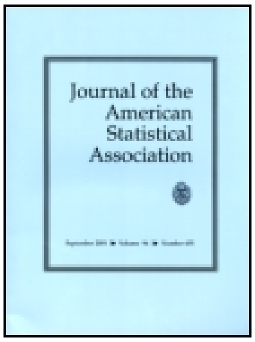


Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Clustering, Spatial Correlations, and Randomization Inference

Thomas Barrios , Rebecca Diamond , Guido W. Imbens & Michal Kolesár

Clustering, Spatial Correlations, and Randomization Inference

Thomas BARRIOS, Rebecca DIAMOND, Guido W. IMBENS, and Michal KOLESÁR

It is a standard practice in regression analyses to allow for clustering in the error covariance matrix if the explanatory variable of interest varies at a more aggregate level (e.g., the state level) than the units of observation (e.g., individuals). Often, however, the structure of the error covariance matrix is more complex, with correlations not vanishing for units in different clusters. Here, we explore the implications of such correlations for the actual and estimated precision of least squares estimators. Our main theoretical result is that with equal-sized clusters, if the covariate of interest is randomly assigned at the cluster level, only accounting for nonzero covariances at the cluster level, and ignoring correlations between clusters as well as differences in within-cluster correlations, leads to valid confidence intervals. However, in the absence of random assignment of the covariates, ignoring general correlation structures may lead to biases in standard errors. We illustrate our findings using the 5% public-use census data. Based on these results, we recommend that researchers, as a matter of routine, explore the extent of spatial correlations in explanatory variables beyond state-level clustering.

KEY WORDS: Clustered standard errors; Confidence intervals; Misspecification; Random assignment.

1. INTRODUCTION

Many economic studies that analyze the causal effects of interventions on economic behavior study interventions or treatments that are constant within clusters whereas the outcomes vary at a more disaggregate level. In a typical example, and the one we focus on in this article, outcomes are measured at the individual level, whereas interventions vary only at the state (cluster) level. Often, the effect of interventions is estimated using least squares regression. Since the mid-1980s (Liang and Zeger 1986; Moulton 1986), empirical researchers in social sciences have generally been aware of the implications of within-cluster correlations in outcomes for the precision of such estimates. The typical approach is to allow for correlation between outcomes in the same state in the specification of the error covariance matrix. However, there may well be more complex correlation patterns in the data. Correlation in outcomes between individuals may extend beyond state boundaries, it may vary in magnitude between states, and it may be stronger in more narrowly defined geographical areas.

In this article, we investigate the implications, for the repeated sampling variation of least squares estimators based on individual-level data, of the presence of correlation structures beyond those that are constant within and identical across states, and vanish between states. First, we address the empirical question whether in census data on earnings with states as clusters such correlation patterns are present. We estimate general spatial correlations for the logarithm of earnings, and find that, indeed, such correlations are present, with substantial correlations within groups of nearby states, and correlations within smaller geographic units (specifically public use microdata areas—pumas) considerably larger than within states. Second, we address whether accounting for such correlations is important for the properties of confidence intervals for the effects of state-level regulations or interventions. We report theoretical results, and demonstrate their relevance using illustrations based on earnings data and state regulations, as well as Monte Carlo evidence. The theoretical results show that if covariate values are as good as randomly assigned to clusters, implying there is no spatial correlation in the covariates beyond the clusters, variance estimators that incorporate only cluster-level outcome correlations remain valid despite the misspecification of the error covariance matrix. Whether this theoretical result is useful in practice depends on the magnitude of the spatial correlations in the covariates. We provide some illustrations that show that, given the spatial correlation patterns we find in the individual-level variables, spatial correlations in state-level regulations can have a substantial impact on the precision of estimates of the effects of interventions.

The article draws on three strands of literature that have largely evolved separately. First, it is related to the literature on clustering, where a primary focus is on adjustments to standard errors to take into account clustering of explanatory variables (see, e.g., Liang and Zeger 1986; Moulton 1986; Bertrand, Duflo, and Mullainathan 2004; Hansen 2007; for textbook discussions, see Diggle et al. 2002; Wooldridge 2002; Angrist and Pischke 2009). Second, the current article draws on the literature on spatial statistics. Here, a major focus is on the specification and estimation of the covariance structure of spatially linked data (for textbook discussions, see Schabenberger and Gotway 2004; Gelfand et al. 2010). In interesting recent works, Bester, Conley, and Hansen (2011) and Ibragimov and Müller (2010) link some of the inferential issues in the spatial and clustering literatures. Finally, we use results from the literature on randomization inference going back to Fisher (1925) and Neyman (1990) (for a recent textbook discussion, see

Thomas Barrios is a Graduate Student, Department of Economics, Harvard University, Cambridge, MA 02138 (E-mail: tbarrios@fas.harvard.edu). Rebecca Diamond is a Graduate Student, Department of Economics, Harvard University, Cambridge, MA 02138 (E-mail: rdiamond@fas.harvard.edu). Guido Imbens is Professor, Department of Economics, Harvard University and Research Associate, National Bureau of Economic Research, Cambridge, MA 02138 (E-mail: imbens@harvard.edu). Michal Kolesár is Graduate Student, Department of Economics, Harvard University, Cambridge, MA 02138 (E-mail: mkolesar@fas.harvard.edu). Financial support for this research was generously provided through NSF grants 0631252, 0820361, and 0961707. The authors thank participants in the econometrics workshop at Harvard University, the referees, the editor, and the associate editor for comments, and, in particular, Gary Chamberlain for helpful discussions.

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Applications and Case Studies
DOI: 10.1080/01621459.2012.682524

Rosenbaum 2002). Although the calculation of Fisher exact p-values based on randomization inference is frequently used in the spatial statistics literature (e.g., Schabenberger and Gotway 2004), and sometimes in the clustering literature (Bertrand, Duflo, and Mullainathan 2004; Abadie, Diamond, and Hainmueller 2010), Neyman's approach to constructing confidence intervals using the randomization distribution is rarely used in these settings. We will argue that the randomization perspective provides useful insights into the interpretation and properties of confidence intervals in the context of spatially linked data.

The article is organized as follows. In Section 2, we introduce the basic set-up. Next, in Section 3, using census data on earnings, we establish the presence of spatial correlation patterns beyond the constant-within-state correlations typically allowed for in empirical work. In Section 4, we discuss randomization-based methods for inference, first focusing on the case with randomization at the individual level. Section 5 extends the results to cluster-level randomization. In Section 6, we present the main theoretical results. We show that if cluster-level covariates are randomly assigned to the clusters, the standard variance estimator based on within-cluster correlations can be robust to misspecification of the error covariance matrix. Next, in Section 7, we show, using Mantel-type tests, that a number of regulations exhibit substantial regional correlations, suggesting that ignoring the error correlation structure may lead to invalid confidence intervals. Section 8 reports the results of a small simulation study. Section 9 concludes. Proofs are collected in the Appendix.

2. FRAMEWORK

Consider a setting where we have information on N units, say individuals in the United States, indexed by $i = 1, \ldots, N$. Associated with each unit is a location Z_i, measuring latitude and longitude for individual i. Associated with a location z are a unique puma $P(z)$ (puma—a Census Bureau defined area with at least 100,000 individuals), a state $S(z)$, and a division $D(z)$ (also a Census Bureau defined concept, with nine divisions in the United States). In our application the sample is divided into nine divisions, which are then divided into a total of 49 states (we leave out individuals from Hawaii and Alaska, and include the District of Columbia as a separate state), which are then divided into 2057 pumas. For individual i, with location Z_i, let P_i, S_i, and D_i, denote the puma, state, and division associated with the location Z_i. The distance $d(z, z')$ between two locations z and z' is defined as the shortest distance, in miles, on the Earth's surface connecting the two points. To be precise, let $z = (z_{\text{lat}}, z_{\text{long}})$ be the latitude and longitude of a location. Then, the formula for the distance in miles between two locations z and z' we use is

$$d(z, z') = 3{,}959 \times \arccos(\cos(z_{\text{long}} - z'_{\text{long}}) \cdot \cos(z_{\text{lat}}) \cdot \cos(z'_{\text{lat}}) + \sin(z_{\text{lat}}) \cdot \sin(z'_{\text{lat}})).$$

In this article, we focus primarily on estimating the slope coefficient β in a linear regression of some outcome Y_i (e.g., the logarithm of individual-level earnings for working men) on a binary intervention or treatment W_i (e.g., a state-level regulation), of the form

$$Y_i = \alpha + \beta \cdot W_i + \varepsilon_i. \tag{2.1}$$

A key issue is that the explanatory variable W_i may be constant within clusters of individuals. In our application, W_i varies at the state level.

Let ε denote the N-vector with typical element ε_i, and let $\mathbf{Y}$, $\mathbf{W}$, $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{D}$, denote the N-vectors with typical elements Y_i, W_i, P_i, S_i, and D_i. Let ι_N denote the N-vector of ones, let $X_i = (1, W_i)$, and let $\mathbf{X}$ and $\mathbf{Z}$ denote the $N \times 2$ matrices with ith rows equal to X_i and Z_i, respectively, so that we can write in matrix notation

$$\mathbf{Y} = \iota_N \cdot \alpha + \mathbf{W} \cdot \beta + \varepsilon = \mathbf{X}(\alpha \quad \beta)' + \varepsilon. \tag{2.2}$$

Let $N_1 = \sum_{i=1}^N W_i$, $N_0 = N - N_1$, $\overline{W} = N_1/N$, and $\overline{Y} = \sum_{i=1}^N Y_i/N$. We are interested in the distribution of the ordinary least squares estimators:

$$\hat{\beta}_{\text{ols}} = \frac{\sum_{i=1}^N (Y_i - \overline{Y}) \cdot (W_i - \overline{W})}{\sum_{i=1}^N (W_i - \overline{W})^2}, \quad \text{and} \quad \hat{\alpha}_{\text{ols}} = \overline{Y} - \hat{\beta}_{\text{ols}} \cdot \overline{W}.$$

The starting point is the following model for the conditional distribution of $\mathbf{Y}$ given the location $\mathbf{Z}$ and the covariate $\mathbf{W}$:

Assumption 1. (Model)

$$\mathbf{Y} | \mathbf{W} = \mathbf{w}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(\iota_N \cdot \alpha + \mathbf{w} \cdot \beta, \Omega(\mathbf{z})).$$

Under this assumption, we can infer the exact (finite sample) distribution of the least squares estimator, conditional on the covariates $\mathbf{X}$, and the locations $\mathbf{Z}$.

Lemma 1. (Distribution of Least Squares Estimator) Suppose Assumption 1 holds. Then $\hat{\beta}_{\text{ols}}$ is unbiased and Normally distributed,

$$\mathbb{E}[\hat{\beta}_{\text{ols}} \mid \mathbf{W}, \mathbf{Z}] = \beta, \quad \text{and} \quad \hat{\beta}_{\text{ols}} | \mathbf{W}, \mathbf{Z} \sim \mathcal{N}(\beta, \mathbb{V}_M(\mathbf{W}, \mathbf{Z})), \tag{2.3}$$

where

$$\mathbb{V}_M(\mathbf{W}, \mathbf{Z}) = \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1 - \overline{W})^2} (\overline{W} \quad -1)(\iota_N \quad \mathbf{W})' \times \Omega(\mathbf{Z})(\iota_N \quad \mathbf{W}) \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}. \tag{2.4}$$

We write the model-based variance $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$ as a function of $\mathbf{W}$ and $\mathbf{Z}$ to make explicit that this variance is conditional on both the treatment indicators $\mathbf{W}$ and the locations $\mathbf{Z}$. This lemma follows directly from the standard results on least squares estimation and is given without proof. Given Assumption 1, the exact distribution for the least squares coefficients $(\hat{\alpha}_{\text{ols}}, \hat{\beta}_{\text{ols}})'$ is Normal, centered at $(\alpha, \beta)'$, and with covariance matrix $(\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\Omega(\mathbf{Z})\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1}$. We then obtain the variance for $\hat{\beta}_{\text{ols}}$ in Equation (2.4) by writing out the component matrices of the joint variance of $(\hat{\alpha}_{\text{ols}}, \hat{\beta}_{\text{ols}})'$.

It is also useful for the subsequent discussion to consider the variance of $\hat{\beta}_{\text{ols}}$, conditional on the locations $\mathbf{Z}$, and conditional on $N_1 = \sum_{i=1}^N W_i$, without conditioning on the entire vector $\mathbf{W}$. With some abuse of language, we refer to this as the unconditional variance $\mathbb{V}_U(\mathbf{Z})$ (although it is still conditional on $\mathbf{Z}$ and N_1). Because the conditional and unconditional expectation of $\hat{\beta}_{\text{ols}}$ are both equal to β, it follows that

Table 1. Summary statistics for census data ($N = 2{,}590{,}190$)

Average log earnings	10.17
Standard deviation of log earnings	0.97
Number of pumas in the sample	2,057
Average number of observations per puma	1,259
Standard deviation of number of observations per puma	409
Number of states (including DC, excluding AK, HI, PR) in the sample	49
Average number of observations per state	52,861
Standard deviation of number of observations per state	58,069
Number of divisions in the sample	9
Average number of observations per division	287,798
Standard deviation of number of observations per division	134,912

the unconditional variance is simply the expected value of the model-based variance:

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\mathbf{W}, \mathbf{Z}) \mid \mathbf{Z}] = \frac{N^2}{N_0^2 \cdot N_1^2} \cdot \mathbb{E}[(\mathbf{W} - N_1/N \cdot \iota_N)'\Omega(\mathbf{Z}) \times (\mathbf{W} - N_1/N \cdot \iota_N) \mid \mathbf{Z}]. \tag{2.5}$$

3. SPATIAL CORRELATION PATTERNS IN EARNINGS

In this section, we provide some evidence for the presence and structure of spatial correlations, that is, how Ω varies with $\mathbf{Z}$. Specifically, we show in our application, first, that the structure is more general than the state-level correlations that are typically allowed for, and, second, that this matters for inference.

We use data from the 5% public use sample from the 2000 census. Our sample consists of 2,590,190 men, at least 20 years and at most 50 years old, with positive earnings. We exclude individuals from Alaska, Hawaii, and Puerto Rico (these states share no boundaries with other states, and as a result spatial correlations may be very different than those for other states), and treat DC as a separate state, for a total of 49 "states." Table 1 presents some summary statistics for the sample. Our primary outcome variable is the logarithm of yearly earnings, in deviations from the overall mean, denoted by Y_i. The overall mean of log earnings is 10.17, and the overall standard deviation is 0.97. We do not have individual-level locations. Instead, we know for each individual only the puma of residence, and so we take Z_i to be the latitude and longitude of the center of the puma of residence.

Let $\mathbf{Y}$ be the variable of interest, in our case log earnings in deviations from the overall mean. Suppose we model the vector $\mathbf{Y}$ as

$$\mathbf{Y} \mid \mathbf{Z} \sim \mathcal{N}(0, \Omega(\mathbf{Z}, \gamma)).$$

If researchers have covariates that vary at the state level, the conventional strategy is to allow for correlation at the same level of aggregation ("clustering by state"), and model the covariance matrix as

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \sigma_\varepsilon^2 \cdot \mathbf{1}_{i=j} + \sigma_S^2 \cdot \mathbf{1}_{S_i=S_j} = \begin{cases} \sigma_S^2 + \sigma_\varepsilon^2 & \text{if } i = j \\ \sigma_S^2 & \text{if } i \neq j,\ S_i = S_j \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\Omega_{ij}(\mathbf{Z}, \gamma)$ is the (i, j)th element of $\Omega(\mathbf{Z}, \gamma)$. The first variance component, σ_ε^2, captures the variance of idiosyncratic errors, uncorrelated across different individuals. The second variance component, σ_S^2, captures correlations between individuals in the same state. Estimating σ_ε^2 and σ_S^2 on our sample of 2,590,190 individuals by maximum likelihood leads to $\hat{\sigma}_\varepsilon^2 = 0.929$ and $\hat{\sigma}_S^2 = 0.016$. The question addressed in this section is whether the covariance structure in Equation (3.1) provides an accurate approximation to the true covariance matrix $\Omega(\mathbf{Z})$. We provide two pieces of evidence that it is not.

The first piece of evidence against the simple covariance matrix structure is based on simple descriptive measures of the correlation patterns as a function of distance between individuals. For a distance d (in miles), define the overall, within-state, and out-of-state covariances as

$$C(d) = \mathbb{E}[Y_i \cdot Y_j | d(Z_i, Z_j) = d],$$
$$C_S(d) = \mathbb{E}[Y_i \cdot Y_j | S_i = S_j, d(Z_i, Z_j) = d],$$

and

$$C_{\overline{S}}(d) = \mathbb{E}[Y_i \cdot Y_j | S_i \neq S_j, d(Z_i, Z_j) = d].$$

If the model in Equation (3.1) was correct, then $C_S(d)$ should be constant (but possibly nonzero) as a function of the distance d, and $C_{\overline{S}}(d)$ should be equal to zero for all d.

We estimate these covariances using averages of the products of individual-level outcomes for pairs of individuals whose distance is within some bandwidth h of the distance d:

$$\widehat{C(d)} = \sum_{i<j} \mathbf{1}_{|d(Z_i,Z_j)-d| \leq h} \cdot Y_i \cdot Y_j \Big/ \sum_{i<j} \mathbf{1}_{|d(Z_i,Z_j)-d| \leq h},$$
$$\widehat{C_S(d)} = \sum_{i<j, S_i=S_j} \mathbf{1}_{|d(Z_i,Z_j)-d| \leq h} \cdot Y_i \cdot Y_j \Big/ \sum_{i<j, S_i=S_j} \mathbf{1}_{|d(Z_i,Z_j)-d| \leq h},$$

and

$$\widehat{C_{\overline{S}}(d)} = \sum_{i<j} \mathbf{1}_{S_i \neq S_j} \cdot \mathbf{1}_{|d(Z_i,Z_j)-d| \leq h} \cdot Y_i \cdot Y_j \Big/ \sum_{i<j, S_i \neq S_j} \mathbf{1}_{|d(Z_i,Z_j)-d| \leq h}.$$

Figure 1(a) and (b) show the covariance functions for $\widehat{C_S(d)}$ and $\widehat{C_{\overline{S}}(d)}$ for the bandwidth $h = 50$ miles for the within-state and out-of-state covariances. (Results based on a bandwidth $h = 20$ are similar.) The main conclusion from Figure 1(a) is that within-state correlations decrease with distance. Figure 1(b) suggests that correlations for individuals in different states are nonzero, also decrease with distance, and are of a magnitude similar to within-state correlations. Thus, these figures suggest that the simple covariance model in Equation (3.1) is not an accurate representation of the true covariance structure.

As a second piece of evidence, we consider various parametric structures for the covariance matrix $\Omega(\mathbf{Z})$ that generalize Equation (3.1). At the most general level, we specify the following

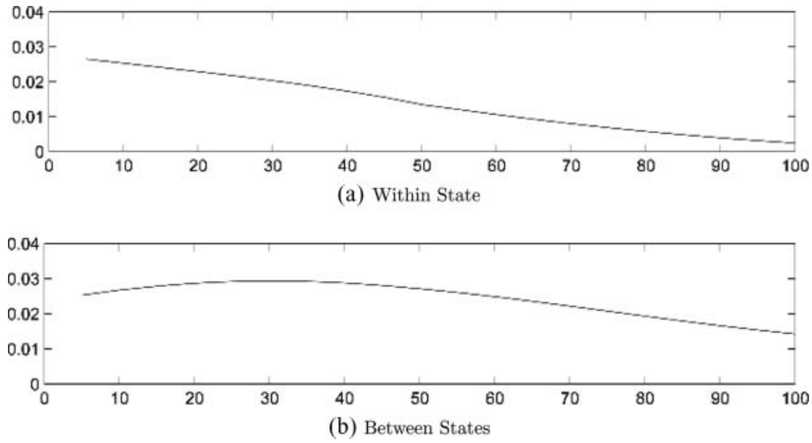


Figure 1. Covariance of demeaned log earnings of individuals as function of distance (in miles). Bandwidth $h = 50$ miles.

form for $\Omega_{ij}(\mathbf{Z}, \gamma)$:

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \begin{cases} \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) \\ \quad + \sigma^2_D + \sigma^2_S + \sigma^2_P + \sigma^2_\varepsilon & \text{if } i = j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) \\ \quad + \sigma^2_D + \sigma^2_S + \sigma^2_P & \text{if } i \neq j, P_i = P_j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) \\ \quad + \sigma^2_D + \sigma^2_S & \text{if } P_i \neq P_j, S_i = S_j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) + \sigma^2_D & \text{if } S_i \neq S_j, D_i = D_j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) & \text{if } D_i \neq D_j. \end{cases} \tag{3.2}$$

Beyond state-level correlations, the most general specification allows for correlations at the puma level (captured by σ^2_P) and at the division level (captured by σ^2_D). In addition, we allow for spatial correlation as a smooth-function geographical distance, declining at an exponential rate, captured by $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(z, z'))$. Although more general than the typical covariance structure allowed for, this model still embodies important restrictions, notably that correlations do not vary by location. A more general model might allow variances or covariances to vary directly by the location z, for example, with correlations stronger or weaker in the Western versus the Eastern United States, or in more densely or sparsely populated parts of the country.

Table 2 gives maximum likelihood estimates for the covariance parameters γ given various restrictions, based on the log earnings data, with standard errors based on the second derivatives of the log-likelihood function. To put these numbers in perspective, the estimated value for α in the most general model, $\hat{\alpha} = 0.0293$, implies that the pure spatial component, $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(z, z'))$, dies out fairly quickly: at a distance of about 25 miles, the spatial covariance due to the $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(z, z'))$ component is half what it is at zero miles. The covariance of log earnings for two individuals in the same puma is $0.080/0.948 = 0.084$. For these data, the covariance between log earnings and years of education is approximately 0.3, so the within-puma covariance is substantively important, equal to about 30% of the log earnings and education covariance. For two individuals in the same state, but in different pumas and ignoring the spatial component, the total covariance

Table 2. Estimates for clustering variances for demeaned log earnings. Standard errors based on the second derivative of log-likelihood in square brackets. Log lik refers to the value of the log-likelihood function evaluated at the maximum likelihood estimates. The last two columns refer to the implied standard errors if the regressor is an indicator for high-state minimum wage (MW) or an indicator for the state being in New England or the East-North-Central Division (NE/ENC)

							$\widehat{\text{s.e.}}(\hat{\beta})$	
σ^2_ε	σ^2_D	σ^2_S	σ^2_P	σ^2_{dis}	α	Log lik	MW	NE/ENC
0.931	0	0	0	0	0	−1213298	0.002	0.002
[0.001]								
0.929	0	0.016	0	0	0	−1200407	0.080	0.057
[0.001]		[0.002]						
0.868	0	0.011	0.066	0	0	−1116976	0.068	0.049
[0.001]		[0.003]	[0.002]					
0.929	0.006	0.011	0	0	0	−1200403	0.091	0.081
[0.001]	[0.002]	[0.002]						
0.868	0.006	0.006	0.066	0	0	−1116972	0.081	0.076
[0.001]	[0.003]	[0.002]	[0.002]					
0.868	0.005	0.006	0.047	0.021	0.029	−1116892	0.074	0.085
[0.001]	[0.005]	[0.001]	[0.002]	[0.003]	[0.005]			

is 0.013. The estimates suggest that much of what shows up as within-state correlations in a model such as Equation (3.1) that incorporates only within-state correlations, in fact captures much more local, within-puma, correlations.

To show that these results are typical for the type of correlations found in individual-level economic data, we calculated results for the same models as in Table 2 for two other variables collected in the census: years of education and hours worked. Results for those variables are reported in an earlier version of the article (Barrios et al. 2010). In all cases, puma-level correlations are an order of magnitude larger than within-state out-of-puma-level correlations, and within-division correlations are of the same order of magnitude as within-state correlations.

The two sets of results, the covariances by distance and the model-based estimates of cluster contributions to the variance, both suggest that the simple model in Equation (3.1) that assumes zero covariances for individuals in different states, and constant covariances for individuals in the same state irrespective of distance, is at odds with the data. Covariances vary substantially within states, and do not vanish at state boundaries.

Now we turn to the second question of this section, whether the magnitude of the correlations we reported matters for inference. To assess this, we look at the implications of the models for the correlation structure for the precision of least squares estimates. To make this specific, we focus on the model in Equation (2.1), with log earnings as the outcome Y_i, and W_i equal to an indicator that individual i lives in a state with a minimum wage that is higher than the federal minimum wage in the year 2000. This indicator takes on the value 1 for individuals living in nine states in our sample—California, Connecticut, Delaware, Massachusetts, Oregon, Rhode Island, Vermont, Washington, and DC—and 0 for all other states in our sample (see Figure 2(a) for a visual impression). (The data come from the website *http://www.dol.gov/whd/state/stateMinWageHis.htm*. To be consistent with the 2000 census, we use the information from 2000, not the current state of the law.) In the second to last column in Table 2, under the label "MW," we report in each row the standard error for $\hat\beta_{\text{ols}}$ based on the specification for $\Omega(\mathbf{Z}, \gamma)$ in that row. To be precise, if $\hat\Omega = \Omega(\mathbf{Z}, \hat\gamma)$ is the estimate for $\Omega(\mathbf{Z}, \gamma)$ in a particular specification, the standard error is

$$\text{s.e.}(\hat\beta_{\text{ols}}) = \left(\frac{1}{N^2\overline{W}^2(1-\overline{W})^2} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}' (\iota_N \ \mathbf{W})'\Omega(\mathbf{Z}, \hat\gamma) \times (\iota_N \ \mathbf{W}) \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix} \right)^{1/2}.$$

With no correlation between units at all, the estimated standard error is 0.002. If we allow only for state-level correlations, Equation (3.1), the estimated standard error goes up to 0.080, demonstrating the well-known importance of allowing for correlation at the level that the covariate varies. There are two general points to take away from the column with standard errors. First, the biggest impact on the standard errors comes from incorporating state-level correlations (allowing σ_S^2 to differ from zero), even though according to the variance component estimates other variance components are substantially more important. Second, among the specifications that allow for $\sigma_S^2 \neq 0$, however, there is still a substantial amount of variation in the implied standard errors. Incorporating only σ_S^2 leads to a standard error around 0.080, whereas also including division-level correlations ($\sigma_D^2 \neq 0$) increases that to approximately 0.091, an increase of 15%. We repeat this exercise for a second binary covariate, with the results reported in the last column of Table 2. In this case, the covariate takes on the value 1 only for the New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) and East-North-Central states (Illinois, Indiana, Michigan, Ohio, and Wisconsin), collectively referred to as the NE/ENC states from here on. This set of states corresponds to more geographical concentration than the set of minimum wage states (see Figure 2(b)). In this case, the impact on the standard errors of misspecifying the covariance structure $\Omega(\mathbf{Z})$ is even larger, with the most general specification leading to standard errors that are almost 50% larger than those based on the state-level correlations specification (Equation (3.1)). In the next three sections, we explore theoretical results that provide some insight into these empirical findings.

4. RANDOMIZATION INFERENCE

In this section, we consider a different approach to analyzing the distribution of the least squares estimator, based on randomization inference (e.g., Rosenbaum 2002). Recall the linear model in Equation (2.1),

$$Y_i = \alpha + \beta \cdot W_i + \varepsilon_i, \quad \text{with} \quad \varepsilon|\mathbf{W}, \mathbf{Z} \sim \mathcal{N}(0, \Omega(\mathbf{Z})).$$

In Section 2, we analyzed the properties of the least squares estimator $\hat\beta_{\text{ols}}$ under repeated sampling. To be precise, the sampling distribution for $\hat\beta_{\text{ols}}$ was defined by repeated sampling in which we keep both the vector of treatments $\mathbf{W}$ and the location $\mathbf{Z}$ fixed on all draws, and redraw only the vector of residuals ε for each sample. Under this repeated sampling thought-experiment, the exact variance of $\hat\beta_{\text{ols}}$ is $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$ as given in Lemma 1.

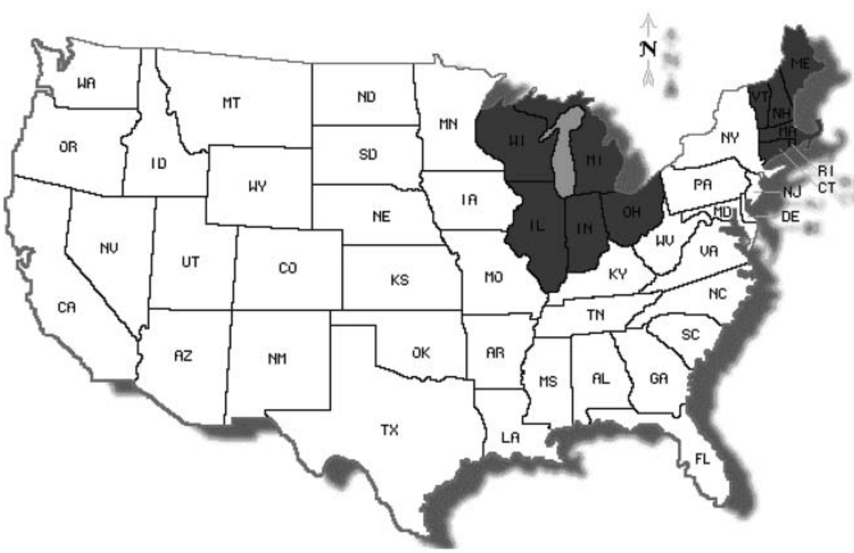
It is possible to construct confidence intervals in a different way, based on a different repeated sampling thought-experiment. Instead of conditioning on the vector $\mathbf{W}$ and $\mathbf{Z}$, and resampling the ε, we can condition on ε and $\mathbf{Z}$, and resample the vector $\mathbf{W}$. To be precise, let $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes under the two levels of the treatment W_i, and let $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ denote the corresponding N-vectors. Then, let $Y_i = Y_i(W_i)$ be the realized outcome. We assume that the effect of the treatment is constant, $Y_i(1) - Y_i(0) = \beta$. Defining $\alpha = \mathbb{E}[Y_i(0)]$, the residual is $\varepsilon_i = Y_i - \alpha - \beta \cdot W_i$. In this section, we focus on the simplest case, where the covariate of interest W_i is completely randomly assigned, conditional on $\sum_{i=1}^N W_i = N_1$.

Assumption 2. Randomization

$$\text{pr}(\mathbf{W} = \mathbf{w} \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = 1 \Big/ \binom{N}{N_1}, \quad \text{for all } \mathbf{w} \text{ s.t.} \sum_{i=1}^N w_i = N_1.$$

Under this assumption, we can infer the exact (finite sample) variance for the least squares estimator for $\hat\beta_{\text{ols}}$ conditional on $\mathbf{Z}$ and $(\mathbf{Y}(0), \mathbf{Y}(1))$:

(a) State minimum wage higher than federal minimum wage (CA, CT, DC, DE, MA, OR, RI, VT, WA).



(b) New England/East North Central states (CT, IL, IN, MA, ME, MI, OH, RI, NH, VT, WI).

Figure 2. Spatial correlation of regressors.

Lemma 2. Suppose that Assumption 2 holds and that the treatment effect $Y_i(1) - Y_i(0) = \beta$ is constant for all individuals. Then (1), $\hat{\beta}_{\mathrm{ols}}$ conditional on $(\mathbf{Y}(0), \mathbf{Y}(1))$ and $\mathbf{Z}$ is unbiased for β,

$$\mathbb{E}[\hat{\beta}_{\mathrm{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}] = \beta, \tag{4.1}$$

and, (2), its exact conditional (randomization-based) variance is

$$\begin{aligned}\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) &= \mathbb{V}(\hat{\beta}_{\mathrm{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \\ &= \frac{N}{N_0 \cdot N_1 \cdot (N-2)} \sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2, \end{aligned} \tag{4.2}$$

where $\overline{\varepsilon} = \sum_{i=1}^{N} \varepsilon_i / N$.

Because this result direct follows from results by Neyman (1990) on randomization inference for average treatment effects, specialized to the case with a constant treatment effect, the proof is omitted. Note that although the variance is exact, we do not have exact Normality, unlike the result in Lemma 1.

In the remainder of this section, we explore two implications of the randomization perspective. First of all, although the model and randomization variances $\mathbb{V}_M$ and $\mathbb{V}_R$ are exact if both Assumptions 1 and 2 hold, they differ because they refer to different repeated sampling thought-experiments, or, alternatively, to different conditioning sets. To illustrate this, let us consider the bias and variance under a third repeated sampling thought-experiment, without conditioning on either $\mathbf{W}$ or ε, just conditioning on the locations $\mathbf{Z}$ and (N_0, N_1), maintaining both the model and the randomization assumption.

Lemma 3. Suppose Assumptions 1 and 2 hold. Then (1), $\hat{\beta}_{\text{ols}}$ is unbiased for β,

$$\mathbb{E}[\hat{\beta}_{\text{ols}}|\mathbf{Z}, N_0, N_1] = \beta, \tag{4.3}$$

(2), its exact unconditional variance is

$$\mathbb{V}_U(\mathbf{Z}) = \left(\frac{1}{N-2}\text{trace}(\Omega(\mathbf{Z})) - \frac{1}{N\cdot(N-2)}\iota_N'\Omega(\mathbf{Z})\iota_N\right)\cdot\frac{N}{N_0\cdot N_1}, \tag{4.4}$$

and (3),

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \mathbb{E}\left[\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})|\mathbf{Z}, N_0, N_1\right] \\ &= \mathbb{E}\left[\mathbb{V}_M(\mathbf{W}, \mathbf{Z})|\mathbf{Z}, N_0, N_1\right].\end{aligned}$$

Thus, in expectation, $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$, is equal to the expectation of $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$.

For the second point, suppose we had focused on the repeated sampling variance for $\hat{\beta}_{\text{ols}}$ conditional on $\mathbf{W}$ and $\mathbf{Z}$, but possibly erroneously modeled the covariance matrix as constant times the identity matrix, $\Omega(\mathbf{Z}) = \sigma^2 \cdot I_N$. Using such a (possibly incorrect) model, a researcher would have concluded that the exact sampling distribution for $\hat{\beta}_{\text{ols}}$ conditional on the covariates would be

$$\hat{\beta}_{\text{ols}}|\mathbf{W}, \mathbf{Z} \sim \mathcal{N}(\beta, \mathbb{V}_{\text{INC}}), \quad \text{where} \quad \mathbb{V}_{\text{INC}} = \sigma^2 \cdot \frac{N}{N_0 \cdot N_1}. \tag{4.5}$$

If $\Omega(\mathbf{Z})$ differs from $\sigma^2 \cdot I_N$, then $\mathbb{V}_{\text{INC}}$ is not in general the correct (conditional) distribution for $\hat{\beta}_{\text{ols}}$. However, in some cases the misspecification need not lead to invalid inferences in large samples. To make that precise, we first need to define precisely how inference is performed. Implicitly, the maximum likelihood estimator for the misspecified variance defines σ^2 as the probability limit of the estimator:

$$\begin{aligned}\hat{\sigma}^2 &= \arg\max\left\{\frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^N(Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}}W_i)^2\right\} \\ &= \frac{1}{N}\sum_{i=1}^N(Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}}W_i)^2.\end{aligned}$$

The probability limit for this estimator $\hat{\sigma}^2$, under Assumptions given in the Lemma below, is $\text{plim}(\text{trace}(\Omega(Z))/N)$. Then, the probability limit of the normalized variance based on the possibly incorrect model is

$$N \cdot \mathbb{V}_{\text{INC}} = \text{plim}(\text{trace}\ (\Omega(Z))/N)\ \text{plim}\left(\frac{N^2}{N_0\cdot N_1}\right).$$

The following result clarifies the properties of this probability limit.

Lemma 4. Suppose Assumption 1 holds with $\Omega(\mathbf{Z})$ satisfying $\text{trace}(\Omega(\mathbf{Z}))/N \to c$ for some $0 < c < \infty$, and $\iota_N'\Omega(\mathbf{Z})\iota_N/N^2 \to 0$, and Assumption 2 holds with $N_1/N \to p$ for some $0 < p < 1$. Then

$$N \cdot (\mathbb{V}_{\text{INC}} - \mathbb{V}_U(\mathbf{Z})) \xrightarrow{p} 0, \quad \text{and} \quad N \cdot \mathbb{V}_{\text{INC}} \xrightarrow{p} \frac{c}{p\cdot(1-p)}.$$

Hence, and this is a key insight of this section, if the assignment $\mathbf{W}$ is completely random, and the treatment effect is constant, one can, at least in large samples, ignore the off-diagonal elements of $\Omega(\mathbf{Z})$, and (mis-)specify $\Omega(\mathbf{Z})$ as $\sigma^2 \cdot I_N$. Although the resulting variance estimator will *not* be estimating the variance under the repeated sampling thought-experiment that one may have in mind (namely $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$), it leads to valid confidence intervals under the randomization distribution. The result that the misspecification of the covariance matrix need not lead to inconsistent standard errors if the covariate of interest is randomly assigned has been noted previously. Greenwald (1983, p. 328) wrote: "when the correlation patterns of the independent variables are unrelated to those across the errors, then the least squares variance estimates are consistent." Angrist and Pischke (2009, p. 311) wrote, in the context of clustering, that: "if the [covariate] values are uncorrelated within the groups, the grouped error structure does not matter for standard errors." The preceding discussion interprets this result formally from a randomization perspective.

5. RANDOMIZATION INFERENCE WITH CLUSTER-LEVEL RANDOMIZATION

Now let us return to the setting that is the main focus of the article. The covariate of interest, W_i, varies only between clusters (states), and is constant within clusters. Instead of assuming that W_i is randomly assigned at the individual level, we now assume that it is randomly assigned at the cluster level. Let M be the number of clusters, M_1 the number of clusters with all individuals assigned to $W_i = 1$, and M_0 the number of clusters with all individuals assigned to $W_i = 0$. The cluster indicator is

$$C_{im} = \mathbf{1}_{S_i=m} = \begin{cases} 1 & \text{if individual } i \text{ is in cluster/state } m, \\ 0 & \text{otherwise,}\end{cases}$$

with $\mathbf{C}$ the $N \times M$ matrix with typical element C_{im}. For randomization inference, we condition on $\mathbf{Z}$, ε, and M_1. Let N_m be the number of individuals in cluster m. We now look at the properties of $\hat{\beta}_{\text{ols}}$ over the randomization distribution induced by this assignment mechanism. To keep the notation precise, let $\tilde{\mathbf{W}}$ be the M-vector of assignments at the cluster level, with typical element $\tilde{W}_m$. Let $\tilde{\mathbf{Y}}(0)$ and $\tilde{\mathbf{Y}}(1)$ be M-vectors, with mth element equal to $\tilde{Y}_m(0) = \sum_{i:C_{im}=1} Y_i(0)/N_m$, and $\tilde{Y}_m(1) = \sum_{i:C_{im}=1} Y_i(1)/N_m$, respectively. Similarly, let $\tilde{\varepsilon}$ be an M-vector with mth element equal to $\tilde{\varepsilon}_m = \sum_{i:C_{im}=1} \varepsilon_i/N_m$, and let $\bar{\tilde{\varepsilon}} = \sum_{m=1}^M \tilde{\varepsilon}_m/M$.

Formally, the assumption on the assignment mechanism is now:

Assumption 3. (Cluster Randomization)

$$\text{pr}(\tilde{\mathbf{W}} = \tilde{\mathbf{w}} \mid \mathbf{Z} = \mathbf{z}) = 1\bigg/\binom{M}{M_1},$$
$$\text{for all } \mathbf{w} \text{ s.t.} \sum_{m=1}^M \tilde{w}_m = M_1, \text{ and 0 otherwise.}$$

We also make the assumption that all clusters are the same size:

Assumption 4. (Equal Cluster Size) $N_m = N/M$ for all $m = 1, \ldots, M$.

Lemma 5. Suppose Assumptions 3 and 4 hold, and the treatment effect $Y_i(1) - Y_i(0) = \beta$ is constant. Then (1), the exact sampling variance of β_{ols}, conditional on $\mathbf{Z}$ and ε, under the cluster randomization distribution is

$$\mathbb{V}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \frac{M}{M_0 \cdot M_1 \cdot (M-2)} \sum_{m=1}^{M} (\tilde{\varepsilon}_m - \overline{\tilde{\varepsilon}})^2, \tag{5.1}$$

and (2) if also Assumption 1 holds, then the unconditional variance is

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}\left[\mathbb{V}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) | \mathbf{Z}, M_1\right] = \frac{M^2}{M_0 \cdot M_1 \cdot (M-2) \cdot N^2} \cdot \left(M \cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \iota'\Omega(\mathbf{Z})\iota\right). \tag{5.2}$$

The unconditional variance is a special case of the expected value of the unconditional variance in Equation (2.5), with the expectation taken over $\mathbf{W}$ given the cluster-level randomization.

6. VARIANCE ESTIMATION UNDER MISSPECIFICATION

In this section, we present the main theoretical result in the article. It extends the result in Section 4 on the robustness of model-based variance estimators under complete randomization to the case where the model-based variance estimator accounts for clustering, but not necessarily for all spatial correlations, and that treatment is randomized at cluster level.

Suppose the model generating the data is the linear model in Equation (2.1), with a general covariance matrix $\Omega(\mathbf{Z})$, and Assumption 1 holds. The researcher estimates a parametric model that imposes a potentially incorrect structure on the covariance matrix. Let $\Omega(\mathbf{Z}, \gamma)$ be the parametric model for the error covariance matrix. The model is misspecified in the sense that there need not be a value γ such that $\Omega(\mathbf{Z}) = \Omega(\mathbf{Z}, \gamma)$. The researcher then proceeds to calculate the variance of $\hat{\beta}_{\text{ols}}$ as if the postulated model is correct. The question is whether this implied variance based on a misspecified covariance structure leads to correct inference.

The example we are most interested in is characterized by a clustering structure by state. In that case, $\Omega(\mathbf{Z}, \gamma)$ is the $N \times N$ matrix with $\gamma = (\sigma_\varepsilon^2, \sigma_S^2)'$, where

$$\Omega_{ij}\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right) = \begin{cases} \sigma_\varepsilon^2 + \sigma_S^2 & \text{if } i = j \\ \sigma_S^2 & \text{if } i \neq j, S_i = S_j, \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

Initially, however, we allow for any parametric structure $\Omega(\mathbf{Z}, \gamma)$. The true covariance matrix $\Omega(\mathbf{Z})$ may include correlations that extend beyond state boundaries, and that may involve division-level correlations or spatial correlations that decline smoothly with distance as in the specification in Equation (3.2).

Under the (misspecified) parametric model $\Omega(\mathbf{Z}, \gamma)$, let $\tilde{\gamma}$ be the pseudo-true value, defined as the value of γ that maximizes the expectation of the logarithm of the likelihood function,

$$\tilde{\gamma} = \arg\max_{\gamma} \mathbb{E}\left[-\frac{1}{2} \cdot \ln\left(\det\left(\Omega(\mathbf{Z}, \gamma)\right)\right) - \frac{1}{2} \cdot \varepsilon'\Omega(\mathbf{Z}, \gamma)^{-1}\varepsilon \,\middle|\, \mathbf{Z}\right].$$

Given the pseudo-true error covariance matrix $\Omega(\tilde{\gamma})$, the corresponding pseudo-true model-based variance of the least squares estimator, conditional on $\mathbf{W}$ and $\mathbf{Z}$, is

$$\mathbb{V}_{\text{INC,CR}} = \frac{1}{N^2 \overline{W}^2 (1-\overline{W})^2} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}' (\iota_N \quad \mathbf{W})' \Omega(\mathbf{Z}, \tilde{\gamma}) \times (\iota_N \quad \mathbf{W}) \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}.$$

Because for some $\mathbf{Z}$, the true covariance matrix $\Omega(\mathbf{Z})$ differs from the misspecified one, $\Omega(\mathbf{Z}, \tilde{\gamma})$, it follows that in general this pseudo-true conditional variance $\mathbb{V}_M(\Omega(\mathbf{Z}, \tilde{\gamma}), \mathbf{W}, \mathbf{Z})$ will differ from the true variance $\mathbb{V}_M(\Omega(\mathbf{Z}), \mathbf{W}, \mathbf{Z})$. Here, we focus on the expected value of $\mathbb{V}_M(\Omega(\mathbf{Z}, \tilde{\gamma}), \mathbf{W}, \mathbf{Z})$, conditional on $\mathbf{Z}$, under assumptions on the distribution of $\mathbf{W}$. Let us denote this expectation by $\mathbb{V}_U(\Omega(\mathbf{Z}, \tilde{\gamma}), \mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\Omega(\mathbf{Z}, \tilde{\gamma}), \mathbf{W}, \mathbf{Z})|\mathbf{Z}]$. The question is under what conditions on the specification of the error covariance matrix $\Omega(\mathbf{Z}, \gamma)$, in combination with assumptions on the assignment process, this unconditional variance is equal to the expected variance with the expectation of the variance under the correct error covariance matrix, $\mathbb{V}_U(\Omega(\mathbf{Z}), \mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\Omega(\mathbf{Z}), \mathbf{W}, \mathbf{Z})|\mathbf{Z}]$.

The following theorem shows that if the randomization of $\mathbf{W}$ is at the cluster level, then solely accounting for cluster-level correlations is sufficient to get valid confidence intervals.

Theorem 1. (Clustering with Misspecified Error Covariance Matrix) Suppose Assumption 1 holds with $\Omega(\mathbf{Z})$ satisfying $\text{trace}(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C})/N \to c$ for some $0 < c < \infty$, and $\iota_N'\Omega(\mathbf{Z})\iota_N/N^2 \to 0$, Assumption 3 holds with $M_1/M \to p$ for some $0 < p < 1$, and Assumption 4 holds. Suppose also that $\Omega(\mathbf{Z}, \gamma)$ is specified as in Equation (6.1). Then

$$N \cdot (\mathbb{V}_{\text{INC,CR}} - \mathbb{V}_U(\mathbf{Z})) \xrightarrow{p} 0, \quad \text{and}$$
$$N \cdot \mathbb{V}_{\text{INC,CR}} \xrightarrow{p} \frac{c}{N_m^2 \cdot p \cdot (1-p)}.$$

This is the main theoretical result in the article. It implies that if cluster-level explanatory variables are randomly allocated to clusters, there is no need to consider covariance structures beyond those that allow for cluster-level correlations. In our application, if the covariate (state minimum wage exceeding federal minimum wage) were as good as randomly allocated to states, then there is no need to incorporate division- or puma-level correlations in the specification of the covariance matrix. It is, in that case, sufficient to allow for correlations between outcomes for individuals in the same state. Formally, the result is limited to the case with equal sized clusters. There are few exact results for the case with variation in cluster size, although if the variation is modest, one might expect the current results to provide useful guidance.

In many econometric analyses, researchers specify the conditional distribution of the outcome given some explanatory variables, and ignore the joint distribution of the explanatory variables. The result in Theorem 1 shows that it may be useful to pay attention to this distribution. Depending on the joint distribution of the explanatory variables, the analyses may be robust to misspecification of particular aspects of the conditional distribution. In the next section, we discuss some methods for assessing the relevance of this result.

7. SPATIAL CORRELATION IN STATE AVERAGES

The results in the previous sections imply that inference is substantially simpler if the explanatory variable of interest is randomly assigned, either at the unit or cluster level. Here, we discuss tests originally introduced by Mantel (1967) (see, e.g., Schabenberger and Gotway 2004) to analyze whether random assignment is consistent with the data, against the alternative hypothesis of some spatial correlation. These tests allow for the calculation of exact, finite sample, p-values. To implement these tests, we use the location of the units. To make the discussion more specific, we test the random assignment of state-level variables against the alternative of spatial correlation.

Let Y_s be the variable of interest for state s, for $s = 1, \ldots, S$, where state s has location Z_s (the centroid of the state). In the illustrations of the tests, we use an indicator for a state-level regulation, and the state-average of an individual-level outcome. The null hypothesis of no spatial correlation in the Y_s can be formalized as stating that conditional on the locations $\mathbf{Z}$, each permutation of the values $(Y_1, \ldots, Y_S)$ is equally likely. With S states, there are $S!$ permutations. We assess the null hypothesis by comparing, for a given statistic $M(\mathbf{Y}, \mathbf{Z})$, the value of the statistic given the actual $\mathbf{Y}$ and $\mathbf{Z}$, with the distribution of the statistic generated by randomly permuting the $\mathbf{Y}$ vector.

The tests we focus on in the current article are based on Mantel statistics (e.g., Mantel 1967; Schabenberger and Gotway 2004). These general form of the statistics we use is Geary's c (also known as a Black-White, or BW, statistic in the case of binary outcomes), a proximity-weighted average of squared pairwise differences:

$$G(\mathbf{Y}, \mathbf{Z}) = \sum_{s=1}^{S-1} \sum_{t=s+1}^{S} (Y_s - Y_t)^2 \cdot d_{st}, \tag{7.1}$$

where $d_{st} = d(Z_s, Z_t)$ is a nonnegative weight monotonically related to the proximity of the states s and t. Given a statistic, we test the null hypothesis of no spatial correlation by comparing the value of the statistic in the actual dataset, G^{obs}, to the distribution of the statistic under random permutations of the Y_s. The latter distribution is defined as follows. Taking the S units, with values for the variable $Y_1, \ldots, Y_S$, we randomly permute the values $Y_1, \ldots, Y_S$ over the S units. For each of the $S!$ permutations g, we re-calculate the Mantel statistic, say G_g. This defines a discrete distribution with $S!$ different values, one for each allocation. The one-sided exact p-value is defined as the fraction of allocations g (out of the set of $S!$ allocations) such that the associated Mantel statistic G_g is less than or equal to the observed Mantel statistic G^{obs}:

$$p = \frac{1}{S!} \sum_{g=1}^{S!} \mathbf{1}_{G^{\text{obs}} \geq G_g}. \tag{7.2}$$

A low value of the p-value suggests rejecting the null hypothesis of no spatial correlation in the variable of interest. In practice, the number of allocations is often too large to calculate the exact p-value and so we approximate the p-value by drawing a large number of allocations, and calculating the proportion of statistics less than or equal to the observed Mantel statistic. In the calculations below, we use 10,000,000 draws from the randomization distribution.

We use six different measures of proximity. First, we define the proximity d_{st} as states s and t sharing a border:

$$d_{st}^{B} = \begin{cases} 1 & \text{if } s, t \text{ share a border,} \\ 0 & \text{otherwise.} \end{cases} \tag{7.3}$$

Second, we define d_{st} as an indicator for states s and t belonging to the same census division of states (recall that the United States is divided into nine divisions):

$$d_{st}^{D} = \begin{cases} 1 & \text{if } D_s = D_t, \\ 0 & \text{otherwise.} \end{cases} \tag{7.4}$$

The last four proximity measures are functions of the geographical distance between states s and t:

$$d_{st}^{GD} = -d\left(Z_s, Z_t\right), \quad \text{and} \quad d_{st}^{\alpha} = \exp\left(-\alpha \cdot d\left(Z_s, Z_t\right)\right), \tag{7.5}$$

where $d(z, z')$ is the distance in miles between two locations z and z', and Z_s is the latitude and longitude of state s, measured as the latitude and longitude of the centroid for each state. We use $\alpha = 0.00138$, $\alpha = 0.00276$, and $\alpha = 0.00693$. For these values, the proximity index declines by 50% at distances of 500, 250, and 100 miles.

We calculate the p-values for the Mantel test statistic based on three variables. First, an indicator for having a state minimum wage higher than the federal minimum wage. This indicator takes on the value 1 in nine out of the 49 states in our sample, with these nine states mainly concentrated in the North East and the West Coast. Second, we calculate the p-values for the average of the logarithm of yearly earnings. Third, we calculate the p-values for the indicator for NE/ENC states. The results for the three variables and six statistics are presented in Table 3. All three variables exhibit considerable spatial correlation. Interestingly, the results are fairly sensitive to the measure of proximity. From these limited calculations, it appears that sharing a border is a measure of proximity that is sensitive to the type of spatial correlations in the data.

8. A SMALL SIMULATION STUDY

We carried out a small simulation study to investigate the relevance of the theoretical results from Section 6. In all cases, the model was

$$Y_i = \alpha + \beta \cdot W_i + \varepsilon_i,$$

Table 3. p-values for Geary's c, one-sided alternatives (10,000,000 draws)

				Distance weights, 50% decline at:		
Proximity	Border	Division	Distance	500 miles	250 miles	100 miles
Minimum wage	<0.0001	0.0028	0.9960	0.0093	0.0365	0.4307
Log wage	0.0005	0.0239	0.0692	0.0276	0.0298	0.1644
NE/ENC	<0.0001	<0.0001	0.0967	0.0877	0.0692	0.0321

with $N = 2{,}590{,}190$ observations to mimic our actual data. In our simulations, every state has the same number of individuals, and every puma within a given state has the same number of individuals. We considered three distributions for W_i. In all cases, W_i varies only at the state level. In the first case, $W_i = 1$ for individuals in nine randomly chosen states. In the second case, $W_i = 1$ for the the nine minimum wage states. In the third case, $W_i = 1$ for the 11 NE/ENC states. The distribution for ε is in all cases Normal with mean zero and covariance matrix Ω. The general specification we consider for Ω is

$$\Omega_{ij}(\mathbf{Z},\gamma) = \begin{cases} \sigma_D^2 + \sigma_S^2 + \sigma_P^2 + \sigma_\varepsilon^2 & \text{if } i = j, \\ \sigma_D^2 + \sigma_S^2 + \sigma_P^2 & \text{if } i \neq j,\, P_i = P_j, \\ \sigma_D^2 + \sigma_S^2 & \text{if } P_i \neq P_j,\, S_i = S_j, \\ \sigma_D^2 & \text{if } S_i \neq S_j,\, D_i = D_j. \end{cases}$$

We look at two different sets of values for $(\sigma_\varepsilon^2, \sigma_P^2, \sigma_S^2, \sigma_D^2)$, (0.929, 0, 0.016, 0) (only state-level correlations, corresponding to the second pair of rows in Table 2), and (0.868, 0.005, 0.005, 0.066) (puma-, state-, and division-level correlations, corresponding to the fifth pair of rows in Table 2).

Given the data, we consider five methods for estimating the variance of the least squares estimator $\hat{\beta}_{\text{ols}}$, and thus for constructing confidence intervals. The first is based on the randomization distribution:

$$\hat{\mathbb{V}}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \frac{M}{M_0 \cdot M_1 \cdot (M-2)} \sum_{m=1}^{M} \hat{\bar{\varepsilon}}_m^2,$$

where $\hat{\bar{\varepsilon}}_m$ is the average value of the residual $\hat{\varepsilon}_i = Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}} \cdot W_i$ over cluster m. The second, third, and fourth variances are model-based:

$$\hat{\mathbb{V}}_M(\hat{\Omega}(\mathbf{Z}), \mathbf{W}, \mathbf{Z}) = \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1-\overline{W})^2} (\overline{W} \quad -1)(\iota_N \quad \mathbf{W})' \times \hat{\Omega}(\mathbf{Z})(\iota_N \quad \mathbf{W}) \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix},$$

using different estimates for $\hat{\Omega}(\mathbf{Z})$. First, we use an infeasible estimator, namely the true value for $\Omega(\mathbf{Z})$. Second, we specify

$$\Omega_{ij}(\mathbf{Z},\gamma) = \begin{cases} \sigma_S^2 + \sigma_\varepsilon^2 & \text{if } i = j, \\ \sigma_S^2 & \text{if } i \neq j,\, S_i = S_j. \end{cases}$$

We estimate σ_P^2 and σ_S^2 using moment-based estimators, and plug that into the expression for the covariance matrix. For the third variance estimator, in this set of three variance estimators, we specify

$$\Omega_{ij}(\mathbf{Z},\gamma) = \begin{cases} \sigma_D^2 + \sigma_S^2 + \sigma_P^2 + \sigma_\varepsilon^2 & \text{if } i = j, \\ \sigma_D^2 + \sigma_S^2 + \sigma_P^2 & \text{if } i \neq j,\, P_i = P_j, \\ \sigma_D^2 + \sigma_S^2 & \text{if } P_i \neq P_j,\, S_i = S_j, \\ \sigma_D^2 & \text{if } S_i \neq S_j,\, D_i = D_j, \end{cases}$$

and again use moment-based estimators.

The fifth and last variance estimator allows for more general variance structures within states, but restricts the correlations between individuals in different states to zero. This estimator assumes Ω is block diagonal, with the blocks defined by states, but does not impose constant correlations within the blocks. The estimator for Ω takes the form

$$\hat{\Omega}_{\text{STATA},ij}(\mathbf{Z}) = \begin{cases} \hat{\varepsilon}_i^2 & \text{if } i = j, \\ \hat{\varepsilon}_i \cdot \hat{\varepsilon}_j & \text{if } i \neq j,\, S_i = S_j, \\ 0 & \text{otherwise,} \end{cases}$$

leading to

$$\hat{\mathbb{V}}_{\text{STATA}} = \frac{1}{N^2 \cdot \overline{W}^2 (1-\overline{W})^2} \cdot (\overline{W} \quad -1)(\iota_N \quad \mathbf{W})' \Omega_{\text{STATA}}(\mathbf{Z}) \times (\iota_N \quad \mathbf{W}) \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}.$$

This is the variance estimator implemented in STATA and widely used in empirical works.

In Table 4, we report the actual level of tests of the null hypothesis that $\beta = \beta_0$ with a nominal level of 5%. First, consider the two columns with random assignment of states to the treatment. In that case, all variance estimators lead to tests that perform well, with actual levels between 5.0% and 7.6%. Excluding the STATA variance estimator, the actual levels are below 6.5%. The key finding is that even if the correlation pattern involves pumas as well as divisions, variance estimators that ignore the division-level correlations do very well.

When we do use the minimum wage states as the treatment group, the assignment is no longer completely random. If the correlations are within state, all variance estimators still perform well. However, if there are correlations at the division level, now only the variance estimator using the true variance matrix does well. The estimator that estimates the division-level correlations does best among the feasible estimators, but, because the data are not informative enough about these correlations to precisely estimate the variance components, even this estimator exhibits substantial size distortions. The same pattern, but even stronger, emerges with the NE/ENC states as the treatment group.

Table 4. Size of t-tests (in %) using different variance estimators for models with only state-level correlations (S), and models with puma-, state-, and division-level correlations (PSD; 500,000 draws)

Treatment type	Random		Min. Wage		NE/ENC	
Shock type	S	PSD	S	PSD	S	PSD
$\hat{\mathbb{V}}_{\text{CR}}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$	5.6	5.6	5.6	16.2	5.6	26.3
$\hat{\mathbb{V}}_M(\Omega(\mathbf{Z}), \mathbf{W}, \mathbf{Z})$	5.0	5.0	5.0	5.0	5.0	5.0
$\hat{\mathbb{V}}_M(\Omega(\hat{\sigma}_\epsilon^2, \hat{\sigma}_S^2), \mathbf{W}, \mathbf{Z})$	6.1	6.1	6.1	17.1	6.1	27.2
$\hat{\mathbb{V}}_M(\Omega(\hat{\sigma}_\epsilon^2, \hat{\sigma}_P^2, \hat{\sigma}_S^2, \hat{\sigma}_D^2), \mathbf{W}, \mathbf{Z})$	6.1	6.5	5.7	9.0	5.4	13.8
STATA	7.6	7.6	8.5	18.5	7.7	30.4

9. CONCLUSION

In empirical studies with individual-level outcomes and state-level explanatory variables, researchers often calculate standard errors allowing for within-state correlations between individual-level outcomes. In many cases, however, the correlations may extend beyond state boundaries. Here, we explore the presence of such correlations, and investigate the implications of their presence for the calculation of standard errors. In theoretical calculations we show that under some conditions, in particular random assignment of regulations, correlations in outcomes between individuals in different states can be ignored. However, state-level variables often exhibit considerable spatial correlation, and ignoring out-of-state correlations of the magnitude found in our application may lead to substantial underestimation of standard errors.

In practice, we recommend that researchers explicitly explore the spatial correlation structure of both the outcomes and the explanatory variables. Statistical tests based on Mantel statistics, with the proximity based on shared borders, or belonging to a common division, are straightforward to calculate and lead to exact p-values. If these test suggest that both outcomes and explanatory variables exhibit substantial spatial correlation, we recommend that one should explicitly account for the spatial correlation by allowing for a more flexible specification than one that only accounts for state-level clustering.

APPENDIX: PROOFS

For the general model, leaving aside terms that do not involve unknown parameters, the log-likelihood function is

$$L(\gamma|\mathbf{Y}) = -\frac{1}{2}\ln(\det(\Omega(\mathbf{Z},\gamma))) - \varepsilon'\Omega^{-1}(\mathbf{Z},\gamma)\varepsilon/2.$$

The matrix $\Omega(\mathbf{Z},\gamma)$ is large in our illustrations, with dimension 2,590,190 by 2,590,190. Direct maximization of the log-likelihood function is therefore not feasible. However, because locations are measured by puma locations, $\Omega(\mathbf{Z},\gamma)$ has a block structure, and calculations of the log-likelihood simplify and can be written in terms of first and second moments by puma. We first give a couple of preliminary results.

Theorem A.1. (Sylvester's Determinant Theorem). Let A and B be arbitrary $M \times N$ matrices. Then:

$$\det(I_N + A'B) = \det(I_M + BA').$$

Proof of Theorem A.1. Consider a block matrix $\left(\begin{smallmatrix} M_1 & M_2 \\ M_3 & M_4 \end{smallmatrix}\right)$. Then:

$$\det\begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} = \det\begin{pmatrix} M_1 & 0 \\ M_3 & I \end{pmatrix}\det\begin{pmatrix} I & M_1^{-1}M_2 \\ 0 & M_4 - M_3M_1^{-1}M_2 \end{pmatrix}$$
$$= \det M_1 \det\left(M_4 - M_3M_1^{-1}M_2\right).$$

Similarly,

$$\det\begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} = \det\begin{pmatrix} I & M_2 \\ 0 & M_4 \end{pmatrix}\det\begin{pmatrix} M_1 - M_2M_4^{-1}M_3 & 0 \\ M_4^{-1}M_3 & I \end{pmatrix}$$
$$= \det M_4 \det\left(M_1 - M_2M_4^{-1}M_3\right).$$

Letting $M_1 = I_M$, $M_2 = -B$, $M_3 = A'$, $M_4 = I_N$ yields the result. □

Lemma A.1. (Determinant of Cluster Covariance Matrix). Suppose $\mathbf{C}$ is an $N \times M$ matrix of binary cluster indicators, with $\mathbf{C}'\mathbf{C}$ equal to a $M \times M$ diagonal matrix, Σ is an arbitrary $M \times M$ matrix, and I_N is the N-dimensional identity matrix. Then, for scalar σ_ε^2, and

$$\Omega = \sigma_\epsilon^2 I_N + \mathbf{C}\Sigma\mathbf{C}' \qquad \Omega_C = \Sigma + \sigma_\epsilon^2(\mathbf{C}'\mathbf{C})^{-1},$$

we have

$$\det(\Omega) = \left(\sigma_\epsilon^2\right)^{N-M}\det(\mathbf{C}'\mathbf{C})\det(\Omega_C).$$

Proof of Lemma A.1. By Sylvester's theorem:

$$\det(\Omega) = \left(\sigma_\epsilon^2\right)^N \det\left(I_N + \mathbf{C}\Sigma/\sigma_\epsilon^2\mathbf{C}'\right)$$
$$= \left(\sigma_\epsilon^2\right)^N \det\left(I_M + \mathbf{C}'\mathbf{C}\Sigma/\sigma_\epsilon^2\right)$$
$$= \left(\sigma_\epsilon^2\right)^N \det\left(I_M + \mathbf{C}'\mathbf{C}\Omega_C/\sigma_\epsilon^2 - I_M\right)$$
$$= \left(\sigma_\epsilon^2\right)^N \det(\mathbf{C}'\mathbf{C})\det\left(\Omega_C/\sigma_\epsilon^2\right)$$
$$= \left(\sigma_\epsilon^2\right)^{N-M}\left(\prod N_p\right)\det(\Omega_C).$$

□

Lemma A.2. Suppose Assumptions 3 and 4 hold. Then for any $N \times N$ matrix Ω,

$$\mathbb{E}[\mathbf{W}'\Omega\mathbf{W}] = \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N'\Omega\iota_N + \frac{M_1\cdot M_0}{M\cdot(M-1)}\cdot\operatorname{trace}\left(\mathbf{C}'\Omega\mathbf{C}\right).$$

Proof of Lemma A.2. We have

$$\mathbb{E}[W_i\cdot W_j] = \begin{cases} M_1/M & \text{if } \forall m, C_{im} = C_{jm}, \\ (M_1\cdot(M_1-1))/(M\cdot(M-1)) & \text{otherwise.} \end{cases}$$

it follows that

$$\mathbb{E}[\mathbf{W}\mathbf{W}'] = \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N\iota_N' + \left(\frac{M_1}{M} - \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\right)\cdot\mathbf{C}\mathbf{C}'$$
$$= \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N\iota_N' + \frac{M_1\cdot M_0}{M\cdot(M-1)}\cdot\mathbf{C}\mathbf{C}'.$$

Thus,

$$\mathbb{E}[\mathbf{W}'\Omega\mathbf{W}]$$
$$= \operatorname{trace}\left(\mathbb{E}[\Omega\mathbf{W}\mathbf{W}']\right)$$
$$= \operatorname{trace}\left(\Omega\cdot\left(\frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N\iota_N' + \frac{M_1\cdot M_0}{M\cdot(M-1)}\cdot\mathbf{C}\mathbf{C}'\right)\right)$$
$$= \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N'\Omega\iota_N + \frac{M_1\cdot M_0}{M\cdot(M-1)}\cdot\operatorname{trace}\left(\mathbf{C}'\Omega\mathbf{C}\right).$$

□

Lemma A.3. Suppose the $N \times N$ matrix Ω satisfies

$$\Omega = \sigma_\varepsilon^2\cdot I_N + \sigma_C^2\cdot\mathbf{C}\mathbf{C}',$$

where I_N is the $N \times N$ identity matrix, and $\mathbf{C}$ is an $N \times M$ matrix of zeros and ones, with $\mathbf{C}\iota_M = \iota_N$ and $\mathbf{C}'\iota_N = (N/M)\iota_M$, so that,

$$\Omega_{ij} = \begin{cases} \sigma_\varepsilon^2 + \sigma_C^2 & \text{if } i = j \\ \sigma_C^2 & \text{if } i \neq j, \forall m, C_{im} = C_{jm}, \\ 0 & \text{otherwise,} \end{cases} \tag{A.1}$$

Then, (1)

$$\ln(\det(\Omega)) = N\cdot\ln\left(\sigma_\varepsilon^2\right) + M\cdot\ln\left(1 + \frac{N}{M}\cdot\frac{\sigma_C^2}{\sigma_\varepsilon^2}\right),$$

and, (2)

$$\Omega^{-1} = \sigma_\varepsilon^{-2}\cdot I_N - \frac{\sigma_C^2}{\sigma_\varepsilon^2\cdot(\sigma_\varepsilon^2 + \sigma_C^2\cdot N/M)}\cdot\mathbf{C}\mathbf{C}'$$

Proof of Lemma A.3. First, consider the first part. Apply Lemma A.1 with

$$\Sigma = \sigma_C^2\cdot I_M, \quad \text{and} \quad \mathbf{C}'\mathbf{C} = \frac{N}{M}\cdot I_M,$$
$$\text{so that} \quad \Omega_C = \left(\sigma_C^2 + \sigma_\varepsilon^2\cdot\frac{M}{N}\right)\cdot I_M.$$

Then, by Lemma A.1, we have

$$\ln\det(\Omega) = (N-M)\cdot\ln\left(\sigma_\varepsilon^2\right) + M\cdot\ln(N/M) + \ln\det(\Omega_C)$$
$$= (N-M)\cdot\ln\left(\sigma_\varepsilon^2\right) + M\cdot\ln(N/M) + M\cdot\ln\left(\sigma_C^2 + \sigma_\varepsilon^2\cdot\frac{M}{N}\right)$$
$$= (N-M)\cdot\ln\left(\sigma_\varepsilon^2\right) + M\cdot\ln\left(\frac{N}{M}\sigma_C^2 + \sigma_\varepsilon^2\right)$$
$$= N\cdot\ln\left(\sigma_\varepsilon^2\right) + M\cdot\ln\left(1 + \frac{N}{M}\cdot\frac{\sigma_C^2}{\sigma_\varepsilon^2}\right).$$

Next, consider part (2). We need to show that

$$\left(\sigma_\varepsilon^2 \cdot I_N + \sigma_C^2 \cdot \mathbf{CC}'\right)\left(\sigma_\varepsilon^{-2} \cdot I_N - \frac{\sigma_C^2}{\sigma_\varepsilon^2 \cdot (\sigma_\varepsilon^2 + \sigma_C^2 \cdot N/M)} \cdot \mathbf{CC}'\right) = I_N,$$

which amounts to showing that

$$\begin{aligned}&- \frac{\sigma_\varepsilon^2 \cdot \sigma_C^2}{\sigma_\varepsilon^2 \cdot \left(\sigma_\varepsilon^2 + \sigma_C^2 \cdot N/M\right)} \cdot \mathbf{CC}' + \sigma_C^2 \cdot \mathbf{CC}'\sigma_\varepsilon^{-2} \\ &- \mathbf{CC}' \cdot \frac{\sigma_C^4}{\sigma_\varepsilon^2 \cdot \left(\sigma_\varepsilon^2 + \sigma_C^2 \cdot N/M\right)} \cdot \mathbf{CC}' = 0.\end{aligned}$$

This follows directly from the fact that $\mathbf{C}'\mathbf{C} = (N/M) \cdot I_M$ and collecting the terms. □

Proof of Lemma 3. The unbiasedness result directly follows from the conditional unbiasedness established in Lemma 2. Next, we establish the second part of the Lemma. By the Law of Iterated Expectations,

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \mathbb{V}(\mathbb{E}[\hat\beta_{\text{ols}} \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}] \mid \mathbf{Z}, N_1) \\ &\quad + \mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \mid \mathbf{Z}, N_1] \\ &= \mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \mid \mathbf{Z}, N_1], \end{aligned} \tag{A.2}$$

where the second line follows since $\hat\beta_{\text{ols}}$ is unbiased. By Lemma 2. we have

$$\begin{aligned}&\mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \mid \mathbf{Z}, N_1] \\ &\quad = \mathbb{E}\left[\frac{N}{N_0 \cdot N_1 \cdot (N-2)} \sum_{i=1}^N (\varepsilon_i - \overline{\varepsilon})^2 \ \mid \mathbf{Z}, N_1\right].\end{aligned}$$

Observe that we can write

$$\begin{aligned}\sum_{i=1}^N (\varepsilon_i - \overline{\varepsilon})^2 &= (\varepsilon - \iota_N \iota_N' \varepsilon/N)'(\varepsilon - \iota_N \iota_N' \varepsilon/N) \\ &= \varepsilon'\varepsilon - 2\varepsilon'\iota_N\iota_N'\varepsilon/N + \varepsilon'\iota_N\iota_N\iota_N\iota_N'\varepsilon/N^2 \\ &= \varepsilon'\varepsilon - \varepsilon'\iota_N\iota_N'\varepsilon/N.\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \frac{N}{N_0 \cdot N_1 \cdot (N-2)} \mathbb{E}[\varepsilon'\varepsilon - \varepsilon'\iota_N\iota_N'\varepsilon/N | \mathbf{Z}, N_0, N_1] \\ &= \frac{N}{N_0 \cdot N_1 \cdot (N-2)} \text{trace}(\mathbb{E}[\varepsilon\varepsilon' - \iota_N'\varepsilon\varepsilon'\iota_N/N | \mathbf{Z}, N_0, N_1]) \\ &= \frac{N}{N_0 \cdot N_1 \cdot (N-2)} (\text{trace}(\Omega(\mathbf{Z})) - \iota_N'\Omega(\mathbf{Z})\iota_N/N)\end{aligned}$$

which establishes Equation (4.4). Finally, we prove the third part of the Lemma. By Lemma 1, $\hat\beta_{\text{ols}}$ is unbiased conditional on $\mathbf{Z}, \mathbf{W}$, so that by argument like in Equation (A.2) above, we can also write:

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \mathbb{V}(\mathbb{E}[\hat\beta_{\text{ols}} \mid \mathbf{Z}, \mathbf{W}] \mid \mathbf{Z}, N_1) + \mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} | \mathbf{Z}, \mathbf{W}) \mid \mathbf{Z}, N_1] \\ &= \mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \mid \mathbf{Z}, N_1]\end{aligned}$$

which equals $\mathbb{E}\left[\mathbb{V}_R\left(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right) \ \mid \mathbf{Z}, N_1\right]$ by (A.2). □

Proof of Lemma 4. We will first show that the second claim in the Lemma holds,

$$N \cdot \mathbb{V}_{\text{INC}} \xrightarrow{p} \frac{c}{p \cdot (1-p)}, \tag{A.3}$$

and then show that

$$N \cdot \mathbb{V}_U \xrightarrow{p} \frac{c}{p \cdot (1-p)}, \tag{A.4}$$

which together prove the first claim in the Lemma.

Consider Equation (A.3). By the conditions in the Lemma, $\hat\alpha_{\text{ols}}$ and $\hat\beta_{\text{ols}}$ are consistent for α and β, and therefore the probability limit of $\hat\sigma^2$ is the probability limit of $\sum_{i=1}^N \sum_{i=1}^N \varepsilon_i^2/N$, which is the probability limit of $\text{trace}(\Omega(\mathbf{Z})/N)$. Then

$$\begin{aligned}\text{plim}\,(N \cdot \mathbb{V}_{\text{INC}}) &= \text{plim}\left(\frac{1}{N}\sum_{i=1}^N \varepsilon_i^2 \cdot \frac{N^2}{N_0 \cdot N - 1}\right) \\ &= \text{plim}\left(\frac{\text{trace}(\Omega(\mathbf{Z}))}{N} \cdot \frac{N^2}{N_0 \cdot N_1}\right) = \frac{c}{p \cdot (1-p)}.\end{aligned}$$

Now consider Equation (A.4). By the conditions in the Lemma,

$$\begin{aligned}N \cdot \mathbb{V}_U &= \left(\frac{1}{N-2}\text{trace}(\Omega(\mathbf{Z})) - \frac{1}{N \cdot (N-2)}\iota_N'\Omega(\mathbf{Z})\iota_N\right) \cdot \frac{N^2}{N_0 \cdot N_1} \\ &= \frac{N}{N-2} \cdot \frac{1}{N}\text{trace}(\Omega(\mathbf{Z})) \cdot \frac{N^2}{N_0 \cdot N_1} - \frac{N}{N-2} \\ &\quad \cdot \frac{1}{N^2}\iota_N'\Omega(\mathbf{Z})\iota_N \cdot \frac{N^2}{N_0 \cdot N_1} \xrightarrow{p} \frac{c}{p \cdot (1-p)}.\end{aligned}$$

□

Proof of Lemma 5. To show the first part of the Lemma, observe that under constant cluster size,

$$\hat\beta_{\text{ols}} = \frac{\sum_{m=1}^M (\tilde Y_m - \bar{\tilde Y})^2 (\tilde W_m - \bar{\tilde W})}{\sum_m (\tilde W_m - \bar{\tilde W})^2}$$

where $\tilde Y_m = (N/M)^{-1}\sum_{i:C_{im}=1} Y_i$, and $\bar{\tilde Y} = M^{-1}\sum_m \tilde Y_m = \bar Y$, and $\bar{\tilde W} = \bar W$. Therefore, we can apply Lemma 2. treating cluster averages $(\tilde Y_m, \tilde W_m, \tilde\varepsilon_m)$ as a unit of observation, which yields the result.

To show the second part, again by Lemma 2. $\hat\beta_{\text{ols}}$ is unbiased, so that by the Law of Iterated Expectations, and the first part of the Lemma,

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \mathbb{V}(\mathbb{E}[\hat\beta_{\text{ols}} \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}] \mid \mathbf{Z}, M_1) \\ &\quad + \mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \mid \mathbf{Z}, M_1] \\ &= \mathbb{E}[\mathbb{V}(\hat\beta_{\text{ols}} \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) \mid \mathbf{Z}, M_1] \\ &= \mathbb{E}\left[\frac{M}{(M-2) \cdot M_0 \cdot M_1} \sum_{m=1}^M \left(\tilde\varepsilon_m - \bar{\tilde\varepsilon}\right)^2 \ \mid \mathbf{Z}, M_1\right]\end{aligned}$$

Hence, it suffices to show that

$$\mathbb{E}\left[\sum_{s=1}^M \left(\tilde\varepsilon_s - \bar{\tilde\varepsilon}\right)^2 \middle| \mathbf{Z}, M_1\right] = \left(\frac{M^2}{N^2} \cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \frac{M}{N^2}\iota'\Omega(\mathbf{Z})\iota\right).$$

Note that in general $\mathbf{C}\iota_M = \iota_N$, and under Assumption 4, it follows that $\mathbf{C}'\mathbf{C} = (N/M) \cdot I_M$. We can write

$$\tilde\varepsilon_m = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\varepsilon = \frac{M}{N}\mathbf{C}'\varepsilon, \quad \text{and} \quad \bar{\tilde\varepsilon} = \frac{1}{M}\iota_M'(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\varepsilon = \frac{1}{N}\iota_N'\varepsilon,$$

so that

$$\begin{aligned}\sum_{m=1}^M (\tilde\varepsilon_m - \bar{\tilde\varepsilon})^2 &= \left(\frac{M}{N}\mathbf{C}'\varepsilon - \frac{1}{M}\iota_M\iota_N'\varepsilon\right)'\left(\frac{M}{N}\mathbf{C}'\varepsilon - \frac{1}{M}\iota_M\iota_N'\varepsilon\right) \\ &= \left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right)'\left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right). \\ &= \varepsilon'\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{E}\left[\sum_{m=1}^M \left(\tilde\varepsilon_s - \bar{\tilde\varepsilon}\right)^2 \middle| \mathbf{Z}, M_1\right] \\ &\quad = \mathbb{E}\left[\varepsilon'\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon \middle| \mathbf{Z}, M_1\right] \\ &\quad = \text{trace}\left(\mathbb{E}\left[\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\varepsilon' \middle| \mathbf{Z}, M_1\right]\right)\end{aligned}$$

$$= \text{trace}\left(\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\Omega(\mathbf{Z})\right)$$
$$= \text{trace}\left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\Omega(\mathbf{Z})\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\right)$$
$$= \frac{M^2}{N^2}\cdot\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \frac{M}{N^2}\cdot\iota_N'\Omega(\mathbf{Z})\iota_N.$$

□

Proof of Theorem 1. We show

$$N\cdot\mathbb{V}_U \xrightarrow{p} \frac{c}{N_m^2\cdot p\cdot(1-p)},$$

and

$$N\cdot\mathbb{V}_{\text{INC,CR}} \xrightarrow{p} \frac{c}{N_m^2\cdot p\cdot(1-p)},$$

which together imply the two claims in the Theorem. First consider the first claim. The normalized variance is

$$N\cdot\mathbb{V}_U(\mathbf{Z}) = \frac{M^2\cdot N}{M_0\cdot M_1\cdot(M-2)\cdot N^2}\cdot(M\cdot\text{trace}(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}) - \iota'\Omega(\mathbf{Z})\iota)$$
$$= \frac{M^2\cdot N}{M_0\cdot M_1\cdot(M-2)}\cdot\left(\frac{M}{N}\cdot\frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{N} - \frac{\iota'\Omega(\mathbf{Z})\iota}{N^2}\right).$$

By the conditions in the Theorem the probability limit of this expression is

$$\text{plim}\left(\frac{M^2\cdot N}{M_0\cdot M_1\cdot(M-2)}\cdot\left(\frac{M}{N}\cdot\frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{N} - \frac{\iota'\Omega(\mathbf{Z})\iota}{N^2}\right)\right)$$
$$= \text{plim}\left(\frac{M^2\cdot N}{M_0\cdot M_1\cdot(M-2)}\right)\cdot\left(\text{plim}\left(\frac{M}{N}\cdot\frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{N}\right)\right.$$
$$\left.-\ \text{plim}\left(\frac{\iota'\Omega(\mathbf{Z})\iota}{N^2}\right)\right)$$
$$= \frac{c}{N_m^2\cdot p\cdot(1-p)}.$$

Next, consider the second claim. Now the probability limit of the model-based variance is

$$\text{plim}(N\cdot\mathbb{V}_{\text{INC,CR}}(\mathbf{Z}))$$
$$= \text{plim}\left(\frac{M^2\cdot N}{M_0\cdot M_1\cdot(M-2)\cdot N^2}\cdot\left(M\cdot\text{trace}\left(\mathbf{C}'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\mathbf{C}\right)\right.\right.$$
$$\left.\left.-\iota'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\iota\right)\right)$$
$$= \text{plim}\left(\frac{M^2\cdot N}{M_0\cdot M_1\cdot(M-2)}\cdot\left(\frac{M}{N}\cdot\frac{\text{trace}\left(\mathbf{C}'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\mathbf{C}\right)}{N}\right.\right.$$
$$\left.\left.-\frac{\iota'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\iota}{N^2}\right)\right)$$
$$= \frac{1}{N_m\cdot p\cdot(1-p)}\cdot\left(\text{plim}\left(\frac{M}{N}\cdot\frac{\text{trace}\left(\mathbf{C}'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\mathbf{C}\right)}{N}\right)\right.$$
$$\left.-\,\text{plim}\left(\frac{\iota'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\iota}{N^2}\right)\right)$$
$$= \frac{1}{N_m^2\cdot p\cdot(1-p)}\cdot\text{plim}\left(\frac{\text{trace}\left(\mathbf{C}'\Omega\left(\mathbf{Z},\tilde\sigma_\varepsilon,\tilde\sigma_S^2\right)\mathbf{C}\right)}{N}\right)$$

Hence, to prove the second claim, it suffices to show that $\text{trace}(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}) = \text{trace}(\mathbf{C}'\Omega(\mathbf{Z},(\tilde\sigma_\varepsilon,\tilde\sigma_S^2))\mathbf{C})$. The log-likelihood function based on the specification in Equation (A.1) is

$$L\left(\sigma_\varepsilon^2,\sigma_S^2\,\middle|\,\mathbf{Y},\mathbf{Z}\right) = -\frac{1}{2}\cdot\ln\left(\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_S^2\right)\right) - \frac{1}{2}\cdot\mathbf{Y}'\Omega\left(\sigma_\varepsilon^2,\sigma_S^2\right)^{-1}\mathbf{Y}.$$

The expected value of the log-likelihood function is

$$\mathbb{E}\left[L\left(\sigma_\varepsilon^2,\sigma_S^2\middle|\mathbf{Y},\mathbf{Z}\right)\,\middle|\,\mathbf{Z}\right]$$
$$= -\frac{1}{2}\ln\left(\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_S^2\right)\right) - \frac{1}{2}\cdot\mathbb{E}\left[\mathbf{Y}'\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_C^2\right)^{-1}\mathbf{Y}\right]$$
$$= -\frac{1}{2}\cdot\ln\left(\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_S^2\right)\right) - \frac{1}{2}\cdot\text{trace}\left(\mathbb{E}\left[\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_S^2\right)^{-1}\mathbf{Y}\mathbf{Y}'\right)\right]$$
$$= -\frac{1}{2}\cdot\ln\left(\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_S^2\right)\right) - \frac{1}{2}\cdot\text{trace}\left(\Omega\left(\mathbf{Z},\sigma_\varepsilon^2,\sigma_S^2\right)^{-1}\Omega(\mathbf{Z})\right).$$

Using Lemma A.3, this is equal to

$$\mathbb{E}\left[L\left(\sigma_\varepsilon^2,\sigma_S^2\middle|\mathbf{Y},\mathbf{Z}\right)\,\middle|\,\mathbf{Z}\right] = -\frac{N}{2}\cdot\ln\left(\sigma_\varepsilon^2\right) - \frac{M}{2}$$
$$\cdot\ln\left(1+N/M\cdot\sigma_S^2/\sigma_\varepsilon^2\right) - \frac{1}{2\cdot\sigma_\varepsilon^2}\cdot\text{trace}(\Omega(\mathbf{Z}))$$
$$+\frac{\sigma_S^2}{2\cdot\sigma_\varepsilon^2\cdot\left(\sigma_\varepsilon^2+\sigma_S^2\cdot N/M\right)}\cdot\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right).$$

The first derivative of the expected log-likelihood function with respect to σ_S^2 is

$$\frac{\partial}{\partial\sigma_S^2}\mathbb{E}\left[L\left(\sigma_\varepsilon^2,\sigma_S^2\,\middle|\,\mathbf{Y},\mathbf{Z}\right)\,\middle|\,\mathbf{Z}\right] = -\frac{N}{2\cdot\left(\sigma_\varepsilon^2+N/M\cdot\sigma_S^2\right)}$$
$$+\frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{(\sigma_\varepsilon^2+\sigma_S^2\cdot(N/M))^2}$$

Hence, the first-order condition for $\tilde\sigma_S^2$ implies that

$$\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) = N\cdot\left(\tilde\sigma_\varepsilon^2+\tilde\sigma_S^2\cdot(N/M)\right).$$

For the misspecified error covariance matrix $\Omega(\mathbf{Z},\tilde\gamma)$, we have

$$\text{trace}(\mathbf{C}'\Omega(\mathbf{Z},\tilde\gamma)\mathbf{C}) = \sum_{m=1}^{M}\left(N_m^2\cdot\tilde\sigma_S^2+N_m\cdot\tilde\sigma_\varepsilon^2\right).$$

By equality of the cluster sizes, this simplifies to

$$\text{trace}(\mathbf{C}'\Omega(\mathbf{Z},\tilde\gamma)\mathbf{C}) = N\cdot\left(\tilde\sigma_\varepsilon^2+\tilde\sigma_S^2\cdot(N/M)\right) = \text{trace}(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}).$$

□

[Received September 2010. Revised December 2011.]

REFERENCES

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105(490), 493–505. [578]

Angrist, J. D., and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press. [578,584]

Barrios, T., Imbens, G. W., Diamond, R., and Kolesár, M. (2010), *Clustering, Spatial Correlations and Randomization Inference*, Working Paper No. 15760, Cambridge, MA: NBER. [582]

Bertrand, M., Duflo, E., and Mullainathan, S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1), 249–275. [578]

Bester, C. A., Conley, T. G., and Hansen, C. B. (2011), "Inference With Dependent Data Using Cluster Covariance Estimators," *Journal of Econometrics*, 165(2), 137–151. [578]

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, Oxford: Oxford University Press. [578]

Fisher, R. A. (1925), *The Design of Experiments* (1st ed.), London: Oliver & Boyd. [578]

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010), *Handbook of Spatial Statistics*, London: Chapman & Hall. [578]

Greenwald, B. C. (1983), "A General Analysis of Bias in the Estimated Standard Errors of Least Squares Coefficients," *Journal of Econometrics*, 22(3), 323–338. [584]

Hansen, C. B. (2007), "Generalized Least Squares Inference in Panel and Multilevel Models With Serial Correlation and Fixed Effects," *Journal of Econometrics*, 140(2), 670–694. [578]

Ibragimov, R., and Müller, U. (2010), "T-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business & Economic Statistics*, 28(4), 453–468. [578]

Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis for Generalized Linear Models," *Biometrika*, 73(1), 13–22. [578]

Mantel, N. (1967), "The Detection of Disease Clustering and a Generalized Regression Approach," *Cancer Research*, 27(2), 209–220. [586]

Moulton, B. R. (1986), "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32(3), 385–397. [578]

Neyman, J. (1990), "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles—Section 9," *Statistical Science*, 5(4), 465–480. [578,583]

Rosenbaum, P. R. (2002), *Observational Studies*, New York: Springer-Verlag. [578,582]

Schabenberger, O., and Gotway, C. A. (2004), *Statistical Methods for Spatial Data Analysis* (1st ed.), London: Chapman & Hall. [578,586]

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press. [578]