Kelly, Morgan

Working Paper

Persistence, randomization, and spatial noise

UCD Centre for Economic Research Working Paper Series, No. WP21/25

Provided in Cooperation with:
UCD School of Economics, University College Dublin (UCD)

Suggested Citation: Kelly, Morgan (2021) : Persistence, randomization, and spatial noise, UCD Centre for Economic Research Working Paper Series, No. WP21/25, University College Dublin, UCD School of Economics, Dublin

This Version is available at:
https://hdl.handle.net/10419/246498

UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2021

Persistence, Randomization, and Spatial Noise

Morgan Kelly,
University College Dublin

WP21/25

November 2021

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Persistence, Randomization, and Spatial Noise

Morgan Kelly*

Abstract

Historical persistence studies and other regressions using spatial data commonly return severely inflated t statistics, and different standard error estimates that attempt to correct for this vary so widely as to be as to be of limited use in practice. This paper proposes a simple randomization inference procedure where the significance level of an explanatory variable is measured by its ability to outperform synthetic noise with the same estimated spatial structure. Spatial noise, in other words, acts as a treatment randomization in an artificial experiment based on correlated observational data. Examining twenty persistence studies, few perform substantially above the level of spatial noise.

1 Introduction

It is only necessary to notice the number of maps that now appear in leading journals to recognize how popular spatial data have become among economists, notably in the area of historical persistence. These studies find that many modern outcomes such as income or social attitudes are strongly correlated with the characteristics of the same places in the more or less distant past, often centuries or millennia previously. Notable examples include how the mortality of European colonists determines the quality of modern institutions; how the slave trade re-

*University College Dublin and CEPR.

tards modern African development; how colonial boundaries still predict poverty in Peru, and how medieval pogroms prefigured Nazi zealotry.[1]

Naturally, such findings are open to various charges of *p* hacking, of publication bias, of answers in search of questions, of scepticism about monocausal and largely atheoretical explanations of complex phenomena, about the mechanisms driving persistence, and so on. However, all of these objections crumble into irrelevance in the face of one blunt fact: the unusual explanatory power of these persistence variables. While a judicious choice of variables or time periods might coax a *t* statistic past 1.96, there would appear to be no way that the *t* statistics of three, four, or even larger that appear routinely in this literature could be the result of massaging regressions, no matter how assiduously. Such persistence results must instead reflect the workings of the deep structural characteristics that underlie historical processes: the enduring legacies of the past.

However, persistence studies are spatial regressions—being based on observations for the same places at different times—and spatial regressions bring us into the domain of Tobler's First Law of Geography: "Everything is related to everything else, but near things are more related than distant things." Spatial data, in other words, tend to be strongly autocorrelated. If you take some towns dotted across a landscape and represent their incomes by elevation on a map, you will usually find a gently rolling landscape where rich areas border on rich areas, and poor areas on poor ones.

Next take some unrelated variable where neighbour again resembles neighbour, say trials for heresy in the middle ages, leading to another rolling landscape. If you regress one variable on the other, peaks in one landscape will often tend either to correspond to peaks in the other, giving positive *t* statistics, or to hollows leading to negative ones.

These spurious *t* statistics are not trivial. The next Section will show regressions where both variables are spatial noise generated with an empirically realistic correlation structure. These regressions return *t* values beyond ± 3.8 (nominal sig-

[1]These are, in turn, Acemoglu, Johnson and Robinson (2001), Nunn (2008), Dell (2010), and Voigtländer and Voth (2012).

nificance $p = 0.0002$) twenty per cent of the time; and to reach five per cent significance requires a t of 6 (nominal $p = 2 \times 10^{-9}$). The high t statistics of persistence regressions may actually be a cause of concern rather than of reassurance.

This may all sound like a lot of fuss about nothing. Inflated t statistics simply mean that standard errors have been underestimated, and there is no shortage of spatial standard error corrections to deal with this. The commonest is simply to cluster residuals at some low level. However for spatial data such clustered standard errors are typically inconsistent and to be avoided as such.

Standard errors that are consistent are easily estimated by methods that include Conley's (1999) kernel weighting; large clusters (Ibragimov and Müller 2010, Bester, Conley and Hansen 2011, and Canay, Romano and Shaikh 2017: IM, BCH, and CRS from now on); or principal components (Müller and Watson 2021, MW). However, as we will see below, these corrections, although all consistent, return estimates that do not differ widely so much as wildly. Kernel standard errors are usually not much larger than uncorrected ones, whereas large clustered ones are often several times as large, with principal components somewhere in between.

There is not only wide variation between consistent estimators but within them. A typical pattern we will find below in regressions on real data is that when Conley returns a t of 3, IM and CRS will give estimates between 0.5 and 2 depending on the number of clusters imposed by the researcher; while MW will return a t of 2.2 and 1.6 depending on whether an average correlation of 0.01 or 0.05 is assumed. The researcher can thus be quite confident that the correct t lies between 0.5 and 3.

Given the uncertainties that surround one size fits all, asymptotic inference based on standard errors, how are we to evaluate spatial regressions? One possible approach, that is tailored to individual samples and places no reliance on idealized population distributions, is Fisher randomization.

The idea of randomization is simple. Given a vector of assigned treatments X and observed outcomes Y for each subject, we have a statistic $T(Y, X)$ giving the treatment effect. If we now shuffle the treatment records into a vector $\tilde{X}$ we can again compute the virtual effect $T\left(Y, \tilde{X}\right)$, and we can keep on shuffling like this to come up with a distribution of these effects. Under the null hypothesis that the

treatment is useless, any of these permutations should be just as effective as the real treatment, so the fraction of permutations where the actual treatment had a more extreme value gives its significance level. This significance level is exact, being based entirely on the data at hand. It does not rely on asymptotic distributions or standard errors which, we know, may be problematic for spatial data.

However, if we simply charge ahead and try to shuffle spatially correlated data we hit an immediate barrier in the form of exchangeability: the requirement that the joint distribution of the data be unchanged by permutation. Exchangeability clearly cannot hold for spatially correlated observations: taking per capita income for instance, if we reassign Mali's income to Argentina and Switzerland's to Brazil, we have shredded the very spatial structure that concerns us. So how are we to proceed?

We saw earlier how regressions of one spatial noise series on another often generate spuriously strong significance. The approach here is to turn this nuisance to our advantage. Rather than permuting the explanatory variable X, what I propose instead is to replace it with drawings of synthetic noise from a distribution that has the same estimated spatial structure. (To be exact, with noise that has the same spatial structure as the component of X that is orthogonal to any control variables Z: otherwise we violate the ancillary principle.)

Because they are drawings from a common spatial distribution, these noise simulations are exchangeable by construction. The significance of X is measured by the fraction of simulations where it outperforms this noise. Spatial noise, in other words, acts as a treatment randomization in an artificial experiment based on correlated observational data.

For this exercise to be meaningful we need our simulated noise to approximate the spatial structure of the explanatory variable as exactly as possible. Fortunately, this estimation problem has already been solved for us, and is in fact the defining problem of geostatistics. Deriving a maximum likelihood estimate of the spatial parameters of an explanatory variable comes down to a textbook exercise in spatial statistics, an operation known as kriging.

The proposed randomization procedure then is extremely simple and involves nothing harder than estimating the spatial parameters of X and simulating noise with the same structure. The significance level of X is the fraction of regressions where the replacement synthetic noise has a more extreme t statistic than the original estimate.[2] In contrast to the fragility of cluster and principal component results, randomized estimates are robust: making quite substantial variations in the spatial parameters around their maximum likelihood estimates leads to only moderate and predictable changes in estimated significance.

I illustrate this randomization using regressions taken from twenty studies of historical persistence that have appeared in the *American Economic Review, Econometrica, Journal of Political Economy,* and *Quarterly Journal of Economics*. In each case I reproduce the regression that used the full set of controls in the original paper. The sole concern of this analysis is with comparing randomized and asymptotic estimates of significance. It is not concerned with issues of data construction. It is not concerned with the plausibility of the mechanism that is said to drive the claimed persistence, or possible alternative explanations, or with the quality of the underlying historical scholarship (although in most cases this is extremely high, especially in regional studies). Above all, and this cannot be emphasized too strongly, it is not concerned with somehow "validating" or "disproving" the findings of particular studies. Although I do identify the studies in the graphs and tables, at no stage do I make reference to any study in the body of the text.

I find that the null randomization distributions of t statistics in most cases have far heavier tails than the normal distributions assumed in the original studies. As a consequence, randomized significance levels are frequently several orders of magnitude larger than the asymptotic ones reported. As mentioned earlier, heavy outliers and strong directional trends are rarely absent from spatial data, and I also examine what happens when we include systematic controls for these.

The contribution of this paper is in bringing together two previously unrelated literatures in randomization inference and geostatistics. Randomization inference

[2] A tutorial giving a one line R command to do this can be found at `https://rpubs.com/Morgan_Kelly`

goes back to Fisher (1935), and Imbens and Rubin (2015, 57–81) provide a recent overview. Its use in economics has been strongly advocated by Imbens and Wooldridge (2009) and Athey and Imbens (2017) and, following Young's (2019) demonstration of its robustness to outliers, randomization has become increasingly popular in experimental work. The systematic analysis of randomization in regressions with observational data starts with Freedman and Lane (1983). Geostatistics originated from the need in mineral prospecting and meteorology to interpolate between observations taken at points (ore concentrations in bore holes or pressure readings at weather stations) to draw maps across a region, an exercise that comes down to estimating the pattern of correlation between these points. A comprehensive survey is Gelfand et al. (2010).

The rest of the paper is as follows. The next Section shows how realistic spatial noise regressions can return spuriously high t statistics. Section 3 analyses the performance of asymptotic standard error corrections. Randomization inference in regressions is outlined in Section 4 and spatial parameter estimation in Section 5. The procedure is illustrated for some persistence studies in Section 6.

2 Fitting Spatial Noise

As we noted at the beginning, regressions of one spatial noise series on another one commonly return the large t statistics that are a hallmark of persistence and other spatial studies. This is illustrated in Figure 1. The parameters used to generate the simulations are the maximum likelihood estimates of the spatial structure of the orthogonalized explanatory variable from Row 4 in Table 1 below, and the values are taken at the white dots which correspond to the locations of historical African tribes.

If we regress the values from the "Modern Outcome" map on those from the "Historical Variable" one, we get a $t = -3.8$ with nominal significance $p = 0.0002$. The randomized significance is, in fact, 0.2 (200 of 1000 simulations had t values at least as extreme); and 5 per cent significance requires a t of 6 with nominal $p = 2 \times 10^{-9}$.

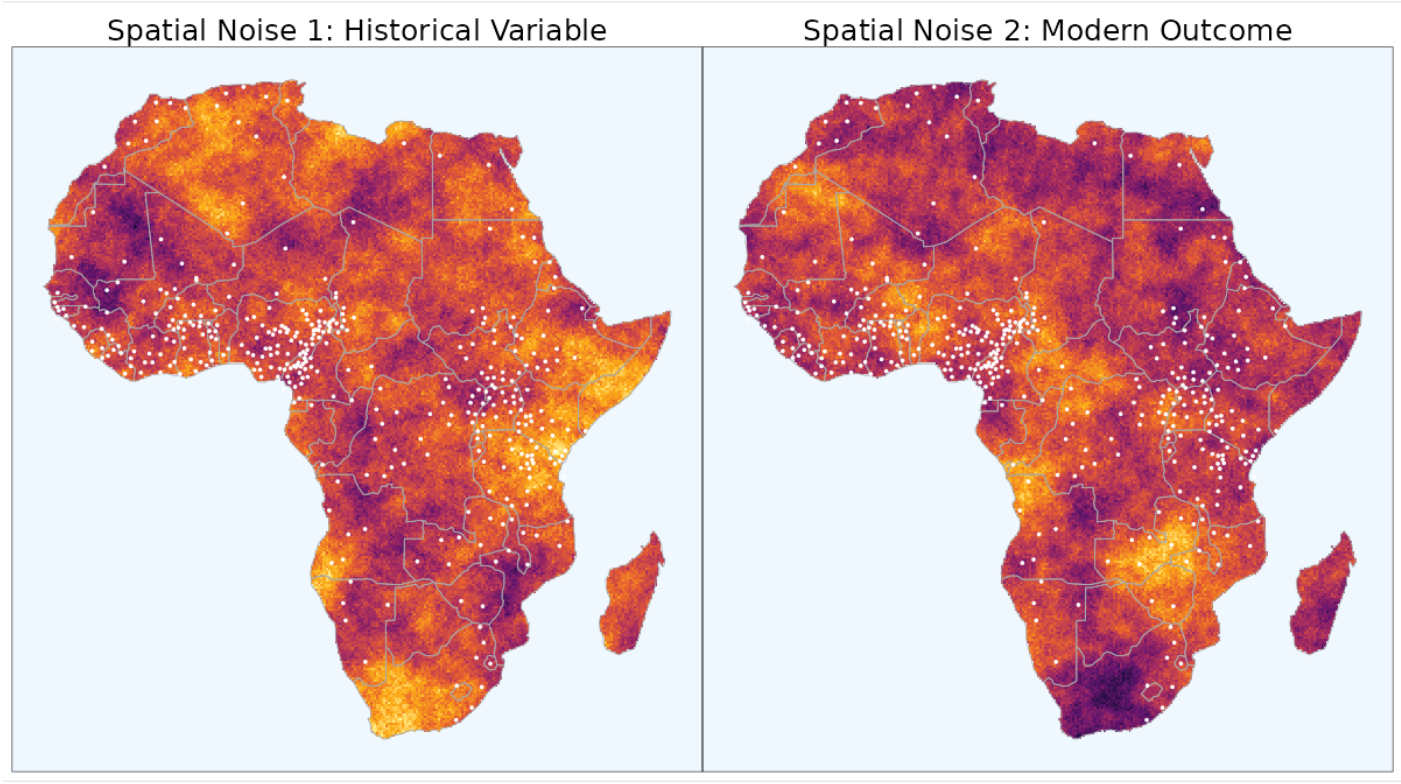


Figure 1: Regressions of one spatial noise series on another can appear highly significant. If we regress values at the white dots (the locations of historical African tribes) on each other we get $t = -3.8$ with nominal significance $p = 0.0002$. The randomized significance level is 0.2 (200 of 1000 simulations had t values at least as extreme); and 5 per cent significance requires a t of 6 with nominal $p = 2 \times 10^{-9}$. The simulations were generated using the maximum likelihood estimates of spatial parameters from the first African study in Table 1 below: $\kappa = 0.5$, $2\theta = 750$, $\rho = 0.99$.

This implies that uncorrected standard errors are one third as large as they should be. Only one ninth of the observations are contributing to improving the precision of the coefficient estimates, while the remainder resemble their neighbours too closely to be of much use.

Statistics is the exercise of extracting structure from data. As Figure 1 illustrates, because spatial noise contains considerable apparent structure it is easy to fit spurious relationships and and mistake them for deep historical truths. We

now examine how well population based inference serves as a check against such a possibility.

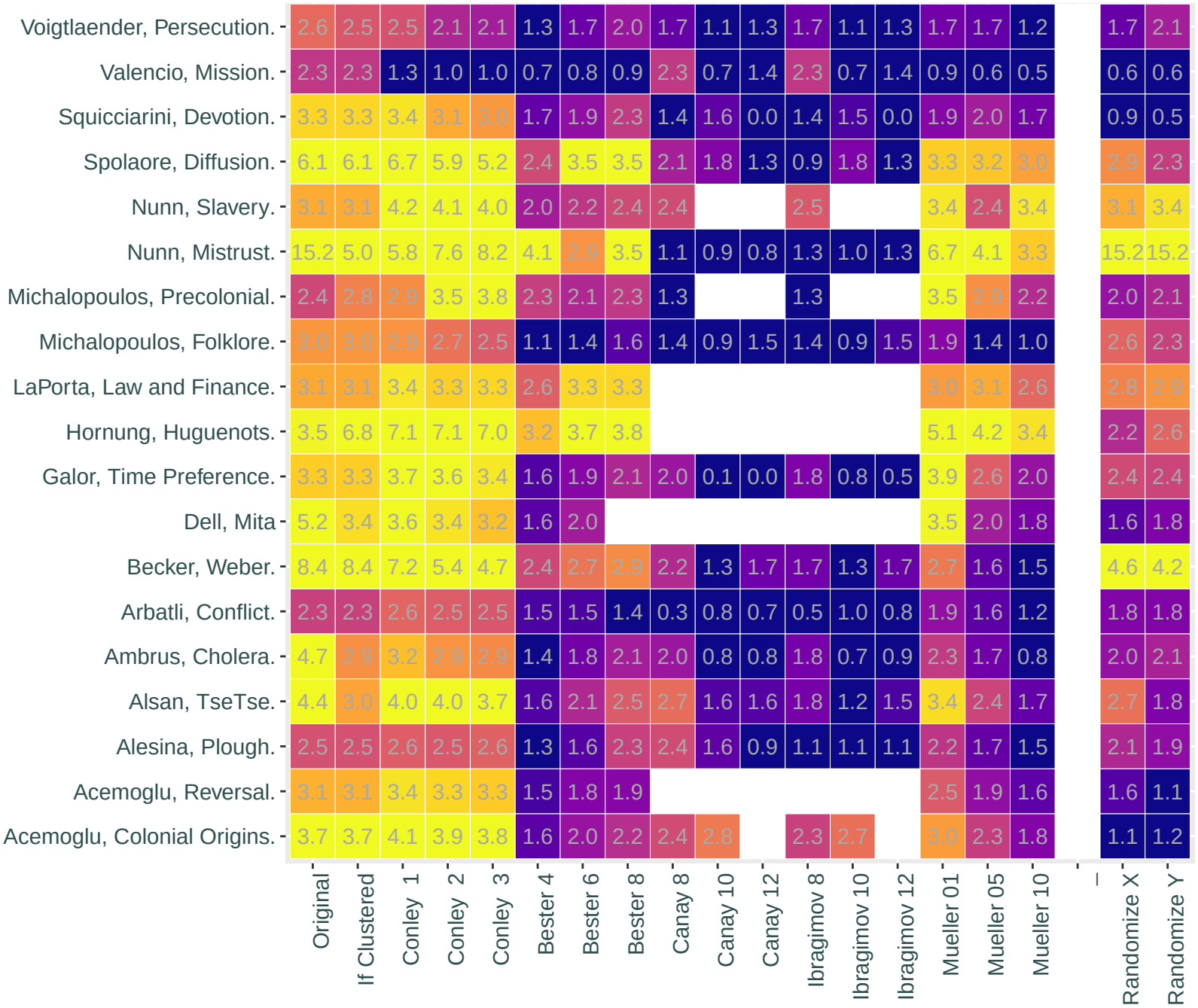
3 Consistent Spatial Standard Errors

When it comes to spatial regressions, the commonest way that researchers try to control for potentially inflated t values is to cluster standard errors at some arbitrary level, typically one administrative level above the original observations. However, for clustered standard errors to be consistent requires that residuals be uncorrelated between clusters, which will usually not be true for spatial data: think of US towns on opposite sides of a state line. Ignoring this requirement leads to the distortions analyzed by Barios et al. (2012).

Turning to consistent estimators, the most popular is the spatial heteroskedasticity and autocorrelation consistent procedure, pioneered by Conley (1999), where the researcher picks a weighting kernel (usually triangular or rectangular) describing how the correlation between regression residuals falls with distance. Although standard errors estimated with many small clusters are typically inconsistent, Ibragimov and Müller (2010), Bester, Conley and Hansen (2011) and Canay, Romano and Shaikh (2017) derive estimators that are consistent under mild assumptions by grouping data into G large clusters. The last adjustment I will examine here is the Müller and Watson (2021) procedure of maximizing the estimated variance of residuals by averaging the first few principal components of a covariance matrix.

Figure 2 shows the t statistics for the persistence studies using different kernel and cluster adjustments (for corrections that return significance levels, these are the pseudo-t values corresponding to estimated significance: a significance of 0.05 is presented as a t of 2). The first column gives unadjusted standard errors, and the next one gives values after naive clustering if that was applied: of the seven studies that used clustering, in two cases t values increased.

Figure 2: Consistent standard error estimates vary widely both between and within estimators. The number in each square is the value of the (pseudo-) t statistic. Conley cutoffs 1, 2 and 3 are equal to 500, 1000 and 1500 km in global and African studies; and 50, 100 and 150 in the others. BCH, IM and CRS are followed by number of clusters, and MW by the average correlation between sites. Randomized estimates are included for comparison.



Conley standard errors were constructed using Bartlett kernels with three cutoffs (uniform kernels gave similar values). For global and African studies the cutoffs are 500, 1000, and 1500 km; for European and Latin American ones, except for two smaller scale ones, they are 50, 100, and 150 km. As a glance at the colours of the heatmap confirms, Conley corrections seldom cause drastic changes in estimates compared with the unadjusted or clustered standard errors.

The opposite holds for large cluster estimates. Clusters were constructed by *k*-means clustering (except in three cases with binary explanatory variables and/or multiple observations at each location, where north-south stripes were used to create variation within clusters). Instances where there were too few spatial points for CRS and IM are blank. It can be seen that BCH t values are somewhat lower than Conley ones and rise quickly and systematically with the number of clusters: t statistics usually rise between 20 and 50 per cent if eight clusters are assumed instead of four.

The pseudo-t statistics of CRS and IM are smaller still, with IM generally below CRS. Both IM and especially CRS are highly volatile, with eight assumed clusters giving the highest t, and larger clusters often being a fraction of this. Last are MW pseudo-t values, with assumed average values of correlation between observations of 0.01, 0.05, and 0.1. It can be seen that these t statistics are noticeably large than CRS and IM, and that they tend to change substantially as the assumed correlation strength varies. If an average correlation of 0.05 is used instead of 0.01, pseudo-t values generally fall by 20–30 per cent; and by 30–50 per cent if a correlation of 0.1 is imposed.

For concreteness, take the study in the third row from the bottom which has an uncorrected t of 2.5, that is unchanged if Conley kernels with cutoffs of 500 to 1500 km are used. Using BCH with four or eight clusters changes this to 1.3 or 2.3; whereas we get a t of 2.4, or 0.9 depending on whether we impose eight, ten or twelve CRS clusters, with IM values being all around 1.1. Finally, MW gives values of 2.2, 1.7, and 1.5 depending on which level of average correlation we assume. We can thus be fairly confident that the t statistic lies in the range of 0.9 to 2.5.

By analogy with time series, it might be imagined that the optimal estimator for all circumstances could be chosen on the basis of Monte Carlo simulations. However, spatial data are inherently messy with each set of data having its own unique layout of observations, often with clumps of substantial outliers. This can be seen in the simulations in Appendix 1 for the German towns in the top row of Figure 2.

Although the trends in the simulations for most estimators resemble the observed results, for CRS and IM the pattern of significance runs in the opposite direction, with eight clusters predicted to generate the smallest t values, rather than the largest ones in the real data. For spatial data, asymptotic inference based on standard errors appears to be inherently unreliable, and we now consider an alternative possible approach through randomization.

4 Randomization Inference in Regressions with Exchangeable Observations

The systematic study of randomization inference in regressions goes back to Freedman and Lane (1983), with developments by Kennedy and Cade (1996), Anderson and Robinson (2001) and others, surveyed by Winkle et al. (2014). We suppose for now that the observations are exchangeable so the that joint distribution of the data is unchanged under permutation. Sufficient conditions for this are that the observations are independent, or are Gaussian with identical variances and covariances (Good, 2006, 268).

We consider the regression

$$Y = \beta X + Z\gamma + \epsilon \tag{1}$$

where Y, X and ϵ are vectors of length n, β is a scalar, Z is an $m \times n$ matrix and γ is a vector of length m. The null hypothesis is $H_0 : \beta = 0$.

In the case where there are no nuisance covariates Z it is possible to permute either Y or X, and to base the test statistic on t statistics or correlation coefficients. However, when there are other variables Z that are correlated with X (something that cannot happen in properly randomized experiments, but will usually be the case with observational data), simply permuting the raw X or Y variables is not permissible.

If X is shuffled, we are changing the relationship between X and Z, causing $\hat{\gamma}$ to vary each time despite the null hypothesis that X has no effect, and so violating the ancillary principle (Anderson and Legendre, 1999). If we permute Y we are testing the far stronger null hypothesis that Y is unrelated to both X and Z, in other words that both β and γ are zero. As a result, tests are based on permuting the components of Y or X that are orthogonal to Z.

The Freedman and Lane (1983) procedure involves estimating the reduced model under the null hypothesis $Y = Z\gamma + \epsilon$. This gives the residuals $\hat{\epsilon}_{Y\perp Z} = Y - Z\hat{\gamma}$ and the fitted value $\hat{Y} = Z\hat{\gamma}$. The permuted value of Y is $\tilde{Y} = \hat{Y} + P\hat{\epsilon}_{Y\perp Z}$ where P is the permutation matrix that reorders $\hat{\epsilon}_{Y\perp Z}$. We then repeatedly carry out the permutation regressions

$$\tilde{Y} = \beta X + Z\gamma + \epsilon \tag{2}$$

and derive the significance level of the null hypothesis as the fraction of permutations where the absolute t value is higher than in estimated absolute value from (1).

The complementary procedure (which Winkle et al. 2014 call the Smith method) is to permute the component of X that is orthogonal to Z. If we have $X = Z\delta + \nu$ then we define $\hat{\nu}_{X\perp Z} = X - Z\hat{\delta}$. The permuted value of X is $\tilde{X} = P\hat{\nu}_{X\perp Z}$ in the regression

$$Y = \beta\tilde{X} + Z\gamma + \epsilon. \tag{3}$$

The validity of randomization rests on the exchangeability of $\hat{\epsilon}_{Y\perp Z}$ or $\hat{\nu}_{X\perp Z}$: the joint distribution of the variables remains the same under permutation so all shuffles are equally likely. A relaxation of this requirement is what Good (2006,

128) calls weak exchangeability. Instead of the joint distribution being invariant under all rearrangements of subscripts, it is only invariant under a subset of them which are used for the permutation test.

This idea underlies the clustered spatial permutations used in economics by Barios et al. (2012) and in neuroscience by Winkle et al. (2014): tiles such as American states are fixed, but observations within each tile, such as workers, can be permuted. However, most observational data lack this hierarchical structure so a different approach is needed.

5 Randomization with Spatial Noise

The approach to generating exchangeable observations for randomization that I propose is to generate synthetic noise that has the same spatial correlation parameters as $\hat{\epsilon}_{Y\perp Z}$ or $\hat{\nu}_{X\perp Z}$. We will let $\tilde{V}_{Y\perp Z}$ denote drawings of synthetic noise with the same correlation structure (as well as mean and standard deviation) as $\hat{\epsilon}_{Y\perp Z}$ and similarly drawings $\tilde{V}_{X\perp Z}$ will have the same correlation structure as $\hat{\nu}_{X\perp Z}$. By construction, these simulations will satisfy exchangeability. We are then looking at the explanatory power of X in the artificial regression

$$\left(\hat{Y} + \tilde{V}_{Y\perp Z}\right) = \beta X + Z\gamma + \epsilon \tag{4}$$

and $\tilde{V}_{X\perp Z}$ in

$$Y = \beta \tilde{V}_{X\perp Z} + Z\gamma + \epsilon. \tag{5}$$

Again, the significance level is the fraction of these artificial regressions where the t statistic has a more extreme value than the one originally estimated in (1). To apply this randomization we need to generate synthetic noise variables $\tilde{V}_{Y\perp Z}$ and $\tilde{V}_{X\perp Z}$ that have the same spatial distribution as $\hat{\epsilon}_{Y\perp Z}$ and $\hat{\nu}_{X\perp Z}$.

It is worth noting that the null hypothesis being tested here differs from the exchangeable case. Whereas with exchangeability we are directly testing the null that $\beta = 0$, here the null is that the explanatory variable X is spatial noise with a

given parametric distribution. As these parameters change, the precise null being tested also changes. However, we will see below that significance levels are robust to quite substantial changes in these parameters.

5.1 Estimation of Spatial Structure

Estimating the spatial distribution of variables is a textbook exercise in geostatistics. Let V denote either $\hat{\epsilon}_{Y\perp Z}$ or $\hat{\nu}_{X\perp Z}$. We have a vector of observations of V at n sites s. The additive spatial model to be estimated is

$$V(\mathbf{s}) = T(\mathbf{s}) + g(\mathbf{s}) + \eta(\mathbf{s}). \tag{6}$$

$T(.)$ is a deterministic spatial trend surface that acts as a mean function, $g(.) \sim N(0, \rho K)$ is a spatial correlation process, and $\eta(.) \sim N(0, \sigma^2 I)$ is idiosyncratic noise (Zimmerman and Stein, 2010). In other words, observations are decomposed into low frequency trends, local correlations, and idiosyncratic noise.

The matrix K with ij-th element $K(s_i, s_j)$ gives the correlation between sites s_i, s_j. It follows that the observations V are Gaussian

$$V(\mathbf{s}) \sim N\left(T(\mathbf{s}), \rho K(\mathbf{s}) + \sigma^2 I\right). \tag{7}$$

The magnitude of ρ relative to σ^2 gives the degree of systematic spatial structure relative to idiosyncratic noise in the observations. We suppose for now that the data have been detrended so $T(s) = 0$.

To estimate the covariance matrix of V requires that a kernel function $K(.)$ be specified. It will be assumed to be isotropic so that correlation between two points depends only on the distance h between them. Because of its adaptable functional form, robust empirical performance, and the fact that it is guaranteed to be positive definite, the workhorse kernel of spatial statistics is the Matérn function.

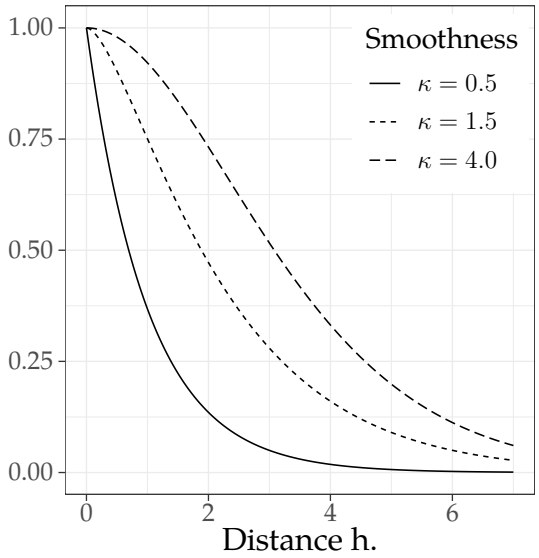


Figure 3: The flexible form of the Matérn function (drawn with range $\theta = 1$) allows it to fit a wide variety of spatial kernels. For the studies analysed here, correlation among residuals tends to fall off exponentially with distance, corresponding to $\kappa = 0.5$.

Correlation between sites s_i, s_j at distance h apart is

$$M\left(h;\theta,\kappa\right) = \frac{2^{1-\kappa}}{\Gamma\left(\kappa\right)}\left(\frac{h}{\theta}\right)^{\kappa} B_{\kappa}\left(\frac{h}{\theta}\right) \quad (\kappa > 0, \theta > 0) \tag{8}$$

where Γ is a gamma function and B_κ is a Bessel function of the second kind (Gneiting and Gutthorp, 2010). The parameter θ is a range parameter controlling how fast correlation decays with distance, and κ is a smoothness parameter. For $\kappa = \frac{1}{2}$, correlation decays exponentially so $M\left(h\right) = \exp\left(-h/\theta\right)$, and as $\kappa \to \infty$, M becomes Gaussian. The flexibility of the Matérn function is illustrated in Figure 3 where range θ is set to 1 and smoothness κ takes on values from 0.5 to 4.

We have then a matrix giving the covariance between observations

$$\Sigma\left(s_i, s_j\right) = \rho\, M\left(h;\theta,\kappa\right) + \sigma^2 \mathbf{1}_{ij} \tag{9}$$

The parameters θ, ρ, κ and σ^2 can be estimated by maximum likelihood or cross validation using standard geostatistical software giving us an estimated covariance matrix $\hat{\Sigma}$.

For this estimated covariance matrix, define its Cholesky decomposition L as

$$LL^{'} = \hat{\Sigma}. \tag{10}$$

If $\phi \sim N(0, I)$ is an $n \times 1$ vector of Gaussian noise, the synthetic vector

$$\tilde{V} = L\phi \tag{11}$$

has a spatial correlation structure identical to $\hat{\Sigma}$, the estimated covariance of V. These simulations $\tilde{V}$ can thus serve as synthetic exchangeable variables that have the same joint distribution as the variables of interest $\hat{\epsilon}_{Y \perp Z}$ or $\hat{\nu}_{X \perp Z}$. They can thus be used in their place in (4) and (5) to carry out randomizations.

Whereas in randomized experiments, the null hypothesis is that the treatment X has no effect, here the null hypothesis is contingent on the randomization process, in other words the synthetic noise used. Specifically, the null hypothesis is that X has no more explanatory power than spatial noise with a given set of generating parameters.[3]

For concreteness of exposition in the empirical analysis that follows, when estimating the parameters of V we can also estimate the parameters when it is scaled to have a standard deviation of one. The covariance matrix of the scaled data is a correlation matrix where $\rho + \sigma^2 = 1$. This normalized ρ is what will be reported in what follows: a value of one means that all observations lie exactly on the spatial correlation surface, and a value of zero implies no spatial correlation in the data.

[3]The dependence of the null on the randomization mechanism is implicit in randomized experiments. If we are testing an experiment with treatments levels doses of, say, 0, 1 and 2, we could in principle randomize on imaginary doses of 0, 1, 2, and 3 and obtain different significance levels.

6 Historical Persistence

To illustrate this spatial randomization inference, I will examine twenty studies on historical persistence that have appeared in leading journals. A single regression from each paper is analyzed which includes the main explanatory variable of interest along with the largest set of controls that were used: this is usually located towards the right of table 2 or 3. Details of each regression are given in Appendix B.

For each case I generate synthetic versions of the components of the dependent and explanatory variable that are orthogonal to any controls used. A property of the Matérn function is that when two sites are separated by a distance $h = \sqrt{8\kappa}\,\theta$, the correlation between them is 0.14: this distance is commonly called the effective range. The estimated effective range and structure of each variable is given in Table 1 for $\kappa = 0.5$: exponential decay of correlation.[4]

The R^2 of the spatial trend surface T for each variable is given based on a quadratic trend in longitude and absolute latitude: higher order terms did not affect the results materially. For a well specified regression, there should be no systematic spatial trends remaining in the orthogonalized variables: R^2 should be zero.

The Table also gives the Moran's I statistic for the regression, a standard measure of the degree of spatial correlation in the residuals that is estimated here using five nearest neighbours. It is not reported for two studies that have multiple observations at each location. While large values of the statistic, which is asymptotically normal, reliably indicate that *t* statistics are inflated, small values do not imply that spatial adjustment is unnecessary.

[4]Given that effective range is $\sqrt{8\kappa}\,\theta$, range θ and smoothness κ cannot be reliably estimated together. The standard procedure is to set κ in increments of 0.5 running from 0.5 to 2.5 and to choose the one that returns the highest likelihood. For the regressions here, using values of κ running from 0.5 to 1.5 return effectively identical likelihoods and randomized significance levels.

Table 1: Maximum likelihood estimates of the spatial structure of orthogonalized explanatory and dependent variables in persistence studies. These parameters are used to generate synthetic noise for randomization inference in Table 2 and Figure 4.

		Explanatory $X_{\perp Z}$			Dependent $Y_{\perp Z}$		
	Moran I	Trend R^2	Range $\sqrt{8\kappa}\theta$	Struct. ρ	Trend R^2	Range $\sqrt{8\kappa}\theta$	Struct. ρ
Global							
Acemoglu, Colonial.	2.19	0.43	4500	0.52	0.37	1000	0.14
Acemoglu, Reversal.	2.57	0.13	9800	0.99	0.33	800	0.86
Alesina, Plough.	6.98	0.23	2100	0.92	0.05	800	0.96
Arbatli, Conflict.	1.87	0.08	6100	0.94	0.01	900	0.48
Ashraf, Africa.	5.65	0.05	12000	0.95	0.01	16	0.00
Galor, Time.	1.48	0.08	3400	0.98	0.09	2800	0.57
La Porta, Law.	0.87	0.11	9400	0.66	0.08	1000	0.96
Michalopoulos, Folklore.	6.56	0.01	400	0.96	0.02	2200	0.65
Spolaore, Diffusion.	8.49	0.14	5900	0.88	0.00	2000	0.68
Africa							
Alsan, TseTse.	9.10	0.10	750	0.99	0.01	1250	0.79
Michalopoulos, Precolonial.	7.29	0.00	300	0.72	0.00	500	0.22
Nunn, Mistrust.	.	0.01	400	0.13	0.00	400	0.00
Nunn, Slavery.	-0.15	0.02	900	0.20	0.06	1200	0.06

Spatial structure of the components of the explanatory $X_{\perp Z}$ and dependent variable $Y_{\perp Z}$ that are orthogonal to any control variables Z. Moran's I is a measure of the spatial correlation of residuals. Trend R^2 gives the explanatory power of a regression of the variable on a quadratic in longitude and latitude. Effective range $\sqrt{8\kappa}\theta$ is the distance in kilometres where the correlation between locations of the detrended variable has fallen to 0.14; and structure ρ is its spatial signal to noise ratio. θ and ρ are computed by maximum likelihood with exponential decay of correlation $\kappa = 0.5$, except in cases where effective range does not end in 0 or 5 when $\kappa = 1.5$ was used.

Continued on next page

Table 1: Maximum likelihood estimates of spatial parameters.(cont.)

		Explanatory $X_{\perp Z}$			Dependent $Y_{\perp Z}$		
	Moran I	Trend R^2	Range $\sqrt{8\kappa}\theta$	Struct. ρ	Trend R^2	Range $\sqrt{8\kappa}\theta$	Struct. ρ
Europe and Latin America							
Ambrus, Cholera.	8.51	0.15	44	0.99	0.03	9	0.97
Becker, Weber.	13.48	0.00	60	0.28	0.00	200	0.37
Dell, Mita	.	0.49	45	0.90	0.02	10	0.01
Hornung, Huguenots.	9.16	0.02	60	0.41	0.05	100	0.01
Squicciarini, Devotion.	5.63	0.23	960	0.87	0.49	180	0.78
Valencio, Mission.	22.62	0.33	207	0.99	0.02	120	0.85
Voigtlaender, Persecution.	12.45	0.09	40	0.97	0.13	207	0.57

Spatial structure of the components of the explanatory $X_{\perp Z}$ and dependent variable $Y_{\perp Z}$ that are orthogonal to any control variables Z. Moran's I is a measure of the spatial correlation of residuals. Trend R^2 gives the explanatory power of a regression of the variable on a quadratic in longitude and latitude. Effective range $\sqrt{8\kappa}\theta$ is the distance in kilometres where the correlation between locations of the detrended variable has fallen to 0.14; and structure ρ is its spatial signal to noise ratio. θ and ρ are computed by maximum likelihood with exponential decay of correlation $\kappa = 0.5$, except in cases where effective range does not end in 0 or 5 when $\kappa = 1.5$ was used.

The Moran statistic looks for autocorrelation in residuals which can be weak even when each variable in the regression has a strong spatial structure, resulting in inflated t values. For instance the Moran statistic in the first row of the Table is only 2.2, but the corresponding asymptotic and randomized significance levels are 0.0002 and 0.3 respectively.

Nearly all the variables in Table 1 display a similar spatial structure. Correlation falls off exponentially with distance and the effective range relative to the study area is usually large, as is the spatial structure ρ. The orthogonalized explanatory variables usually show more structure than dependent variables. In several cases, spatial trends have high explanatory power for the persistence variable, suggesting that these variables are acting as proxies for omitted directional trends, something that will be examined below.

Once we have estimates of these spatial parameters we can construct synthetic dependent and explanatory variables with the same joint distribution as the original variables, and use drawings from these in the randomization regressions (4) and (5). The significance level of the explanatory variable is the fraction of these drawings there the original t statistic is larger in absolute value than the artificial ones.

Figure 4 gives the null randomization distributions for the regressions (5) when the X variable is replaced by spatial noise based on 10,000 simulations. The red line denotes the original, unadjusted t value (in one case in row 2 the explanatory variable enters quadratically, so the null hypothesis is that quadratic terms jointly have no explanatory power, giving an F distribution). It can be seen immediately that, with few exceptions, the randomization distribution has heavy tails compared with the assumed asymptotic normal distribution on which population based inference was originally made.

Figure 4: Null randomization distributions of t-statistics derived by replacing the explanatory variable X with noise that has the same spatial structure. Light grey denotes a 95% confidence interval. Red lines are the nominal t or F statistics, and significance P_X is the tail area beyond these.

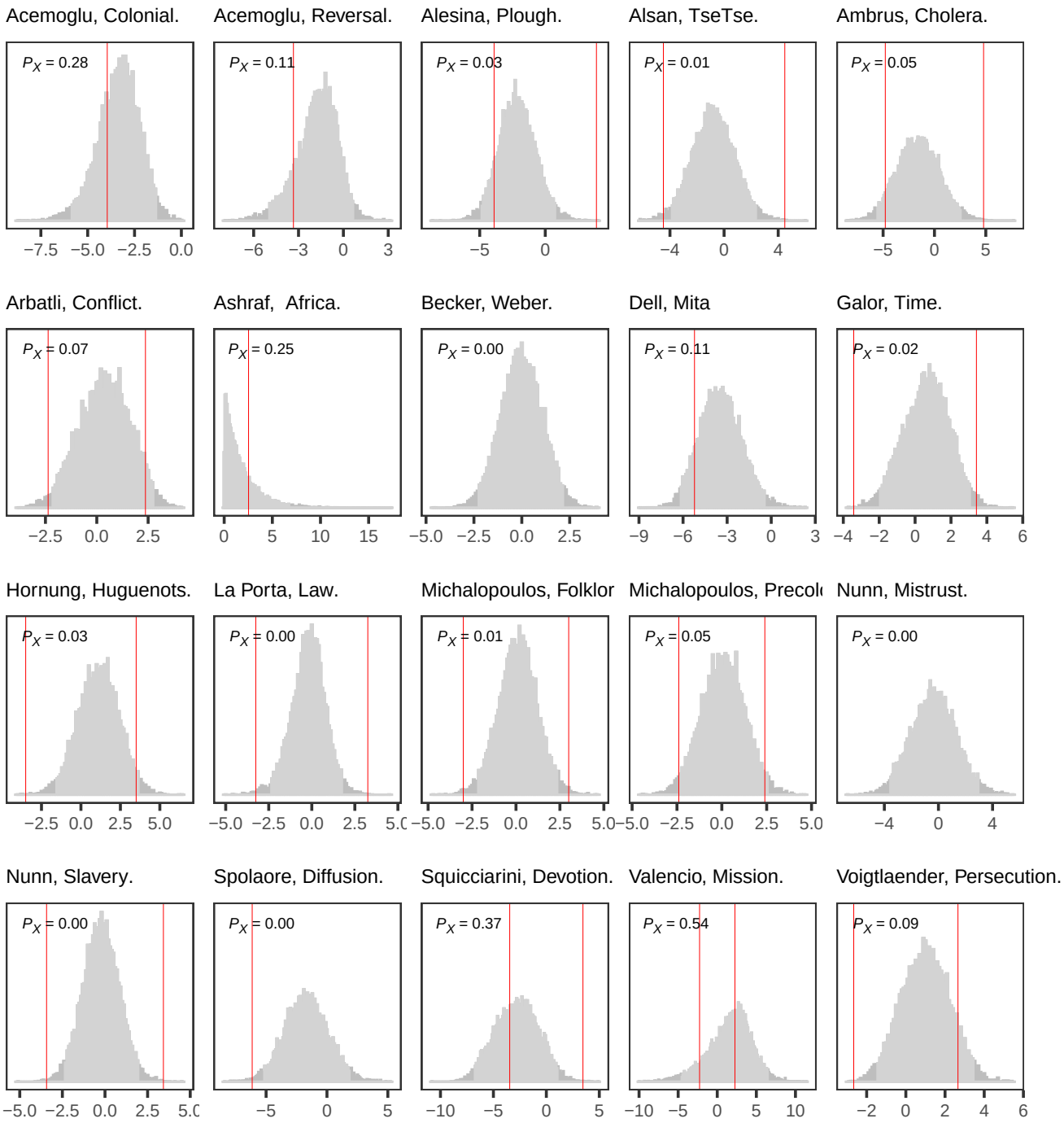


Table 2: Randomized significance levels of persistence variables using orthogonalized explanatory and dependent variables.

		Randomized p	
	Asymptotic p	$X_{\perp Z}$	$Y_{\perp Z}$
Acemoglu, Colonial.	1.9×10^{-4}	0.280	0.239
Acemoglu, Reversal.	2.1×10^{-3}	0.114	0.252
Alesina, Plough.	1.5×10^{-4}	0.033	0.055
Alsan, TseTse.	9.7×10^{-6}	0.007	0.070
Ambrus, Cholera.	2.5×10^{-6}	0.051	0.037
Arbatli, Conflict.	2.0×10^{-2}	0.066	0.067
Ashraf, Africa.	8.4×10^{-2}	0.246	0.094
Becker, Weber.	1.1×10^{-6}	0.000	0.000
Dell, Mita	2.0×10^{-7}	0.107	0.069
Galor, Time.	1.0×10^{-3}	0.016	0.018
Hornung, Huguenots.	5.0×10^{-4}	0.031	0.009
La Porta, Law.	2.2×10^{-3}	0.005	0.003
Michalopoulos, Folklore.	3.2×10^{-3}	0.009	0.020
Michalopoulos, Precolonial.	1.7×10^{-2}	0.050	0.034
Nunn, Mistrust.	1.9×10^{-65}	0.000	0.000
Nunn, Slavery.	1.8×10^{-3}	0.002	0.001
Spolaore, Diffusion.	9.6×10^{-9}	0.004	0.021
Squicciarini, Devotion.	9.1×10^{-4}	0.374	0.627
Valencio, Mission.	2.4×10^{-2}	0.538	0.534
Voigtlaender, Persecution.	8.2×10^{-3}	0.092	0.040

Asymptotic and randomized significance levels for persistence regressions. The synthetic noise for the orthogonalized explanatory ($X_{\perp Z}$) variable and dependent ($Y_{\perp Z}$) variables was generated using the maximum likelihood estimates of spatial parameters from Table 1.

In two cases the original t values of 8 and 15 lie, literally, off the chart. In the first case the randomized significance of 6×10^{-6} is the same as the original one but does not appear in the diagram which is based on 10,000 simulations. In the second case, the fact that the estimated spatial surface fails to come remotely close

to matching the spatial pattern of the original data suggests that the reported regression is fitting some heavy outliers that are not being mapped by our kriging estimation. I will return to this below.

Table 2 compares the reported results of the original studies with the significance levels generated with noise based on the explanatory and dependent variables. In most cases the new significance levels differ from the original ones by several orders of magnitude. The two artificial regressions based on the $X_{\perp Z}$ and $Y_{\perp Z}$ variables tend to return similar significance levels, with the simulated explanatory variable usually being somewhat higher. These results are graphed in Figure 5: to be conservative it reports the higher significance level of the two calculated. The changes in significance levels of the persistence studies after randomization are shown in Figure 5

Naturally, in contrast to a designed experiment, there is no such thing here as *the* significance level, any more than there is in a regression where coefficients and standard errors change as covariates are added or removed. The goal of the exercise is to be roughly right rather than precisely wrong. It is increasingly realized in economics that sharp "significant-insignificant" distinctions are unhelpful at best: an elasticity estimate of 0.15 with standard error of 0.025 is "highly significant" whereas one of 2 with a standard error of 1.5 is "insignificant" even though it has a 90 per cent chance of being larger. The randomized significance levels reported here mainly serve as a guide to the likely precision of the coefficient estimates, not necessarily to their importance in any meaningful sense.

6.1 Robustness.

Having seen the fragility of consistent spatial standard error estimates in Figure 3, the natural concern is that the randomization results are equally lacking in robustness, that small changes in the assumed range or structure of the synthetic noise will change estimated significance levels markedly.

To examine the robustness of randomization estimators, Figure 6 shows how significance of three studies changes in responses to changes in the structure and

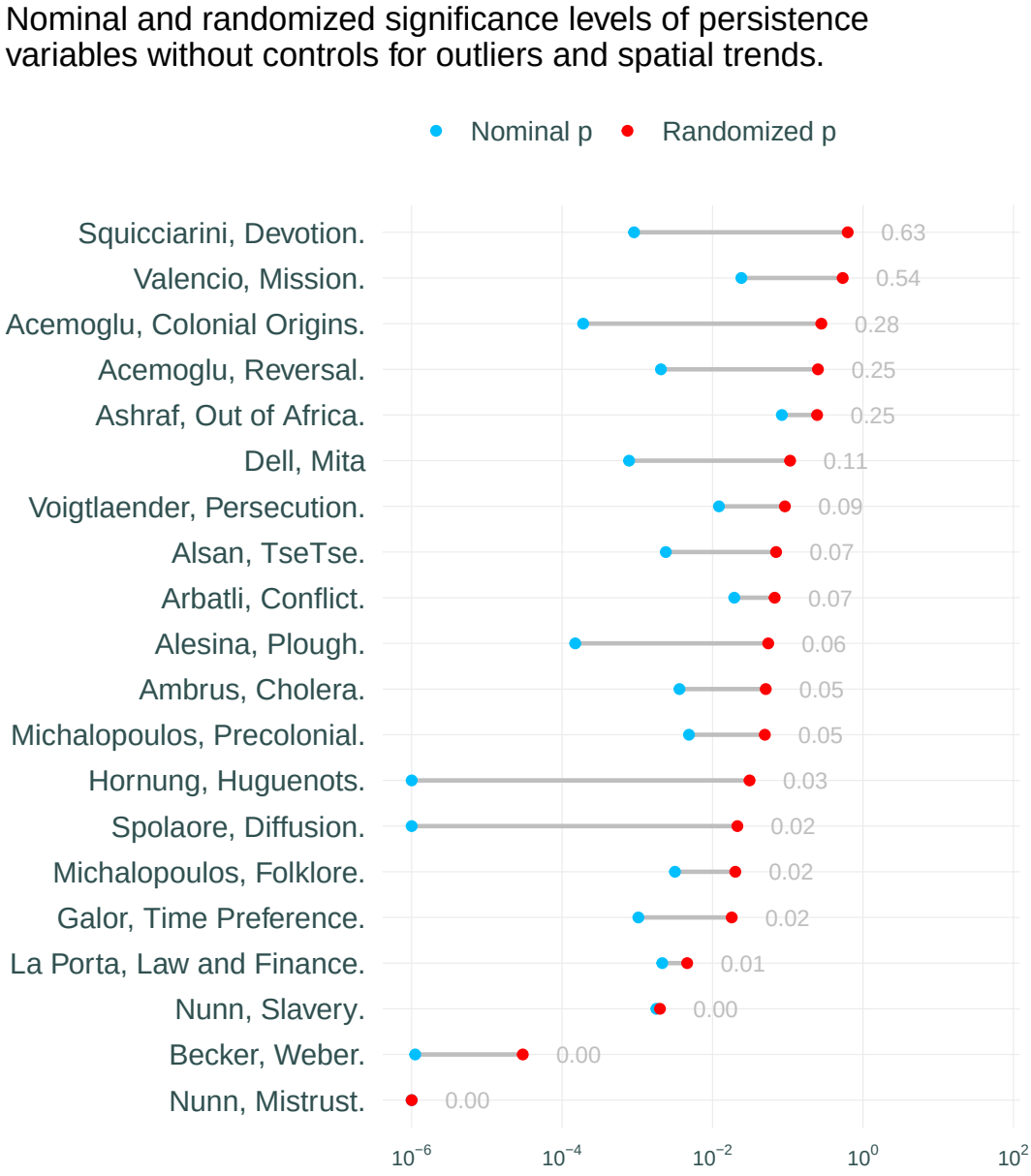


Figure 5: Nominal and randomized significance levels for persistence variables. Logarithmic axis.

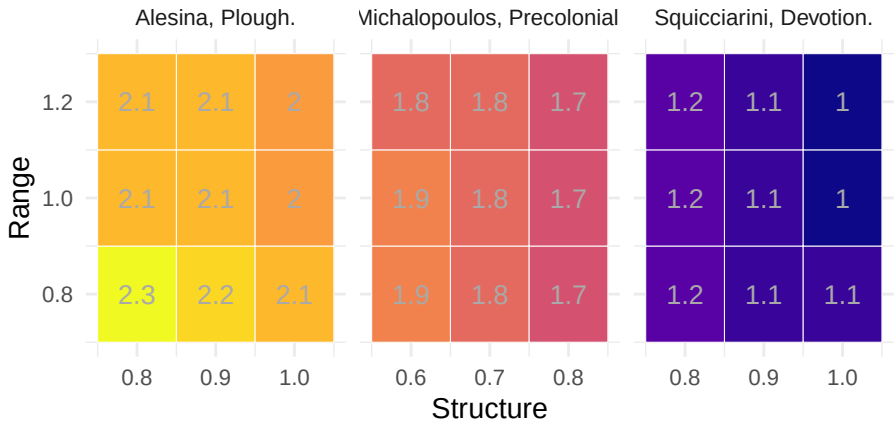


Figure 6: Robustness of spatial randomization estimates to changes in the range and structure of synthetic noise. The centre square gives the original pseudo-*t* value based on the maximum likelihood estimates of the parameters of the explanatory variable reported in Table 1, and the surrounding squares show how this alters if the assumed range changes by 20 per cent, or structure changes by 0.1.

range used to generate the spatial noise underlying the null randomization distribution. These patterns are representative of the other studies. The central square gives the pseudo-*t* statistic for the original randomization based on the maximum likelihood parameters of the explanatory variable in Table 1. The surrounding squares show how this changes if the range is increased or decreased by 20 per cent, or if the structure rises or falls by 0.1.

It can be seen that in all cases, these quite substantial changes alter the result by less than five per cent, with one exception where it changes by 10 per cent in response to a simultaneous change in both parameters. The relative explanatory power of spatial noise increases as its structure or range rises towards the north-east of each box, and the significance level of the explanatory variable increases accordingly.

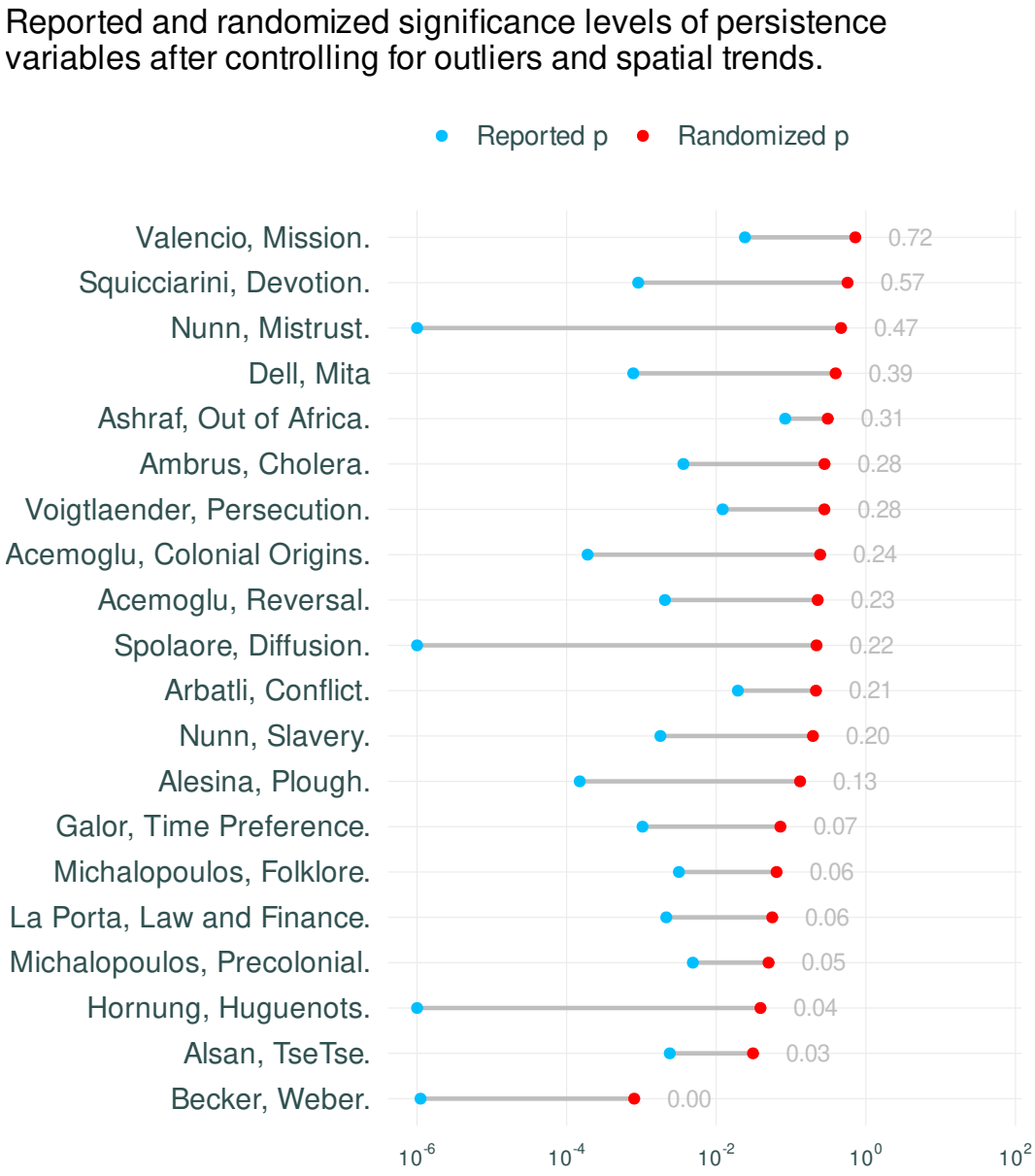


Figure 7: Significance levels originally reported (clustered if applicable), and randomized ones computed after controlling for outliers and spatial trends. Logarithmic axis.

6.2 Controlling for Outliers and Spatial Trends

Besides inflated t statistics, the findings of spatial regressions may be distorted by failing to control for spatial trends or outliers. The high explanatory power of spatial trends for orthogonalized explanatory variables we saw earlier in Table 1 suggest that some results may be artefacts of failing to control for directional trends in the data. At the same time, the results of any cross-sectional regression, spatial or otherwise, may be driven by substantial outliers. This section analyses the robustness of persistence findings when these are systematically controlled for.

The robustness checks for spatial trends are simple, and commonly used in practice. For regressions on a global scale, I add distance from the equator and dummies for World Bank regions. For all other regressions I add a quadratic in longitude and latitude. Regarding outliers, in one case I examine the impact of omitting eight of 157 locations, and in another six of 325, that lie far from the other observations on a scatterplot of the historical and outcome variables. Malaria prevalence is a routine control for African regressions and I add it to the one African regression where it was omitted. To the extent that any proposed check appears unreasonable or unduly *ad hoc* it should, naturally, be ignored.

Figure 7 shows the original, reported significance levels of the studies (after clustering if applied), and the randomized significance levels after the directional and outlier controls were added.

7 Extensions and Conclusions

This paper began with the observation that spatial regressions often return inflated t values, and that standard error adjustments that try to correct for this return widely different estimates. Given the uncertainties around asymptotic estimates, this paper considered an approach to inference tailored to each sample, Fisher randomization. The approach was to circumvent the problem of exchangeability by basing inference on spatial noise that has the same estimated structure as the variable of interest. The parameters of this spatial noise are straightforwardly

estimated by kriging the original variable and estimated significance is robust to quite large changes in their values.

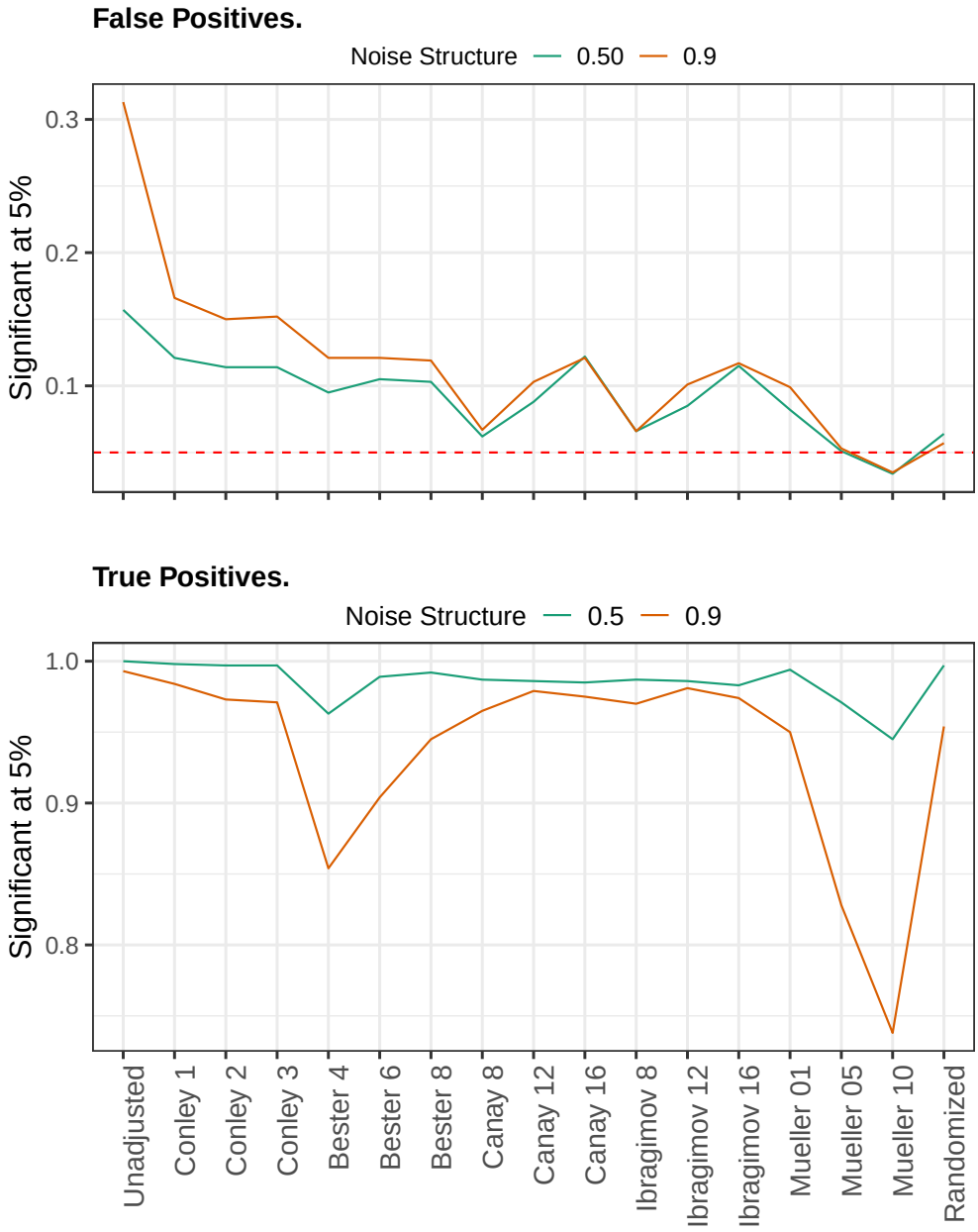
The focus here has been on OLS, but it can be straightforwardly extended. Instrumental variables can be approached by analyzing the first stage regression (in the presence of spatial correlation, instruments can appear spuriously strong), and by randomizing on the dependent variable in the second stage. Just as we analyzed cross-sectional models here by applying spatial correlation techniques, panel models can be randomized by first estimating their parameters using spatio-temporal methods.

Appendix A Size and Power of Spatial Standard Error Corrections

Figure A.1 shows the size and power of standard error corrections using simulated data with an empirically realistic structure. The points correspond to the German towns used by Voigtländer and Voth (2012) and correlation between sites falls off exponentially, having an effective range where it reaches 0.1 at 75 km: varying this does not affect results materially. I give results for cases where observations have little idiosyncratic variation around their predicted correlation value and where the variation is large. (In the notation of Section 5, $\kappa = 0.5, \theta = 37.5, \rho = 0.5$ or 0.9.)

The top panel shows the percentage of regressions that are significant at five per cent when one spatial noise series is regressed on another. For high structure ($\rho = 0.9$) we can see that about 30 per cent of unadjusted regressions are significant at 5 per cent, and if Conley corrections are applied (using a Bartlett kernels with ranges of 50, 75, and 100 km) this is halved. When spatial structure is reduced to $\rho = 0.5$, unadjusted and Conley standard errors are significant about 15 and 10 per cent of the time. BCH by contrast are significant in about ten per cent of cases with little change according to cluster number of the structure of the noise. Both CRS and IM overfit spatial noise almost identically increasing from 8 per cent with 8 clusters increasing to 12 per cent with 16 clusters. MW falls from 10 per cent at

Figure A.1: Standard error adjustments for spatial noise simulations based on the locations of German towns. Conley cutoffs equal 50, 75, and 100 km. BCH, IM and CRS are followed by number of clusters, and MW by average spatial correlation.



0.01 correlation, to 5 per cent at 0.05, and slightly below 5 at 0.10. Randomization is significant in 6 per cent of cases.

By contrast, CRS and IM have accurate size. The observations are placed successively in six to eight clusters using k-means clustering, and regardless of spatial structure, five per cent are significant at five per cent. By contrast, BCH clustering returns too many false positives. Müller-Watson has correct size when average correlation is set to what they consider to be a high value of 0.05 (the true average correlation of these simulations ranged from 0.04 for $\rho = 0.5$ to 0.07 for $\rho = 0.95$).

The lower panel takes two regressions where the dependent variable $y = bx + e$ where $b = 0.4$ and x and e are spatial noise with the same parameters as previously. For low spatial structure $\rho = 0.5$ all adjustments return significance in almost all simulations. However, for high structure $\rho = 0.9$ BCH has low power for small numbers of clusters. Most notable, however, is the poor performance of MW. At 0.05 correlation where the test has the right size in the top panel, it finds significance in only 82 per cent of cases, falling to 75 with assumed correlation of 0.1. CRS, IM, and randomization find significance in over 95 per cent of simulations.

It is notable that the patterns of significance in these artificial simulations are sometimes quite different than those for the real regressions in Figure 2, where CRS and IM are significant far less frequently than other adjustments, and where *t* values fall as the number of assumed clusters rise. This highlights the way that, in contrast to time series, with different locations of observations and, often, strong outliers, each set of spatial data is unique.

Appendix B Studies Examined.

Here we give details of the regressions we examined from the papers analysed above. We group them into three categories by their geographical focus: global; Africa; and Europe and the Americas. In every case, I chose the column of the table with the maximum number of controls applied, and regional or country dummies if applicable. Longitudinal clusters indicates that large cluster estimators in Figure

2 were constructed using north-south stripes instead of nearest neighbours to give variation in the explanatory variable within clusters.

B.1 Global

Acemoglu, Johnson and Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation

Table 3.1. Regress protection against expropriation risk on estimated settler mortality

Acemoglu, Johnson and Robinson (2002). Reversal of Fortune

Table 3.1. Regress GDP per capita on urbanization in 1500.

Alesina, Giuliano and Nunn (2013). On the Origin of Gender Roles: Women and the Plough.

Table 3.1. Regress women's labour force participation on plough adoption.

Arbatli et al. (2020) Diversity and Conflict

Table 1.8. Regress civil conflict on diversity.

Ashraf and Galor (2013). The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development

Table 6.3. Regress per capita GDP on quadratic diversity. Because the estimated significance levels of individual terms in a polynomial have no statistical interpretation, the hypothesis that they have no joint effect is tested.

Galor and Özak (2016). The Agricultural Origins of Time Preference

Table 1.2. Regress long term orientation on crop yield.

La Porta et al. (1998). Law and Finance

Table 6.2. Regress efficiency of judicial system on a civil law dummy.

Michalopoulos and Xue (2021) Folklore

Table 5.6. Regress GDP per capita on punishment of anti-social behaviour.

Spolaore and Wacziarg (2009). The Diffusion of Development

Table 1.3. Regress GDP per capita on genetic distance from the US.

B.2 Africa

Alsan (2015). The Effect of the TseTse Fly on African Development

Table 1.4. Regress historical population density on tsetse fly suitability.

Michalopoulos and Papaioannou (2013). Pre-Colonial Ethnic Institutions

Table 3.8. Regress nighttime illumination on binary political centralization.

Nunn (2008). The Long Term Effect of the Slave Trade

Table 3.5. Regress GDP per capita on slave exports. The robustness check is to add the share of population at risk of malaria and a dummy for the outlier of Democratic Republic of the Congo.

Nunn and Wantchekon (2011). The Slave Trade and the Origins of Mistrust in Africa

Table 2.3 Regress trust of neigbours on slave exports. The robustness check is to omit eight (of 157) districts with extreme exports. Longitudinal clusters.

B.3 Europe and the Americas

Ambrus, Field and Gonzalez (2020). Loss in the Time of Cholera: Long Run Impact of a Disease Epidemic on the Urban Landscape

Table 3.4. Regress 1936 rental prices on dummy for catchment area of the Broad Street cholera pump. Conley cutoffs in Figure 2 are set at 10, 15 and 20 per cent of the study range. Longitudinal clusters.

Becker and Woessmann (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History

Table 3.4. Regress literacy on the percentage Protestant.

Dell (2010). The Persistent Effects of Peru's Mining *Mita*

Table 2.1, panel three. Regress household consumption on Mita dummy. Conley cutoffs in Figure 2 are set at 20, 40 and 60 km. Longitudinal clusters.

Hornung (2014) Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia

Table 3.6. Regress textile productivity on Huguenot population share.

Squicciarini (2020)Devotion and Development: Religiosity, Education, and Economic Progress in Nineteenth-Century France

Table 3.1. Regress industrial employment on refractory clergy.

Valencia Caicedo (2019). The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America

Table 2.2. Regress modern literacy rates on distance from a Jesuit mission.

Voigtländer and Voth (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany.

Table 4.2. Regress Nazi vote share on pogroms. The robustness check is to leave out six constituencies (of 325) with extreme vote share.

References

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 95:1369–1401.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117:1231–1294.

Alesina, Alberto, Paola Giuliano and Nathan Nunn. 2013. "On the Origin of Gender Roles: Women and the Plough." *Quarterly Journal of Economics* 128:469–530.

Alsan, Marcella. 2015. "The Effect of the TseTse Fly on African Development." *American Economic Review* 105:382–410.

Ambrus, Attila, Erica Field and Robert Gonzalez. 2020. "Loss in the Time of Cholera: Long Run Impact of a Disease Epidemic on the Urban Landscape." *American Economic Review* 110:475–525.

Anderson, Marti J. and John Robinson. 2001. "Permutation Tests for Linear Models." *Australian and New Zealand Journal of Statistics* 43:75–88.

Anderson, Marti J. and Pierre Legendre. 1999. "An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model." *Journal of Statistical Computation and Simulation* 62:271–303.

Arbatli, Cemal Eren, Quamrul H. Ashraf, Oded Galor and Marc Klemp. 2020. "Diversity and Conflict." *Econometrica* 88:727–797.

Ashraf, Quamrul and Oded Galor. 2013. "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review* 103:1–46.

Athey, Susan and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31:3–32.

Barios, Thomas, Rebecca Diamond, Guido W. Imbens and Michal Kolesár. 2012. "Clustering, Spatial Correlations, and Randomization Inference." *Journal of the American Statistical Association* 107:578–591.

Becker, Sascha O. and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *Quarterly Journal of Economics* 124:531–596.

Bester, C. Alan, Timothy G. Conley and Christian B. Hansen. 2011. "Inference with Dependent Data Using Cluster Covariance Estimators." *Journal of Econometrics* 165:137–151.

Canay, Ivan M., Joseph P. Romano and Azeem M. Shaikh. 2017. "Randomization Inference under an Approximate Symmetry Assumption." *Econometrica* 85:1013–1030.

Conley, Timothy. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92:1–45.

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining *Mita*." *Econometrica* 78:1863–1903.

Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Freedman, David and David Lane. 1983. "A Nonstochastic Interpretation of Reported Significance Levels." *Journal of Business and Economic Statistics* 1:292–298.

Galor, Oded and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106:3064–3103.

Gelfand, Alan E., Peter Diggle, Peter Guttorp and Montserrat Fuentes, eds. 2010. *Handboook of Spatial Statistics*. Boca Raton: CRC Press.

Gneiting, Tilmann and Peter Gutthorp. 2010. Continuous Parameter Stochastic Process Theory. In *Handbook of Spatial Statistics*, ed. Alan E. Gelfand, Peter Diggle, Peter Guttorp and Montserrat Fuentes. Boca Raton: CRC Press.

Good, Phillip I. 2006. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. New York: Springer.

Hornung, Erik. 2014. "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia." *American Economic Review* 104:84–122.

Ibragimov, Rustam and Ulrich K. Müller. 2010. "t-Statistic Based Correlation and Heterogeneity Robust Inference." *Journal of Business and Economic Statistics* 28:453–468.

Imbens, Guido W. and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.

Imbens, Guido W. and Jeffrey M. Wooldridge. 2009. "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature* 47:5–86.

Kennedy, Peter E. and Brian S. Cade. 1996. "Randomization tests for multiple regression." *Communications in Statistics—Simulation and Computation* 25:923–926.

La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer and Robert W. Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106:1113–1155.

Michalopoulos, Stelios and Elias Papaioannou. 2013. "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81:113–152.

Michalopoulos, Stelios and Melanie Meng Xue. 2021. "Folklore." *Quarterly Journal of Economics* .

Müller, Ulrich K. and Mark W. Watson. 2021. Spatial Correlation Robust Inference. Working paper. Department of Economics Princeton University.

Nunn, Nathan. 2008. "The Long-term Effects of Africa's Slave Trades." *Quarterly Journal of Economics* 123:139–176.

Nunn, Nathaniel and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101:3221–3252.

Spolaore, Enrico and Romain Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics* 124:469–529.

Squicciarini, Mara. 2020. "Devotion and Development: Religiosity, Education, and Economic Progress in Nineteenth-Century France." *American Economic Review* 110:3454–3491.

Valencia Caicedo, Felipe. 2019. "The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America." *Quarterly Journal of Economics* 134:507–556.

Voigtländer, Nico and Hans-Joachim Voth. 2012. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *Quarterly Journal of Economics* 127:1339–1392.

Winkle, Anderson M., Gerard R. Ridgwa, Matthew A. Webster, Stephen M. Smith and Thomas E. Nichols. 2014. "Permutation inference for the general linear model." *Neuroimage* 92:381–397.

Young, Alwyn. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Quarterly Journal of Economics* 134:557–598.

Zimmerman, Dale L. and Michael Stein. 2010. Classical Geostatistical Methods. In *Handbook of Spatial Statistics*, ed. Alan E. Gelfand, Peter Diggle, Peter Guttorp and Montserrat Fuentes. Boca Raton: CRC Press.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP20/31 Bernardo Buarque, Ronald Davies, Ryan Hynes and Dieter Kogler: 'Hops, Skip & a Jump: The Regional Uniqueness of Beer Styles' December 2020
WP21/01 Kevin Devereux: 'Returns to Teamwork and Professional Networks: Evidence from Economic Research' January 2021
WP21/02 K Peren Arin, Kevin Devereux and Mieszko Mazur: 'Taxes and Firm Investment' January 2021
WP21/03 Judith M Delaney and Paul J Devereux: 'Gender and Educational Achievement: Stylized Facts and Causal Evidence' January 2021
WP21/04 Pierluigi Conzo, Laura K Taylor, Juan S Morales, Margaret Samahita and Andrea Gallice: 'Can ❤️s Change Minds? Social Media Endorsements and Policy Preferences' February 2021
WP21/05 Diane Pelly, Michael Daly, Liam Delaney and Orla Doyle: 'Worker Well-being Before and During the COVID-19 Restrictions: A Longitudinal Study in the UK' February 2021
WP21/06 Margaret Samahita and Leonhard K Lades: 'The Unintended Side Effects of Regulating Charities: Donors Penalise Administrative Burden Almost as Much as Overheads' February 2021
WP21/07 Ellen Ryan and Karl Whelan: 'A Model of QE, Reserve Demand and the Money Multiplier' February 2021
WP21/08 Cormac Ó Gráda and Kevin Hjortshøj O!Rourke: 'The Irish Economy During the Century After Partition' April 2021
WP21/09 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' April 2021
WP21/10 Benjamin Elsner, Ingo E Isphording and Ulf Zölitz: 'Achievement Rank Affects Performance and Major Choices in College' April 2021
WP21/11 Vincent Hogan and Patrick Massey: 'Soccer Clubs and Diminishing Returns: The Case of Paris Saint-Germain' April 2021
WP21/12 Demid Getik, Marco Islam and Margaret Samahita: 'The Inelastic Demand for Affirmative Action' May 2021
WP21/13 Emmanuel P de Albuquerque: 'The Creation and Diffusion of Knowledge - an Agent Based Modelling Approach' May 2021
WP21/14 Tyler Anbinder, Dylan Connor, Cormac Ó Gráda and Simone Wegge: 'The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants' June 2021
WP21/15 Morgan Kelly: 'Devotion or Deprivation: Did Catholicism Retard French Development?' June 2021
WP21/16 Bénédicte Apouey and David Madden: 'Health Poverty' July 2021
WP21/17 David Madden: 'The Dynamics of Multidimensional Poverty in a Cohort of Irish Children' August 2021
WP21/18 Vessela Daskalova and Nicolaas J Vriend: 'Learning Frames' August 2021
WP21/19 Sanghamitra Chattopadhyay Mukherjee: 'A Framework to Measure Regional Disparities in Battery Electric Vehicle Diffusion in Ireland' August 2021
WP21/20 Karl Whelan: 'Central Banks and Inflation: Where Do We Stand and How Did We Get Here?' August 2021
WP21/21 Tyler Anbinder, Cormac Ó Gráda and Simone Wegge: '"The Best Country in the World": The Surprising Social Mobility of New York's Irish Famine Immigrants' August 2021
WP21/22 Jane Dooley and David Madden: 'Ireland's Post Crisis Recovery, 2012-2019: Was It Pro-Poor?' September 2021
WP21/23 Matthew Shannon: 'The Impact of Victimisation on Subjective Well-Being' September 2021
WP21/24 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' October 2021

UCD Centre for Economic Research
Email economics@ucd.ie