NBER WORKING PAPER SERIES

CLUSTERING, SPATIAL CORRELATIONS AND RANDOMIZATION INFERENCE

Thomas Barrios
Rebecca Diamond
Guido W. Imbens
Michal Kolesar

Working Paper 15760
http://www.nber.org/papers/w15760

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2010

Financial support for this research was generously provided through NSF grant 0820361. We are grateful for comments by participants in the econometrics lunch seminar at Harvard University, and in particular for discussions with Gary Chamberlain. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Clustering, Spatial Correlations and Randomization Inference
Thomas Barrios, Rebecca Diamond, Guido W. Imbens, and Michal Kolesar
NBER Working Paper No. 15760
February 2010
JEL No. C01,C1,C31

ABSTRACT

It is standard practice in empirical work to allow for clustering in the error covariance matrix if the explanatory variables of interest vary at a more aggregate level than the units of observation. Often, however, the structure of the error covariance matrix is more complex, with correlations varying in magnitude within clusters, and not vanishing between clusters. Here we explore the implications of such correlations for the actual and estimated precision of least squares estimators. We show that with equal sized clusters, if the covariate of interest is randomly assigned at the cluster level, only accounting for non-zero covariances at the cluster level, and ignoring correlations between clusters, leads to valid standard errors and confidence intervals. However, in many cases this may not suffice. For example, state policies exhibit substantial spatial correlations. As a result, ignoring spatial correlations in outcomes beyond that accounted for by the clustering at the state level, may well bias standard errors. We illustrate our findings using the 5% public use census data. Based on these results we recommend researchers assess the extent of spatial correlations in explanatory variables beyond state level clustering, and if such correlations are present, take into account spatial correlations beyond the clustering correlations typically accounted for.

Thomas Barrios
Department of Economics
Harvard University
tbarrios@fas.harvard.edu

Rebecca Diamond
Department of Economics
Harvard University
rdiamond@fas.harvard.edu

Guido W. Imbens
Department of Economics
Littauer Center
Harvard University
1805 Cambridge Street
Cambridge, MA 02138
and NBER
imbens@fas.harvard.edu

Michal Kolesar
Department of Economics
Harvard University
kolesarm@nber.org

1 Introduction

Many economic studies analyze the effects of interventions on economic behavior, using outcome data that are measured on units at a more disaggregate level than that of the intervention of interest. For example, outcomes may be measured at the individual level, whereas interventions or treatments vary only at the state level, or outcomes may be measured at the firm level, whereas regulations apply at the industry level. Often the program effects are estimated using least squares regression. Since the work by Moulton (Moulton, 1986, 1990; Moulton and Randolph, 1989), empirical researchers in economics are generally aware of the potential implications of within-cluster correlations in the outcomes on the precision of such estimates, and often incorporate within-cluster correlations in the specification of the error covariance matrix. However, there may well be more complex correlations patterns in the data. Correlations in outcomes between individuals may extend beyond state boundaries, or correlations between firm outcomes may extend beyond firms within the same narrowly defined industry groups.

In this paper we investigate the implications, for the sampling variation in least squares estimators, of the presence of correlation structures beyond those which are constant within clusters, and which vanish between clusters. Using for illustrative purposes census data with states as the clusters and individuals as the units,[1] we allow for spatial correlation patterns that vary in magnitude within states, as well as allow for positive correlations between individual level outcomes for individuals in different, but nearby, states. The first question addressed in this paper is whether such correlations are present to a substantially meaningful degree. We address this question by estimating spatial correlations for various individual level variables, including log earnings, years of education and hours worked. We find that, indeed, such correlations are present, with substantial correlations within divisions of states, and correlations within puma's (public use microdata area) considerably larger than within-state correlations. Second, we investigate whether accounting for correlations of such magnitude is important for the properties of confidence intervals for the effects of state-level regulations. Note that whereas clustering corrections in settings where the covariates vary only at the cluster level always increase standard errors, general spatial correlations can improve precision. In fact, in settings where smooth spatial correlations in outcomes are strong, regression discontinuity designs can exploit the presence of covariates which vary only at the cluster level. See Black (1999) for an application of regression discontinuity designs in such a setting, and Imbens and Lemieux (2008) and Lee and Lemieux (2009) for recent surveys. In our discussion of the second question we report both theoretical results, as well as demonstrate their relevance using illustrations based on earnings data and state regulations. We show that if regulations are as good as randomly assigned to clusters, implying there is no spatial correlation in the regulations beyond the clusters, some 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 actual spatial correlation patterns of the regulations. We

[1]We are grateful to Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander for making the PUMS data avaiable. See Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2008) for details.

provide some illustrations that show that, given the spatial correlation patterns we find in the individual level variables, spatial correlations in regulations may or may not have a substantial impact on the precision of estimates of treatment effects.

The paper draws on three literatures that have largely evolved separately. First, the literature on clustering and difference-in-differences estimation, where one focus is on adjustments to standard errors to take into account clustering of explanatory variables. See, e.g., Moulton (1986, 1990), Bertrand, Duflo, and Mullainathan (2004), Donald and Lang (2007), Hansen (2009), and the textbook discussions in Angrist and Pischke (2009) and Wooldridge (2002). Second, the literature on spatial statistics. Here a major focus is on the specification and estimation of the covariance structure of spatially linked data. For discussions in the econometric literature see Conley (1999), and for textbook discussions see Schabenberger and Gotway (2004), Arbia (2006), Cressie (1993), and Anselin, Florax, and Rey (2004). In interesting recent work Bester, Conley and Hansen (2009) and Ibragimov and Müller (2009) link some of the inferential issues in the spatial and clustering literatures. Finally, we use results from the literature on randomization inference, most directly on the calculation of exact p-values going back to Fisher (1925), and results on the exact bias and variance of estimators under randomization (Neyman, 1923, reprinted, 1990). For a recent general discussion of randomization inference see Rosenbaum (2002). Although the calculation of 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, 2009; Small, Ten Have, and Rosenbaum, 2008), 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 of confidence intervals in the context of spatially linked data.

The paper is organized as follows. In Section 2 we introduce the basic setting. Next, in Section 3, we establish, using census data on earnings, education and hours worked, the presence of spatial correlation patterns beyond the within-state correlations typically allowed for. Specifically, we show that for these variables the constant-within-state component of the total variance that is often the only one explicitly taken into account, is only a small part of the overall covariance structure. Nevertheless, we find that in some cases standard confidence intervals for regression parameters based on incorporating only within-state correlations are quite similar to those based on more accurate approximations to the actual covariance matrix. In Section 4 we introduce an alternative approach to constructing confidence intervals, based on randomization inference and originating in the work by Neyman (1923), that sheds light on these results. Initially we focus on the case with randomization at the unit level. Section 5 extends this to the case with cluster-level randomization. In Section 6, we present some theoretical results on the implications of correlation structures. We show that if cluster-level covariates are as good as randomly assigned to the clusters, the standard variance estimator based on within-cluster correlations is robust to misspecification of the error-covariance matrix, as long as the clusters are of equal size. Next, in Section 7 we investigate the presence of spatial correlations of some state regulations using Mantel-type tests from the spatial statistics literature. We find that a number of regulations exhibit substantial regional correlations. As a practical recommendation we suggest that researchers carry out such tests to investigate whether ignoring between state correlations in the specification of covariance structures may be misleading. If substantial spatial correlations between covariates are present, the error correlation structure must be modelled carefully. Section 9 concludes.

2 Set Up

Consider a setting where we have information on N units, say individuals, indexed by $i = 1, \ldots, N$. Associated with each unit is a location Z_i. For most of the paper we focus on exact repeated sampling results, keeping the number of units as well as their locations fixed. We think of Z_i as the pair of variables measuring latitude and longitude for individual i. Associated with the locations z is a clustering structure. In general we can think of z taking on values in a set $\mathcal{Z}$, and the clustering forming a partition of the set $\mathcal{Z}$. In our application to data on individuals, the clustering structure comes from the partition of the United States into areas, at various levels of disaggregation. Specifically, in that case we consider, in increasing order of aggregation, pumas, states, and divisions.[2] Thus, associated with a location z is a unique puma $P(z)$. Associated with a puma p is a unique state $S(p)$, and, finally, associated with a state s is a unique division $D(s)$. 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.[3]

In this paper we focus primarily on estimating the slope coefficient β in a linear regression of of some outcome Y_i (e.g., individual level earnings) on an intervention W_i (e.g., a state level regulation), of the form

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

The explanatory variable W_i may vary only at a more aggregate level than the outcome Y_i. For ease of exposition, and for some of the theoretical results, we focus on the case with W_i binary, and varying at the state level, and we abstract from the presence of additional covariates.

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

$$\mathbf{Y} = \iota \cdot \alpha + \mathbf{W} \cdot \beta + \varepsilon = \mathbf{X}\theta + \varepsilon. \tag{2.2}$$

Let $N_1 = \sum_{i=1}^{N} W_i$, $N_0 = N - N_1$, and $\overline{W} = N_1/N$.

[2] There are 2,057 pumas, 49 states, and 9 divisions in our sample. See http://www.census.gov/geo/www/us_regdiv.pdf for the definitions of the divisions.

[3] 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

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

We are interested in the distribution of the ordinary least squares estimator for $\hat{\beta}$, equal to:

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

where $\overline{Y} = \sum_{i=1}^{N} Y_i/N$. For completeness let $\hat{\alpha}_{\text{ols}}$ be the least squares estimator for α, $\hat{\alpha}_{\text{ols}} = \overline{Y} - \hat{\beta}_{\text{ols}} \cdot \overline{W}$.

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)

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

Alternatively, we can write this assumption as

$$\varepsilon \,\Big|\, \mathbf{W} = \mathbf{w}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(0, \Omega(\mathbf{z})).$$

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, (i), $\hat{\beta}_{\text{ols}}$ is unbiased,*

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}} \Big| \mathbf{W}, \mathbf{Z}\right] = \beta, \tag{2.3}$$

and, (ii), its exact distribution is,

$$\hat{\beta}_{\text{ols}} \big| \mathbf{W}, \mathbf{Z} \sim \mathcal{N}\left(\beta, \mathbb{V}_M(\mathbf{W}, \mathbf{Z})\right), \tag{2.4}$$

where

$$\begin{aligned}\mathbb{V}_M(\mathbf{W}, \mathbf{Z}) &= \mathbb{V}_M(\hat{\beta}_{\text{ols}} | \mathbf{W}, \mathbf{Z}) \\ &= \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1 - \overline{W})^2} \begin{pmatrix} \overline{W} & -1 \end{pmatrix} \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \Omega(\mathbf{Z}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}.\end{aligned}$$

We write the 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 $\mathbf{Z}$. We refer to this variance $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$ as the model-based variance for the least squares estimator.

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 full vector of least squares coefficients $(\hat{\alpha}_{\text{ols}}, \hat{\beta}_{\text{ols}})'$ is

$$\begin{pmatrix} \hat{\alpha}_{\text{ols}} \\ \hat{\beta}_{\text{ols}} \end{pmatrix} \Bigg| \mathbf{X}, \mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\mathbf{X}'\Omega(\mathbf{Z})\mathbf{X}\right) \left(\mathbf{X}'\mathbf{X}\right)^{-1} \right). \tag{2.5}$$

We can then obtain (2.4) by writing out the component matrices of the joint variance of $(\hat{\alpha}_{\rm ols}, \hat{\beta}_{\rm ols})'$. The uncertainty in the estimates $\hat{\beta}_{\rm ols}$ arises from the randomness of the residuals ε_i postulated in Assumption 1. One interpretation of this randomness, perhaps the most natural one, is to think of a large (infinite) population of individuals, so that we can resample from the set of individuals with exactly the same location Z_i and the same values for W_i. Alternatively we can view the randomness arising from variation in the measures of the individual level outcomes, *e.g.*, measurement error.

Finally, it will be useful to explicitly consider the variance of $\hat{\beta}_{\rm ols}$, conditional on the locations $\mathbf{Z}$, and conditional on $\sum_{i=1}^{N} W_i = N_1$, but not 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_0 and N_1). Because the conditional and unconditional expectation of $\hat{\beta}_{\rm ols}$ are identical (and equal to β), it follows that the marginal (over $\mathbf{W}$) variance is simply the expected value of the conditional variance. Thus:

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\mathbf{W},\mathbf{Z})|\mathbf{Z}] = \frac{N^2}{N_0^2 \cdot N_1^2} \cdot \mathbb{E}\left[(\mathbf{W} - N_1/N)'\Omega(\mathbf{W} - N_1/N) \middle| \mathbf{Z}\right]. \tag{2.6}$$

There is typically no reason to use the unconditional variance. In fact, there are two reasons what it would be inappropriate to use $\mathbb{V}_U(\mathbf{Z})$ instead of $\mathbb{V}_M(\mathbf{W},\mathbf{Z})$ to construct confidence intervals. First, one would lose the exact nature of the distribution, because it is no longer the case that the exact distribution of $\hat{\beta}_{\rm ols}$ is centered around β with variance equal to $\mathbb{V}_U(\mathbf{Z})$. Second, $\mathbf{W}$ is ancillary, and there are general arguments that inference should be conditional on ancillary statistics (e.g., Cox and Hinkley, 1974).

3 Spatial Correlation Patterns in Individual Level Variables

In this section we show some evidence for the presence and structure of spatial correlations in various individuals level variables, that is, how Ω varies with $\mathbf{Z}$. We use data from the 5% public use sample from the 2000 census. Our sample consists of 2,590,190 men at least 20, 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 treats DC as a separate state, for a total of 49 "states". Table 1 and 2 present 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, the overall standard deviation is 0.97. We do not have individual level locations. Instead we know for each individual only puma (public use microdata area) of residence, and so we take Z_i to be the latitude and longitude of the center of the puma of residence.

First we look at simple descriptive masures of the correlation patterns separately within-state, and out-of-state, as a function of distance. For a distance d (in miles), define the overall, within-state, and out-of-state covariances as

$$C(d) = \mathbb{E}\left[Y_i \cdot Y_j \middle| d(Z_i, Z_j) = d\right],$$

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

and

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

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

$$\widehat{C(d)} = \frac{1}{N_{d,h}} \cdot \sum_{i<j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h} \cdot Y_i \cdot Y_j,$$

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

and

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

where

$$N_{d,h} = \sum_{i<j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h}, \qquad \text{and} \quad N_{S,d,h} = \sum_{i<j} \mathbf{1}_{S_i=S_j} \cdot \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h},$$

and

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

Figures 1abc and 2abc show the covariance functions for two choices of the bandwidth, $h = 20$ and $h = 50$ miles, for the overall, within-state, and out-of-state covariances. If the standard assumption of zero correlations between individuals in different states were true, one would have expected the curve in out-of-state figures to be close to zero with no upward or downward trend. The main conclusion from these figures is that out-of-state correlations are substantial, and of a magnitude similar to the within-state correlations. Moreover, these correlations appear to decrease with distance, as one would expect if there were out-of-state spatial correlations.

Next, we consider various parametric structures that can be imposed on the covariance matrix $\Omega(\mathbf{Z})$. Let $\mathbf{Y}$ be the variable of interest, log earnings in deviations from the overall mean. Here we model the vector $\mathbf{Y}$ as

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

At the most general level, we specify the following form for the (i, j)th element of Ω, denoted by Ω_{ij}:

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \sigma_\varepsilon^2 \cdot \mathbf{1}_{i=j} + \sigma_P^2 \cdot \mathbf{1}_{P_i=P_j} + \sigma_S^2 \cdot \mathbf{1}_{S_i=S_j} + \sigma_D^2 \cdot \mathbf{1}_{D_i=D_j} + \sigma_{\text{dist}}^2 \cdot \exp(-\alpha \cdot d(Z_i, Z_j))$$

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

In other words, we allow for clustering at the puma, state, and division level, with in addition spatial correlation based on geographical distance, declining at an exponential rate. Note that pumas and divisions are generally not associated with administrative jurisdictions. Findings of correlations within such groupings are therefore not to be interpreted as arising from institutional differences between pumas or divisions. As a result, it is unlikely that such correlations in fact sharply decline at puma or division borders. The estimates based on models allowing for within-puma and within-division correlations are to be interpreted as indicative of correlation patterns that extend beyond state boundaries. What their precise interpretation is, for example arising from geographical features, or from local labor markets, is a question not directly addressed here.

In this specification $\Omega(\mathbf{Z}, \gamma)$ is a function of $\gamma = (\sigma^2_D, \sigma^2_S, \sigma^2_P, \sigma^2_{\text{dist}}, \alpha, \sigma^2_\varepsilon)'$. As a result the log likelihood function, leaving aside terms that do not involve the unknown parameters, is

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

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.

Table 3 gives maximum likelihood estimates for the covariance parameters γ 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, $\hat{\alpha} = 0.0468$, implies that the pure spatial component, $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d)$ dies out fairly quickly: at a distance $d(z, z')$ of about fifteen 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 correlation for two individuals in the same puma is $0.0888/0.9572 = 0.0928$. To put this in perspective, the covariance between log earnings and years of education is approximately 0.3, so the within-pum covariance is about a third of the log earnings and education covariance. For two individual in the same state, but in different puma's and ignoring the spatial component, the total covariance is 0.0093. The estimates suggest that much of what shows up as within-state correlations in a model 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 report in Tables 4, 5, and 6 results for the same model, and the same set of individuals, for three other variables. First, in Table 4, we look at correlation patterns in earnings residuals, based on a regression of log earnings on years of education, experience and the square of experience, where experience is defined as age minus six minus years of education. Next, in Table 5 we report results for years of education. Finally, in Table 6 we report results for hours worked. In all cases puma-level correlations are a 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 that assumes zero covariances between states is at odds with the data. Covariances vary substantially within states, and do not vanish at state boundaries. The results do raise some questions regarding the interpretation of the covariance structure. Whereas there are many regulations and institutions that vary by state, there are no institutions associated with divisions of states, or with pumas. It is possible that positive covariances between states within divisions arise from spatial correlations in institutions. States that are close together, or states that share borders, may have similar institutions. This could explain covariance structures that have discontinuities at state boundaries, but which still exhibit positive correlations extending beyond state boundaries. Within puma correlations may arise from variation in local labor market conditions. Both explanations make it unlikely that covariances change discontinuously at division or puma boundaries, and therefore the models estimated here are at best approximations to the covariance structure. We do not explore the interpretations for these correlations further here, but focus on their implications for inference.

To put these results in perspective, we look at the implications of these models for the precision of least squares estimates. To make this specific, we focus on the model in (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, California, Connecticut, Delaware, Massachusetts, Oregon, Rhode Island, Vermont, Washington, and DC (the state minimum wage is also higher than the federal minimum wage in Alaska and Hawaii, but these states are excluded from our basic sample).[4] In the second to last column in Tables 3-6, under the label "Min Wage," 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 specific, if $\hat{\Omega} = \Omega(\mathbf{Z},\hat{\gamma})$ is the estimate for $\Omega(\mathbf{Z},\gamma)$ in a particular specification, the standard error is

$$\sqrt{\widehat{\mathbb{V}_M}} = \sqrt{\mathbb{V}_M}(\mathbf{Z},\hat{\gamma})$$

$$= \left(\frac{1}{N^2 \cdot \overline{W}^2 \cdot (1-\overline{W})^2} (\overline{W} \quad -1) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \Omega(\mathbf{Z},\hat{\gamma}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix} \right)^{1/2}.$$

Despite the fact that the spatial correlations are quite far from being consistent with the conventional specification of constant within-state and zero out-of-state correlations, the standard errors for $\hat{\beta}_{\text{ols}}$ implied by the various covariance specifications are fairly similar for the set of models that incorporate state level correlations. For example, using the earnings variable, the one specification that does not include state level clustering leads to a standard error of 0.001,

[4]The data come from the website http://www.dol.gov/whd/state/stateMinWageHis.htm. Note that to be consistent with the 2000 census, we use the information from 2000, not the current state of the law.

and the five specifications that do include state level clustering lead to a range of standard errors from 0.068 to 0.091. Using the years of education variable, the one specification that does not include state level clustering leads to a standard error of 0.004, and the five specifications that do include state level clustering lead to a range of standard errors from 0.185 to 0.236. In the next three sections we develop a theoretical argument to provide some insights into this finding.

4 Randomization Inference with Complete Randomization

In this section we consider a different approach to analyzing the distribution of the least squares estimator, based on randomization inference. See for a general discussion Rosenbaum (2002). Recall the linear model given in equation (2.1),

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

combined with Assumption 1,

$$\varepsilon | \mathbf{W}, \mathbf{Z} \sim \mathcal{N}(0, \Omega(\mathbf{Z})).$$

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, as given in Lemma 1:

$$\mathbb{V}_M(\mathbf{W}, \mathbf{Z}) = \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1 - \overline{W})^2} (\overline{W} \;\; -1) \left(\iota_N \quad \mathbf{W} \right)' \Omega(\mathbf{Z}) \left(\iota_N \quad \mathbf{W} \right) \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}. \quad (4.1)$$

We can use this variance in combination with normality to construct confidence intervals for $\hat{\beta}$. The standard 95% confidence interval is

$$\text{CI}^{0.95}(\beta) = \left(\hat{\beta}_{\text{ols}} - 1.96 \cdot \sqrt{\mathbb{V}_M}, \hat{\beta}_{\text{ols}} + 1.96 \cdot \sqrt{\mathbb{V}_M} \right).$$

With the true value for Ω in this expression for the confidence interval, this confidence interval has the actual coverage equal to the nominal coverage. If we do not know Ω, we would substitute a consistent estimator $\hat{\Omega}$ into this expression. In that case the confidence interval is approximate, with the approximation becoming more accurate in large samples.

It is possible to do 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 N-vectors of these potential outcomes. 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 residuals 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 $\sum_{i=1}^N W_i = N_1$. In the next section we look at the case where W_i is randomly assigned to clusters. Formally, we assume

Assumption 2. RANDOMIZATION

$$\text{pr}(\mathbf{W} = \mathbf{w}|\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = 1 \Big/ \begin{pmatrix} N \\ N_1 \end{pmatrix}, \qquad \text{for all } \ \mathbf{w}, \text{ s.t. } \sum_{i=1}^{N} w_i = N_1.$$

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 $(\mathbf{Y}(0), \mathbf{Y}(1))$:

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

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}} \middle| \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right] = \beta, \tag{4.2}$$

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

$$\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \mathbb{V}\left(\hat{\beta}_{\text{ols}} \middle| \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right) = \frac{1}{N-1} \sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2 \cdot \left(\frac{1}{N_0} + \frac{1}{N_1}\right), \tag{4.3}$$

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

This follows from results by Neyman (1923) on randomization inference for average treatment effects. In the Appendix we provide some details. 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 the differences between this conditional variance and the conditional variance given $\mathbf{W}$ and $\mathbf{Z}$ and its interpretation. We make four specific commments, interpreting the model and randomization based variances $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$ and $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$, clarifying their relation, and the relative merits of both. First of all, although both variances $\mathbb{V}_M$ and $\mathbb{V}_R$ are exact, the two are valid under different, complementary assumptions. To see 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 assumption and the randomization assumption. In other words, we assume the existence of a large population. Each unit in the population is characterized by two potential outcomes, $(Y_i(0), Y_i(1))$, with the difference constant, $Y_i(1) - Y_i(0) = \beta$, and with the mean of $Y_i(0)$ in this infinite population equal to α. Then, conditional on the N locations $\mathbf{Z}$, we repeatedly randomly draw one unit from each the set of units in each location, randomly assign each unit a value for W_i, and observe W_i, Z_i, and $Y_i(W_i)$ for each unit.

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

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}} \middle| \mathbf{Z}, N_0, N_1\right] = \beta, \tag{4.4}$$

and, (*ii*), *its exact unconditional variance is*

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{V}\left(\hat{\beta}_{\text{ols}} \middle| \mathbf{Z}, N_0, N_1\right) = \left(\frac{1}{N-1}\text{trace}(\Omega(\mathbf{Z})) - \frac{1}{N \cdot (N-1)} \iota_N' \Omega(\mathbf{Z}) \iota_N\right) \cdot \left(\frac{1}{N_0} + \frac{1}{N_1}\right). \tag{4.5}$$

Given Assumption 1, but without Assumption 2, the unconditional variance is given in expression (2.6). Random assignment of $\mathbf{W}$ simplifies this to (4.5).

By the law of interated expectations, it follows that

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}\left[\mathbb{V}(\hat{\beta}_{\text{ols}}|\mathbf{Z}, \mathbf{Y}(0), \mathbf{Y}(1))\Big|\mathbf{Z}, N_0, N_1\right] + \mathbb{V}\left(\mathbb{E}\left[\hat{\beta}_{\text{ols}}|\mathbf{Z}, \mathbf{Y}(0), \mathbf{Y}(1)\right]\Big|\mathbf{Z}, N_0, N_1\right)$$

$$= \mathbb{E}\left[\mathbb{V}(\hat{\beta}_{\text{ols}}|\mathbf{Z}, \mathbf{Y}(0), \mathbf{Y}(1))\Big|\mathbf{Z}, N_0, N_1\right],$$

where the last equality follows because $\hat{\beta}_{\text{ols}}$ is unbiased conditional on $(\mathbf{Y}(0), \mathbf{Y}(1))$ and $\mathbf{Z}$. By the same argument

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}\left[\mathbb{V}(\hat{\beta}_{\text{ols}}|\mathbf{Z}, \mathbf{W})\Big|\mathbf{Z}\right] + \mathbb{V}\left(\mathbb{E}\left[\hat{\beta}_{\text{ols}}|\mathbf{Z}, \mathbf{W}\right]\Big|\mathbf{Z}\right) = \mathbb{E}\left[\mathbb{V}(\hat{\beta}_{\text{ols}}|\mathbf{Z}, \mathbf{W})\Big|\mathbf{Z}\right],$$

because, again, $\hat{\beta}_{\text{ols}}$ is unbiased, now conditional on $\mathbf{Z}$ and $\mathbf{W}$. Hence, in expectation the two variances, one based on randomization of $\mathbf{W}$, and one based on resampling the $(\mathbf{Y}(0), \mathbf{Y}(1))$, are both equal to the unconditional variance:

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

In finite samples, however, $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$ will generally be different from $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$.

The second point is that there is no reason to expect one of the confidence intervals to have better properties. Because of ancillarity arguments one would expect the confidence interval based on either of the conditional variances, either $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$ or $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$, to perform better than confidence intervals based on the unconditional variance $\mathbb{V}_U(\mathbf{Z})$, but we are not aware of a general repeated sampling argument that would support a preference for $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$ over $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$, or the other way around.

Third, as a practical matter, the randomization variance $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$ is much easier to estimate than the model-based variance $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$. Given the point estimates $\hat{\beta}_{\text{ols}}$ and $\hat{\alpha}_{\text{ols}}$, the natural estimator for $\sum_i (\varepsilon_i - \overline{\varepsilon})^2/(N-1)$ is

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}} \cdot W_i\right)^2.$$

This leads to

$$\hat{\mathbb{V}}_R = \hat{\sigma}_\varepsilon^2 \cdot \left(\frac{1}{N_0} + \frac{1}{N_1}\right), \tag{4.6}$$

with corresponding confidence interval. Estimating the variance conditional on $\mathbf{W}$ and $\mathbf{Z}$, on the other hand is not easy. One would need to impose some structure on the covariance matrix Ω in order to estimate it (or at least on $\mathbf{X}'\Omega\mathbf{X}$) consistently. The resulting estimator for the variance will be sensitive to the specification unless W_i is randomly assigned. If, therefore, one relies on random assignment of W_i anyway, there appears to be little reason for using the model-based variance at all.

For the fourth point, suppose we had focused on the repeated sampling variance for $\hat{\beta}_{\text{ols}}$ conditional on $\mathbf{W}$ and $\mathbf{Z}$, but under a model with independent and homoskedastic errors,

$$\varepsilon|\mathbf{W},\mathbf{Z} \sim \mathcal{N}(0, \sigma^2 \cdot I_N). \tag{4.7}$$

Under that independent and homoskedastic model (4.7), the exact sampling distribution for $\hat{\beta}_{\text{ols}}$ would be

$$\hat{\beta}_{\text{ols}}\Big| \mathbf{W},\mathbf{Z} \sim \mathcal{N}\left(\beta, \mathbb{V}_H(\mathbf{W},\mathbf{Z})\right), \tag{4.8}$$

where

$$\mathbb{V}_H(\mathbf{W},\mathbf{Z}) = \sigma^2 \cdot \left(\frac{1}{N_0} + \frac{1}{N_1}\right).$$

Given this model, the natural estimator for the variance is

$$\hat{\mathbb{V}}_H = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}} \cdot W_i\right)^2 \cdot \left(\frac{1}{N_0} + \frac{1}{N_1}\right). \tag{4.9}$$

This estimated variance $\hat{\mathbb{V}}_H$ is identical to the estimated variance under the randomization distribution, $\hat{\mathbb{V}}_R$ given in (4.6). Hence, and this is the main insight of this section, if the assignment $\mathbf{W}$ is completely random, and the treatment effect is constant, one can 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 $\hat{\mathbb{V}}_H$ will *not* be estimating the variance under the repeated sampling thought experiment that one may have in mind, conditional on $\mathbf{Z}$ and $\mathbf{W}$, $\hat{\mathbb{V}}_H$ <u>is</u> consistent for the variance under the randomization distribution, conditional on $\mathbf{Z}$ and $(\mathbf{Y}(0), \mathbf{Y}(1)$, and for the unconditional one.

The result that the mis-specification 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) writes: "when the correlation patterns of the independent variables are unrelated to those across the errors, then the least squares variance estimates are consistent," and points out that the interpretation changes slightly, because the variance is no longer conditional on the explanatory variables. Angrist and Pischke (2009) write, 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 preceeding discussion interprets this result 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 current paper. The covariate of interest, W_i, varies only between clusters, 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 $W_i = 1$, and M_0 the number of clusters with all individuals assigned to $W_i = 0$. The cluster indicator is

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

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 clear, 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 m-th element equal to $\tilde{Y}_m(0) = \sum_{i:C_{ij}=m} Y_i(0)/N_m$, and $\tilde{Y}_s(1) = \sum_{i:C_{ij}=1} Y_i(1)/N_m$ respectively. Similarly, let $\tilde{\varepsilon}$ be an M-vector with m-th element equal to $\tilde{\varepsilon}_m = \sum_{i:C_{ij}=1} \varepsilon_i/N_m$, and let $\overline{\tilde{\varepsilon}} = \sum_{m=1}^{M} \tilde{\varepsilon}_m/M$.

Formally the assumption on the assignment mechanism is now:

Assumption 3. (Cluster Randomization)

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

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

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

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

$$\mathbb{V}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \mathbb{V}(\hat{\beta}_{\text{ols}} | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \frac{1}{M-1} \sum_{m=1}^{M} \left(\tilde{\varepsilon}_m - \overline{\tilde{\varepsilon}}\right)^2 \cdot \left(\frac{1}{M_0} + \frac{1}{M_1}\right). \quad (5.1)$$

The proof is given in the Appendix. If we also make the model assumption (Assumption 1), we can derive the unconditional variance:

Lemma 5. (Marginal Variance Under Cluster Randomization)
Suppose that Assumptions 1, 3, and 4 hold. Then

$$\mathbb{V}_U(\Omega(\mathbf{Z})) = \left(\frac{M^2}{N^2 \cdot (M-1)} \cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \frac{M}{N^2 \cdot (M-1)} \iota'\Omega(\mathbf{Z})\iota\right) \cdot \left(\frac{1}{M_0} + \frac{1}{M_1}\right). \quad (5.2)$$

This unconditional variance is a special case of the expected value of the unconditional variance in (2.6), with the expectation taken over $\mathbf{W}$ given the cluster-level randomization. A special case of the result in this Lemma is that with each unit its own cluster, so that $M = N$, $M_0 = N_0$, $M_1 = N_1$, and $\mathbf{C} = I_N$, in which case the unconditional variance under clustering, (5.2), reduces to the unconditional variance under complete randomization, (4.5).

6 Variance Estimation Under Misspecification

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

Suppose the model generating the data is the linear model in (2.1), with a general covariance matrix Ω, and Assumption 1 holds. The researcher estimates a parametric model that imposes a potentially incorrect structure on the covariance matrix. Let $\Sigma(\mathbf{Z}, \gamma)$ be the parametric model for the error covariance matrix. The example we are most interested in is characterized by a clustering structure. In that case $\Sigma(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_C^2)$ is the $N \times N$ matrix with $\gamma = (\sigma_\varepsilon^2, \sigma_C^2)'$, where

$$\Sigma_{ij}(\sigma_\varepsilon^2, \sigma_C^2) = \begin{cases} \sigma_\varepsilon^2 + \sigma_C^2 & \text{if } i = j \\ \sigma_C^2 & \text{if } i \neq j, C_{im} = C_{jm}, \text{ for all } m = 1, \ldots, M, \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

Initially, however, we allow for any parametric structure $\Sigma(\mathbf{Z}, \gamma)$.

Under the parametric model $\Sigma(\mathbf{Z}, \gamma)$, 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,

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

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

$$\mathbb{V}_M(\Sigma(\mathbf{Z}, \tilde{\gamma}), \mathbf{W}, \mathbf{Z}) = \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1 - \overline{W})^2} (\overline{W} \;\; -1) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \Sigma(\mathbf{Z}, \tilde{\gamma}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}.$$

In general this pseudo-true conditional variance will differ from $\mathbb{V}_M(\mathbf{W}, \mathbf{Z}) = \mathbb{V}_M(\Omega, \mathbf{W}, \mathbf{Z})$, based on the correct error-covariance matrix $\Omega(\mathbf{Z})$. However, we need not have equality for every value of $\mathbf{W}$, if we have equality on average. Here we focus on the expected value of $\mathbb{V}_M(\Sigma(\mathbf{Z}, \tilde{\gamma}), \mathbf{W}, \mathbf{Z})$, conditional on $\mathbf{Z}$, under assumptions on the distribution of $\mathbf{W}$. Let us denote this expectation by $\mathbb{V}_U(\Sigma(\mathbf{Z}, \tilde{\gamma}), \mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\Sigma(\mathbf{Z}, \tilde{\gamma}), \mathbf{W}, \mathbf{Z})|\mathbf{Z}]$. The question is under what conditions on the specification of the error-covariance matrix $\Sigma(\mathbf{Z}, \gamma)$, in combination with assumptions on the assignment process, this unconditional variance is is equal to the expected variance with the expectation of the variance under the correct error-covariance matrix, $\mathbb{V}_U(\Omega, \mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\Omega, \mathbf{W}, \mathbf{Z})|\mathbf{Z}]$.

The following lemma 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 that Assumptions 1, 3, and 4 hold, and suppose that that Σ is specified as in (6.1) Then

$$\mathbb{V}_U(\Sigma(\mathbf{Z}, \tilde{\gamma}), \mathbf{Z}) = \mathbb{V}_U(\Omega, \mathbf{Z}).$$

For the proof, see the Appendix. This is the main theoretical result in the paper. 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. The limitation of the result to equal sized clusters does not play a major conceptual role, although it is important for the specific results. One could establish similar results if the estimator was the simple average of cluster averages, instead of the average of all individuals. The weighting of cluster averages induced by the focus on the average over all individuals creates complications for the Neyman framework. See Dylan, Ten Have and Rosenbaum (2008) for some discussion of cluster randomized experiments.

In many econometric analyses we specify the conditional distribution of the outcome given some explanatory variables, and we pay no attention to the joint distribution of the explanatory variables. The result in Theorem 1 shows that it may be useful to do so. Depending on the joint distribution of the explanatory variables, the analyses may be robust to mis-specification 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, or the average of individual level outcomes, e.g., the logarithm of earnings, years of education, or hours worked per week. 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 paper are based on Mantel statistics (e.g., Mantel, 1967; Schabenberger and Gotway, 2004). These general form of the statistics we use is a proximity-weighted average of squared pairwise differences:

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

where $d_{st} = d(Z_s, Z_t)$ is a non-negative 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 data set, M^{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 m we re-calculate the Mantel statistic, say M_m. 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 m (out of the set of $S!$ allocations) such that the associated Mantel statistic M_m is less than or equal to the observed Mantel statistic M^{obs}:

$$p = \frac{1}{S!} \sum_{m=1}^{S!} \mathbf{1}_{M^{\text{obs}} \geq M_m}. \tag{7.2}$$

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. In that case 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 US is divided into 9 divisions):

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

Third, we define proximity d_{st} as minus the geographical distance between states s and t:

$$d_{st}^{G} = -d\left(Z_s, Z_t\right), \tag{7.5}$$

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.

The last three proximity measures are based on transformations of geographical distance:

$$d_{st}^{\alpha} = \exp\left(-\alpha \cdot d\left(Z_s, Z_t\right)\right), \tag{7.6}$$

for $\alpha = 0.00138$, $\alpha = 0.00276$, 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 five variables. First, an indicator for having a state minimum wage higher than the federal minimum wage. This indicator takes on the value 1 in eleven out of the forty nine states in our sample, with these eleven states mainly concentrated in the North East. Second, the average of the logarithm of yearly earnings. Third, average years of education. Fourth, average hours worked. Fifth, average weeks worked. The results for the five variables and three statistics are presented in Table 7.

All five 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 Does the Spatial Correlation Matter?

The theoretical results in Section 6 show that spatial correlations beyond state borders do not matter if the interventions of interest are randomly assigned to states. The empirical results in Section 7 suggest that there are in fact statistically significant spatial correlation patterns in state-level regulations. Hence the results in Section 6 need not apply in practice. Here we assess the practical implications of the combination of these findings.

We focus on two examples. First, we return to the explanatory variable we looked at in Section 3, the indicator whether a state has a minimum wage exceeding the federal level. Nine states in our sample fit that criteria, and in the previous section it was shown that this indicator variable exhibits a statistically significant degree of spatial clustering. Second, we consider an artificial intervention, applying only to the six states making up the New England division and the five states making up the East North Central division (so that we again have 11 states with the intervention). The last two columns in Tables 3-6 present standard errors under different specifications for the error covariance matrix. We find that although the minimum wage variable exhibits substantial spatial correlation, allowing for within-state correlation leads to standard errors that are fairly close to those for more general models. On the other hand, with the artificial regulation applying only to NE/ENC states, ignoring division level correlations leads to underestimation of the standard errors almost by a substantial amount. For the earnings outcome, the range of values for standard errors within specifications for the variance that all allow for state-level clustering is $[0.494, 0.854]$, and for years of education it the range of standard errors $[0.136, 0.227]$. Thus, if the state level regulation is clustered to a substantial degree, taking into account between state correlations for the outcomes is important for inference.

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, including 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 depending on the form of that correlation, using a more flexible specification of the error covariance structure may be important.

In practice we recommend that researchers explicitly explore the spatial correlation structure of both the outcomes as well as 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, one should explicitly account for the spatial correlation, either by allowing for a more flexible specification than one that only accounts for state level clustering, or for using robust variance estimators that allow for general spatial correlation structures by relying more heavily on large sample approximations, using, for exampe, the methods discussed in Conley (1999), Bester, Conley and Hansen (2009), and Ibragimov and Müller (2009).

Appendix

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:*

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

For a proof see XXXX

Proof of Theorem A.1:

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

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

similarly

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

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

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

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

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

□

Lemma A.2. (Neyman) *Suppose we have N triples $(W_i, Y_i(0), Y_i(1)$, with $W_i \in \{0,1\}$. Define*

$$\beta = \frac{1}{N}\sum_{i=1}^N \left(Y_i(1) - Y_i(0)\right),$$

$$\overline{Y}(0) = \frac{1}{N}\sum_{i=1}^N Y_i(0), \qquad \overline{Y}(1) = \frac{1}{N}\sum_{i=1}^N Y_i(1),$$

$$S_0^2 = \frac{1}{N-1}\sum_{i=1}^N \left(Y_i(0) - \overline{Y}(0)\right)^2 \qquad S_1^2 = \frac{1}{N-1}\sum_{i=1}^N \left(Y_i(1) - \overline{Y}(1)\right)^2,$$

$$S_{01}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right)^2,$$

$$N_1 = \sum_{i=1}^{N} W_i, \qquad N_0 = \sum_{i=1}^{N}(1 - W_i)$$

and

$$\hat{\beta} = \frac{1}{N_1}\sum_{i=1}^{N} W_i \cdot Y_i(1) - \frac{1}{N_0}\sum_{i=1}^{N}(1 - W_i)\cdot Y_i(0).$$

Suppose that W_i is randomly assigned to the N units, subject to $\sum_{i=1}^{N} W_i = N_1$. Then

$$\mathbb{E}\left[\hat{\beta}|\mathbf{Y}(0), \mathbf{Y}(1)\right] = \beta,$$

and

$$\mathbb{V}(\hat{\beta}|\mathbf{Y}(0), \mathbf{Y}(1)) = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{01}^2}{N}.$$

Lemma A.3. (i), *Suppose Assumption 2 holds, then for any $N \times N$ matrix Ω,*

$$\mathbb{E}\left[\mathbf{W}'\Omega\mathbf{W}\right] = \frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)} \cdot \iota_N' \Omega \iota_N + \frac{N_1 \cdot N_0}{N \cdot (N-1)} \cdot \text{trace}(\Omega),$$

and (ii), suppose Assumptions 3 and 4 hold, then for any $N \times N$ matrix Ω,

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

Proof of Lemma A.3: First, consider part (i): Because

$$\mathbb{E}[W_i \cdot W_j] = \begin{cases} \frac{N_1}{N} & \text{if } i = j, \\ \frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)} & \text{if } i \neq j, \end{cases}$$

it follows that

$$\begin{aligned}\mathbb{E}[\mathbf{W}\mathbf{W}'] &= \frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)} \cdot \iota_N \iota_N' + \left(\frac{N_1}{N} - \frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)}\right) \cdot I_N \\ &= \frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)} \cdot \iota_N \iota_N' + \frac{N_1 \cdot N_0}{N \cdot (N-1)} \cdot I_N.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}[\mathbf{W}'\Omega\mathbf{W}] &= \text{trace}\left(\mathbb{E}[\Omega\mathbf{W}\mathbf{W}']\right) \\ &= \text{trace}\left(\Omega \cdot \left(\frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)} \cdot \iota_N \iota_N' + \frac{N_1 \cdot N_0}{N \cdot (N-1)} \cdot I_N\right)\right) \\ &= \frac{N_1 \cdot (N_1 - 1)}{N \cdot (N-1)} \cdot \iota_N' \Omega \iota_N + \frac{N_1 \cdot N_0}{N \cdot (N-1)} \cdot \text{trace}(\Omega).\end{aligned}$$

Next, consider part (ii). Now

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

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

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

□

Lemma A.4. *Suppose the $N\times N$ matrix Ω satisfies*

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

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, (*i*)

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

and, (*ii*)

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

or,

$$\left(\Omega^{-1}\right)_{ij} = \begin{cases} \sigma_\varepsilon^{-2} - \frac{\sigma_C^2}{\sigma_\varepsilon^2\cdot(\sigma_\varepsilon^2+\sigma_S^2\cdot N/M)} & \text{if } i=j \\ -\frac{\sigma_C^2}{\sigma_\varepsilon^2\cdot(\sigma_\varepsilon^2+\sigma_C^2\cdot N/M)} & \text{if } i\neq j, \forall m, C_{im}=C_{jm}, \\ 0 & \text{otherwise,}\end{cases}$$

Proof of Lemma A.4: First, consider the first part. Apply Lemma A.1 with

$$\Sigma = \sigma_C^2\cdot I_M, \qquad \text{and}\ \ \mathbf{C}'\mathbf{C} = \frac{N}{M}\cdot I_M,$$

so that

$$\Omega_S = \left(\sigma_S^2+\sigma_\varepsilon^2\cdot\frac{M}{N}\right)\cdot I_M.$$

Then, by Lemma A.1, we have

$$\ln\det(\Omega) = (N-M)\cdot\ln(\sigma_\varepsilon^2) + M\cdot\ln(N/M) + \ln\det(\Omega_C)$$

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

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

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

Next, consider part (ii). 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

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

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

Proof of Lemma 2: By Assumption 2, we can apply Lemma A.2. This directly implies the unbiasedness result. It also implies that the conditional variance is

$$\mathbb{V}\left(\hat\beta_{\text{ols}}\middle|\,\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right) = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{01}^2}{N}.$$

The assumption that the treatment effect is constant implies that $S_{01}^2 = 0$, and that $\varepsilon_i - \overline{\varepsilon} = Y_i(0) - \overline{Y(0)} = Y_i(1) - \overline{Y(1)}$, which in turn implies that

$$\mathbb{V}\left(\hat\beta_{\text{ols}}\middle|\,\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right) = \frac{\sum_{i=1}^N(\varepsilon_i-\overline{\varepsilon})^2/(N-1)}{N_0} + \frac{\sum_{i=1}^N(\varepsilon_i-\overline{\varepsilon})^2/(N-1)}{N_1}$$

$$= \frac{1}{N-1}\sum_{i=1}^N(\varepsilon_i-\overline{\varepsilon})^2\cdot\left(\frac{1}{N_0}+\frac{1}{N_1}\right).$$

□

Proof of Lemma 3: The unbiasedness result directly follows from the conditional unbiasedness established in Lemma 2. To establish the second part of the Lemma, we need to show that the expectation of the conditional variance is equal to the marginal variance, or:

$$\mathbb{E}\left[\frac{1}{N-1}\sum_{i=1}^N(\varepsilon_i-\overline{\varepsilon})^2\cdot\left(\frac{1}{N_0}+\frac{1}{N_1}\right)\middle|\,\mathbf{Z},N_0,N_1\right] = \left(\frac{1}{N-1}\text{trace}(\Omega(\mathbf{Z})) - \frac{1}{N\cdot(N-1)}\iota_N'\Omega(\mathbf{Z})\iota_N\right)\cdot\left(\frac{1}{N_0}+\frac{1}{N_1}\right),$$

or, equivalently,

$$\mathbb{E}\left[\sum_{i=1}^N(\varepsilon_i-\overline{\varepsilon})^2\middle|\,\mathbf{Z},N_0,N_1\right] = \text{trace}(\Omega) - \frac{1}{N}\iota_N'\Omega(\mathbf{Z})\iota_N, \tag{A.2}$$

where $\mathbb{E}[\varepsilon\varepsilon'|\mathbf{Z}] = \Omega(\mathbf{Z})$. First,

$$\sum_{i=1}^N(\varepsilon_i-\overline{\varepsilon})^2 = (\varepsilon - \iota_N\iota_N'\varepsilon/N)'(\varepsilon - \iota_N\iota_N'\varepsilon/N)$$

$$= \varepsilon'\varepsilon - 2\varepsilon'\iota_N\iota_N'\varepsilon/N + \varepsilon'\iota_N\iota_N\iota_N\iota_N'\varepsilon/N^2$$

$$= \varepsilon'\varepsilon - \varepsilon'\iota_N\iota_N'\varepsilon/N.$$

Thus

$$\mathbb{E}\left[\sum_{i=1}^{N}(\varepsilon_i-\overline{\varepsilon})^2\middle|\mathbf{Z},N_0,N_1\right]=\mathbb{E}\left[\varepsilon'\varepsilon-\varepsilon'\iota_N\iota_N'\varepsilon/N\right|\mathbf{Z},N_0,N_1]$$

$$=\text{trace}\left(\mathbb{E}\left[\varepsilon\varepsilon'-\iota_N'\varepsilon\varepsilon'\iota_N/N\right|\mathbf{Z},N_0,N_1]\right)$$

$$=\text{trace}\left(\Omega(\mathbf{Z})\right)-\iota_N'\Omega(\mathbf{Z})\iota_N/N,$$

which proves (A.2), and thus the result in the Lemma. □

Proof of Lemma 4: Applying Lemma A.2 to the clusters implies that

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}}\middle|\tilde{\mathbf{Y}}(0),\tilde{\mathbf{Y}}(1)\right]=\beta,$$

and

$$\mathbb{V}\left(\hat{\beta}_{\text{ols}}\left|\tilde{\mathbf{Y}}(0),\tilde{\mathbf{Y}}(1)\right.\right)=\frac{\tilde{S}_0^2}{S_0}+\frac{\tilde{S}_1^2}{S_1}-\frac{\tilde{S}_{01}^2}{S},$$

where

$$\tilde{S}_0^2=\frac{1}{S-1}\sum_{s=1}^{S}\left(\tilde{Y}_s(0)-\overline{\tilde{Y}(0)}\right)^2,\qquad \tilde{S}_1^2=\frac{1}{S-1}\sum_{s=1}^{S}\left(\tilde{Y}_s(1)-\overline{\tilde{Y}(1)}\right)^2,$$

and

$$\tilde{S}_{01}^2=\frac{1}{S-1}\sum_{s=1}^{S}\left(\tilde{Y}_s(1)-\tilde{Y}_s(0)-\left(\overline{\tilde{Y}(1)}-\overline{\tilde{Y}(0)}\right)\right)^2.$$

Under the constant treatment effect assumption, it follows that $\tilde{S}_{01}^2=0$, and that $\tilde{S}_0^2=\tilde{S}_1^2=\sum_{s=1}^{S}(\tilde{\varepsilon}_s-\overline{\tilde{\varepsilon}_s})^2/(S-1)$, and the result follows. □

Proof of Lemma 5: For proving the claim in the Lemma it is sufficient to show that the expectation of the conditional variance is equal to the marginal variance, or:

$$\mathbb{E}\left[\frac{1}{M-1}\sum_{m=1}^{M}\left(\tilde{\varepsilon}_m-\overline{\tilde{\varepsilon}}\right)^2\cdot\left(\frac{1}{M_0}+\frac{1}{M_1}\right)\middle|\mathbf{Z},\mathbf{C},N_0,N_1\right]$$

$$=\left(\frac{M^2}{N^2\cdot(M-1)}\cdot\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)-\frac{M}{N^2\cdot(M-1)}\iota'\Omega(\mathbf{Z})\iota\right)\cdot\left(\frac{1}{M_0}+\frac{1}{M_1}\right),$$

or, equivalently,

$$\mathbb{E}\left[\sum_{s=1}^{M}\left(\tilde{\varepsilon}_s-\overline{\tilde{\varepsilon}}\right)^2\middle|\mathbf{Z},\mathbf{C},N_0,N_1\right]=\left(\frac{M^2}{N^2}\cdot\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)-\frac{M}{N^2}\iota'\Omega(\mathbf{Z})\iota\right).\qquad\text{(A.3)}$$

Note that in general

$$\mathbf{C}\iota_M=\iota_N,$$

and under Assumption 4, it follows that

$$\mathbf{C}'\mathbf{C}=\frac{N}{M}\cdot I_M.$$

We can write

$$\tilde{\varepsilon}_s=\left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\varepsilon=\frac{M}{N}\mathbf{C}'\varepsilon,$$

and

$$\overline{\tilde{\varepsilon}} = \frac{1}{M}\iota_M' \left(\mathbf{C}'\mathbf{C}\right)^{-1} \mathbf{C}'\varepsilon = \frac{1}{N}\iota_N'\varepsilon,$$

so that

$$\sum_{m=1}^{M} \left(\tilde{\varepsilon}_s - \overline{\tilde{\varepsilon}}\right)^2 = \left(\frac{M}{N}\mathbf{C}'\varepsilon - \frac{1}{M}\iota_M\iota_N'\varepsilon\right)' \left(\frac{M}{N}\mathbf{C}'\varepsilon - \frac{1}{M}\iota_M\iota_N'\varepsilon\right)$$

$$= \left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right)' \left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right).$$

$$= \varepsilon' \left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)' \left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon.$$

Thus

$$\mathbb{E}\left[\left.\sum_{m=1}^{M} \left(\tilde{\varepsilon}_s - \overline{\tilde{\varepsilon}}\right)^2 \right| \mathbf{Z}, \mathbf{C}, N_0, N_1\right] = \mathbb{E}\left[\left.\varepsilon' \left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)' \left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right| \mathbf{Z}, \mathbf{C}, N_0, N_1\right]$$

$$= \text{trace}\left(\mathbb{E}\left[\left.\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)' \left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\varepsilon'\right| \mathbf{Z}, \mathbf{C}, N_0, N_1\right]\right)$$

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

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

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

□

Proof of Theorem 1: For equality of $\mathbb{V}_U(\Sigma)$ and $\mathbb{V}_U(\Omega)$ we need equality of trace($\mathbf{C}'\Omega\mathbf{C}$) and trace($\mathbf{C}'\Sigma(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\mathbf{C}$).
The log likelihood function based on the specification (A.1) is

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

The expected value of the log likelihood function is

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

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

$$= -\frac{1}{2} \cdot \ln\left(\Sigma\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_C^2\right)\right) - \frac{1}{2} \cdot \text{trace}\left(\Sigma\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_C^2\right)^{-1}\Omega\right).$$

Using Lemma A.4, this is equal to

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

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

$$\frac{\partial}{\partial \sigma_C^2} \mathbb{E}\left[L(\sigma_\varepsilon^2, \sigma_C^2 | \mathbf{Y}, \mathbf{Z}) \middle| \mathbf{Z} \right] = -\frac{S \cdot N_s / \sigma_\varepsilon^2}{2 \cdot (1 + N_s \cdot \sigma_C^2 / \sigma_\varepsilon^2)} + \text{trace}\left(\mathbf{C}'\Omega\mathbf{C}\right) \cdot \left(\frac{1}{\sigma_\varepsilon^2 \cdot (\sigma_\varepsilon^2 + \sigma_C^2 \cdot N_m)} - \frac{(N/M) \cdot \sigma_C^2 \cdot \sigma_\varepsilon^2}{\sigma_\varepsilon^4 \cdot (\sigma_\varepsilon^2 + \sigma_C^2 \cdot (N/M))^2} \right)$$

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

Hence the first order condition for $\tilde{\sigma}_C^2$ implies that

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

For the misspecified error-covariance matrix Σ we have

$$\text{trace}\left(\mathbf{C}'\Sigma\mathbf{C}\right) = \sum_{m=1}^{M} \left(N_m^2 \cdot \sigma_C^2 + N_m \cdot \sigma_\varepsilon^2 \right).$$

By equality of the cluster sizes this simplifies to

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

□

References

Abadie, A., A. Diamond, and J. Hainmueller, (2009), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," forthcoming, *Journal of the American Statistical Association.*

Angrist, J., and S. Pischke, (2009), *Mostly Harmless Econometrics*, Princeton University Press, Princeton, NJ.

Anselin, L., R. Florax, and S. Rey, (2004), *Advances in Spatial Econometrics*, Springer.

Arbia, G., (2006), *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Springer.

Bester, C., T. Conley, and C. Hansen, (2009), "Inference with Dependent Data Using Clustering Covariance Matrix Estimators," Unpublished Manuscript, University of Chicago Business School.

Bertrand, M., E. Duflo, and S. Mullainathan, (2004), "How Much Should We Trust Difference-in-Differences Estimates," *Quarterly Journal of Economics*, 119: 249-275.

Black, S., (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, Vol. CXIV, 577.

Conley, T., (1996) "Econometric Modelling of Cross-Sectional Dependence," Ph.D. Dissertation, University of Chicago.

Conley, T., (1999), "GMM Estimation with Cross-sectional Dependence," *Journal of Econometrics*, 92: 1-45.

Cox, D., and D. Hinkley, (1974), *Theoretical Statistics*, Chapman and Hall.

Cressie, N., (1993), *Statistics for Spatial Data*, Wiley.

Donald, S. and K. Lang, (2007), "Inference with Difference in Differences and Other Panel Data," *Review of Economics and Statistics*, Vol. 89(2): 221-233.

Fisher, R. A., (1925), *The Design of Experiments*, 1st ed, Oliver and Boyd, London.

Greenwald, B.., (1983), "A General Analysis of the Bias in Estimated Standard Errors of Least Squares Coefficients," *Journal of Econometrics*, 22, 323-338.

Hansen, C., (2009), "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects," *Journal of Econometrics*, 140(2), 670-694.

Ibragimov, R., and U. Müller, (2009), "t-statistic based correlation and heterogeneity robust inference" Unpublished Manuscript, Department of Economics, Harvard University.

Imbens, G., and T. Lemieux, (2008) "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* Vol 142(2):615-635.

Kloek, T., (1981), "OLS Estimation in a Model where a Microvariable is Explained by Aggregates and Contemporaneous Disturbances are Equicorrelated," *Econometrica*, Vol. 49, No. 1, 205-207.

Lee, D., and T. Lemieux, (2009), "Regression Discontinuity Designs in Economics," Working Paper, Dept of Economics, Princeton University.

Liang, K., and S. Zeger, (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73(1): 13-22.

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

Moulton, B., (1986), "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385-397.

Moulton, B., (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 334-338.

Moulton, B., and W. Randolph, (1989) "Alternative Tests of the Error Component Model," *Econometrica, Vol. 57, No. 3, 685-693.*

Neyman, J., (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9,"translated in *Statistical Science*, (with discussion), Vol 5, No 4, 465–480, 1990.

Rosenbaum, P., (2002), *Observational Studies*, 2nd edition, Springer Verlag, New York.

Ruggles, S., M. Sobek, T. Alexander, C. Fitch, R. Goeken, P. Hall, M. King, and C. Ronnander, (2008), "Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database]" Minneapolis, MN: Minnesota Population Center [producer and distributor].

Schabenberger, O., and C. Gotway, (2004), *Statistical Methods for Spatial Data Analysis*, Chapman and Hall.

Small, D., T. Ten Have, and P. Rosenbaum., (2008), "Randomization Inference in a Group-Randomized Trial of Treatments for Depression: Covariate Adjustment, Noncompliance, and Quantile Effects," *Journal of the American Statistical Association*, 103(481), 271-279.

William S. (1998), *Comparison of standard errors for robust, cluster, and standard estimators, StataCorp* http://www.stata.com/support/faqs/stat/cluster.html

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

Table 1: Summary Statistics

	log earnings	years of educ	hours worked
Average	10.17	13.05	43.76
Stand Dev	0.97	2.81	11.00
Average of PUMA Averages	10.17	13.06	43.69
Stand Dev of PUMA Averages	0.27	0.95	1.63
Average of State Averages	10.14	13.12	43.94
Stand Dev of State Averages	0.12	0.33	0.75
Average of Division Averages	10.17	13.08	43.80
Stand Dev of Division Averages	0.09	0.31	0.48

Table 2: Sample Sizes

Number of observation in the sample	2,590,190
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 (incl DC, excl AK, HA, 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

Table 3: Maximum likelihood Estimates for Clustering Variances for Demeaned Log Earnings

σ^2_ε	σ^2_D	σ^2_S	σ^2_P	σ^2_{dis}	a	LLH	$\widehat{\text{s.e.}}(\hat{\beta})$ Min Wage	NE/ENC
0.9388 [0.0008]	0	0	0	0	0	1213298.132	0.00145	0.0015
0.9294 [0.0008]	0	0.01610 [0.0018]	0	0	0	1200406.963	0.07996	0.0568
0.8683 [0.0008]	0	0.0111 [0.0029]	0.0659 [0.0022]	0	0	1116976.379	0.06794	0.0494
0.9294 [0.0008]	0.0056 [0.0020]	0.0108 [0.0020]	0	0	0	1200403.093	0.09091	0.0810
0.8683 [0.0008]	0.0056 [0.0033]	0.0058 [0.0021]	0.0660 [0.0021]	0	0	1116971.990	0.08054	0.0760
0.8683 [0.0008]	0.0080 [0.0049]	0.0008 [0.0012]	0.0331 [0.0021]	0.0324 [0.0030]	0.0468 [0.0051]	1603400.922	0.08602	0.0854

Table 4: Maximum likelihood Estimates for Clustering Variances for Log Earnings Residuals

$\hat{\sigma}^2_\epsilon$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_S$	$\hat{\sigma}^2_P$	$\hat{\sigma}^2_{\text{dist}}$	$\hat{\alpha}$	Log Lik	$\widehat{\text{s.e.}}(\hat{\beta})$ Min Wage	NE/ENC
0.7125 [0.0006]	0	0	0	0	0	-856182.0271	0.0013	0.0013
0.7071 [0.0006]	0	0.0124 [0.0025]	0	0	0	-846374.7763	0.0702	0.0499
0.6810 [0.0006]	0	0.0087 [0.0022]	0.0284 [0.0009]	0	0	-801566.9926	0.0597	0.0430
0.7071 [0.0006]	0.0031 [0.0020]	0.0091 [0.0023]	0	0	0	-846371.8921	0.0765	0.0655
0.6810 [0.0006]	0.0028 [0.0020]	0.0059 [0.0018]	0.0284 [0.0009]	0	0	-801563.9898	0.0661	0.0585
0.6810 [0.0006]	0.0027 [0.0020]	0.0060 [0.0021]	0.0183 [0.0011]	0.0120 [0.0015]	0.0271 [0.0057]	-801460.8862	0.0710	0.0619

Table 5: Maximum likelihood Estimates for Clustering Variances for Demeaned Years of Education

$\hat{\sigma}^2_\epsilon$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_S$	$\hat{\sigma}^2_P$	$\hat{\sigma}^2_{\text{dist}}$	$\hat{\alpha}$	Log Lik	$\widehat{\text{s.e.}}(\hat{\beta})$	
							Min Wage	NE/ENC
8.0400 [0.0071]	0	0	0	0	0	-3994627.262	0.0043	0.0043
7.9337 [0.0070]	0	0.1068 [0.0219]	0	0	0	-3977534.893	0.2060	0.1464
7.1264 [0.0063]	0	0.0792 [0.0218]	0.8319 [0.0264]	0	0	-3843542.281	0.1846	0.1361
7.9337 [0.0070]	0.0393 [0.0257]	0.0702 [0.0162]	0	0	0	-3977530.761	0.2363	0.2121
7.1264 [0.0063]	0.0531 [0.0328]	0.0362 [0.0142]	0.8313 [0.0264]	0	0	-3843537.315	0.2339	0.2271
7.1264 [0.0063]	0.0466 [0.0158]	0.0321 [0.0332]	0.6956 [0.0307]	0.1355 [0.0289]	0.0343 [0.0123]	-3843496.376	0.2334	0.2222

Table 6: Maximum likelihood Estimates for Clustering Variances for Demeaned Hours Worked Per Week

$\hat{\sigma}^2_\epsilon$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_S$	$\hat{\sigma}^2_P$	$\hat{\sigma}^2_{\text{dist}}$	$\hat{\alpha}$	Log Lik	$\widehat{\text{s.e.}}(\hat{\beta})$	
							Min Wage	NE/ENC
120.9749 [0.1036]	0	0	0	0	0	-7505830.603	0.0165	0.0165
120.5942 [0.1061]	0	0.5697 [0.1176]	0	0	0	-7501871.878	0.4759	0.3384
118.4649 [0.1040]	0	0.2556 [0.0758]	2.2410 [0.0743]	0	0	-7482236.836	0.3298	0.2421
120.5942 [0.1060]	0.1161 [0.1022]	0.4573 [0.1056]	0	0	0	-7501870.39	0.5140	0.4279
118.4650 [0.1038]	0.1177 [0.0823]	0.1335 [0.0595]	2.2451 [0.0745]	0	0	-7482235.194	0.3794	0.3557
118.5308 [0.1042]	0.0701 [0.0631]	0.1257 [0.0601]	1.6442 [0.1282]	0.6658 [0.1321]	0.0496 [0.0172]	-7482652.492	0.3574	0.3165

Table 7: p-values for Mantel Statistics, based on 10,000,000 draws and one-sided alternatives

Proximity $\longrightarrow$	Border	Divison	$-d(Z_s, Z_t)$	$\exp(-\alpha \cdot d(Z_s, Z_t))$		
				$\alpha = 0.00138$	$\alpha = 0.00276$	$\alpha = 0.00693$
Minimum wage	0.0002	0.0032	0.0087	0.2674	0.0324	0.0033
Log wage	0.0005	0.0239	0.0692	0.0001	< 0.0001	< 0.0001
Education	< 0.0001	0.0314	0.0028	< 0.0001	< 0.0001	< 0.0001
Hours Worked	0.0055	0.8922	0.0950	0.0243	0.0086	0.0182
Weeks Worked	0.0018	0.5155	0.1285	0.0217	0.0533	0.3717

Figure 1: Covariance of Demeaned Log(Earnings) by Distance Between Individuals.

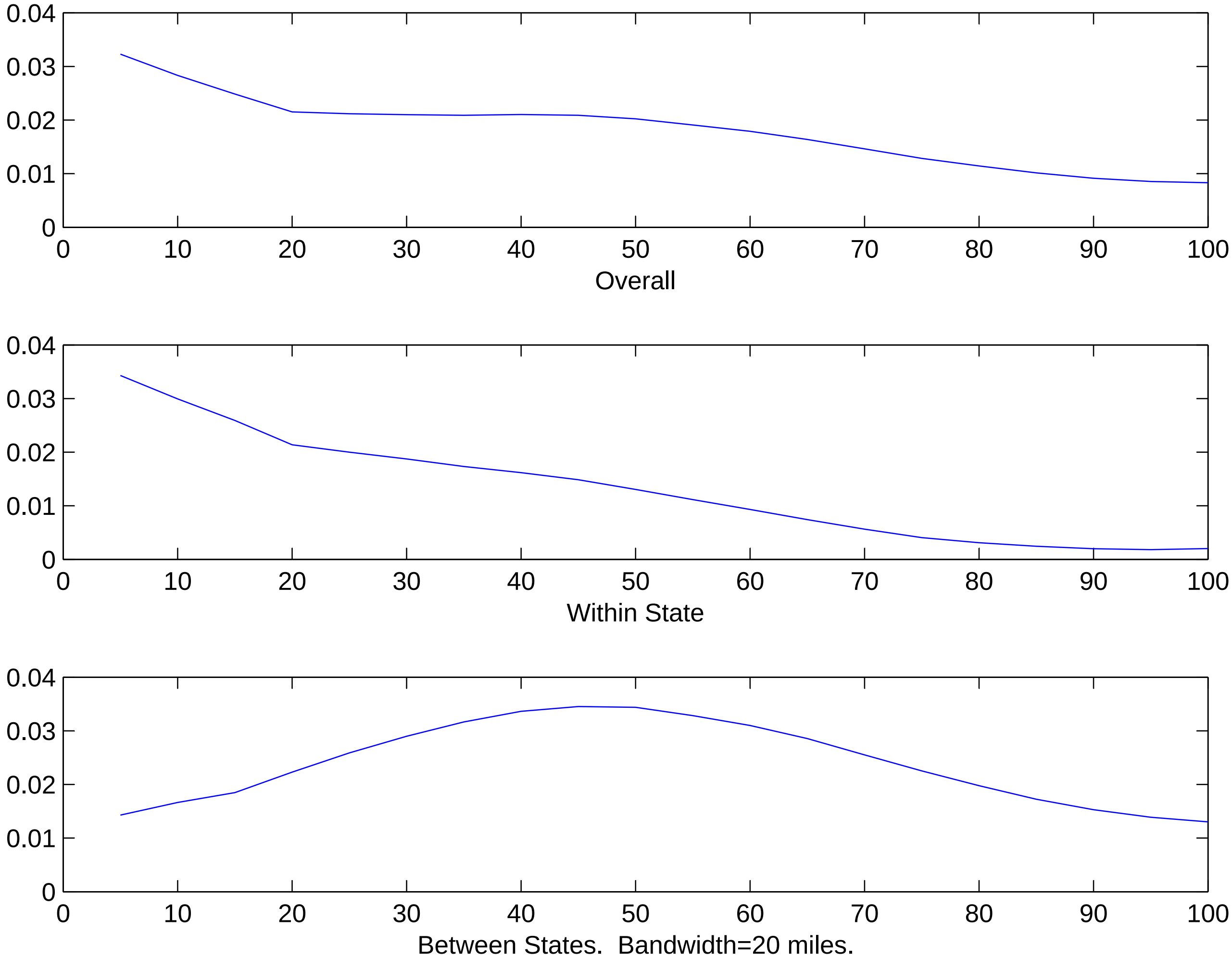

Figure 2: Covariance of Demeaned Log(Earnings) by Distance Between Individuals.