we have just seen, part of the problem associated with nonzero measurement error is remedied by including an intercept in the regression model. Unfortunately, this will not alleviate similar problems when misspecifying $\mathbf{W}$ in a spatial regression model. Consider the spatial lag of an independent variable (SLX) model:

$$\mathbf{y} = \boldsymbol{\iota}\alpha + \mathbf{x}^*\beta + \mathbf{W}^*\mathbf{x}^*\theta + \boldsymbol{\varepsilon} \tag{7}$$

in which $\boldsymbol{\varepsilon}$ is a well-behaved error term. Suppose a scholar would like to model the relationship between $\mathbf{x}^*$ and $\mathbf{y}$ but does not have access to $\mathbf{W}_*$. Instead, the scholar has access to $\mathbf{W}$, a representation of $\mathbf{W}^*$ measured with error:

$$\mathbf{W} \equiv \mathbf{W}^* + \mathbf{U} + \mathbf{C} \tag{8}$$

in which $\mathbf{U}$ is a weights matrix where all nondiagonal elements are drawn from variable $\mathbf{u}$, which has a mean of zero, and $\mathbf{C}$ is a weights matrix where all nondiagonal elements are equal to a constant c. Note how equation (8) mirrors equation (3). By substituting equation (8) into equation (7), we observe unique problems created by measurement error within $\mathbf{W}$:

$$\mathbf{y} = \boldsymbol{\iota}\alpha + \mathbf{x}^*\beta + (\mathbf{W} - \mathbf{U} - \mathbf{C})\mathbf{x}^*\theta + \boldsymbol{\varepsilon} \tag{9}$$

$$\Leftrightarrow \mathbf{y} = \boldsymbol{\iota}\alpha + \mathbf{x}^*\beta + \mathbf{W}\mathbf{x}^*\theta + (\boldsymbol{\varepsilon} - \mathbf{U}\mathbf{x}^*\theta - \mathbf{C}\mathbf{x}^*\theta). \tag{10}$$

As before, there are two sources of endogeneity. The first is from the mean-zero measurement error. Because of the correlation between the elements of $\mathbf{W}$ and $\mathbf{U}$, $\mathbf{W}\mathbf{x}^*$ and $\mathbf{U}\mathbf{x}^*$ will also be correlated. As mentioned before, this attenuation bias is unlikely to be solved due to the problems previously noted in solving mean-zero measurement error in simpler regression contexts. The second source of endogeneity is derived from the matrix $\mathbf{C}$ and its nondiagonal elements c. Unlike in the simple regression model, $\mathbf{C}$ is postmultiplied by $\mathbf{x}^*$ to create a new variable. Any change to the expectation of the error term resulting from $\mathbf{C}\mathbf{x}_*$ is dependent upon the expectation of $\mathbf{x}^*$. As in the previous context, the expectation of $-\mathbf{C}\mathbf{x}^*\theta$ will be absorbed into the constant.

But the intercept will not absorb the entirety of $-\mathbf{C}\mathbf{x}^*\theta$, however, because of the variance associated with $\mathbf{x}^*$. Further, this variation will necessarily be correlated with $\mathbf{x}^*$ since $-\mathbf{C}\mathbf{x}^*\theta$ is a linear transformation of the total of $\mathbf{x}_*$ within a cross-section minus the influence of x_i^*. Thus, there will be additional bias in β and the coefficients of all other variables correlated with $\mathbf{x}^*$ that is unique to the spatial regression context.

It is important here to note that measurement error in $\mathbf{W}$ is not unique to the SLX context. Spatial autoregressive and spatial error models (SAR and SEM, respectively) also use $\mathbf{W}$ to account for spatial dependence. If $\mathbf{W}$ is measured with error, it is highly likely that estimates from these models will also exhibit bias in some way. The additional econometric difficulties associated with these models, however, prevents an analytic demonstration of this bias at the present.

2.2 A Test and Control for the Misspecification of $\mathbf{W}$

The misspecification of $\mathbf{W}$ introduces additional endogeneity that is not a problem in simpler regression contexts and will cause inconsistent estimates. This additional endogeneity, however, is also an opportunity to develop a specification test of $\mathbf{W}$ by attempting to model that

endogeneity. For the SLX data-generating process in equation (7), the model is equivalent to:

$$y = \iota\alpha + x^*\beta + W^*x^*\theta + Kx^*0 + \varepsilon \tag{11}$$

in which K is a W matrix where every nondiagonal element equals one. Because the new term Kx^* has a coefficient of zero, it does not influence y and is thus equivalent to the original model. But if a scholar only had W instead of W^* and modeled this relationship, one would get the following result:

$$y = \iota\alpha + x^*\beta + (W - U - C)x^*\theta + Kx^*0 + \varepsilon \tag{12}$$

$$\Leftrightarrow y = \iota\alpha + x^*\beta + Wx^*\theta + Kx^*0 + (\varepsilon - Ux^*\theta - Cx^*\theta) \tag{13}$$

$$\Leftrightarrow y = \iota\alpha + x^*\beta + Wx^*\theta + Kx^*0 + (\varepsilon - Ux^*\theta - Kx^*c\theta) \tag{14}$$

$$\Leftrightarrow y = \iota\alpha + x^*\beta + Wx^*\theta + Kx^*(-c\theta) + (\varepsilon - Ux^*\theta). \tag{15}$$

A substantially changed result appears in equation (16). As before, there will still be bias resulting from mean-zero measurement error in W. Now, however, the additional bias resulting from $-Cx^*\theta$ is transferred to the coefficient of Kx^* rather than staying in the error term and affecting inferences in the rest of the model. Thus, scholars concerned about the misspecification of W should estimate a full model:

$$y = \iota\alpha + x^*\beta + Wx^*\theta_1 + Kx^*\theta_2 + \varepsilon. \tag{16}$$

Kx^*, hereafter referred to as Kx or the K test, can serve as a specification test of W within the regression model. If one assumes that Kx—or its counterparts Ky and $K\varepsilon$ in other models—is not a part of the data-generating process, then the K test should be zero in expectation and distinguishable from zero only by chance.[1] If the elements of W have been misspecified by a constant in expectation, then the K test should be nonzero and distinguishable from zero at rates greater than expected by chance. Thus, a K test that is statistically distinguishable from zero is evidence that W is misspecified and that the scholar has failed to adequately model spatial relationships within the data.[2]

Beyond functioning as a specification test, the K test also controls for a part of the endogeneity resulting from the misspecification of W. By absorbing the effect of $-Cx^*\theta$, the only endogeneity remaining from the misspecification of W is $-Ux^*\theta$. This is similar to how the intercept controls for part of the endogeneity resulting from measurement error in an additive variable. While the leftover endogeneity will result in bias in the rest of the regression model, it now makes spatial regression models comparable to other regression models in that nonzero measurement error is modeled explicitly in the regression.[3]

While the K test can be used outside of the SLX case, it is unclear how the K test will perform in the SAR and SEM cases. It is highly possible that the properties of the K test are different given their nonrecursive nature. I do not have an analytic proof on the point, and the Monte Carlo results presented later provide evidence that the additional complexities associated with these models need to be taken into account. Future research should further examine the K test's properties in these settings.

---

1 This should not be a controversial assumption, as scholars rarely, if ever, hypothesize a data-generating process containing Kx, Ky, or $K\varepsilon$; no such model could be found in my search of the literature.

2 In applied settings, the K test cannot distinguish from a W matrix that is simply misspecified by a constant and an omitted spatial lag (Hays, Kachi and Franzese 2010). Thus, it can only determine that the spatial processes have been inadequately modeled.

3 For more on the relationship between the K test and mean-zero measurement error, see the appendix.

2.3 Limits and Opportunities in Cross-Sectional and Panel Data Settings

While the K test is theoretically appealing, it suffers from crippling limitations in cross-sectional settings. In a SLX model of a single cross-section, the sum of Kx and x results in a constant that will be perfectly multicollinear with the intercept. The researcher is left with an untenable choice between forgoing the test and omitting a variable and causing bias. The problem is not limited to the SLX case; the use of the K test alongside a constant in the SAR case results in inconsistent estimates (Kelejian and Prucha 2002).[4]

This limitation is mitigated when using panel data (Franzese and Hays 2007). If x varies over both time and space and Wx is a purely spatial lag[5], then the sum of x and Kx results in a time-invariant variable. This allows for the estimation of x, Kx, and the constant in the same regression model. The same is true in the SAR setting (Kelejian, Prucha and Yuzefovich 2006).

But the use of the test in these instances also has unique considerations associated with it. When a spatial lag is taken of x, then the sum of x and Kx results in a new variable, x_{total}, that only varies over time. This makes the test potentially sensitive to temporal unit effects, both that are uncorrelated and correlated with x_{total}. I offer suggestions for each case in turn. If there are temporal unit effects that are uncorrelated with x_{total} but are not accounted for in the model, the K test may report false positives. Fortunately, it is relatively easy to model these effects using random effects.

More problematic is the possibility of temporal unit effects correlated with x_{total}. A scholar could normally combat this problem using fixed effects. The lack of spatial variation in x_{total}, however, prevents the inclusion of temporal fixed effects in SLX models. SAR models have similar problems: estimates are inconsistent with the inclusion of fixed effects and the K test (Kelejian, Prucha and Yuzefovich 2006). As present, there is no readily available solution to employ both techniques. Scholars must assume an otherwise well-specified regression model, at least on the temporal dimension, in order to make inferences about whether the spatial relationships specified in the data are valid. Alternatively, temporal fixed effects should control from the endogeneity resulting from W being inflated by a constant (though it would be impossible to infer whether W is misspecified).

3 Simulation

To demonstrate the K test's utility, I employ two sets of Monte Carlo simulations: one featuring cross-sectional data and the other panel data.[6] One thousand data points are generated that either follow a SAR data-generating process or a SLX data-generating process:

$$y = x\beta + Wy\rho + Ky0 + \varepsilon$$
$$y = x\beta + Wx\theta + Kx0 + \varepsilon$$

where $\beta = 5$ and $\rho = \theta = 0.75$. ε is drawn from a standard normal distribution, while x is generated from a uniform distribution between -1 and 1. W is randomly generated. As per the above discussion, the intercept is omitted in cross-sectional data. In panel data, an intercept $\iota\alpha$ is included with $\alpha = 10$.[7]

A spatial regression model is then estimated on the data that also employs my test of the specification of W, which includes either Ky or Kx depending on the regression model. OLS is used to estimate the SLX models, while 2SLS is used to estimate the SAR models using Wx and Kx as instruments. This process is repeated one thousand times. In the analysis, I see how the test fairs

---

4   While unsubstantiated here, this limitation may also apply to the SEM case.
5   As opposed to a lag over time and space.
6   Replication materials for this article are available from the Political Analysis Dataverse, cited at the end of this article.
7   For panel data, $n = 100$ and $t = 10$, resulting in 1,000 observations.

Table 1. The K test and bias in SLX Models using cross-sectional and panel data when **W** is misspecified by a constant (Coverage Probabilities in parentheses).

		−45	−25	0	+25	+45
Cross-Section	$\hat{\beta}$	0.00	0.00	0.00	−0.00	0.00
		(0.94)	(0.94)	(0.96)	(0.95)	(0.95)
	$\hat{\theta}$	0.00	−0.00	−0.00	0.00	−0.00
		(0.95)	(0.95)	(0.95)	(0.95)	(0.94)
	K test	33.75	18.75	0.00	−18.75	−33.75
		(0.00)	(0.00)	(0.95)	(0.00)	(0.00)
Panel	$\hat{\alpha}$	−0.00	0.00	−0.00	−0.00	−0.00
		(0.96)	(0.94)	(0.95)	(0.95)	(0.95)
	$\hat{\beta}$	−0.00	0.00	0.00	0.00	0.00
		(0.95)	(0.94)	(0.95)	(0.95)	(0.96)
	$\hat{\theta}$	0.00	−0.00	0.00	0.00	0.00
		(0.95)	(0.94)	(0.96)	(0.95)	(0.96)
	K test	33.75	18.75	−0.00	−18.75	−33.75
		(0.00)	(0.00)	(0.96)	(0.00)	(0.00)

when a model is estimated using the correct **W** matrix inflated by a constant. When the matrix is inflated by zero, this reduces to a model with a correctly specified **W**. I subsequently report the average bias for all coefficient estimates and their coverage probabilities using 95% confidence intervals. These quantities are reported for two reasons. First, these quantities demonstrate both the size and power of the K test's ability to detect misspecifications of **W**. Second, these quantities show the K test's ability to control for the bias caused by the misspecification of **W** when estimating a model. If the K test is a valid way for controlling for the misspecification of **W**, then the bias of all other coefficients should be zero and their coverage probabilities should be 95%. This should be true regardless of whether **W** is correctly or incorrectly specified.

Tables 1 and 2 present the results for misspecifying a **W** matrix by a constant for SLX and SAR processes, respectively. As we can see in Table 1, the K test works as a specification test in the SLX context. When **W** is correctly specified, the test's coefficient is centered on zero and correctly sized (in that it is only statistically distinguishable from zero about five percent of the time). But when **W** is inflated by a constant, either positive or negative, the test is consistently biased in a singular direction and almost always statistically distinguishable from zero. The K test, then, has the expected size and power in the SLX context.

The K test also works as a control for endogeneity in the SLX context. Across all model specifications, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ are centered on their true values and their confidence intervals regularly include them. In the instances in which **W** is misspecified, these estimates would normally be biased. The only reason they are not is due to the inclusion of the K test in the model, thereby demonstrating its utility as a control for this form of endogeneity.

In Table 2, the K test's utility as a specification test in the SAR case is strong in applied settings. Using both cross-sectional and panel data, the K test has the same size in the SAR context as it does in the SLX context: it is centered around zero when **W** is correctly specified. The test has slightly lower power than its SLX counterpart when using panel data. When using it with cross-sectional data, however, the test has little power. When **W** is inflated by a positive constant, the test appears to have no power at all. While concerning on the surface, this limitation will have little relevance in applied work given that the K test cannot be consistently estimated in a single cross-section.

The K test's utility as a control in the SAR case is not as strong as it is in the SLX case. In both the cross-sectional and panel case, $\hat{\beta}$ is not biased when the K test is estimated, even if **W** is misspecified. In contrast, though, $\hat{\rho}$ is biased when **W** is misspecified even when the K test

Table 2. The K test and bias in SAR Models using cross-sectional and panel data when **W** is misspecified by a constant (Coverage Probabilities in parentheses).

		−45	−25	0	+25	+45
Cross-Section	$\hat{\beta}$	−0.00	−0.00	0.00	−0.00	−0.00
		(0.95)	(0.96)	(0.94)	(0.95)	(0.95)
	$\hat{\rho}$	−0.45	−0.22	0.31	0.82	0.97
		(0.72)	(0.91)	(0.95)	(0.95)	(0.96)
	K test	0.34	0.35	−0.46	−0.70	−0.99
		(0.74)	(0.92)	(0.95)	(0.95)	(0.96)
Panel	$\hat{\alpha}$	0.19	0.11	0.17	0.20	0.17
		(0.95)	(0.97)	(0.96)	(0.97)	(0.96)
	$\hat{\beta}$	−0.00	0.00	−0.00	−0.00	−0.00
		(0.97)	(0.97)	(0.97)	(0.98)	(0.97)
	$\hat{\rho}$	−0.52	−0.29	0.00	0.29	0.53
		(0.00)	(0.12)	(1.00)	(0.69)	(0.39)
	K test	0.54	0.29	−0.01	−0.30	−0.54
		(0.00)	(0.14)	(1.00)	(0.70)	(0.38)

is included. This result is likely due to the nonrecursive nature of SAR models compared to SLX models. Interestingly, the coefficient of the K test is biased in the exact opposite direction when **W** is misspecified. Additional analysis in the replication materials reveal that a linear combination of $\hat{\rho}$ and the K test's coefficient is equal to ρ and has the correct coverage probability of 95%. This curious result does not have a readily available explanation. Regardless, additional consideration needs to be given to the SAR case given its complexity.

4 Discussion

This letter presents the uncertainty surrounding the specification of **W** as a measurement error problem. The approach clearly demonstrates the bias that results when **W** is misspecified in regression models and reveals a theoretically appealing test and control for the misspecification of **W** by a constant, which I call the K test. But while the K test has its uses, it is not a silver bullet for measurement error in **W**. The test's utility as a control in the SAR case is limited. The test also does not identify or control for mean-zero measurement error. Finally, this letter does not consider the possibility that measurement error in **W** is correlated with measurement error in other variables in the model. Future research should consider these possibilities.

Scholars must exercise care when using the K test to adjudicate between multiple **W** matrices. If the test is used multiple times, the results must be appropriately weighted in order to avoid spurious inferences that result in repeated hypothesis testing. Alternatively, this problem can be avoided by using the test in conjunction with other methods of adjudicating between multiple **W** matrices. This would result in a two-step procedure: identify the best choice of **W** among a set of alternatives and then test the validity of that choice.

Appendix

This letter focuses on the ability for the K test to control for measurement error resulting from the nondiagonal elements of **W** being inflated by a constant. It implicitly assumes that the K test will not be able to account for mean-zero measurement error. But as one reviewer noted, the K test, or **Kx**, may in fact be correlated with mean-zero measurement error in **W**, **Ux**. If true, then the K test would be able to detect and possibly control for mean-zero measurement error. This implications of this correlation are too important to ignore, and for that reason I using Monte Carlo analysis to investigate whether these two quantities of interest are correlated.
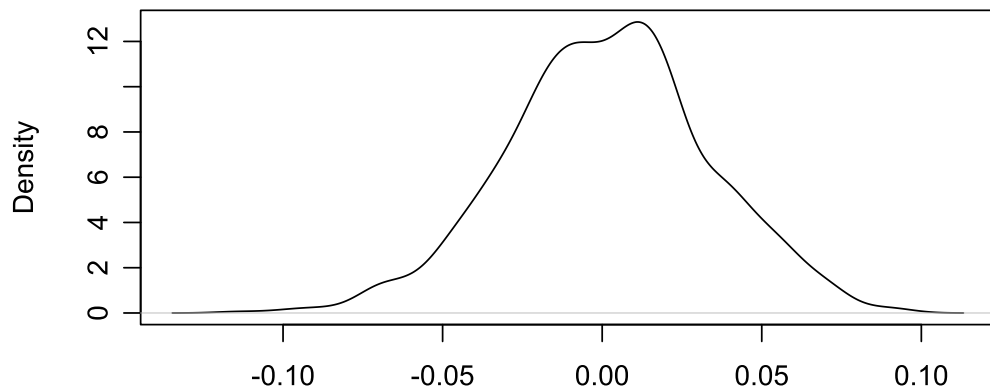
Figure 1. Density Plot of the Correlation between Kx and Ux.

I randomly generate one thousand observations of a variable x, $x \sim U(-a + c, a + c)$. I use this variable to create two more variables, Kx and Ux. I create Kx by premultiplying x by K, a weights matrix where all the nondiagonal elements are equal to 1. I create Ux by premultiplying x by U, a weights matrix where all the nondiagonal elements are randomly drawn from a variable u, $u \sim U(-b, b)$. Parameters a and b are identically distributed, $U(0, 1,000)$, while parameter c is uniformly distributed with mean zero, $U(-1,000, 1,000)$. After both Kx and Ux are generated, I calculate and record their correlation and its 95% confidence interval. I repeat this process one thousand times, with the underlying distributions of x and u varying as well as the set of observations drawn.

A density plot of these one thousand correlations is presented in Figure 1. The density appears to be normally distributed with a mean of zero and low variance. This is exactly what one would expect if Kx and Ux were uncorrelated. In addition, I investigated what proportion of these correlations were statistically distinguishable from zero. The proportion is 0.06, close to the proportion of statistically significant results one would expect when repeatedly calculating a 95% confidence interval on truly uncorrelated data.

There is a possibility that the elements of U could be correlated with measurement error in x, should x be measured with error. In that case, Kx and Ux may be correlated. Similarly, if the elements of U can be predicted by x, then there is a possibility of correlation. But if you assume that the error in W is orthogonal to everything else in the model, as I do throughout, then Kx and Ux will not be correlated.

References

Bhattacharjee, A., and C. Jensen-Butler. 2013. "Estimation of the Spatial Weights Matrix Under Structural Constraints." *Regional Science and Urban Economics* 43(4):617–634.

Buonaccorsi, J. P. 2010. *Measurement Error: Models, Methods, and Applications*. Chapman and Hall.

Corrado, L., and B. Fingleton. 2012. "Where is the Economics in Spatial Econometrics? *Journal of Regional Science* 52(2):210–239.

Franzese, R., and J. Hays. 2007. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis* 15:140–164.

Greene, W. H. 2017. *Econometric Analysis*, 8th edn. Pearson.

Hays, J. C., A. Kachi, and R. Franzese. 2010. "A Spatial Model Incorporating Dynamic, Endogenous Network Interdependence: A Political Science Application." *Statistical Methodology* 7(3):406–428.

Harris, R., J. Moffat, and V. Kravtsova. 2011. "In Search of 'W'." *Spatial Economic Analysis* 6(3):249–270.

Kelejian, H. H., and I. R. Prucha. 2002. "2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights." *Regional Science and Urban Economics* 32(6):691–707.

Kelejian, H. H., I. R. Prucha, and Y. Yuzefovich. 2006. "Estimation Problems in Models with Spatial Weighting Matrices Which Have Blocks of Equal Elements." *Journal of Regional Science* 46(3):507–515.

Leenders, R. T. A. 2002. "Modeling Social Influence Through Network Autocorrelation: Constructing the Weight Matrix." *Social Networks* 24(1):21–47.

LeSage, J. P., and R. K. Pace. 2014. "The Biggest Myth in Spatial Econometrics." *Econometrics* 2(4):217–249.

Gibbons, S., and H. G. Overman. 2012. "Mostly Pointless Spatial Econometrics? *Journal of Regional Science* 52(2):172–191.

Neumayer, E., and T. Plumper. 2016. "W." *Political Science Research and Methods* 4(1):175–193.

Stakhovych, S., and T. Bijmolt. 2009. "Specification of Spatial Models: A Simulation Study on Weights Matrices." *Papers in Regional Science* 88(2):389–408.

Vande Kamp, G. N. 2019. "Replication Data for: Measurement Error and the Specification of the Weights Matrix in Spatial Regression Models," https://doi.org/10.7910/DVN/W9DGRR, Harvard Dataverse, V1.