UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2020

Understanding Persistence

Morgan Kelly, University College Dublin

WP20/23

September 2020

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Understanding Persistence

Morgan Kelly*

Abstract

A large literature on persistence finds that many modern outcomes strongly reflect characteristics of the same places in the distant past. These studies typically combine unusually high t statistics with severe spatial autocorrelation in residuals, suggesting that some findings may be artefacts of underestimating standard errors or of fitting spatial trends. For 25 studies in leading journals, I apply three basic robustness checks against spatial trends and find that effect sizes typically fall by over half, leaving most well known results insignificant at conventional levels.

Turning to standard errors, there is currently no data-driven method for selecting an appropriate HAC spatial kernel. The paper proposes a simple procedure where a kernel with a highly flexible functional form is estimated by maximum likelihood. After correction, standard errors tend to rise substantially for cross sectional studies but to fall for panels. Overall, credible identification strategies tend to perform no better than naive regressions. Although the focus here is on historical persistence, the methods apply to regressions using spatial data more generally.

*University College Dublin and CEPR. Some of the approach developed here circulated in a very preliminary form under the title "The Standard Errors of Peristence."

1 Introduction

A substantial literature on deep origins or persistence finds that many modern outcomes such as income or social attitudes strongly reflect 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 settlers determines the quality of modern institutions; how countries that inherited common law enjoy better legal systems; how medieval pogroms prefigured Nazi zealotry; how the slave trade still retards modern African development; and how colonial boundaries drive poverty in Peru and civil conflict in Africa.[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 one's regressions, no matter how assiduously.[2] 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 regressions are spatial regressions: the values today of some variable in a given set of places are regressed on another variable for the same places in the past. Now, Tobler's (1970) First Law of Geography states that "everything is related to everything else, but near

[1]These are, in turn, Acemoglu, Johnson and Robinson (2001), La Porta, de Silanes and Shleifer (2008), Voigtländer and Voth (2012), Nunn (2008), Dell (2010), and Michalopoulos and Papaioannou (2016).

[2]Of the 25 studies examined below, 14 report a t above 3.3 ($p = 10^{-3}$) and six above 5.1 ($p = 10^{-7}$).

things are more related than distant things." Spatial data, in other words, tend to be highly autocorrelated and, moreover, to display strong directional trends. This creates two potential difficulties for estimation.

First, as with time series, fitting spatial trends can lead to spurious correlation. Given that income is higher in Europe than in Africa, any variable that is high in one and low in the other will appear to explain progress and poverty. This is less of a truism than it sounds, as we will see in a moment.

Second, the fact that economic variables tend to show high spatial autocorrelation means that places resemble not only their immediate neighbours but also quite distant places as well. The result is that many observations do not add much to the precision of coefficient estimates, so that standard errors may be considerably larger than might be expected given the nominal sample size. If you fail to compensate for this it can be easy to mistake spatial noise regressions for deep, world-historical relationships.

Starting with spatial trends, nobody wants their persistence regression to be merely a roundabout way of saying something to the effect that rich countries tend to be richer than poor countries. It is therefore routine to add some extra variables such as continental dummies, distance from the equator, or terrain ruggedness by way of controls. In this paper I systematically apply three geographical controls as robustness checks, each of them very simple.

The first, for studies of countries around the globe, is to add a dummy for World Bank regions. Whereas continents are fairly arbitrary groupings, these regions are more informative: Sub-Saharan Africa, the Middle East and North Africa, and so on. The impact of this basic control turns out to be substantial.

For instance, adding World Bank dummies to Acemoglu et al's (2001) regression of security of property rights on European settler mortality reduces the effect size from -0.6 (t=-4.0) to -0.2 (t=-1.4). Regional dummies have similar impacts on results such as Alesina, Giuliano and Nunn (2013) on the impact of plough adoption on gender roles; Ashraf and Galor (2013) on genetic diversity; and Acemoglu, Johnson and Robinson (2002) on reversals of fortune. Adding a dummy for rich countries similarly attenuates the finding of La Porta et al. (1998) that common law countries enjoy higher judicial efficiency; as does a dummy for Europe for the Nunn and Qian (2011) claim that the potato drove population growth.

What makes these global studies fragile is the fact that their persistence variables show such long range spatial correlation that they act for all purposes as regional proxies: see Figure 1 below. This means that once explicit regional controls are added their apparent explanatory power falls markedly.

The second control, for studies on smaller geographical scales, is to add longitude and latitude to control for directional gradients. These are especially important for studies of historical frontiers: if a hillside runs north-south, then any east-west line will separate higher regions from lower ones. Figure 2 below illustrates how a strong north-south gradient in living standards underlies the apparent role of the Peruvian Mita (Dell, 2010) in driving modern consumption.

The final robustness check is to examine how the results change after we omit regions with extreme values, that usually lie towards the edge of the study area. For the impact of medieval pogroms on Nazi support (Voigtländer and Voth, 2012) this control region is Hitler's adopted home of Bavaria; for the way the modern levels of trust mirror historical slave exports (Nunn and Wantchekon, 2011) this is the "Slave Coast" along the Bight of Benin; and looking at how colonial frontiers aggravate civil conflict (Michalopoulos and Papaioannou, 2016) the control region is Somalia. In all cases the decrease in effect sizes is considerable.

On top of the three basic robustness checks, in three instances I add a potentially important omitted variable: malaria. Malaria is routinely included in studies of Africa (such as Alsan, 2015 or Nunn and Wantchekon,

2011) but if it were omitted, as in Nunn (2008), then some variable that happened to be high in malarial regions, such as historical slave exports, might appear to explain modern income.

Moving on from robustness, the second potential difficulty with persistence studies, and spatial regressions more generally, arises from underestimated standard errors. Because many observations closely resemble each other, many contribute little to sharpening the precision of estimates, so that useful sample sizes may be a good deal lower than they appear.

The usual way that researchers try to control for this 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) and illustrated in Section 3 below. Moreover, as Abadie et al. (2017) demonstrate, standard error estimates vary substantially according to the assumed level of clustering, and even in experimental settings it can be hard to determine what this should be.

A theoretically well grounded approach to standard error estimation is to apply a heteroskedasticity and autocorrelation consistent (HAC) procedure, pioneered for spatial regressions by Conley (1999). However, estimates can be very sensitive to the choice of kernel: we will see below how small changes in the assumed cutoff distance in the rectangular kernel recommended by Conley and Molinari (2007) return widely different standard errors, and there is no way of knowing which, if any, is the correct one. Combining this sensitivity with the fact there exists no automatic and data-dependent procedure for selecting a kernel, means that spatial HAC standard errors are currently not entirely operational.

However, the estimation of reliable spatial kernels is the defining problem of spatial statistics.[3] Its canonical question is how to interpolate between observations taken at a few locations—such as mineral concentrations in test boreholes, or barometric readings at weather stations—to infer the values at other locations, usually on a grid to allow maps to be drawn. This exercise revolves around estimating a kernel that reliably describes how correlation changes with distance. Because of its extremely adaptable function form—that ranges from exponential to Gaussian as a single smoothness parameter increases—and the fact that is it guaranteed to be positive definite, the workhorse kernel of spatial statistics is the Matérn function (Gneiting and Gutthorp, 2010), and that is the form that will be applied here.

The proper concern with any kernel is that its assumed functional form is inappropriate for the data at hand leading to large biases in standard errors. For spatial data a further complication is that economic space, which is what we care about, may not coincide with geographical space, which is what we get to observe, so that the relevant locations of places are observed with error. Appendix A presents extensive Monte Carlo simulations which find that the downward bias of standard errors estimated with a Matérn kernel is moderate even when residuals are generated by very different (Cauchy and spatial autoregressive) processes, and locations are observed with substantial Gaussian errors.

The residuals from the regressions examined below share a characteristic form with correlation falling off at a slow exponential rate with distance. Comparing adjusted standard errors with those originally reported, consistent patterns emerge.

When no correction was originally applied, adjusted standard errors can be a large multiple of reported ones—up to three times in one case—

[3] This is analogous to the way that Andrews (1991) was able to draw on existing kernels in developing HAC standard errors for time series.

that is roughly proportional to the degree of spatial autocorrelation in the residuals. For clustered standard errors the distortion varies considerably by study. However, in papers where multiple observations are clustered at locations that are close together—such as households in towns from Dell (2010)—the standard errors originally reported are large underestimates.

For longitudinal data the corrected standard errors tend to be markedly *smaller* than the clustered ones that are routinely calculated. The reason is that fixed effects have already soaked up a good deal of the spatio-temporal structure of residuals, with the result that clustering corrections tend to be too aggressive at clearing out whatever correlation remains.

I address the two issues of robustness to spatial trends and standard errors in turn by looking at their impact on a variety of published studies. The analysis is limited to 25 papers, that appeared in the *American Economic Review* (10), *Quarterly Journal of Economics* (8), and *Econometrica* (2), with one each taken from the *American Economic Journal: Macroeconomics, Journal of Political Economy, Journal of Politics*, *Review of Economics and Statistics*, and *Science*. Studies were chosen either because they are well known or because they struck me as interesting and well executed, and to allow a variety that included longitudinal data, discrete dependent variables, difference in differences, instrumental variables, and regression discontinuities.

The sole concerns of this paper are with spatial robustness checks and the computation of reliable standard errors. 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). It is not concerned with, and does not remark on, any econometric issues in the original regressions that it replicates, although in a few cases these are not trivial.

Above all, and this cannot be emphasized too strongly, this paper is not concerned with somehow "validating" or "disproving" the findings of any particular study. In fact, I am not interested in any individual result except insofar as it helps to illustrate the broader issues of spatial trends and standard errors. The fact, moreover, that the single regression analysed here performs poorly does not in any way imply that later regressions in the paper, that often use different dependent variables, are equally problematic. Rather than the negative goal of trying to disparage anyone's research, the purpose of this study is the positive one of marking out two potentially serious difficulties in persistence regressions, and in spatial studies more widely, and showing how they can be remedied straightforwardly.

In terms of existing literature, there appears to have been no previous attempt to systematically analyze the effects of spatial trends and underestimated standard errors on persistence results. Persistence studies themselves fall into two broad groups. On one side there is what can be called the "Attitudes and Institutions" literature reviewed, for example, by Cantoni and Yuchtman (2020) and Nunn (2020); and on the other there are studies of "Genes and Geology" surveyed by Spolaore and Wacziarg (2013). Although these two literatures tend not to cite one another, I fold them together because of their common statistical structure.

For spatial standard errors, the literature is small, especially compared with time series. The HAC approach was pioneered by Conley (1999). Alternative approachs, based on partitioning the data into a small number of large groups, have been developed by Bester, Conley and Hansen (2011) and Ibragimov and Müller (2010) but we find below that estimates are sensitive to the assumed clusters.[4]

[4]Inference based on specific autoregressive models of spatial dependence has been developed by Kelejian and Prucha (1999, 2007), Lee (2004, 2007*a*, 2007*b*) and Kim and Sun (2011). However, because space unlike time is symmetric, these face potential issues of endogeneity.

The rest of the paper is as follows. In the following Section I illustrate how simple robustness checks cause the effect sizes of many well known persistence studies to fall substantially. Section 3 presents simulations to show underestimated standard errors can make regressions of one spatial noise series on another appear highly significant, and Section 4 proposes a simple kernel to deal with this. The impact of these standard error corrections is discussed in Section 5. Appendix A provides simulation results to examine how robust the estimated standard errors are to errors in the assumed functional form of the kernel and the spatial location of the observations, Appendix B considers the performance of some existing standard error corrections, and details of the studies examined are given in Appendix C.

2 Persistence Studies: Robustness Checks.

We begin with some robustness checks of persistence results. These checks fall into three groups: For global studies they are World Bank region. For studies on a smaller scale, the controls are either direction (longitude and latitude); or omitting areas that have extreme values of the dependent or explanatory variable. The controls applied to each regression are summarized in Table 1 and exact details are given in Appendix C. Typically, the regression examined is the lead regression of the paper including the additional robustness variables added by the authors.

For two global studies, robustness checks differ somewhat from World Bank regions. For La Porta et al. (1998) a rich country (income over $10,000) dummy is applied; and in Nunn and Qian (2011) a dummy for Europe after 1700 is used, to control for Europe's rapid growth after this time and the fact that it is also a good place for potatoes.

Figure 1: The long range correlation of many persistence variables makes them act as regional proxies which means that their impact diminishes when explicit regional dummies are added. Each tile represents a country and observations are shaded by decile with bright colours highest. Each of the first three variables gives Europe or its offshoots high values and Africa low ones, whereas with female employment Africa scores high and the Middle East low.

Table 1: Summary of studies and robustness checks.

	Regression	Robustness Check[a]
	Global	
Acemoglu, Colonial Origins.	Property rights on settler mortality.	WB Regions.
Acemoglu, Reversal.	Income on AD 1500 popn.	WB Regions. Malaria.
Alesina, Plough.	Female employment on plough adoption.	WB Regions.
Ashraf, Malthusian.	AD 1500 popn on neolithic transition.	WB Regions.
Ashraf, Out of Africa.	Income on genetic diversity.	WB Regions.
Comin, 1000 BC.	Income on technology in 1000 BC.	WB Regions.
Galor, Time Preference.	Patience on soil fertility.	WB Regions.
La Porta, Law Finance.	Judicial efficiency on Common Law.	Rich country dummy.
Nunn, Ruggedness.	Income on ruggedness $\times$ Africa.	Malaria.
Nunn, Potato.	Population on potato suitability.	Europe dummy.
Schulz, Kinship.	Individualism on kinship intensity.	WB Regions.
Spolaore, Diffusion.	Income on genetic distance from US.	WB Regions.
	Europe and the Americas	
Acharya, American Slavery	Republican support on 1861 slavery	.

[a] Blanks denote cases where the results were unaffected by the standard robustness checks applied to similar studies.

Continued on next page

Table 1: Summary of studies and robustness checks. (cont.)

	Regression	Robustness Check[a]
Ambrus, Cholera.	Rent on cholera pump boundary.	Direction.
Becker, Anti-Semitism	Pogroms on Protestantism	Ex. Brandenburg.
Becker, Weber.	Literacy on percentage Protestant.	Direction.
Caicedo, Mission.	Literacy on mission distance.	.
Dell, Mita.	Consumption on Mita boundary.	Direction.
Voigtlaender, Persecution.	Nazi vote on pogroms.	Ex Bavaria.
Africa and India		
Alsan, Tsetse.	Slavery on tsetse suitability.	.
Banerjee, Land Tenure.	Crop yield on British tax.	Ex North India.
Michalopoulos, Pre-Colonial.	Light density on political complexity.	Direction.
Michalopoulos, Scramble.	Civil conflict on border split.	Ex Somalia.
Nunn, Mistrust.	Mistrust on slave exports.	Ex Bight of Benin.
Nunn, Slavery.	Income on slave exports.	Malaria.

[a] Blanks denote cases where the results were unaffected by the standard robustness checks applied to similar studies.

For studies on a smaller geographical scale, longitude and latitude are applied to boundary studies like Dell (2010) and Ambrus, Field and Gonzalez (2020).[5] Distance from the equator is used for Michalopoulos and Papaioannou (2013) to capture the fact that northern and southern Africa are the richest parts of the continent; and longitude is used for Becker and Woessmann (2009) as a control for a strong eastward fall in German literacy.

The third class of robustness check is to analyze the impact of removing a control region with unusually high values. So, for Nunn and Wantchekon (2011) who look at the impact of slave exports on interpersonal trust, the three modern states that border on the Slave Coast of the Bight of Benin are omitted; while for Voigtländer and Voth (2012) the control is six constituencies in Hitler's adopted home of Bavaria that supported him disproportionately. Becker and Pascali (2019) look at the role of Protestantism in driving Anti-Semitism after 1500. However, Brandenburg accounts for 5 per cent of pogroms in the sample, but one fifth in the sixteenth century, and Hesse accounts for 7 per cent in total but one sixth in the seventeenth century. The control is to examine the impact of removing these two regions.

Given the importance of malaria, in three cases it is added as a control. These are in Acemoglu, Johnson and Robinson (2002), Nunn and Puga (2012) and, more importantly, for Nunn's (2008) investigation of the connection between historical slave exports and modern income which, unusually among studies of Africa, does not include it.

[5]One other notable boundary study, which examines how trust in institutions changes across the frontier of the former Habsburg empire, is Becker et al. (2014). For the first regression of Table 2 showing trust in courts, the coefficient of 0.14 (SE=0.07) reduces to 0.03 (SE=0.08) after including longitude and latitude: there is a steady downward gradient heading east. However, because the study uses ordered logit regressions where residuals are not well defined, the HAC corrections here are not applicable and the study is not included in the Tables.

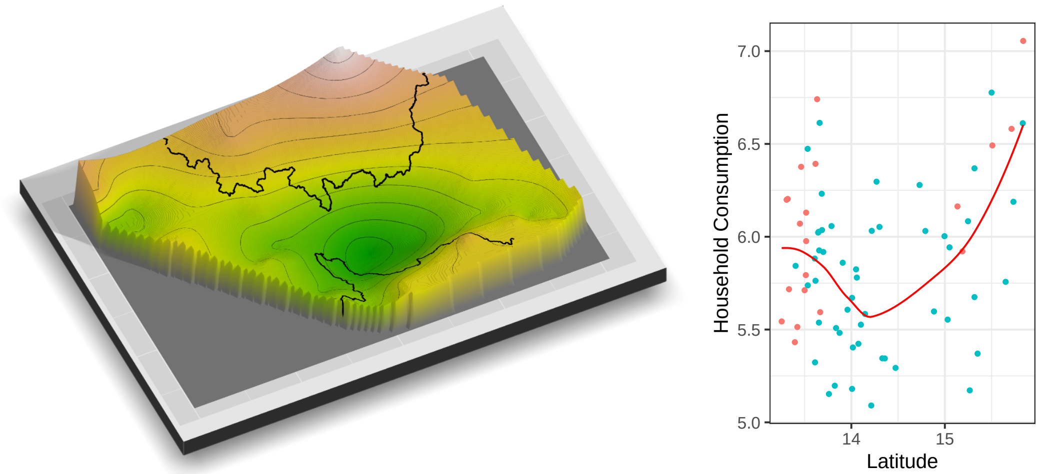

Figure 2: The left panel maps median household consumption, looking south, from Dell (2010) with dark areas indicating low consumption. Black lines are the Mita frontiers. It can be seen that although consumption is indeed lower beneath the southern frontier, given the strong north-south gradient this will be also be true for any east-west line nearby. The U-shaped pattern of consumption going north-south is shown in the right panel where each dot gives a town's median consumption, and blue ones lie within the Mita. Given the latitude of a town, knowing whether it lay within the frontier adds little information about its household consumption.

The first column of Table 2 details the impact of the robustness checks. For each study, successive rows give the coefficient of the main explanatory variable before and after applying the check: three studies with identical rows were those left unchanged by the robustness checks applied to similar studies. The changes in effect sizes are summarized in Figure 3 which gives the value of the coefficient on the main explanatory variable after the robustness check relative to its original value. The falls tend in most cases to be considerable.

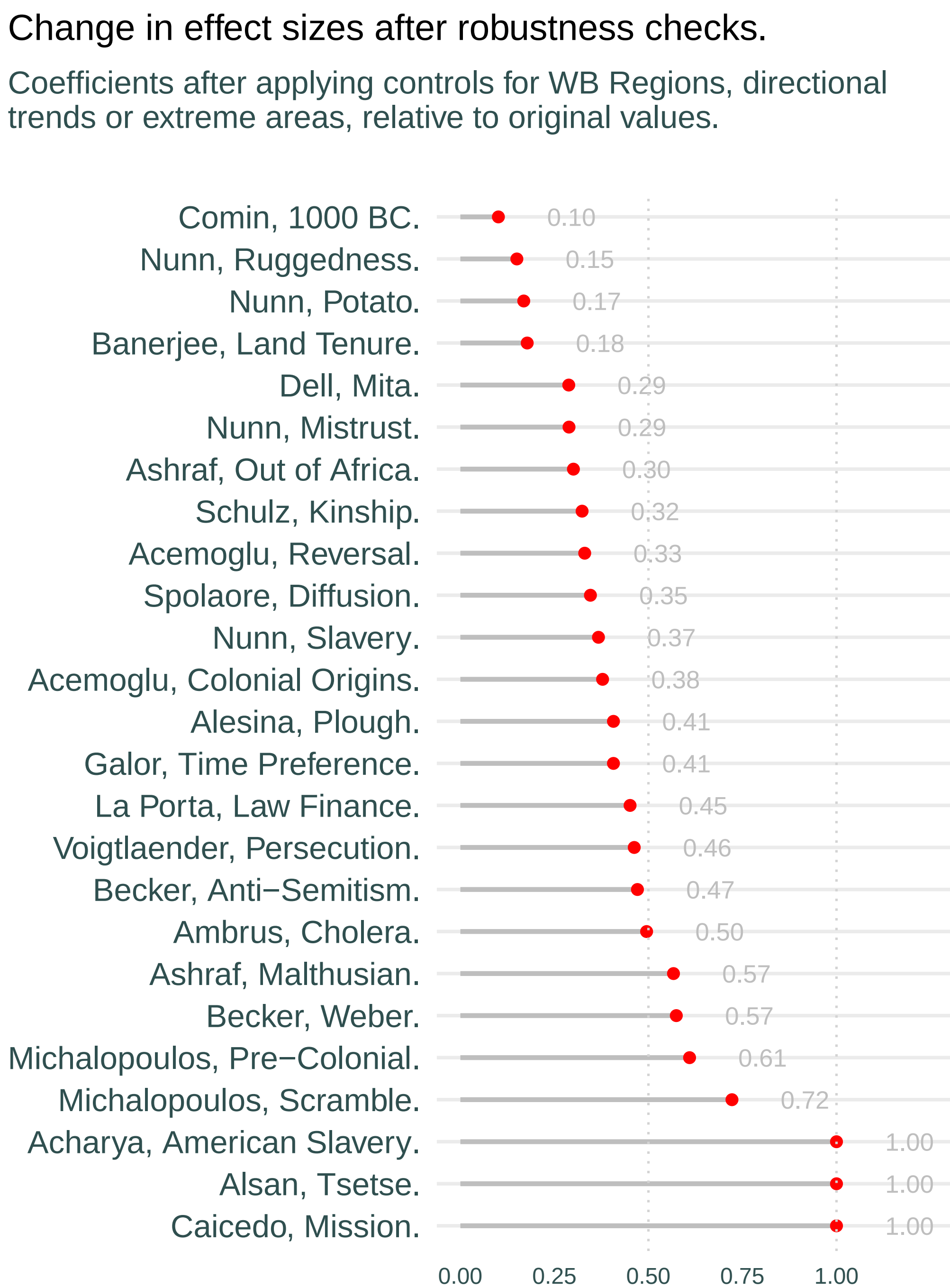

Figure 3: Regression coefficients after applying robustness checks, relative to their original values.

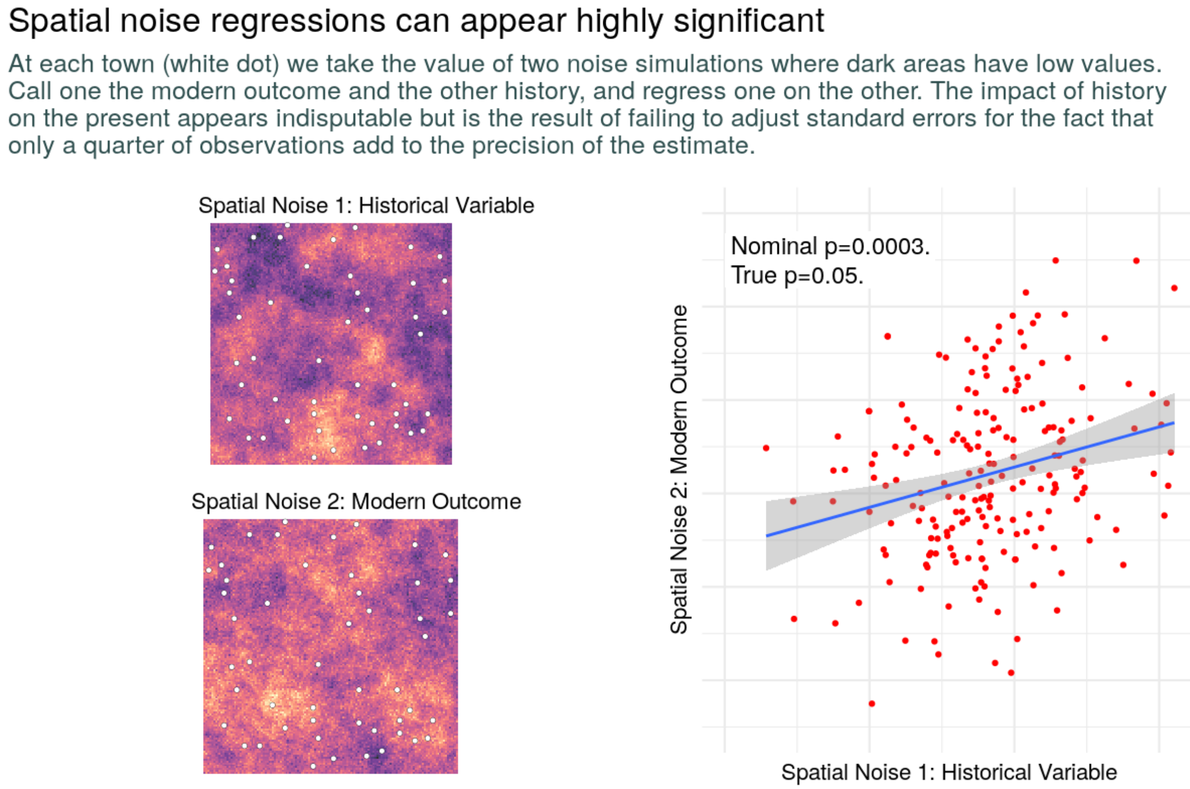

Figure 4: Regressions of one spatial noise series on another can appear highly significant if one fails to correct the standard error for the fact that most observations contribute little to the precision of coefficient estimates.

3 Fitting Spatial Noise

Economic variables tend to show strong spatial autocorrelation: places not only resemble their immediate neighbours but quite distant places as well. This autocorrelation means that many observations do not add much to the precision of coefficient estimates so that standard errors may be considerably larger than might be expected given the nominal sample size. A naive spatial regression that spuriously matches high points in one variable with high (or low) points in another will often return what looks like a strong relationship unless standard errors are corrected appropriately.

Figure 4 illustrates the problem. It takes two simulations of spatial noise, each on a square with sides of length 100. Across the square towns are scattered at random, represented by white dots. The spatial noise has an empirically realistic pattern: correlation falls off exponentially and has

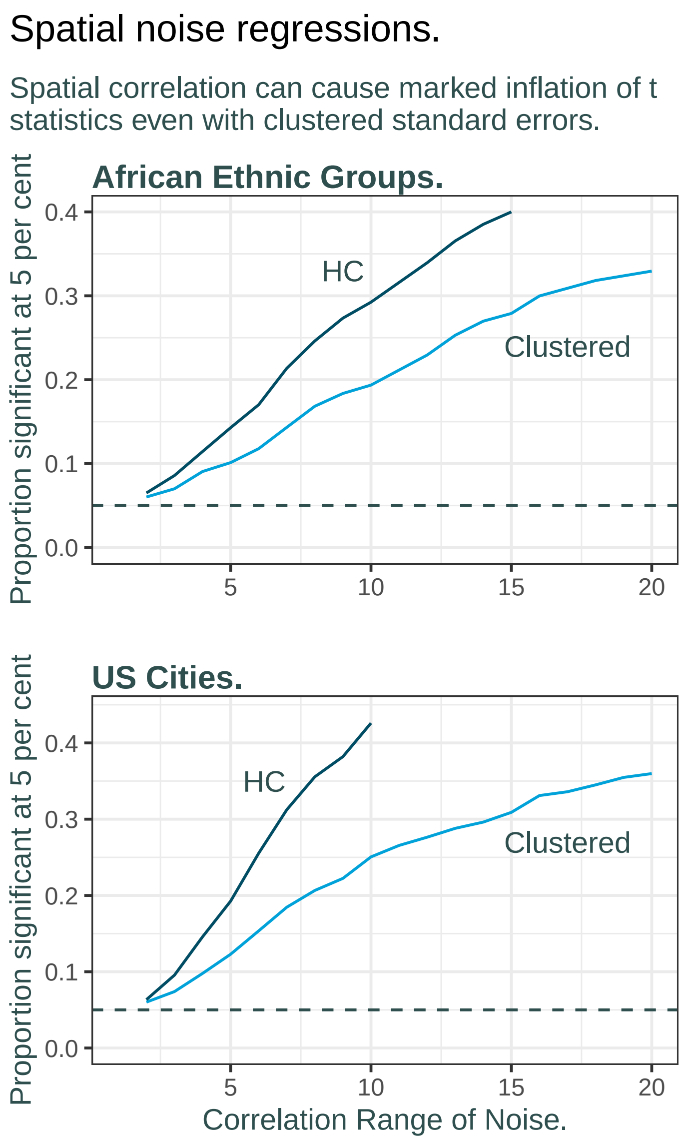

Figure 5: Without proper standard error corrections, regressions of one spatial noise series on another return inflated t statistics that vary by dataset. This inflation occurs both with heteroskedasticity consistent (HC) and clustered standard errors.

largely disappeared after a distance of 20; and only one quarter of the variance of observations is due idiosyncratic noise. (In the notation of the next Section $\kappa = 0.5, \theta = 10, \rho = 0.75$). Light areas denote regions with high values. Suppose now that we take two noise processes and evaluate them at each town, and then regress one on the other. In the example here, this leads to a regression with *t* of 3.7 and a nominal significance of $p = 0.0003$. In fact the empirical significance level is 5 per cent (out of 1000 random noise regressions, its *p* value was at the 5th percentile): the estimated *t* statistic is twice its correct value. The inflated *t* statistic is the result of our failure to adjust standard errors for the fact that only around one quarter of observations in this case contribute anything useful to the precision of the coefficient estimate.

This inflation of *t* values is shown systematically in Figure 5 where the points are now based on the African ethnic groups used by Michalopoulos and Papaioannou (2013) and US commuting zones from Chetty et al. (2014), where the original coordinate axes have been changed to make each a 100×100 square. Noise follows the same exponential falloff as before, and the correlation range varies between 1 to 20. Figure 5 shows the percentage of regressions of one noise series on another that are significant at 5 per cent, using either heteroskedasticity consistent (HC) standard errors or clustered (by district and state respectively) ones. The *t* statistics are noticeably inflated even for moderate ranges of spatial correlation, in a way that differs by dataset. At a range of 10, nearly one third of African regressions are significant using HC standard errors, and 40 per cent of US ones.

As noted earlier, by clustering one is using a procedure to safeguard against spatial correlation that should not be used in the presence of spatial correlation. When standard errors are clustered, the proportion of significant regressions is roughly halved, but the inflation is still considerable.

4 Spatial Kernel Estimation.

We have observations of data at N sites s_i and estimate the regression $y_{s_i} = \beta x_{s_i} + u_{s_i}$. This leads to the estimate $\hat{\beta} = \left(X'X\right)^{-1} X'y$ with variance

$$\begin{aligned} \mathrm{Var}\left(\hat{\beta}\right) &= \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1} \\ &= \left(\tfrac{1}{N}X'X\right)^{-1} \Phi \left(\tfrac{1}{N}X'X\right)^{-1} \end{aligned} \tag{1}$$

The spectral approach, pioneered in spatial regression by Conley (1999), is to estimate Φ as a weighted sum of cross products

$$\hat{\Phi} = \frac{1}{N} \sum_{s_i, s_j} K\left(s_i, s_j\right) x_{s_i} \hat{u}_{s_i} x'_{s_j} \hat{u}_{s_j} \tag{2}$$

where $K\left(s_i, s_j\right)$ is a weighting kernel that must be chosen.

It is useful to decompose the kernel into a systematic spatial component and idiosyncratic noise

$$K\left(s_i, s_j\right) = \rho\, C\left(s_i, s_j\right) + (1-\rho)\, \mathbf{1}_{ij} \tag{3}$$

where the indicator $\mathbf{1}_{ij} = 1$ when $i = j$ and 0 otherwise, and $0 \leq \rho \leq 1$. The parameter ρ reflects the ratio of spatial signal to noise in the residuals, and in the limit where spatial structure ρ goes to zero, we return in (2) to standard HC covariance estimation.

What is required is a kernel C that accurately reflects the systematic correlation structure of the residuals. Because of its adaptable functional form, and the fact that it is guaranteed to be positive definite, the most widely used kernel in spatial statistics is based on the Matérn function. Correlation between sites s_i, s_j at distance h apart is

$$M\left(h; \theta, \kappa\right) = \frac{2^{1-\kappa}}{\Gamma\left(\kappa\right)} \left(\frac{h}{\theta}\right)^{\kappa} B_{\kappa}\left(\frac{h}{\theta}\right) \quad (\kappa > 0, \theta > 0) \tag{4}$$

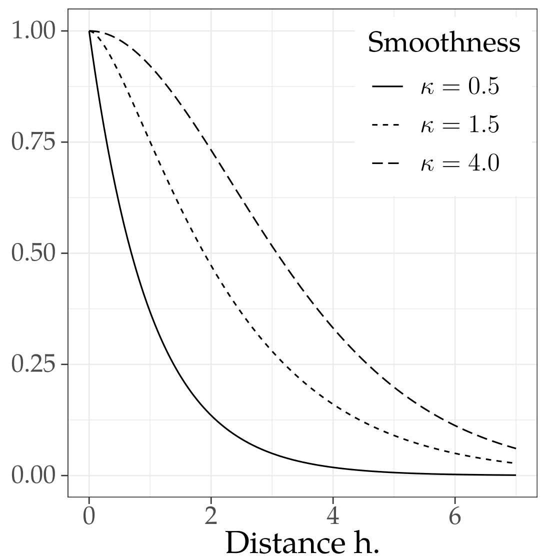

Figure 6: The flexible form of the Matérn function (drawn with range $\theta = 1$) allows it to fit a wide variety of spatial kernels. For the studies analysed here, correlation among residuals tends to fall off exponentially with distance, corresponding to $\kappa = 0.5$.

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 $M(h) = \exp(-h/\theta)$, and as $\kappa \to \infty$, M becomes Gaussian. The flexibility of the Matérn function is illustrated in Figure 6 where range θ is set to 1 and smoothness κ takes on values from 0.5 to 4.

We have then a weighting kernel giving the correlation between the residuals at every location

$$K(s_i, s_j) = \rho\, M(h; \theta, \kappa) + (1 - \rho)\, I \tag{5}$$

whose three parameters θ, ρ and κ can be estimated by maximum likelihood from the estimated residuals. K is then substituted into (2) to estimate Φ.

Because the Matérn function is monotonic, a distance can be chosen beyond which correlation is negligible and can be set to zero. This gives us a compact support for K and, assuming that this cutoff distance is of order $o\left(L^{1/3}\right)$ where L is the length of the study space, allows the kernel to satisfy Conley's (1999) sufficient conditions for the estimated standard errors (1) to be consistent.

The approach extends to nonlinear models in the usual way: see, for instance, Andrews (1991). Similarly for panels, if A is the correlation matrix between residuals across different time periods, and K is the spatial correlation each period, then the longitudinal kernel is the Kronecker product of the two.

The proper concern about any such exercise is that the estimated standard errors will be substantially biased if the spatial correlation of the residuals differs from the assumed functional form, if the relevant economic locations of the observations differ from their geographic ones, or if the strength of correlation varies with direction. Appendix A presents Monte Carlo simulations which indicate that even substantial departures from these assumptions lead to standard error estimates that are biased downwards by under five per cent.

5 Persistence Studies: Adjusted Standard Errors.

The main results are presented in Table 2. For each study there are two rows displaying results first for the original specification, and then after the robustness checks detailed in Table 1. Each row first gives the regression coefficient for the main variable, alongside its original and adjusted standard errors. The next two columns give the maximum likelihood estimates of the spatial range and structure parameters for the regression residuals.

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. What this means is that range θ and smoothness κ cannot be reliably estimated together, because as one rises the other tends to fall. Instead, a grid search is conducted, with parameters θ and ρ estimated for each value of κ, and the parameters with the highest likelihood are selected.

In the regressions here, as in most areas of spatial statistics, the maximum likelihood value of κ is low, usually 0.5—exponential correlation—and increasing the value of κ simply led to compensating falls in θ and almost identical standard errors. For instance, for Schulz et al. (2019) the maximum likelihood value occurred at $\kappa = 1$, but using the exponential correlation reported in the Table reduced the standard error from 0.179 to 0.173.

The final column reports the degree of spatial autocorrelation in neighbouring residuals given by the Moran statistic, $I = W \sum_{i \neq j} w_{ij} \hat{u}_i \hat{u}_j / \sum_i \hat{u}_i^2$ where w_{ij} are weights and $W = N / \sum_{i,j} w_{ij}$. Here we assigned a weight of one to the 5 nearest neighbours, and zero otherwise: changing this number did not alter the results materially. The statistic has an asymptotic normal distribution and in most cases is markedly above 2.

It can be seen that, except in cases where the Moran statistic is low, most regressions display a common spatial pattern where correlation among residuals dies away at a slow exponential rate, and there is a strong spatial structure ρ. In the second row, after the robustness checks have often removed considerable spatial structure, the difference between the original and adjusted standard errors is often lower, reflecting the smaller kernel parameters and reduced Moran statistics.

Figure 7: Standard errors estimated with exponential kernel relative to original values, before robustness checks were carried out.

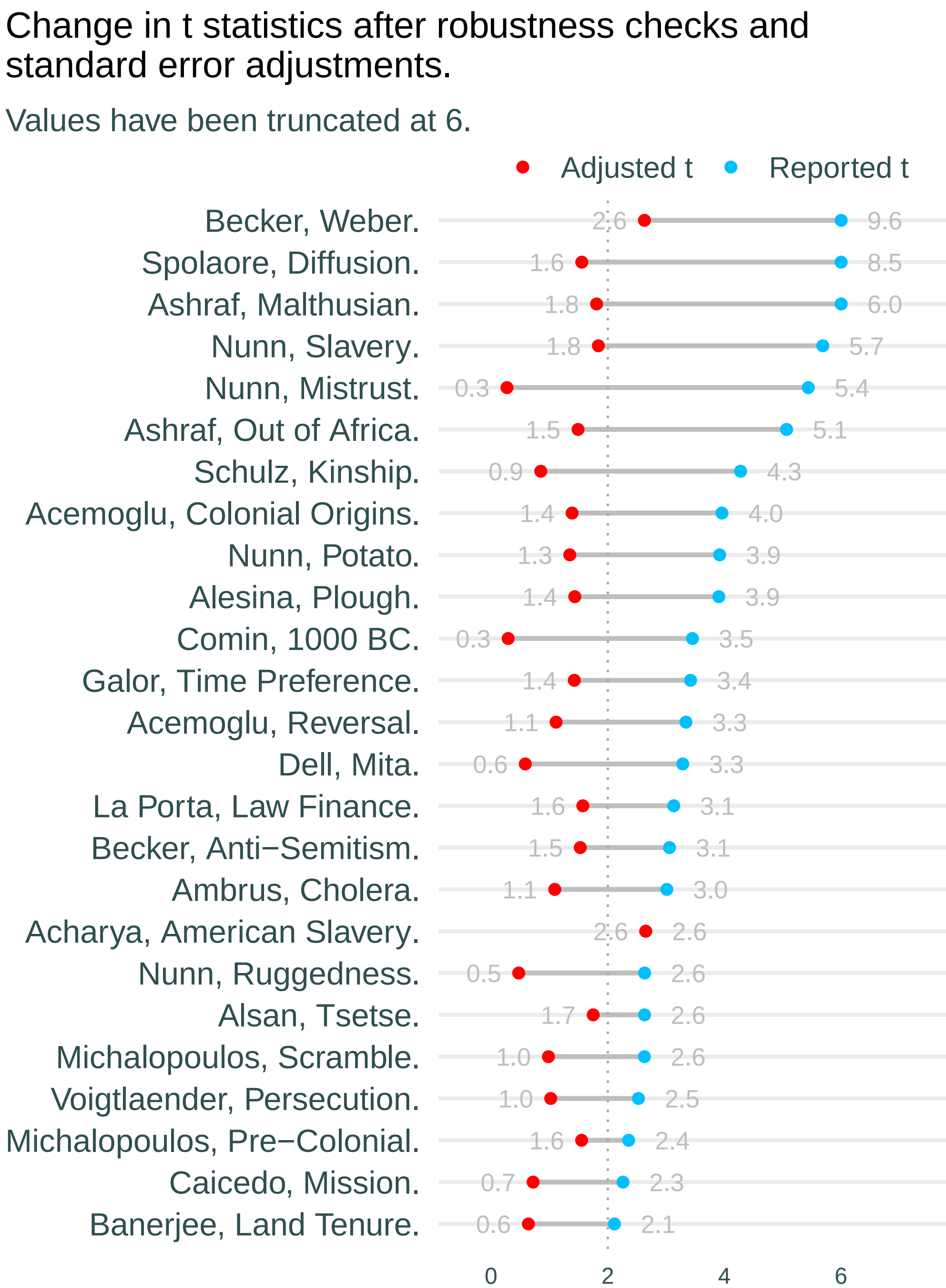

Figure 8: Change in t statistics after standard error adjustments and robustness checks. Values have been truncated at 6.

Table 2 breaks down the studies into three groups depending on how their standard errors were originally estimated.[6] For HC corrections, where no account was taken of the possibility that the residuals might be autocorrelated, the rise in standard errors in the first row is often substantial, more than doubling in several cases. The exceptions are cases where the spatial autocorrelation in the residuals is small, reflected by low Moran statistics, so the standard error is unchanged.

For clustered standard errors, in most cases the kernel correction leads to a rise of about one third. There are two exceptions however where the change is considerably larger: Dell (2010) and Nunn and Wantchekon (2011). In both studies, individual households or survey respondents were clustered by town or district, but most of these places tend to clump near each other. In such cases where there were multiple observations in each location, the spatial kernel parameters and Moran statistics (which are low for both studies reflecting their unusual residual structure) were calculated based on the average residual at each site: changing the values of range and structure had a small effect on the standard error estimates.

[6] In some cases decimal places have been moved to give all coefficients a similar order of magnitude.

Table 2: Summary of regression results and spatial kernel parameters.

		SEs		Spatial Kernel		
	Coef	**Orig.**	**HAC**	**Range**	**Struct.**	**Moran**
Heteroskedasticity Robust SEs						
Acemoglu, Colonial Origins.	-0.59	0.15	0.19	5240	0.19	2.19
	-0.22	0.16	0.16	780	0.05	0.48
Acemoglu, Reversal.	-0.78	0.23	0.27	2220	0.86	2.57
	-0.26	0.21	0.23	860	1.00	1.34
Acharya, American Slavery.	-2.77	1.04	1.04	420	0.13	2.29
	-2.77	1.04	1.04	420	0.13	2.29
Alesina, Plough.	-1.52	0.39	0.58	2200	0.72	6.98
	-0.62	0.41	0.43	1290	0.57	2.54
Ashraf, Malthusian.	1.37	0.23	0.37	2630	1.00	5.03
	0.78	0.25	0.43	3660	1.00	6.56
Ashraf, Out of Africa.	5.42	1.07	2.12	4540	0.82	13.12
	1.63	0.70	1.09	2230	0.69	7.70
Becker, Weber.	0.11	0.01	0.03	190	0.85	18.19
	0.06	0.01	0.02	140	0.94	18.83
Caicedo, Mission.	1.12	0.50	1.56	90	0.88	22.62
	1.12	0.50	1.56	90	0.88	22.62
Galor, Time Preference.	9.84	2.88	2.99	320	0.43	1.48
	4.01	2.81	2.81	10	0.00	1.58
La Porta, Law Finance.	-1.38	0.44	0.44	930	0.64	0.87
	-0.62	0.40	0.40	480	0.41	0.07
Nunn, Ruggedness.	3.28	1.25	1.35	1900	0.83	6.98
	0.49	1.02	1.04	2440	0.61	7.19
Nunn, Slavery.	-0.13	0.02	0.02	810	0.36	1.89
	-0.05	0.03	0.03	150	0.74	0.03
Schulz, Kinship.	-0.39	0.09	0.18	7670	0.73	8.29
	-0.13	0.09	0.15	3450	0.63	4.47
Spolaore, Diffusion.	-4.53	0.53	1.35	4090	0.89	8.29
	-1.57	0.65	1.01	2670	0.81	5.32

Successive rows give results before and after robustness checks. Each row reports the original and adjusted standard errors along with estimated kernel parameters—effective range 2θ and spatial structure ρ—and Moran statistic.

Continued on next page

Table 2: Summary of regression results. (*cont.*)

		SEs		Spatial Kernel		
	Coef	**Orig.**	**HAC**	**Range**	**Struct.**	**Moran**
		Clustered SEs				
Alsan, Tsetse.	0.11	0.04	0.06	3460	1.00	11.34
	0.11	0.04	0.06	3460	1.00	11.34
Ambrus, Cholera.	-0.44	0.15	0.20	1	1.00	7.71
	-0.22	0.17	0.20	1	1.00	6.77
Comin, 1000 BC.	1.60	0.46	0.86	7580	0.91	10.28
	0.16	0.55	0.55	2310	0.89	6.58
Dell, Mita.	-2.89	0.88	2.34	10	0.72	0.80
	-0.83	1.15	1.42	10	0.10	-0.16
Michalopoulos, Pre-Colonial.	0.21	0.09	0.09	870	0.59	11.81
	0.13	0.08	0.08	680	0.59	10.51
Michalopoulos, Scramble.	0.45	0.17	0.33	680	0.50	32.84
	0.33	0.20	0.33	480	0.45	32.04
Nunn, Mistrust.	-1.83	0.34	1.46	210	0.57	1.19
	-0.53	0.36	1.94	390	0.40	1.88
Voigtlaender, Persecution.	1.44	0.57	0.81	250	0.55	12.45
	0.67	0.47	0.65	420	0.53	12.32
		Fixed Effects				
Banerjee, Land Tenure.	1.48	0.70	0.29	240	0.91	7.60
	0.26	0.73	0.41	400	0.66	5.09
Becker, Anti-Semitism.	0.50	0.16	0.14	80	0.11	4.53
	0.24	0.18	0.15	80	0.09	3.57
Nunn, Potato.	4.11	1.05	0.44	4800	0.91	6.44
	0.69	1.36	0.51	5360	0.91	6.11

Successive rows give results before and after robustness checks. Each row reports the original and adjusted standard errors along with estimated kernel parameters—effective range 2θ and spatial structure ρ—and Moran statistic.

For longitudinal studies, adjusted standard errors are considerably *lower* than the clustered ones originally calculated. The reason is that fixed effects have already absorbed a good deal of the spatio-temporal structure of the residuals so that clustering is an aggressive solution to a problem that has substantially dissipated. The spatial correlation parameters θ and ρ for panels, as well as Moran statistics, were calculated as the mean of the values estimated for each period, and temporal autocorrelation was similarly an average of the autocorrelation between residuals in each period. Once again, large changes in the assigned kernel parameters values did not affect standard errors materially.

The changes in t values after robustness checks and standard error adjustments are shown in Figure 8. It is increasingly appreciated that because t statistics conflate size of effects with how precisely those effects were estimated they are not a useful metric of the importance of a variable. A coefficient is "significant" if it has a 95 per cent confidence interval of $[0.1, 0.2]$ but "insignificant" if the confidence interval is $[-1, 5]$ even though the latter effect is as likely to be above 4 as below 0. Nevertheless, given the importance that many of these studies seem to attach to significance levels it is perhaps useful to see how robust they are.

6 Conclusions

This paper considered two potentially important issues that arise from the fact that the data underlying the historical persistence literature usually show strong spatial autocorrelation. The first was that explanatory variables may be proxying for spatial trends in the data; and the second was that standard errors may be underestimated. For 25 persistence studies, controlling for these two possibilities usually had a noticeable impact on the estimated results, both on effect sizes and standard errors.

As for practical advice, it is straightforward to compute fairly reliable standard errors. However, spatial data, which clump together in different ways, are inherently messy and it is naive to look for estimates accurate to four decimal places. In the light of the Monte Carlo results in Appendix A, it seems prudent to expect that confidence intervals are perhaps five to ten per cent wider than calculated.

The risk of chasing spatial trends can be reduced by following the robustness checks applied above. Quadratics in longitude and latitude should be added to regressions as a matter of routine. Areas with particularly high values of the explanatory or dependent variables should be highlighted and the effect of removing them should be made explicit. Given their unusual fragility, global studies based on country level data should be undertaken with considerable caution.

The simplest and most important step however is simply to graph your data. It is always advisable to be skeptical of any claimed regression result where a scatter plot of the main variables is not provided. But with spatial data it is equally important to see simple coloured maps of the dependent and explanatory variables along with residuals to understand quickly whether a regression is fitting anything more profound than spatial trends.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens and Jeffrey Wooldridge. 2017. When Should You Adjust Standard Errors for Clustering? Working paper MIT.

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.

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "The Political Legacy of American Slavery." *Journal of Politics* 78:621–641.

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.

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–858.

Ashraf, Quamrul and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review* 101:2003–2041.

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

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95:1190–1213.

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., Katrin Boeckh, Christa Hainz and Ludger Woessmann. 2014. "The Empire is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy." *Economic Journal* 126:40–74.

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.

Becker, Sascha O. and Luigi Pascali. 2019. "Religion, Division of Labor and Conflict: Anti-Semitism in German Regions over 600 Years." *American Economic Review* 109:1764–1804.

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.

Cantoni, Davide and Noam Yuchtman. 2020. Historical Natural Experiments: Bridging Economics and Economic History. Working Paper 26754 NBER.

Chetty, Raj, Nathaniel Hendren, Patrick Kline and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129:1553–1623.

Comin, Diego, William Easterly and Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics* 2:65–97.

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

Conley, Timothy G. 2010. Spatial Econometrics. In *New Palgrave Dictionary of Economics*, ed. Steven N. Durlauf and Lawrence E. Blume. London: Palgrave Macmillan.

Conley, Timothy G. and Francesca Molinari. 2007. "Spatial Correlation Robust Inference with Errors in Location or Distance." *Journal of Econometrics* 140:76–96.

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

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

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.

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

Kelejian, Harry H. and Ingmar R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40:509–533.

Kelejian, Harry H. and Ingmar R. Prucha. 2007. "HAC Estimation in a Spatial Framework." *Journal of Econometrics* 140:131–154.

Kim, Min Seong and Yixiao Sun. 2011. "Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix." *Journal of Econometrics* 160:349–371.

La Porta, Rafael, Florencio Lopez de Silanes and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46:285–322.

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

Lee, Lung-Fei. 2004. "Asymptotic Distributions of Quasi Maximum Likelihood Estimators for Spatial Autoregressive Models." *Econometrica* 72:1899–1925.

Lee, Lung-Fei. 2007*a*. "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models." *Journal of Econometrics* 137:489–514.

Lee, Lung-Fei. 2007*b*. "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects." *Journal of Econometrics* 140:333–374.

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

Michalopoulos, Stelios and Elias Papaioannou. 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106:1802–1848.

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

Nunn, Nathan. 2020. "The Historical Roots of Economic Development." *Science* 367:eaaz9986.

Nunn, Nathan and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics* 94(20-36).

Nunn, Nathan and Nancy Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." *Quarterly Journal of Economics* 126:593–650.

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

Schulz, Jonathan, Duman Bahrami-Rad, Jonathan Beauchamp and Joseph Henrich. 2019. "The Church, Intensive Kinship, and Global Psychological Variation." *Science* 366. Issue 6466.

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

Spolaore, Enrico and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature* 51:325–369.

Tobler, W. R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46:234–240.

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.

Appendix A Robustness of Standard Error Estimates

In estimating spatial standard errors above we made three assumptions about the structure of the spatial structure of the data being analysed: that the spatial correlation of residuals decays as a Matérn function; that the relevant economic distance between points corresponds to their observed, geographical distance; and that the correlation between residuals was stationary and isotropic. This Appendix presents Monte Carlo simulations to assess the degree of bias in standard error estimates that arises when these assumptions are violated. All simulations are for 200 points scattered randomly on a 100×100 square.

A.1 Errors in the Assumed Kernel

To assess the bias in standard errors when the underlying residuals do not follow the assumed Matérn function we analyze two cases where the the spatial structure is extremely different. The first is where the true correlation of the residuals follows a Cauchy (power law) so that instead of (4) the actual kernel describing correlation between sites s_i, s_j at distance h apart is

$$C\left(s_i, s_j; \theta, \alpha\right) = \left(1 + (h/\theta)^2\right)^{-\alpha} \quad (\alpha > 0, \theta > 0) \tag{6}$$

Compared with a Matérn, the falloff in correlation with distance is extremely slow as the parameter α falls.

The second case is where the residuals have a spatially autoregressive (SAR) structure. Specifically, for a vector of residuals u

$$u = \lambda W u + \epsilon \tag{7}$$

where ϵ are iid standard normal variables and W is a weighting matrix with diagonal elements of zero.

To simulate in each case we generate a vector of residuals u and an explanatory variable x that both obey the same Cauchy or SAR process, and then a dependent variable $y = x + u$. The goal then is to see how far the standard error for the regression of y on x estimated with a Matérn kernel differs from the the correct value.

Both the Cauchy and SAR processes imply considerably greater smoothness than the exponential falloff ($\kappa = 0.5$) that gave the best fit for the persistence regressions examined above. Specifically, in all cases the maximum likelihood estimate occurred with a smoothness parameter $\kappa = 4$ (increasing κ further not increase the likelihood materially) that is tending towards a Gaussian falloff: see Figure 6 above.

Starting with the Cauchy case in Table A1, the Matérn standard error is biased downwards, reflecting the considerably greater spatial structure of Cauchy residuals. Nevertheless, for shorter ranges and/or lower spatial structure its performance is not hugely in error, with a downward bias of under 10 per cent even with a slow falloff of $\alpha = 0.5$. The bias of heteroskadisticity robust standard errors, where no steps are taken to correct for spatial correlation, is also included for comparison.

We can see the same behaviour in Table A2 where the data follow a SAR process. It is assumed that the weighting matrix W gives equal weights to the five nearest neighbours of each point: increasing this to 10 had no material effect on the results. The Matérn kernel again performs well so long as the degree of spatial structure, this time controlled by λ, is not excessive.

Table A.1: Bias in standard errors when a Matérn kernel is applied to residuals that have a Cauchy correlation structure.

		$\alpha = 1$				$\alpha = 0.75$				$\alpha = 0.25$			
Corr Range	**Struct-ure**	**Bias**	**RMSE**	**95% Cover**	**HC Bias**	**Bias**	**RMSE**	**95% Cover**	**HC Bias**	**Bias**	**RMSE**	**95% Cover**	**HC Bias**
5	0.25	-0.00	0.10	0.98	-0.03	-0.01	0.09	0.96	-0.04	-0.02	0.11	0.90	-0.07
	0.50	-0.03	0.11	0.94	-0.13	-0.02	0.13	0.96	-0.18	-0.05	0.13	0.94	-0.23
	0.75	-0.09	0.11	0.88	-0.26	-0.06	0.13	0.92	-0.30	-0.13	0.17	0.94	-0.39
	1.00	-0.15	0.16	0.94	-0.38	-0.12	0.16	0.88	-0.43	-0.15	0.16	0.86	-0.48
10	0.25	-0.03	0.11	0.96	-0.10	-0.05	0.10	0.92	-0.11	-0.08	0.10	0.86	-0.12
	0.50	-0.07	0.14	0.92	-0.28	-0.07	0.15	0.88	-0.30	-0.11	0.15	0.94	-0.33
	0.75	-0.12	0.16	0.96	-0.44	-0.13	0.13	0.88	-0.48	-0.16	0.20	0.92	-0.53
	1.00	-0.10	0.16	0.86	-0.57	-0.19	0.16	0.84	-0.61	-0.23	0.17	0.78	-0.64
15	0.25	-0.04	0.13	0.98	-0.13	-0.05	0.13	0.96	-0.17	-0.05	0.15	0.98	-0.15
	0.50	-0.10	0.15	0.94	-0.38	-0.09	0.20	0.92	-0.40	-0.16	0.17	0.84	-0.40
	0.75	-0.16	0.15	0.94	-0.55	-0.17	0.20	0.84	-0.55	-0.21	0.19	0.92	-0.59
	1.00	-0.19	0.16	0.90	-0.67	-0.27	0.15	0.86	-0.71	-0.33	0.16	0.84	-0.74

Bias, Root MSE, and 95% CI coverage probabilities level when a Matérn kernel is applied to residuals with Cauchy correlation. α controls the rate of falloff in correlation. HC bias denotes bias when robust SEs are used. 1000 Monte Carlo replications. 200 random points on 100×100 grid. In all cases a Matérn smoothing parameter $\kappa = 4$ is applied.

λ	Bias	RMSE	Coverage	HC Bias
0.1	-0.02	0.08	0.93	-0.01
0.3	-0.03	0.09	0.94	-0.05
0.5	-0.05	0.11	0.94	-0.19
0.7	-0.13	0.15	0.90	-0.43
0.9	-0.30	0.20	0.79	-0.72

Bias, Root MSE and 95% CI coverage probabilities, when a Matérn kernel is applied to data with spatial autoregressive structure with coefficient λ. HC bias denotes bias when robust SEs are used. 1000 Monte Carlo replications. 200 random points on 100×100 grid. For all estimates a Matérn smoothing parameter $\kappa = 4$ is applied.

Table A.2: Bias in standard errors when a Matérn kernel is applied to residuals that have a spatial autoregressive (SAR) structure.

The fairly robust behaviour of the Matérn kernel when applied to residuals with extremely slow, and empirically unrealistic, falloff of correlation indicates that it should be reliable in cases where the data have a more realistic spatial structure. If, for example, the true kernel follows a power exponential distribution (stable law) where $C\left(s_i, s_j; \theta, \gamma\right) = \exp\left((h/\theta)^{-\gamma}\right)$, the bias of the Matérn kernel is minor and the results are not reported.

A.2 Errors in the Assumed Location of Observations.

As Conley (1999) and Conley and Molinari (2007) have emphasized, economic distance does not always coincide with geographical distance. To assess robustness we assume that each point is moved in a random direction by a distance that is normally distributed with mean zero and standard deviation τ. On average, then, each location is moved an average distance of $\tau\sqrt{2/\pi}$ and we examine the what happens when this average equals 1, 2, or 5. For concreteness, for the United States which is 5,000 km across, a distance of 2 implies that each town lies an average of 100 km away from its position on the map.

Table A3 reports simulations where points are an average distance of 1, 2 or 5 from their observed positions, and correlation is exponential. It can be seen that overall the bias is small when the distance is 1 or 2, but becomes substantial at 5 when the spatial structure of residuals is above 0.5. As well as bias, the Table gives the average range and structure calculated from the observed points: structure is estimated fairly accurately but range tends to be overestimated: this works to mitigate the downward bias of the estimates.

A.3 Anisotropy

It has been assumed so far that data are isotropic: correlation between sites depends only on the distance h between them, independently of direction. We now consider cases where correlation is geometrically anisotropic: instead of the isocorrelation contours about each point being circles they are ellipses. Table A4 shows simulations of the bias caused by ignoring anisotropy, displaying cases where the ratio of major to minor axis is 1.5 or 2, and the main axis of correlation is rotated by zero or 45 degrees. We apply exponential correlation with a range of 10.

It can be seen from the Table that the bias induced by neglecting isotropy is not substantial for a ratio of 1.5, and even when the ratio rises to 2 and the residual have strong spatial structure, the downward bias is in the region of 3 per cent.

Table A.3: Bias in standard errors when the true position of points is observed with Gaussian error.

		Range =5				Range=10				Range=15			
Location Error	**Struct-ure**	**Bias**	**RMSE**	**Est. Range**	**Est. Struct**	**Bias**	**RMSE**	**Est. Range**	**Est. Struct**	**Bias**	**RMSE**	**Est. Range**	**Est. Struct**
1	0.25	-0.00	0.01	5.58	0.36	-0.00	0.03	7.60	0.33	-0.00	0.02	16.80	0.32
	0.50	-0.00	0.01	5.78	0.52	-0.02	0.04	8.40	0.51	0.00	0.02	17.06	0.51
	0.75	-0.01	0.02	5.64	0.71	-0.03	0.05	8.63	0.73	0.01	0.05	16.16	0.74
	1.00	-0.01	0.02	5.91	0.88	-0.04	0.04	9.35	0.93	-0.01	0.03	17.62	0.95
2	0.25	-0.00	0.01	6.01	0.32	-0.01	0.03	8.00	0.31	-0.00	0.02	17.94	0.32
	0.50	-0.01	0.02	6.45	0.46	-0.03	0.05	8.81	0.47	-0.00	0.04	17.91	0.49
	0.75	-0.02	0.03	6.65	0.60	-0.04	0.06	9.12	0.66	0.00	0.07	18.19	0.70
	1.00	-0.04	0.03	6.91	0.75	-0.07	0.05	9.57	0.84	-0.02	0.05	20.16	0.90
5	0.25	-0.01	0.02	6.24	0.28	-0.02	0.03	8.27	0.25	-0.01	0.03	19.33	0.25
	0.50	-0.04	0.03	7.75	0.31	-0.07	0.06	9.28	0.36	-0.04	0.06	21.38	0.40
	0.75	-0.07	0.04	8.36	0.38	-0.12	0.06	9.66	0.48	-0.05	0.09	23.18	0.56
	1.00	-0.11	0.04	8.64	0.47	-0.18	0.06	9.85	0.60	-0.10	0.08	24.91	0.72

Bias, RMSE, estimated structure and range when observations differ from their true location by Gaussian noise with an average distance of 1, 2 and 5. Location error gives the average displacement of each point. 1000 Monte Carlo replications with exponential kernel. 200 random points on 100×100 grid.

Appendix B Alternative Standard Error Corrections

We noted earlier that when data are spatially autocorrelated, standard errors clustered at some arbitrary level will in general be inconsistent. However Bester, Conley and Hansen (2011) and Ibragimov and Müller (2010) derive estimators that are consistent under mild assumptions by grouping data into G large clusters. Intuitively, most observations lie in the interior of a cluster and will be uncorrelated with most observations in other clusters. The Bester, Conley and Hansen (2011) approach is to estimate familiar clustered standard errors, whereas Ibragimov and Müller (2010) take averages of the regression t statistics estimated for each cluster.

To evaluate these estimators we will use data taken from two of the studies analysed earlier. Real world observations usually clump together geographically—most cities are on coasts or large rivers, for example—which suggests that the small changes in G may cause large changes in the estimates.

Table A5 gives results for two regressions analysed above: others behave similarly. Estimates are calculated for two to eight clusters where the clusters are parallel strips running either north to south, or east to west. To make the results easily interpretable given the low degrees of freedom, the table takes the relevant p values and converts them into a normal variable that has the same significance level.

Both clustering procedures return a wide spread of estimates. For instance, if seven clusters are assumed, BCH will give a value of 2.2 or 1.6 depending on the direction chosen compared with 1.7 or 1.9 for IM; and across all cluster sizes estimates range from 0.8 to 2.4.

Moving to the rectangular kernels suggested by Conley (2010), Table A6 shows how standard errors for global studies vary as the assumed cutoff distance move from 500 to 3000km, where entries are reported as a mul-

Anisotropy					
Ratio	Angle	Structure	Bias	RMSE	Coverage
1.5	0	0.25	0.00	0.01	0.95
		0.50	0.00	0.02	0.92
		0.75	-0.00	0.02	0.90
		1.00	-0.01	0.02	0.89
1.5	45	0.25	0.00	0.01	0.94
		0.50	0.01	0.02	0.94
		0.75	0.01	0.02	0.94
		1.00	0.01	0.02	0.94
2.0	0	0.25	-0.00	0.01	0.95
		0.50	-0.00	0.03	0.94
		0.75	-0.01	0.03	0.92
		1.00	-0.03	0.03	0.91
2.0	45	0.25	0.00	0.01	0.93
		0.50	0.01	0.02	0.95
		0.75	0.03	0.03	0.92
		1.00	0.03	0.03	0.89

Bias, Root MSE and 95% CI coverage probabilities, when the residuals are geometrically anisotropic with a ratio of major to minor axis of 1.5 or 2; and the major axis running at 0 or 45 degrees to horizontal. The residuals have exponential correlation with a range of 10. 1000 Monte Carlo replications. 200 random points on 100×100 grid.

Table A.4: Bias in standard errors when the residuals are anisotropic, according to the ratio and direction of the anisotropy.

tiple of original estimates and tend to vary considerably as assumed cutoffs change.

Appendix C 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 India; and Europe and the Americas. Because the original studies were conducted in Stata and we use R to take advan-

	Study 1				Study 2			
Groups	**BCH**		**IM**		**BCH**		**IM**	
2	0.99	1.22	1.28	1.34	1.36	3.40	1.35	3.33
3	2.01	2.27	2.40	0.82	2.11	2.36	2.10	1.86
4	1.73	2.04	1.83	1.56	1.92	2.75	1.73	1.43
5	2.33	1.71	1.47	1.32	2.28	2.81	1.97	1.58
6	1.96	1.67	1.26	1.66	2.52	3.10	2.36	1.77
7	2.17	1.55	1.71	1.91	2.89	3.48	2.87	2.03
8	2.19	1.93	1.44	1.58	2.95	3.75	3.21	1.71

Estimated t statistics for Alsan (2015) and Spolaore and Wacziarg (2009) regressions in Table 2 above, estimated using Bester, Conley and Hansen (2011) and Ibragimov and Müller (2010) clustering. Groups denotes the assumed number of clusters, and for each estimate a value is given first for when the clusters are assumed to be parallel strips running north to south, and then east to west.

Table A.5: Change in Bester, Conley and Hansen (2011) and Ibragimov and Müller (2010) *t* statistics as the assumed number and shape of clusters varies.

Cutoff (kilometres)	**500**	**1000**	**1500**	**2000**	**2500**	**3000**
Acemoglu, Colonial Origins.	1.1	1.5	2.0	2.2	2.3	2.5
Acemoglu, Reversal.	1.1	1.2	1.4	1.3	1.3	1.2
Alesian, Plough.	1.3	1.6	1.8	2.1	2.3	2.4
Ashraf, Malthusian.	1.0	1.4	1.7	1.9	2.1	2.1
Ashraf, Out of Africa.	1.2	1.7	2.2	2.5	2.7	2.9
Comin, 1000 BC.	1.2	1.5	1.8	2.0	2.1	2.3
Galor, Time Preference.	1.1	1.3	1.6	1.4	0.9	.
La Porta, Law Finance.	0.9	1.1	1.4	1.6	1.6	1.8
Schulz, Kinship.	1.2	1.8	2.1	2.4	2.5	2.5
Spolaore, Diffusion.	1.3	1.9	2.5	3.1	3.5	3.7

Conley standard errors computed with rectangular kernel with assumed cutoffs from 500 to 3000km for global studies. Each estimate is reported as a multiple of the original, heteroskedastic consistent standard error. Blank entries are cases where estimated variance was negative.

Table A.6: Change in Conley standard errors as the assumed cutoff distance varies.

tage of its strong geostatistical capabilities, we have taken care to ensure that we could exactly replicate the original results, notably their standard errors.

C.1 Global

World Bank regions and share of population at risk of malaria are taken from the online replication data accompanying Ashraf and Galor (2013).

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

We replicate the Acemoglu, Johnson and Robinson (2001) regression of average protection against expropriation risk on estimated settler mortality, both in logs from Table 3. Robustness control is WB regions.

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

We replicate Column 1 of Table 3 in Acemoglu, Johnson and Robinson (2002), regressing GDP per capita in 1995 on estimated urbanization in 1500. The additional variables added are WB regions and a dummy for high malarial prevalence where the share of population at risk of malaria exceeds 0.05.

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

We take Alesina *et al*'s (2013) Table 3, column 1 regression of women's labour force participation on plough adoption, with a variety of geographical and historical controls. WB regions and absolute latitude are added as robustness checks and two substantial outliers—Iceland and the Solomon Islands—are omitted.

Ashraf and Galor (2011). Dynamics and Stagnation in the Malthusian Epoch.

Here we analyse Table 2, column 5 where the log of population density in the year 1500 is regressed on the number of years since the neolithic and geographical controls. The robustness check involves a dummy for countries below 10 degrees south, and one for Europe, Central and South Asia.

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

We reproduce the first column of Table 5 where per capita GDP in 2000 is regressed on a measure of genetic diversity based on migratory distance from East Africa and adjusted to take account of settler ancestry. We apply robust standard errors in all regressions, rather than the bootstrapped errors of the original study which were somewhat larger. Additional variables added are distance from the Equator as in the original study, and a dummy for the country being in Sub-Saharan Africa or South Asia.

Comin, Easterly and Gong (2010). Was the Wealth of Nations Determined in 1000 BC?

I reproduce Table 8A, Column 1 regressing log income per capita in 2002 on migration-adjusted technology level in 1000 BC. Robustness check is WB Regions.

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

This replicates column 2 of Table 1, regressing long term orientation on crop yield, continent dummies, absolute latitude, mean elevation, terrain roughness, distance to coast, landlocked and island dummies. The robustness check is to replace continents with WB regions, and to add dummies for Australia, Finland, Japan, and Korea.

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

This replicates column 2 of Table 6, a regression of Efficiency of Judicial System on a civil law dummy, controlling for GDP per capita. The robustness check is add a rich country dummy for whether GDP per capita in 1998 was above $10,000.

Nunn and Puga (2012) Ruggedness: The Blessing of Bad Geography in Africa

We analyse the final column of Table 1 where income is regressed on the interaction between terrain ruggedness and a dummy for African countries with geographical controls added. Malaria is added as a control.

Nunn and Qian (2011). The Potato's Contribution to Population and Urbanization

Here we replicate the regression in Table 4 Column 1 of Nunn and Qian (2011) which regresses a country's population from 1100 to 1900 on the area of land suitable for potato cultivation multiplied by a dummy for years after 1700, the assigned start of potato cultivation. As a robustness check Europe is interacted with a post 1700 dummy.

Schulz et al. (2019) The Church, Intensive Kinship, and Global Psychological Variation

This replicates the regression of individualism on kinship intensity from the top line of Table 2. The robustness check is a dummy for Europe and Central Asia, East Asia and Pacific, and North America.

Spolaore and Wacziarg (2009). The Diffusion of Development

We examine the baseline regression of per capital income on a measure of a country's genetic distance from the United States in the first column of Table 1. In doing this we use their updated measure of genetic difference, and GDP per capita for 2000: the regression results are more or less unchanged from those reported for the original study although the sample size is somewhat larger. The robustness check is a dummy for Sub-Saharan Africa and South Asia, and for 41 degrees above the equator, the latitude of Italy.

C.2 Africa and India

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

We consider the regression in Table 1 on how local suitability for tsetse flies affects the presence of slavery, controlling for climate variables and clustering by district. The results here were unchanged by the robustness checks applied to other studies.

Banerjee and Iyer (2005). History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India

We take the regression in column 1 of Table 3 of log yield of 15 major crops on the share of land controlled by landlords, alongside geographical controls and how long the area was under British rule. The control region is the most northerly 33 (out of 166) districts, above 27 degrees of latitude.

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

Michalopoulos and Papaioannou (2013) examine the extent to which modern regional development, measured by satellite images of night time luminosity, is affected by the degree of pre-colonial political centralization which ranges from stateless societies at zero to large states at four. We examine the baseline regression of night time luminosity on pre-colonial political centralization in Column 1 of Table 2. As a robustness check, a quadratic in absolute latitude is added.

Michalopoulos and Papaioannou (2016). The Long Term Effects of the Scramble for Africa

We consider the negative binomial regression in Column 1 of Table 2 which regresses the number of all violent incidents on an indicator variable of whether the homeland is split alongside political and geographical controls. The control region here is Somalia.

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

We consider a regression of GDP 2000 on intensity of slave exports, relative to a country's area from Table 3, with log of coastline length relative to country area, original colonial power, and log average oil per capita included as controls. As in the paper's original checks, North Africa and small offshore islands are omitted. For robustness, a dummy is added for where the share of population at risk of malaria exceeds 0.5 and also for the heavy outlier of Democratic Republic of the Congo.

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

Column 1 of Table 2 regresses individual's trust of neigbours on slave exports relative to geographic area with controls for individual and district factors and a country fixed effect. The control region is the three modern states—Benin, Ghana, and Nigeria—that lie along the Slave Coast of the Bight of Benin.

C.3 Europe and the Americas

Acharya, Blackwell and Sen (2016) The Political Legacy of American Slavery

We reproduce column 2 of Table 2 which is a regression of percentage Democratic supporters on 1861 slavery, using suitability for cotton as an instrument for slavery, adding state dummies and geographical controls. Controls applied to similar studies had no impact on the regression results.

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

I reproduce the regression of 1936 rental prices on whether the area lay within the catchment area of the Broad Street cholera pump using a wide bandwidth (this had the highest explanatory power in the study) in column 4 of Table 3, including the full set of controls originally used. The robustness check is longitude plus latitude and an interaction term.

Becker and Pascali (2019): Religion, Division of Labor and Conflict: Anti-Semitism in German Regions over 600 Years

We analyse column 2 of Table 2 which is a panel regression, by century, of expulsions or killings of Jews on the interaction between being Protestant in 1546 and post-Reformation centuries, with controls for the presence of Jews. The control is to remove Brandenburg and Hesse.

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

We analyse column 2 of Table 2 where literacy across Prussian counties in 1871 is regressed on the percentage of the population that is Protestant with a variety of demographic controls. Across most of Germany there appears to be a weak relationship between religion and literacy, except in the extreme east where literacy was low among all groups, but Catholics especially. To deal with this, the control variable adds a regional dummy, dividing Germany into four east west groups at 7, 12, and 16 degrees of longitude.

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

We examine the regression in column 1 of Table 2, which compares equivalent household consumption in a hundred kilometre strip on either side of the Mita boundary with controls for distance to the boundary, elevation, slope and household characteristics. As a robustness check, latitude and latitude squared are added to control for the strong north-south trend in consumption.

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

We analyse column 2 of Table 2 which regresses modern literacy rates on distance from a Jesuit mission, with geographic controls and state fixed effects. None of the robustness checks affected the results materially.

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

Here we examine the regression in column two of Table 4 of Nazi vote share in 1928 on pogroms, with controls for city population, and the percentage of Protestants and Jews. The locational control is to exclude 5 constituencies in Bavaria and one in neighbouring Baden where vote share was above 20 per cent.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP19/31 Judith M Delaney and Paul J Devereux: 'The Effect of High School Rank in English and Math on College Major Choice' December 2019
WP19/32 Tensay Hadush Meles, Lisa Ryan and Sanghamitra Mukherjee: 'Heterogeneity in Preferences for Renewable Home Heating Systems' December 2019
WP20/01 Cormac Ó Gráda: 'Economic History: 'An Isthmus Joining Two Great Continents'?' January 2020
WP20/02 Margaret Samahita and Håkan J Holm: 'Mining for Mood Effect in the Field' January 2020
WP20/03 Vincent Hogan and Patrick Massey: 'Professional Rugby on the Celtic Fringe' January 2020
WP20/04 Sanghamitra Chattopadhyay Mukherjee, Séin Healy, Tensay Hadush Meles, Lisa Ryan, Robert Mooney, Lindsay Sharpe and Paul Hayes: 'Renewable Energy Technology Uptake: Public Preferences and Policy Design in Early Adoption' February 2020
WP20/05 Ronald B Davies and James R Markusen: 'The Structure of Multinational Firms' International Activities' February 2020
WP20/06 Morgan Kelly: 'Direct Standard Errors for Regressions with Spatially Autocorrelated Residuals' March 2020
WP20/07 Anna Bayona and Oana Peia: 'Financial Contagion and the Wealth Effect: An Experimental Study' March 2020
WP20/08 Orla Doyle: 'Can Early Intervention have a Sustained Effect on Human Capital?' April 2020
WP20/09 Supriya Kapoor and Oana Peia: 'The Impact of Quantitative Easing on Liquidity Creation' April 2020
WP20/10 Judith M Delaney and Paul J Devereux: 'Choosing Differently? College Application Behaviour and the Persistence of Educational Advantage' May 2020
WP20/11 Jonathan Briody: 'Parental Unemployment During the Great Recession and Childhood Adiposity' May 2020
WP20/12 Tensay Meles, Lisa Ryan, and Joe Wheatley: 'COVID-19 and EU Climate Targets: Going Further with Less?' May 2020
WP20/13 Xidong Guo and Sarah Parlane: 'Addressing Private Practice in Public Hospitals' May 2020
WP20/14 Patrick Massey and Moore McDowell: 'EU Competition Law: An Unaffordable Luxury in Times of Crisis?' May 2020
WP20/15 Judith M Delaney and Paul J Devereux: 'How Gender and Prior Disadvantage predict performance in College' May 2020
WP20/16 Morgan Kelly, Joel Mokyr and Cormac Ó Gráda: 'The Mechanics of the Industrial Revolution' June 2020
WP20/17 Morgan Kelly and Cormac Ó Gráda: 'Connecting the Scientific and Industrial Revolutions: The Role of Practical Mathematics' June 2020
WP20/18 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' June 2020
WP20/19 Ronald B Davies, Yutao Han, Kate Hynes, and Yong Wang: 'Competition in Taxes and IPR' June 2020
WP20/20 David Madden: 'The Socioeconomic Gradient of Cognitive Test Scores: Evidence from Two Cohorts of Irish Children' June 2020
WP20/21 Deirdre Coy and Orla Doyle: 'Should Early Health Investments Work? Evidence from an RCT of a Home Visiting Programme' July 2020
WP20/22 Cormac Ó Gráda: 'On Plague and Ebola in a Time of COVID-19' August 2020

UCD Centre for Economic Research
Email economics@ucd.ie