The Standard Errors of Persistence

Preprint · June 2019
DOI: 10.13140/RG.2.2.26903.21922

Morgan Kelly
University College Dublin

The Standard Errors of 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. However, alongside unusually high t statistics, these regressions display severe spatial autocorrelation in residuals, and the purpose of this paper is to examine whether these two properties might be connected. We start by running artificial regressions where both variables are spatial noise and find that, even for modest ranges of spatial correlation between points, t statistics become severely inflated leading to significance levels that are in error by several orders of magnitude. We analyse 27 persistence studies in leading journals and find that in most cases if we replace the main explanatory variable with spatial noise the fit of the regression commonly improves; and if we replace the dependent variable with spatial noise, the persistence variable can still explain it at high significance levels. We can predict in advance which persistence results might be the outcome of fitting spatial noise from the degree of spatial autocorrelation in their residuals measured by a standard Moran statistic. Our findings suggest that the results of persistence studies, and of spatial regressions more generally, might be treated with some caution in the absence of reported Moran statistics and noise simulations.
Keywords: Persistence, Deep Origins, Spatial Noise.

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 showing how the adoption of plough determines

*University College Dublin, CAGE and CEPR.

current women's rights; how varying patterns of administration under the British Raj affect regional development in India; how medieval pogroms prefigured Nazi zealotry; how the slave trade still retards modern African development; and how colonial boundaries still drive poverty in Peru and 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 *t* statistics towards two, there would appear to be no way that the *t* statistics of four, five, 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 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 things are more related than distant things." Spatial data, in other words, tend to be autocorrelated. In fact, for the persistence regressions considered here the degree of spatial autocorrelation in residuals—measured by a Moran test, the two dimensional analogue of Durbin-Watson—tends in most cases to be extreme. Our goal is to examine whether these two properties of persistence regressions—unusually high *t* statistics combined with severe spatial autocorrelation of residuals—might in fact be connected.

To understand whether such a connection exists, we begin by simply carrying out synthetic regressions of spatial noise variables. We generate two sets of spatial noise, take their values at a common set of sites, and regress these values on each other. What we find is that even for modest ranges of spatial correlation in the spatial noise, the empirical distribution of regression *t* statistics becomes severely distorted.

Take 200 sites spread at random over a unit square. If we generate noise where the correlation of values between points largely disappears after a distance of only 0.1, then about 20 per cent of synthetic regressions will return a *t* statistic above 2, a nominal significance level of five per cent. The correct five per cent significance

[1]These are, in turn, Alesina, Giuliano and Nunn (2013), Banerjee and Iyer (2005), Voigtländer and Voth (2012), Nunn (2008), Dell (2010), and Michalopoulos and Papaioannou (2016).

level is 2.8, which has a nominal p value of 0.005: the significance level is already in error by one order of magnitude.

If we extend the correlation range to 0.3, nearly half of noise regressions will return a *t* statistic above 2, and a quarter above 3. In order to reach the correct five per cent significance level requires a *t* 5.5, something with a nominal significance level of $p = 4 \times 10^{-8}$.

When places are not scattered randomly across the landscape but instead clump together, as they tend to do in the real world, the distortions are more severe yet. If we take historical African tribal areas and suppose that correlation has disappeared in only 1,000 kilometres (about the distance across Nigeria), almost one quarter of noise regressions will return a *t* statistic above five. So too will regressions using interwar German towns, when the noise variables have a correlation range around 300 kilometres: see Figures 2 and 3 below.

The intuition behind what causes these inflated *t* statistics is simple. Take some towns dotted across a landscape and represent their average income levels as elevation on a map. If there is little correlation between neighbours, rich will border poor, leading to a jagged topography. However, as Tobler observed, usually the correlation is long range so affluent places are surrounded by other affluent ones, leading to a gently rolling landscape. Now take some other unrelated variable from the past, say trials for heresy in the middle ages. Again neighbour will resemble neighbour, leading to another rolling landscape. If we regress these variables on one another, 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. When the towns cluster together in a few geographical areas the probability of coincidence will be corresponding greater.

Statistics is the exercise of extracting structure from data. The difficulty with spatial noise is that because the value of each point is correlated with nearby points it can appear to display a lot of order (see the simulations in Figure 1 below) making it deceptively easy to uncover structure in spatial noise and mistake it for a deeply informative signal.

In positive terms, what can be done to rule out the possibility that a high *t* statistic in a persistence regression arises not because the study has unearthed a deep historical truth, but because it happened to fit some spatial noise instead? One approach, adopted by several studies, is to adjust standard errors along the lines suggested by Conley (1999). We will see that although the Conley procedure does lead to substantial falls in estimated *t* statistics (so long as the cutoff radius is set at considerably longer levels than are typically used in the literature), it still tends to return excessively large values when observations are spatially clustered, and the empirical significance levels vary according to the spatial pattern of each set of observations.[2]

To remedy potential difficulties arising from spatial correlation, we outline instead a simple two step procedure, that we then apply to existing persistence studies. The first step is to compute a Moran statistic, the spatial analogue of the Durbin-Watson test. Like Durbin-Watson, this statistic turns out to be an useful indicator of potential misspecificiation: in studies where the Moran statistic is insignificant at conventional levels regression results tend to be robust; whereas large Moran statistics are reliable warnings that nominal significance levels differ substantially from true ones.

The second step is to employ standard methods from geostatistics to generate synthetic spatial noise that roughly matches the correlation structure of the variables of interest. These noise variables can then be used as artificial explanatory or dependent variables in persistence regressions in place of the original variables.

To simply permute observations at random—a familiar placebo generation procedure in econometrics—is to lose the spatial correlation structure that lies at the root of inflated t values. Suppose, for instance, that we have some variable that strongly explains GDP per capita across the world, and to test its robustness, we take Brazil and Argentina and randomly reassign Switzerland's income to one and Rwanda's to the other. The exercise is pointless because all of the spatial correlation driving potential spurious significance has been washed out. Instead of randomly reordering variables we need to swirl their values around in such a way that each point gets a new value similar to its neighbours' ones, something that happens when we generate spatially correlated noise.

As well as replacing the dependent variable, we can go the other way and replace the explanatory variable to see how often the persistence variable is outperformed by noise. Although this two step approach may be applied to any spatial data, our concern here is with studies of persistence.

We examine 27 persistence studies published in four journals: *The American Economic Review*, *Econometrica*, *The Journal of Political Economy*, and *The Quarterly Journal of Economics.* Each of these papers is a careful and lengthy statistical exercise that we do not attempt to replicate in full. Instead we analyse the "leading" regression of the paper, usually reported in the first column of Table 2 or 3, that establishes the strength of the relationship between the modern outcome and the historical variable, before control variables are added in subsequent columns.

[2]At the same time the common procedure of clustering standard errors is simply invalid in the presence of non-negligible spatial autocorrelation: residuals in neighbouring clusters will tend to be correlated, violating the basic assumption underlying the procedure.

The sole concern of this paper is with possible distortions to significance levels arising from fitting spatial noise. 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 substantial.

Above all, and this cannot be emphasized this too strongly, this paper is not concerned with somehow validating or "disproving" the findings of individual studies. In fact, we are not interested in any individual result except insofar as it illustrates the broader contours of the literature. The fact that the persistence variable performs poorly against spatial noise in the first regression of a paper does not imply that later regressions in the paper necessarily suffer from the same difficulties. Rather than the negative and rather pointless goal of discrediting existing research, the purpose of this study is the positive one of marking out one potential pitfall in persistence regressions, and in spatial studies more widely, and to outline simple measures to avoid it.

The focus throughout is on significance levels. It is increasingly recognised that, by diverting attention from effect sizes and the uncertainty attaching to them, the traditional econometric approach of hypothesis testing where a variable is either "significant" or "insignificant" can be extremely misleading. Coefficients with several stars may have impacts that are negligible, whereas insignificant variables can have large effects (if a confidence interval runs from -1 to 5 the coefficient is as likely to be 4 as zero) but their size cannot be tied down with the data at hand. However, because the focus of the persistence literature is largely on p values, our concern here will be almost entirely with significance levels.

We find that only about one quarter of the persistence results that we examine are robust after we take account of possibility that their regressions might be fitting spatial noise. We can predict how robust a result is to spatial noise from the degree of spatial autocorrelation in its regression residuals, measured by a Moran statistic.

In terms of existing literature, our findings that a class of results are not always robust are most closely related to Bertrand, Duflo and Mullainathan (2004) on difference-in-differences; Young (2018) on instrumental variables; and Gelman and Carlin (2014) on effect sizes in social psychology; and has clear parallels with Granger and Newbold (1974) on spurious regressions in time series. Several of the studies considered here are analysed separately, in terms of their robustness to omitted variable bias by Oster (2019). However, this appears to the first study to look at distortions of regression significance levels arising from spatially autocorrelated data.

The rest of the paper is as follows. The next section presents simulations to illustrate the distortion of regression *t* statistics that occurs at even modest ranges of spatial autocorrelation, especially when observations cluster geographically, and the section after analyses differences across historic frontiers. Having seen the large inflation in *t* statistics that can occur by fitting spatial noise we outline positive steps to remedy these difficulties. The remainder of the paper analyses some persistence studies in leading journals in the light of our findings.

2 The Significance of Spatial Noise

As we stated above, our approach to investigating whether *t* statistics are inflated in persistence regressions is straightforward. We simply generate two spatial noise patterns, take their values at a common set of points and regress these values on each other. Repeating these simulations we can derive the empirical distribution of the regression *t* statistic.

The statistical study of spatially correlated processes originated in mineral prospecting where samples taken at several points are interpolated to map the overall concentrations of ore and determine the best place to excavate a mine. At its simplest, spatial interpolation or Kriging assumes that value of observations $(X(s_1), X(s_2), \ldots, X(s_N))$ taken at sites $(s_1, s_2, \ldots, s_N)$ are normally distributed with covariance matrix Σ whose elements are controlled by a correlation function ρ: for points i and j a distance $\|h\|$ apart

$$\Sigma_{ij} = \operatorname{cov}(x_i x_j) = \sigma^2 \rho(\|h\| / \phi)$$

where the range parameter ϕ determines how fast correlation decays with distance. So for exponential correlation we have $\rho(h, \phi) = \exp[-(\|h\| / \phi)]$ and for Gaussian we have $\rho(h, \phi) = \exp\left[-(\|h\| / \phi)^2\right]$.

The most widely used covariance matrix is the Matérn function which has a shape parameter κ that determines how fast correlation falls off with the range: low values correspond to exponential falloff, and high values to Gaussian: see Brown (2015, 23). For that function ϕ is the distance where $\rho(\phi) \approx 0.14$, irrespective of κ. For a given set of empirical observations these parameters, along with σ, can be estimated by maximum likelihood.

The Matérn function can be extended to allow for anisotropy where the degree of correlation changes with direction. Specifically one can estimate an anisotropy

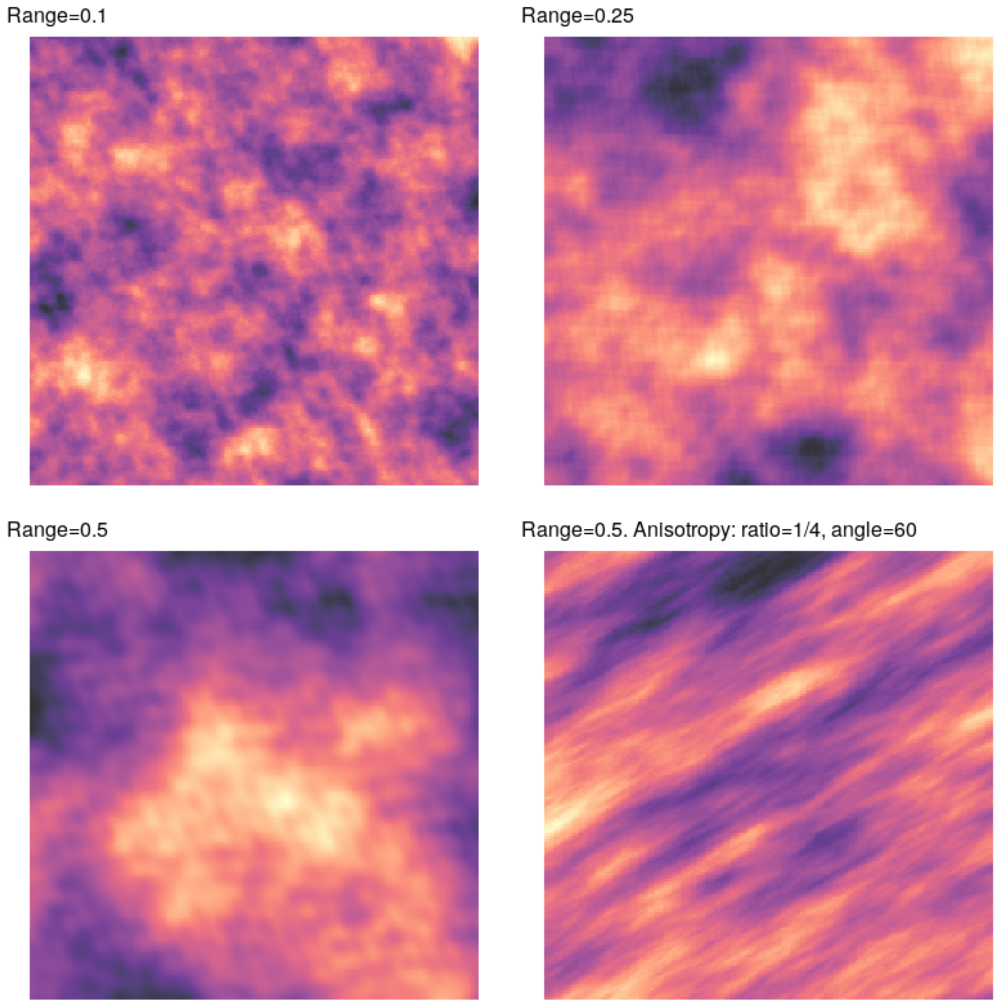

Figure 1: Spatial noise with correlation ranges of 0.1, 0.2, 0.5. In the last panel, the correlation range in the x direction is 0.5, but one quarter of this in the y direction, and the main axis of correlation is at 60 degrees clockwise.

ratio giving the strength of north-south correlation relative to east-west correlation, and the angle where correlation is maximized.

A useful way to get an idea of how patterns of spatial correlation change with range is by looking at simulations of landscapes as in Figure 1 : we can think of bright zones as rich areas and dark zones as poor ones, for example. In all cases we set the variance σ^2 and shape parameter κ to one. The figure shows simulations of noise over a unit square with ranges of 0.1, 0.25, and 0.5.[3] To put these ranges in perspective, modern Germany is roughly 600 kilometres square, so a range of 0.25 implies that correlation between places has largely disappeared after only 150 kilometres.

We can see that already at a range of 0.1 a good deal of spatial structure has already emerged, and by 0.5 the large light and dark patches have a high chance of matching the patterns in another simulation. In the fourth panel the landscape is anisotropic, with a correlation range of 0.5 in the x direction but only one quarter of this in the y direction, and the main axis of correlation runs at 60 degrees clockwise. Naturally, the range can be extended past one but the resulting pattern will be almost entirely bright or dark. The final panel shows a case of anisotropic noise where correlation strength vertically is twice that horizontally, and the main axis of correlation runs southeast to northwest.

2.1 Empirical *t* Statistics

The substantial distortions in regression t statistics induced by even moderate ranges of spatial correlation are shown in Figure 2. We start with a set of 200 random sites and generate pairs of noise patterns. We take the value of each at every point and regress these values on each other. The Figure displays the absolute values of t statistics of these regressions as the correlation range of the Matérn function runs from 0.01 to 0.6. Each line traces an empirical significance level of 0.5, 0.25, 0.1, and 0.05. For a correlation range of 0.01 these significance levels are only slightly above the nominal levels: 5 percent of observations will have a t statistic of 2, and so on (the slightly greater value reflects the fact that there is some small correlation beyond the range).

However, as the range of correlation increases, regression t statistics inflate rapidly. At a range of 0.1 about 20 per cent of noise regressions will return a t statistic above 2, and the true five per cent significance level is 2.8. This has a nominal significance level of 0.005: the nominal significance level already errs by one order of magnitude.

At a range of 0.3 it can be seen from Figure 2 that nearly half of spatial noise regressions will return a t statistic above 2, and a quarter will have a t statistic above 3 while, for a regression to be significant at 5 per cent calls for a t value of 5.5 (nominal $p = 3.8 \times 10^{-8}$).

[3] All spatial analysis here relies on the excellent `geostatsp` R package of Brown (2015).

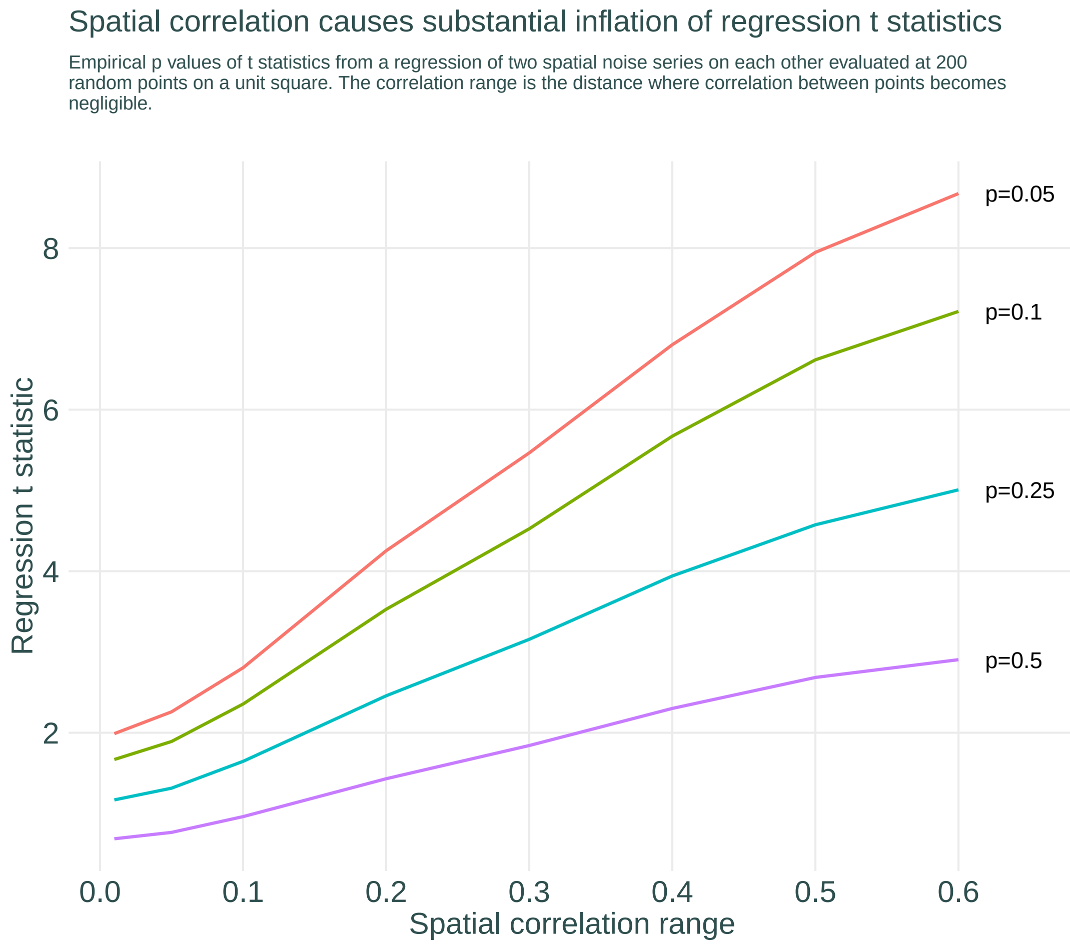

Figure 2: Empirical significance levels of t statistics from spatial noise regressions for different maximum ranges of spatial correlation. It can be seen that at a correlation range of 0.3, one quarter of regressions will return a t statistic above 3.3 (p=0.001), but that a regression requires a t statistic of 5.5 to be significant at 5 per cent.

2.2 Spatially Clustered Observations

In carrying out these simulations we assumed that the sites where we take measurements are randomly distributed in space, whereas in the real world places tend to cluster geographically. For instance, half of the capital cities in the world are within 400 kilometres of another capital, and two thirds within 500, whereas if they were randomly scattered across the globe (oceans included) only one eight and one quarter respectively should lie within these ranges.

The simplest way to model and estimate clustered observations is with a Thomas process where imaginary "parent" points are randomly distributed with Poisson parameter μ and each of these generates a number of observable "child" points with Poisson parameter λ. The distance of each child from the parent follows a normal distribution with standard deviation σ. As λ rises relative to μ, and as σ falls, clustering will rise.

For example, the 229 towns in interwar Germany analysed by Satyanath, Voigtländer and Voth (2017) lie in a roughly 8 by 7 degree rectangular box with observations clustered along the western edge. The estimated parameters of a Thomas process for these points are $\mu = 14.48, \lambda = 0.27, \sigma = 0.42$. Generating locations at random using these parameters, and regressing noise variables with a correlation range of 0.3 times the box width gives t statistics above 3 in 55 per cent of spatial noise regressions—twice as many as when the observations are randomly scattered—and the 5 per cent significance level corresponds to a t statistic of 9.94 compared with 5.5 for unclustered data.

Naturally, with three parameters there is an unlimited number of potential Thomas processes we can simulate, so it is more informative to look at t statistics of noise regressions carried out with real data, specifically three frequently used datasets: countries of the world, historical African tribal areas; and interwar German towns.

Countries of the world are located by capital city in 2001 taken from the Correlates of War database. Only observations between -90 and 155 degrees longitude are considered (so that neighbouring places with longitudes of -179 and 179 are not treated as distant: this turns out not to matter in practice) to give 178 observations. The locations of 522 historical African tribal areas are taken from Alsan (2015); and the German towns previously mentioned are from Satyanath, Voigtländer and Voth (2017).

Figure 3 shows empirical 25 per cent significance levels of spatial noise regressions over these locations, where the range of the noise process is given in degrees. One degree at the equator is approximately 110 kilometres so degrees here can be thought of as roughly 100 kilometres. For world capitals spread over 240 degrees of longitude, 40 degrees in the diagram corresponds to 0.17 for for the randomly distributed locations on a unit square in Figure 2 but gives a t statistic above 3.5 compared with 2.5 earlier: a significance level of $p = 0.0005$ compared with $p = 0.01$. For the other two datasets the inflation in t statistics is still more extreme.

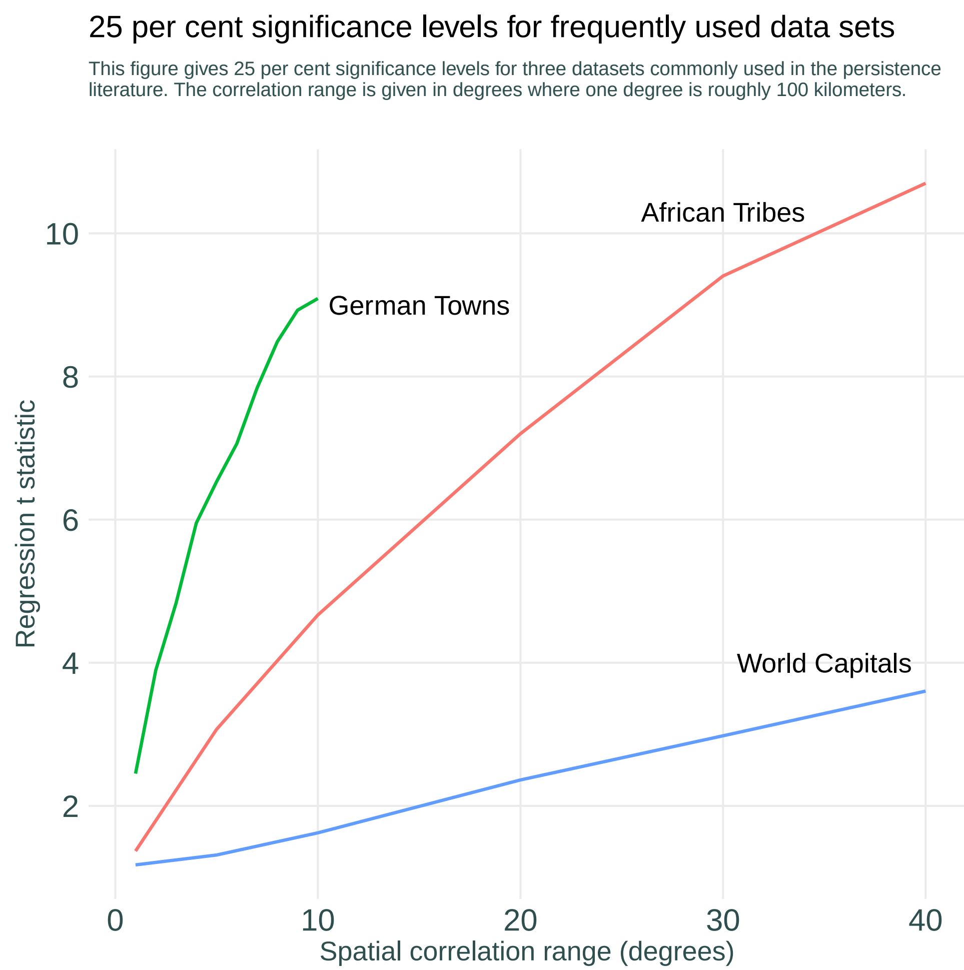

Figure 3: Empirical 25 per cent significance levels of t statistics of spatial noise regressions measured at 178 world capitals, 229 interwar German towns, and 225 African tribal areas. Correlation range is measured in degrees.

For Africa, we can see that at a correlation range of only 10 degrees (about the distance across Nigeria, or one fifth of the trip from Nairobi to Lagos), one quarter of spatial noise regressions will return a t statistic above 4.7 (nominal $p = 8 \times 10^{-6}$). For Germany, as the Thomas process above suggested, the distortion is more marked. At a correlation range of 200 kilometres (Frankfurt to Cologne), almost one quarter of noise regressions will generate t statistics above 4 (nominal $p = 1 \times 10^{-4}$).

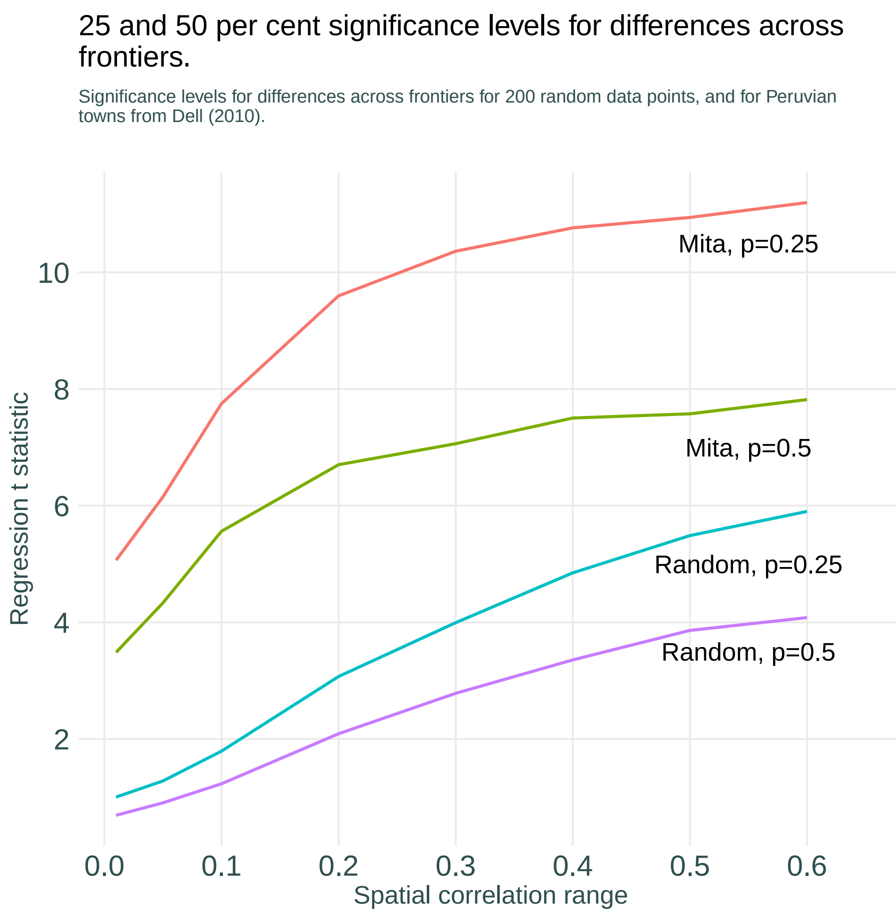

Figure 4: Empirical significance values for differences across a frontier, both for random points, and Peruvian households across the Mita boundary. Generating artificial consumption data with spatial noise, 50 per cent of regressions for differences across the Peruvian Mita boundary will return t statistics above 7.5.

3 Institutional Boundaries.

An approach to examining the enduring importance of historical institutions is to examine modern outcomes on either side of an earlier border. Naturally, the break must be arbitrary and not correspond to natural boundaries where soil fertility, altitude and other geographical characteristics may differ.

Again we begin by generating 200 sites on a unit square, with a vertical border drawn halfway across. We can then run a regression of the value at each point, including a dummy for one side of the border. From Figure it can be seen that the distortions are even more severe than in standard persistence regressions. For a correlation range of 0.3, half of all noise regressions will return t statistics above 2, and one quarter above 3.

The other lines in Figure 4 use as observations the town's on either side of the Peruvian Mita boundary from Dell (2010). These towns are inside a 4 by 2.5 degree box, which is renormalized to be one unit across. In this case, if spatial noise with a range of 0.3 is used for household consumption, half of all regressions will detect a difference across the frontier that has a t statistic above 7.5.

4 Remedies for Spatial Distortions

Given the severe inflation of t statistics associated with even modest ranges of spatial correlation demonstrated we have seen in Figures 2 and 3, how can we assure ourselves that a persistence regression (or any other sort of spatial regression for that matter) is not merely fitting noise? We outline a simple two-step procedure to handle potential spurious fits but first examine the Conley adjustment to standard errors.

4.1 Conley Standard Errors

Several papers considered below make a Conley (1999) adjustment to their standard errors to control for spatial correlation. This adjustment, analogous to an adjustment for heteroskedasticity and autocorrelation, consists of calculating a covariance matrix with spatial weights chosen from some kernel, typically a uniform kernel where points within a given distance are assigned a weight of unity, and zero otherwise. Conley (2010) notes how such simple non-parametric kernels typically out-perform parametric ones, and that is the case here: although the spatial noise is generated as a Matérn process, using a kernel with the same parameters as the generating process typically returned more distorted t statistics than Conley's uniform kernel with the same range.[4] However, despite the extreme inflation

[4] A common estimation technique in the spatial statistics literature (see, for example, Bivand, Pebesma and Gomez-Rubio 2008, 274–289) is through spatial autoregression (the analogue of a Cochrane-Orcutt procedure in time series). Each regression residual is a weighted average of its neighbours: $e_i = \theta \sum_j w_{ij} e_{ij} + \epsilon_i$ where $w_{ii} = 0$ and $\epsilon \sim N(0, \Sigma)$, leading to a covariance matrix $(1 - \theta W)^{-1} \Sigma (1 - \theta W)^{-1}$. For a variety of weighting schemes the results were similar to applying a uniform kernel and are not reported here. There are in addition a variety of spatial lag models such as $Y = \theta W Y + \beta X$ but their estimates tend to vary substantially depending on the weighting

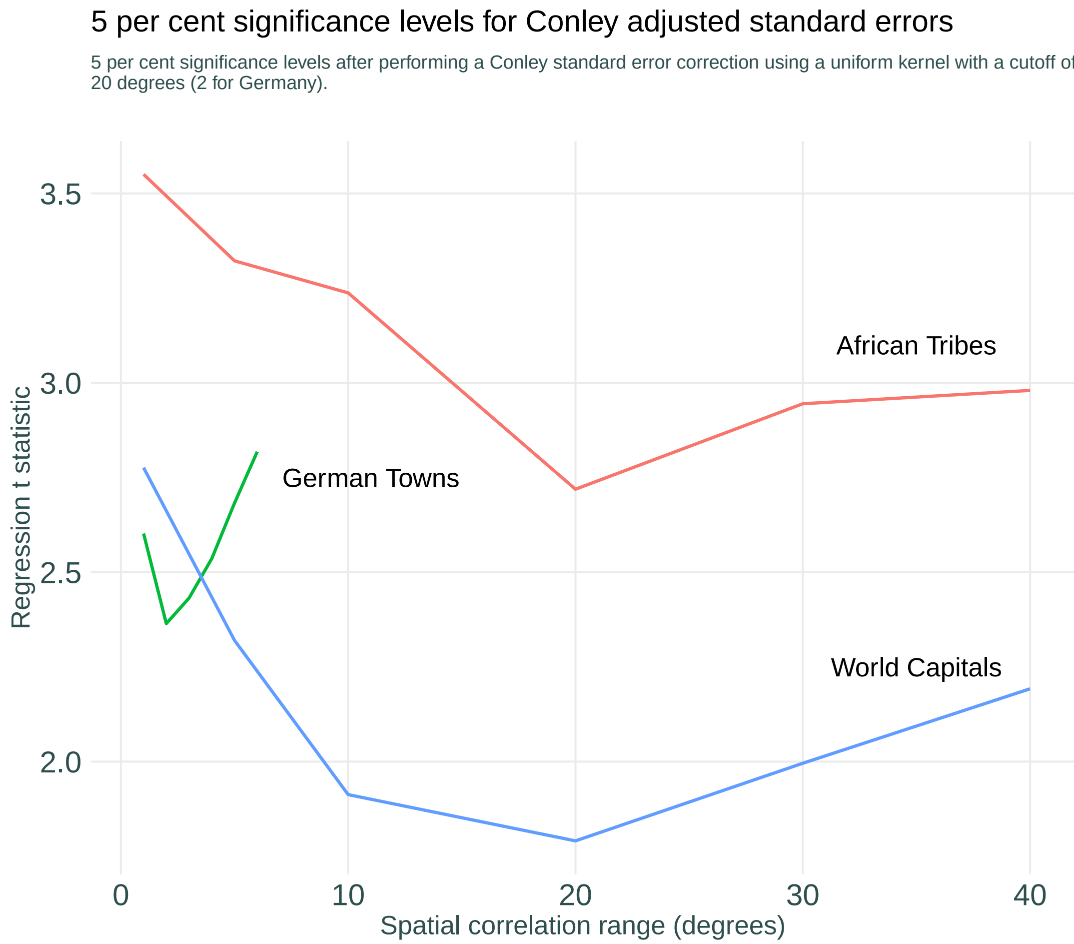

Figure 5: Empirical 5 per cent significance levels for Conley adjusted t statistics of spatial noise regressions measured at 178 world capitals, 229 interwar German towns, and 225 African tribal areas.

in t statistics we saw above, in no persistence study does the Conley adjustment increase the standard errors materially.

The fact that adjusted standard errors are largely unchanged is a consequence of setting an extremely low cutoff radius beyond which observations are given a weight of zero. For example Nunn and Wantchekon (2011) set a radius of three degrees for Africa, which spans about 70 degrees each way; while the global analysis

matrix imposed, and the value of any regression whose specification is driven primarily by a desire to cause potentially informative patterns in its residuals to vanish is unclear.

of Ashraf and Galor (2013) uses a limit of 500 kilometres: less than the distance from London to Edinburgh. Figure 5 shows the 5 per cent significance level of t statistics for the three empirical datasets used in Figure 3 where the cutoff radius is set at 20 degrees for the world and Africa, and one tenth of that for Germany. Increasing the radius to 30 or 40 degrees increased the t statistics somewhat but not by a large amount.

Several things appear from the diagram. The first is that a Conley adjustment, so long as we assign it a sufficiently long range, does cause t statistics to fall substantially compared with the original, uncorrected regressions. For the case of world capitals, the new t statistics are fairly close to the nominal level of 2. However, for Germany and Africa, although the the t statistics do fall considerably, they remain high: around 2.5 for Germany and 3 for Africa. So although Conley corrections provide a useful warning of potential distortions by causing a large reduction in estimated t statistics, the significance levels they return for finite samples are not entirely reliable and vary according to the spatial clustering pattern of the data.

4.2 A Two-Step Procedure

To analyse potential spatial distortions we use here instead a simple two step procedure. First, the simplest warning of misleading of t statistics turns out, as we will see below, to be Moran's I statistic: the spatial analogue of the Durbin-Watson statistic. This is proportional to a weighted sum of the covariance between every pair of residuals, where the weighting scheme to assign distant observations less importance than nearby ones can be chosen in a variety of ways.

A "true" Moran statistic does not exist: for different choices of spatial weighting schemes its value can vary considerably. However, for most of the studies considered here the spatial autocorrelation is so extreme that the choice of weighting scheme is immaterial. We therefore only report results for the case where a point's five closest neighbours (using geodesic distance) are given equal weight and other points have zero weight.

The second step to ensure that a spatial regression is not spurious is to generate artificial noise variables to test the robustness of the claimed results. This is done in two ways. The first is to replace the dependent variable with noise, to test the regression's ability to explain what it should not be able to explain. These placebo procedures are commonplace, but involve permuting existing variables. Such random permutations destroy the pattern of spatial correlation between neighbouring observations that are at the root of the difficulties we have seen here. By generating

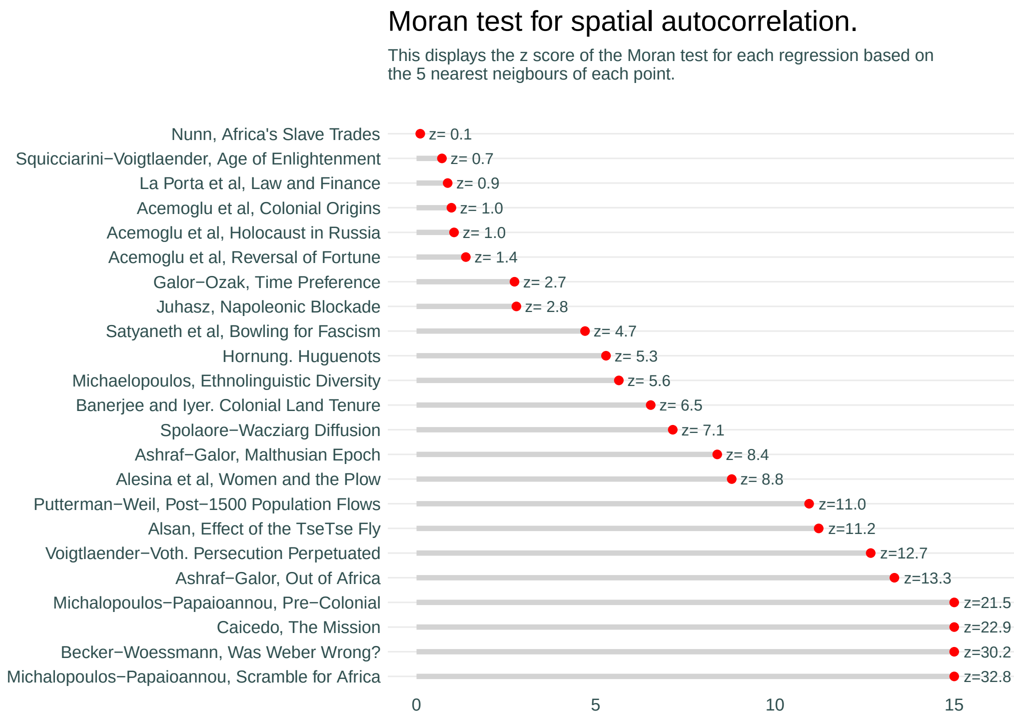

Figure 6: Z scores of Moran tests for spatial autocorrelation in regression residuals.

noise that is spatially correlated, neighbours still have similar values, sometimes high or low, depending on the particular simulation.

The second use of artificial spatial noise is to replace the explanatory variable of interest. By seeing how often random noise outperforms the original persistence variable, we can get some rough idea of its empirical significance level which, as we have seen, can differ by several orders of magnitude from its nominal one. We now apply this procedure to some existing persistence studies.

5 Persistence versus Spatial Noise

Having seen the severe distortions in regression results that stem from spatial correlation among observations, along with some simple steps to check how strong these distortions are, we now analyse some persistence studies that have appeared in major journals. The regression we choose for analysis is usually the first regression of the paper, typically reported in the first column of Table 2 or 3, which includes the explanatory variable of interest along with a few or no control variables. These control variables are added in subsequent columns and typically, but not invariably, cause the significance of the main variable to fall.

We first report Moran statistics to test for the presence of spatial autocorrelation in regression residuals. For large samples the Moran statistic follows a normal distribution, and the associated z scores for each regression are graphed in Figure 6. The diagram shows how only about one quarter of studies return a Moran statistic above two (remember this cutoff is not intended to be an arbitrary "significant/insignificant" one but only suggestive), and in most cases the degree of autocorrelation is severe: a Z score of 13 corresponds to a p value of order 10^{-41}. For analyses using large numbers of people at each site (Dell 2010; Nunn and Wantchekon 2011) Moran statistics are not computed.

5.1 Generating Spatial Noise

Having seen that most of the studies considered here appear to show substantial spatial autocorrelation in residuals, we now examine how robust their results are when we account for this, successively replacing their explanatory and dependent variables with noise. The question then arises of what parameters to use when generating noise.

Naturally, there are no "correct" simulation parameters, only more or less plausible ones: we do not want to base our findings on ridiculous patterns of noise that would never be encountered in the wild. To choose plausible parameters for global studies we compute maximum likelihood estimates of the parameters generating GDP per capita in 2000 (using the figures in Ashraf and Galor (2013)). Correlation range ϕ only appears in the Matérn function in terms of $\sqrt{2\kappa/\phi}$, meaning that correlation range and shape κ cannot be reliably estimated together. Instead, parameters are estimated after setting shape. In carrying out simulations we use a value of 16, associated with a Gaussian pattern of falloff. For income data this led to a 99 per cent confidence interval for range of between 20 and 80 degrees east-west, and half of this north-south. Setting $\kappa = 1$ led to a range between 20 and 240 degrees. In simulating spatial noise we therefore report values for correlation ranges of 30, 60, and 90 degrees.

Another gauge of the plausibility of the noise pattern is to look at typical simulations to see if clusters of high and low values appear too large or small on a map compared with something familiar such as GDP. Figure 7 shows simulations with correlation ranges running from 30 to 120 degrees east-west, and half this north-south, where each country is plotted as an equal sized tile. It can be seen that at 30 degrees, the map is split into small clusters with light areas frequently border-

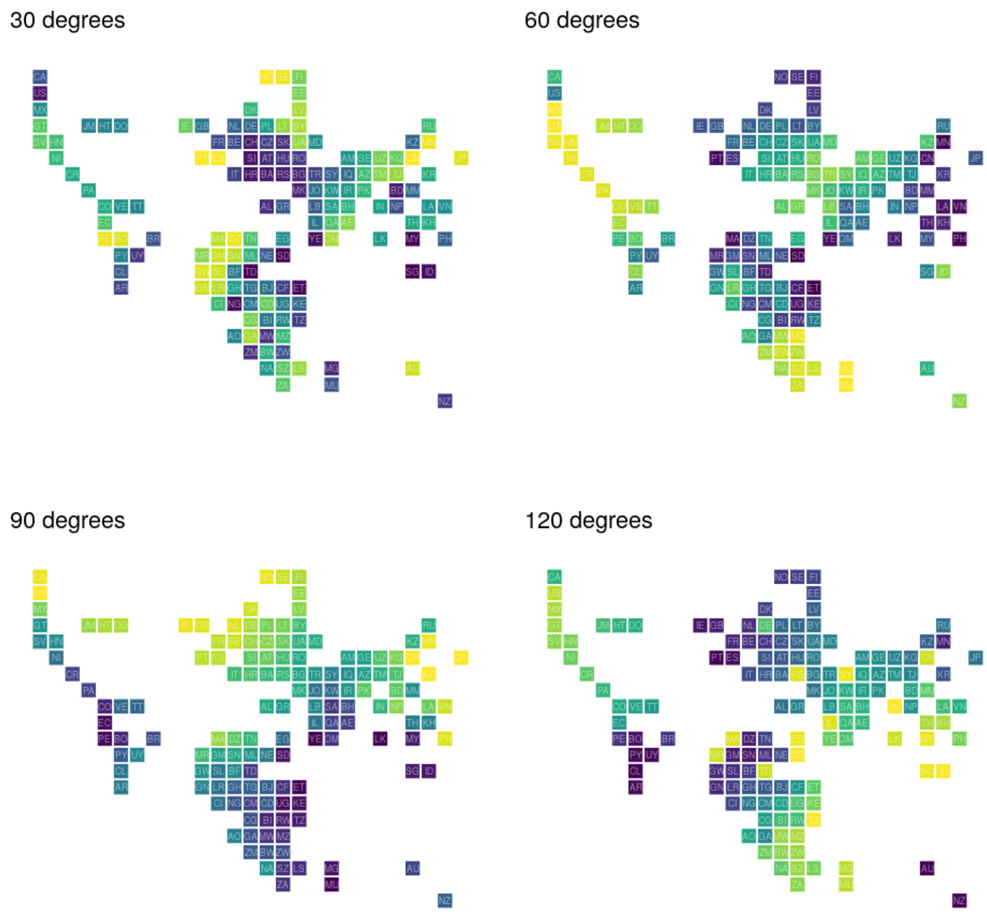

Figure 7: A tile map of simulated world income where synthetic income is generated as spatial noise with correlation range going from 30 to 120 degrees east-west and half this north-south.

ing dark ones. At 60 degrees more homogeneous clusters have emerged, while the particular simulation at 90 degrees came out looking like the actual world distribution with an affluent north and impoverished south. At 120 degrees the clusters are becoming very large but still do not look unrealistic.

Africa spans 70 degrees in both directions. We therefore report results for ranges of 10, 20, and 30 degrees (with north-south ranges again half this). For France, Germany, and South America we give results for 3, 4, and 5 degrees, but for the Peruvian Mita of Dell (2010) we use ranges of 0.5, 1, and 1.5 reflecting the smaller study area. To repeat, we are agnostic about parameter values but the ranges here do not seem too implausible empirically.

In simulating noise we did not include spatial trends: we know for example that income rises as one moves away from the equator. This means that in any one simulation Chad is as likely to be rich as is Switzerland. As a result the performance of noise variables will be attenuated, but we do not wish to introduce the extra complication of choosing spatial trends and causing confusion about whether the results might be driven by these.

5.2 Noise as the Explanatory Variable.

Tables A1–A3 summarize the performance of noise in persistence regressions for global studies, Africa and India, and Europe and South America, using the correlation ranges already mentioned. Each is split in two parts. One is for the case where noise is the explanatory variable, and this tells in what fraction of simulations the noise variable is significant at standard significance levels from 0.05 to 0.0001 as well as how often the noise variable outperforms the original persistence variable by having a higher significance level. The other columns are for the case where noise is the dependent variable, and show how often the persistence variable explains noise at various standard significance levels.

It is more informative if we graph the data, using in each case the results for the middle value of the correlation range in each table: 60 for global studies, 20 for Africa, and 4 for Europe, with the north-south ranges being half this. The performance of noise explanatory variables compared against the original persistence variable is shown in Figure 8. This gives the original significance level of the persistence variable (or their joint significance level if there are several as in Ashraf and Galor, 2013 and Michalopoulos, 2012). A logarithmic axis truncated at 10^{-9} is used to cope with the extremely high significance levels of some studies in this literature.

Beside it is graphed the proportion of simulations where the noise variable had higher explanatory power than the original persistence one. In some rough sense this measure of outperformance corresponds to an empirical p value of the sort we had in Section 2, however in this case the noise has no claim to be the "true" value, especially because we have omitted any spatial trends in generating it.

The pattern of results is revealing. Studies that had low Moran statistics in Figure 6 all cluster towards the top of the graph, rarely being out-performed by spatial noise. Interspersed with them are some studies with high Moran statistics

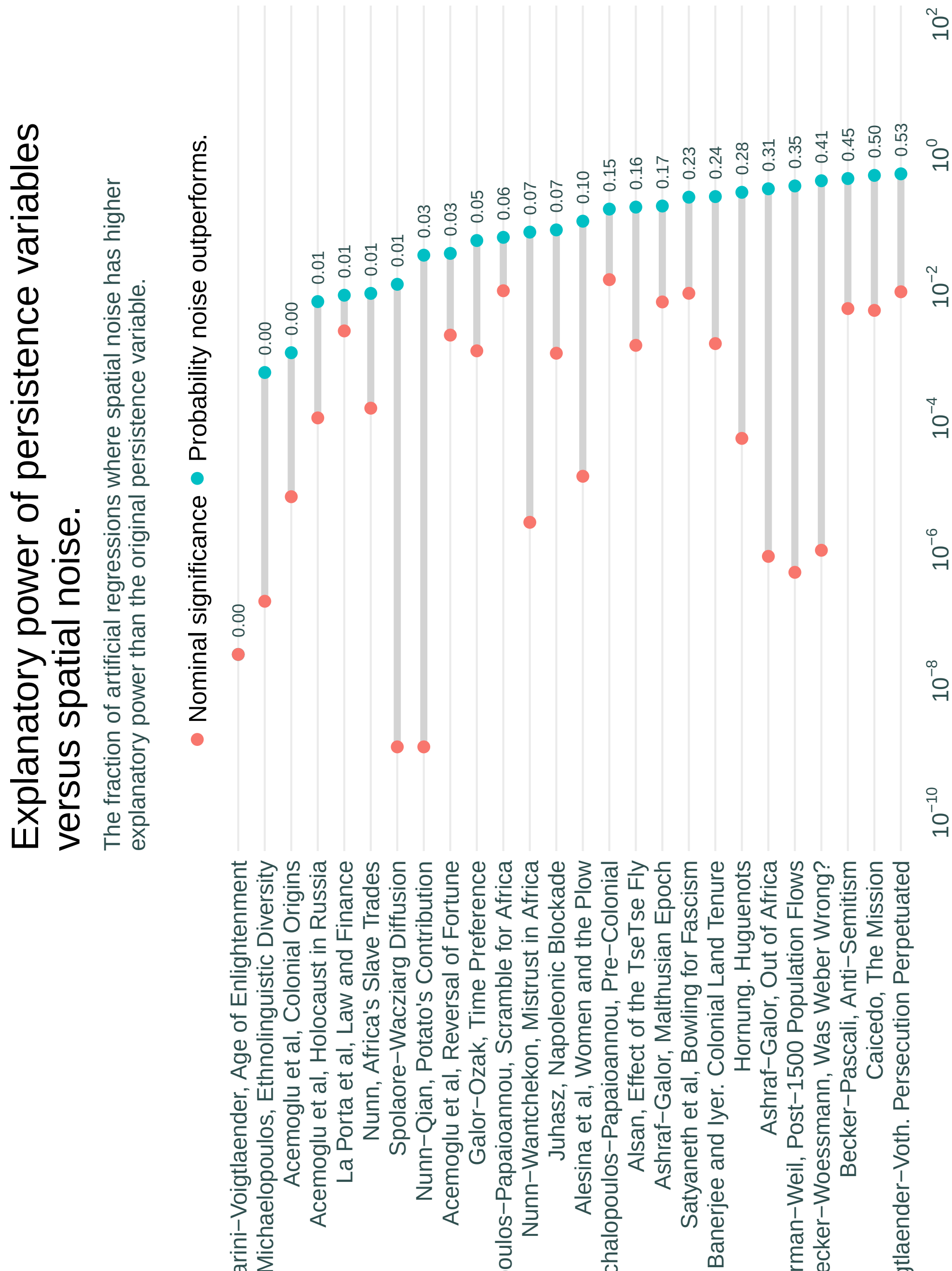

Figure 8: Nominal significance levels of persistence studies and the proportion of simulations where they have lower explanatory power than spatial noise.

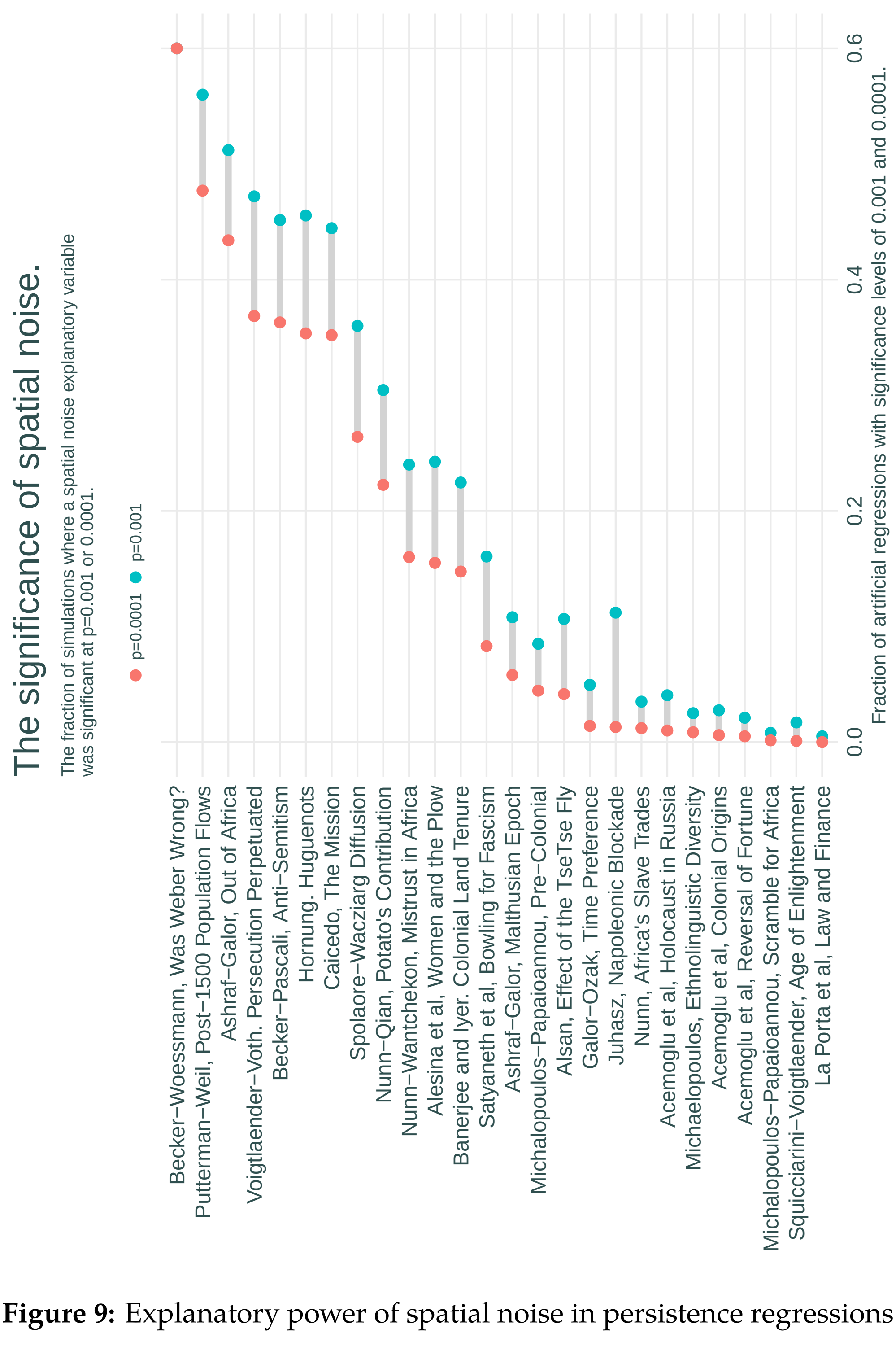

Figure 9: Explanatory power of spatial noise in persistence regressions.

but which began with extreme levels of significance. As one moves downwards into studies with high Moran statistics we can see that it is increasingly common for noise to perform better than the original persistence variable. In particular, several studies with nominal significance levels of order 10^{-6} are outperformed by noise in over thirty per cent of simulations, and this would be higher if the noise were superimposed on spatial trends.

Figure 9 looks at how often a spatial noise explanatory variable, used in place of the original persistence one, is significant at $p = 0.001$ or $p = 0.0001$. It can be seen for one third of studies, spatial noise will be significant at 0.0001 over 20 per cent of the time, and for two thirds of them, noise will be significant at 0.001 in ten per cent of cases. It is interesting that although several studies look at the determinants of modern GDP per capita, the explanatory power of noise varies considerably reflecting the different samples and methodology being used. Acemoglu, Johnson and Robinson (2001) and Acemoglu, Johnson and Robinson (2002) consider fairly small samples, whereas Spolaore and Wacziarg (2009) use most countries in existence. By contrast, Ashraf and Galor (2013) and Putterman and Weil (2010) weight their explanatory variables by a matrix of settler origins, as we therefore do with spatial noise, leading to predictive power that is higher still.

5.3 Noise as the Dependent Variable

Having examined the ability of spatial noise to explain different outcomes, we now turn things around and see how well the persistence variable can explain a dependent variable which is spatial noise. This is shown in Figure 10 which gives the fraction of simulations that returned significance levels of $p = 0.001$ and $p = 0.0001$. Again the tendency is for studies with lower Moran statistics to have less ability to predict noise, alongside regressions whose initial explanatory power given in Figure 8 is comparatively low. At the other end, many persistence variables have unusual power to predict spatial noise, especially those with high explanatory power in the original studies.

One thing that is notable however is that quite a few regressions that returned low Moran statistics have substantial explanatory power for noise. However, in all of these simulations the estimated Moran statistics have now become extremely high: these are noise regressions after all. Moran statistics again differentiate robust relationships from specious ones.

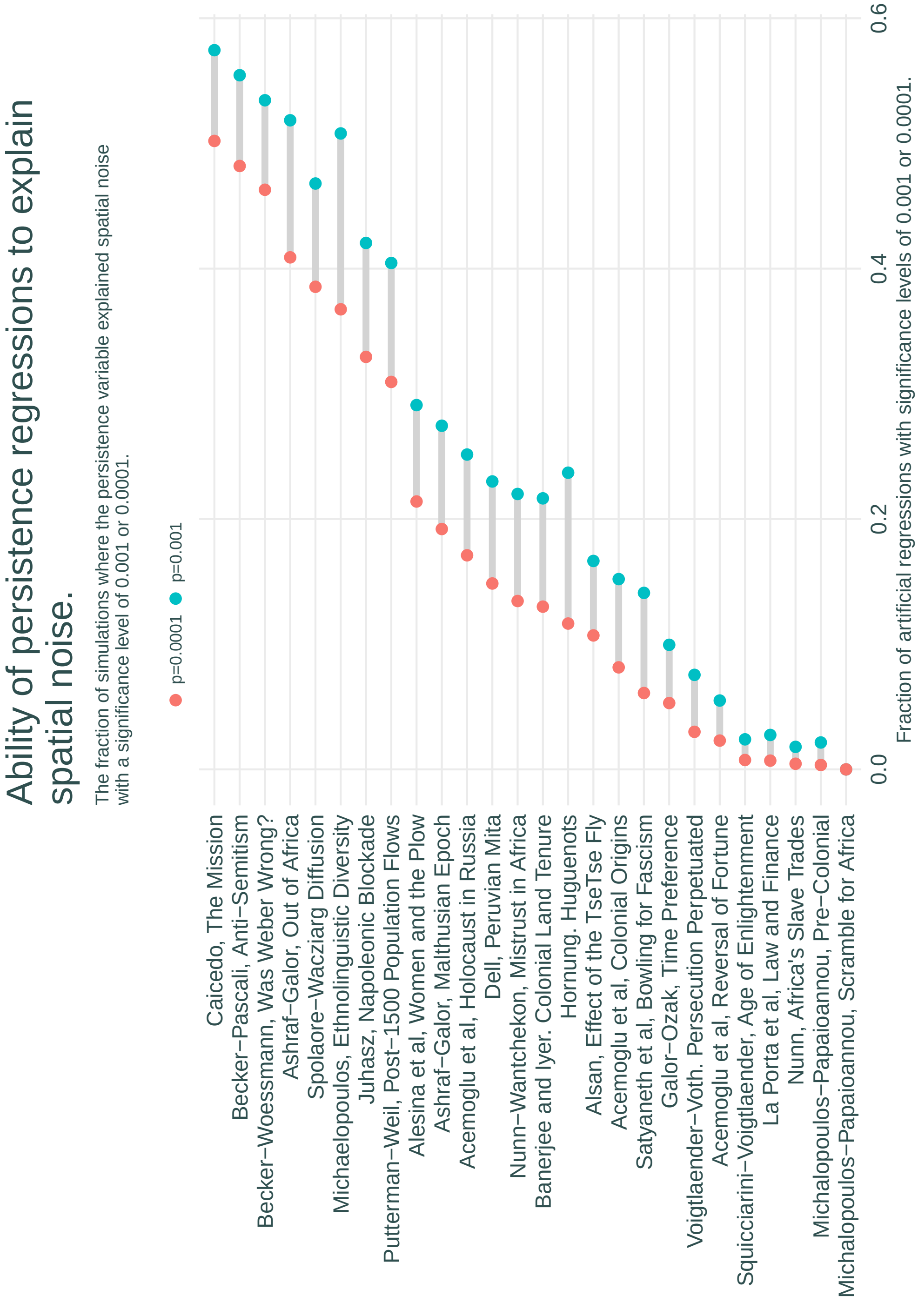

Figure 10: Ability of persistence variables to explain spatial noise.

6 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 Latin America. Because the original studies were conducted in Stata and we use R to take advantage of its strong geostatistical capabilities, we have taken care to ensure that we could exactly replicate the original results, notably their standard errors.

6.1 Global Studies.

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, with mortality truncated at 250 per thousand.

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.

Alesina, Giuliano and Nunn (2013): The Plow and Women's Rights.

We take Alesina *et al*'s (2013) Table 3, column 1 regression of women's labour force participation on plow adoption, controlling for agricultural suitability, tropical climate, large animals, political hierarchies, and economic complexity.

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

The timing of the neolithic transition is used as source of exogenous technological progress in a Malthusian model. Here we analyse the first regression of Table 2, where the log of population density in the year 1500 is regressed on the number of years since the neolithic and a continental dummy.

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. Noise was first generated then weighted according to the matrix of settler origins. We apply robust standard errors in all regressions, rather than the bootstrapped errors of the original study which are closer to non-adjusted ones.

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

This paper looks at the impact of pre-industrial agro-climatic characteristics on a measure of long term orientation. Our results are based on the first column of Table 2 which uses crop yield as the explanatory variable, along with dummies for continent.

LaPorta et al. (1998) Law and Finance

This paper studies of the impact of legal origins on institutional quality. We analyse column 2 of Table 6 where judicial efficiency is regressed on a dummy for civil law, and per capita income.

Michalopoulos (2012) The Origins of Ethnolinguistic Diversity

This paper examines the connection between a measure of ethnolinguistic diversity and geographical factors, in particular the mean and variance of elevation and land quality. We examine column 2 of table 1 where the log of the number of languages within a country is regressed on these four variables, with absolute distance from the equator as a control. We report joint significance levels for the hypothesis that the coefficients on all four are zero.

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 country 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. The regression includes controls for the area suited to old world crops, elevation, ruggedness, percentage tropical, all interacted by date. Potato suitability is the variable that we simulate.

The dependent variable in this study is country population for every time period. Reliably simulating such a spatio-temporal process, even if the population of neighbouring countries did not vary hugely, with a short number of time periods is not straightforward and we do not attempt it here so no results for regressions with artificial dependent variables are reported.

Putterman and Weil (2010). Post 1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality

Historical societies are assigned a score depending on their level of political development at points through history, and modern societies are then given weighted scores depending on the ancestral scores of their current populations, and this is used as an explanatory variable for modern income levels. We reproduce column 2 of Table 2. In simulating the explanatory variable we generate an observation for each country in the sample and then multiply it by the ancestry weighting matrix.[5]

Spolaore and Wacziarg (2009). The Diffusion of Development

This paper looks at the ability of a measure of a country's genetic difference from the United States to explain its per capita GDP. We examine the baseline regression of these two variables 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.

6.2 Africa and India

In terms of persistence, there are two immediately salient facts about Africa: the prevalence of the slave trade, particularly in West Africa, and the wide variety of tribal cultures in existence before European colonization, catalogued in Murdock's (1967) *Ethnographic Atlas*. In India, attention has centred on the long run impact of the British Raj. For Africa the correlation range in simulations was in the range of 15–30 degrees, relative to the 70 length and width of the continent.

[5]We use updated scores and migration matrices from `https://sites.google.com/site/econolaols/extended-state-history-index` and `https://www.brown.edu/Departments/Economics/Faculty/Louis_Putterman/world%20migration%20matrix.htm` which give slightly stronger effect sizes that those of the original study.

Alsan (2015). The effect of the TseTse fly on African development

Alsan (2015) analyses how the distribution of the tsetse fly, which transmits trypanosome parasites to animals and humans, has retarded the African development. We consider the regression in Table 1, on how local suitability for tsetse flies affects urbanization rates, controlling for climate variables and clustering by province.

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

Banerjee and Iyer (2005) examine the long run impact of colonial land revenue regimes on modern India. Their Table 3 regresses various aspects of modern agriculture on the share of land controlled by landlords, alongside geographical controls and how long the area was under British rule. We focus on fertilizer use because this is the regression where landlord control has the strongest explanatory power. In carrying out the placebo regressions, we repeat the same generated noise variables for yield in each year in the panel. To compute a Moran statistic we used data for one year, 1982: using other periods gave effectively identical results.

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 ethnic political centralization reported by Murdock (1967) which ranges from stateless societies at zero to large states at four. We examine the baseline regression in Column 1 of Table 2.

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

The legacy of the arbitrary straight line borders drawn across much of Africa by European colonial powers on modern levels of political violence is considered by Michalopoulos and Papaioannou (2016). Specifically they analyse whether violence is higher in the traditional homelands of tribes that found themselves partitioned by such borders. 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, as well as the share of adjacent groups that are partitioned. Controls include country population at independence, distance to the national border and dummies for rivers and lakes. This was the sole case where the standard error estimated using R differed somewhat from the original study: 0.17 rather than 0.16 but the same R procedure is applied over all simulations.

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

Nunn (2008) examines whether the intensity of slave exports, relative to a country's area, is negatively correlated with its GDP in 2000. We replicate the regression in the first column of Table 3 in the paper, where a dummy for each colonial power is included.

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

The impact of the slave trade on modern levels of trust using individual level survey data are examined by Nunn and Wantchekon (2011). In column 1 of their Table 2 they report the impact of slave exports relative to geographic area on trust of neighbours, including as control variables individual factors (such as age, sex, education, and urban or rural location), district controls for ethnic fractionalization, and a country fixed effect.

6.3 Europe and Latin America.

The large amounts of detailed data available for Germany, especially Prussia, since the late eighteenth century have led to several notable studies of persistence. By contrast there are fewer studies for France where data are sparser.

Acemoglu, Hassan and Robinson (2011). Social Structure and Development: A Legacy of the Holocaust in Russia

Acemoglu, Hassan and Robinson (2011) examine the impact of Nazi occupation on subsequent urban growth by examining the interaction between pre-war Jewish population and whether an area was conquered by the Wehrmacht. Their results are distorted by one severe outlier (the town of Derbent) and if this is excluded or given a dummy, the effect size rises markedly, residuals become more normal, and the Moran statistic falls substantially. We therefore report results for the first regression of Table 2 after excluding this observation.

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

Against the Weber thesis that Protestant economic success reflected an ingrained work ethic, Becker and Woessmann (2009) argue that it originated instead in higher literacy rates associated with Bible reading. To test this they examine the connection between the percentage of a Prussian county that was Protestant in 1871 and the literacy of its population and we analyse the first regression of Table 2.

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

This paper considers whether the transition in some cities to Lutheranism, which had no ban on usury, reduced tolerance of Jews. We analyse the first column 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. For both dependent and explanatory variables we simulate two sets of spatial noise, and assign one to the first two pre-Reformation centuries, and the other to the later centuries.

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

This study examines differences in household consumption and child stunting on either side of Peru's *Mita* boundary. It finds that areas which traditionally had to provide conscripted mine labour have household consumption almost 30 per cent lower than on the other side of the boundary.

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 boundary with controls for distance to the boundary, elevation, slope and household characteristics. The variable of interest is a dummy for being inside the boundary. We examine here how well the regression explains arbitrary patterns of consumption generated as spatial noise. To do this we take the locations where households live and simulate consumption levels based on median consumption at the points.

The original study found a 28 per cent difference in consumption levels across the historic boundary. If we normalize the noise variables to have the same mean and standard deviation as the original consumption data, we get a difference of at least 28 per cent (positive or negative) in 70 per cent of cases.

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

In 1685 Louis XIV revoked the Edict of Nantes, which had granted religious toleration to French Protestants, leading many to settle in western Prussia. Whether their skills stimulated the textile industry in their new homes is examined by Hornung (2014) who analyses the output of enterprises recorded in 1802 compared with the share of Huguenot population in the same towns in 1700.[6] Because we are interested in the spatial properties of this result we focus on towns rather than firms by aggregating all firms in a town together. Hornung's findings of the strong impact of Huguenots are essentially unchanged after this aggregation.

Juhász (2018). Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade

This study examines the long run impact on the French cotton industry of Napoleon's Continental System which aimed to exclude English products from Europe: specifically areas along the north coast were more tightly protected in relative terms than before. We replicate the first column of Table 1 by regressing changes in spindles per capita between 1803 and 1812 on the change in the effective distance from London.

Satyanath, Voigtländer and Voth (2017). Bowling for Fascism: Social Capital and the Rise of the Nazi Party

The impact of social capital on the rise of Hitler is considered by Satyanath, Voigtländer and Voth (2017) who examine the links between membership of associations and subsequent Nazi membership. We consider the regression in column 1 of Table 3 which links numbers joining the Nazi party per capita between 1925 and 1933 and membership of all associations, with controls for population, and numbers of Catholics and blue collar workers.

Squicciarini and Voigtländer (2015). Human Capital and Industrialization: Evidence from the Age of Enlightenment

This study examines how growth in French cities in the late eighteenth century can be predicted by the number of inhabitants who subscibed to Diderot's *Encyclopédie*. We consider their Table 3 column 2 which is a convergence regression of urban

[6] I would like to thank Erik Hornung for kindly providing the location data used in the study.

growth from 1750 to 1800 on 1750 population, to which they add subscribers and some geographical controls.

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

This examines how distance from a Jesuit mission affects modern illiteracy rates in Argentina, Brazil, and Paraguay. We analyse column 1 of Table 2.

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

Studying how deep the roots of violent antisemitism in Germany might reach, Voigtländer and Voth (2012) compared towns that recorded pogroms after the Black Death of 1348–49 with those that did not, and found that the former had stronger support for National Socialism. Here we examine their 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.

7 Conclusions

The regressions in studies of historical persistence combine notably high t statistics with severe levels of spatial autocorrelation of residuals, and the goal of this paper was to determine whether these two properties are related.

We found that in the presence of even short range autocorrelation the t statistics of regressions develop extremely thick tails, and outlined a simple procedure to handle this. The first step was to estimate a Moran statistic as a useful diagnostic against the possibility that the regression is simply fitting spatial noise. We went on to simulate spatial noise with an empirically plausible structure: either to be used as the dependent variable or to replace the original persistence explanatory variable. Applying this procedure to 27 studies in leading journals we found that in about three quarters of cases, corresponding to large Moran statistics, the persistence variable had strong ability to predict noise, and frequently had lower predictive as an explanatory variable than did spatial noise. Our findings suggest that the results of persistence studies, and of spatial regressions more generally, might be treated with some caution in the absence of reported Moran statistics and noise simulations.

Appendix: Simulation Results.

Here we present tabulations of how persistence studies performed relative to spatial noise. Specifically, for each area we generate noise at three, hopefully plausible, correlation ranges, and report how it performed as an explanatory variable, replacing the original persistence variable; and how well the persistence variable was able to explain noise as a dependent variable.

Table A1: Persistence versus spatial noise: Global studies.

Correlation Range[c]	Explanatory Noise[a]					Dependent Noise[b]			
	Out-performs[d]	$p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	$p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
Acemoglu et al, Colonial Origins									
30	0.00	0.14	0.05	0.01	0.00	0.30	0.17	0.08	0.04
60	0.00	0.22	0.10	0.03	0.01	0.41	0.26	0.15	0.08
90	0.00	0.27	0.13	0.05	0.01	0.44	0.29	0.17	0.10
Acemoglu et al, Reversal of Fortune									
30	0.01	0.13	0.05	0.01	0.00	0.21	0.10	0.03	0.01
60	0.03	0.16	0.07	0.02	0.01	0.26	0.14	0.06	0.02
90	0.04	0.19	0.09	0.03	0.01	0.27	0.14	0.05	0.02
Alesina et al, Women and the Plow									
30	0.02	0.35	0.21	0.10	0.04	0.37	0.23	0.12	0.07
60	0.10	0.53	0.38	0.24	0.15	0.53	0.41	0.29	0.21
90	0.15	0.55	0.43	0.30	0.22	0.62	0.50	0.38	0.29
Ashraf-Galor, Malthusian Epoch									
30	0.15	0.30	0.18	0.09	0.04	0.39	0.24	0.13	0.07
60	0.17	0.34	0.21	0.11	0.06	0.54	0.40	0.27	0.19

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Continued on next page

Table A1: Global studies. (cont.)

Correlation Range[c]	Explanatory Noise[a] Out-performs[d]	$p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	Dependent Noise[b] $p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
90	0.21	0.37	0.25	0.13	0.08	0.60	0.48	0.36	0.25
Ashraf-Galor, Out of Africa									
30	0.14	0.56	0.45	0.32	0.24	0.55	0.40	0.25	0.17
60	0.31	0.70	0.61	0.51	0.43	0.77	0.66	0.52	0.41
90	0.38	0.75	0.67	0.58	0.51	0.83	0.75	0.63	0.52
Galor-Ozak, Time Preference									
30	0.02	0.19	0.07	0.02	0.00	0.33	0.20	0.11	0.06
60	0.05	0.27	0.14	0.05	0.01	0.31	0.20	0.10	0.05
90	0.05	0.32	0.16	0.05	0.01	0.37	0.22	0.11	0.06
La Porta et al, Law and Finance									
30	0.01	0.10	0.03	0.01	0.00	0.12	0.04	0.01	0.00
60	0.01	0.07	0.02	0.01	0.00	0.20	0.09	0.03	0.01
90	0.00	0.05	0.01	0.00	0.00	0.26	0.11	0.03	0.01
Michaelopoulos, Ethnolinguistic Diversity									
30	0.00	0.17	0.08	0.03	0.01	0.61	0.45	0.29	0.17

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Continued on next page

Table A1: Global studies. (cont.)

Correlation Range[c]	Explanatory Noise[a]					Dependent Noise[b]			
	Out-performs[d]	$p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	$p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
60	0.00	0.21	0.09	0.03	0.01	0.80	0.67	0.51	0.37
90	0.00	0.18	0.07	0.02	0.00	0.82	0.71	0.56	0.45
Nunn-Qian, Potato's Contribution									
30	0.01	0.40	0.27	0.17	0.10	.	.	.	.
60	0.03	0.55	0.42	0.30	0.22	.	.	.	.
90	0.03	0.56	0.44	0.31	0.22	.	.	.	.
Putterman-Weil, Post-1500 Population Flows									
30	0.17	0.60	0.49	0.39	0.30	0.47	0.33	0.21	0.13
60	0.35	0.74	0.66	0.56	0.48	0.62	0.51	0.40	0.31
90	0.41	0.76	0.70	0.61	0.54	0.67	0.57	0.46	0.39
Spolaore-Wacziarg Diffusion									
30	0.00	0.47	0.32	0.19	0.11	0.51	0.38	0.26	0.17
60	0.01	0.60	0.48	0.36	0.26	0.67	0.58	0.47	0.39
90	0.02	0.65	0.53	0.40	0.32	0.72	0.64	0.56	0.50

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Table A2: Persistence versus spatial noise: Africa and India.

Correlation Range[c]	Explanatory Noise[a]					Dependent Noise[b]			
	Out-performs[d]	$p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	$p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
Alsan, Effect of the TseTse Fly									
10	0.14	0.37	0.24	0.13	0.06	0.53	0.42	0.29	0.20
20	0.16	0.43	0.28	0.15	0.08	0.41	0.28	0.17	0.11
30	0.12	0.40	0.24	0.11	0.04	0.22	0.12	0.05	0.02
Michalopoulos-Papaioannou, Pre-Colonial									
10	0.07	0.17	0.06	0.02	0.01	0.12	0.04	0.01	0.00
20	0.15	0.25	0.14	0.06	0.02	0.19	0.08	0.02	0.00
30	0.20	0.31	0.19	0.09	0.04	0.24	0.11	0.03	0.00
Michalopoulos-Papaioannou, Scramble for Africa									
10	0.13	0.26	0.14	0.06	0.03	0.05	0.01	0.00	0.00
20	0.06	0.15	0.06	0.02	0.01	0.05	0.01	0.00	0.00
30	0.03	0.12	0.04	0.01	0.00	0.05	0.01	0.00	0.00
Nunn-Wantchekon, Mistrust in Africa									
10	0.02	0.36	0.23	0.11	0.06	0.43	0.29	0.17	0.11
20	0.07	0.47	0.34	0.21	0.13	0.51	0.36	0.22	0.13

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Continued on next page

Table A2: Africa and India. (cont.)

Correlation Range[c]	Explanatory Noise[a]					Dependent Noise[b]			
	Out-performs[d]	$p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	$p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
30	0.09	0.51	0.37	0.24	0.16	0.54	0.40	0.25	0.14
Nunn, Africa's Slave Trades									
10	0.00	0.10	0.04	0.01	0.00	0.12	0.03	0.01	0.00
20	0.01	0.17	0.07	0.02	0.01	0.19	0.07	0.02	0.00
30	0.01	0.21	0.10	0.04	0.01	0.25	0.12	0.03	0.01
Banerjee and Iyer. Colonial Land Tenure									
8	0.20	0.47	0.32	0.18	0.10	0.46	0.31	0.19	0.10
10	0.24	0.49	0.36	0.22	0.15	0.49	0.34	0.22	0.13
12	0.28	0.54	0.40	0.27	0.18	0.51	0.36	0.24	0.15

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Table A3: Persistence versus spatial noise: Europe and South America.

Correlation Range[c]	Out-performs[d]	Explanatory Noise[a] $p=0.05$[e]	$p=0.01$	$p=0.001$	$p=0.0001$	Dependent Noise[b] $p=0.05$[f]	$p=0.01$	$p=0.001$	$p=0.0001$
Acemoglu et al, Holocaust in Russia									
3	0.00	0.18	0.07	0.02	0.00	0.47	0.34	0.22	0.14
4	0.01	0.22	0.09	0.03	0.01	0.51	0.38	0.25	0.17
5	0.01	0.23	0.11	0.04	0.01	0.52	0.39	0.26	0.18
Becker-Pascali, Anti-Semitism									
3	0.40	0.55	0.44	0.32	0.23	0.71	0.61	0.51	0.44
4	0.45	0.62	0.50	0.37	0.27	0.72	0.65	0.55	0.48
5	0.54	0.67	0.57	0.45	0.36	0.73	0.65	0.56	0.49
Becker-Woessmann, Was Weber Wrong?									
3	0.31	0.68	0.58	0.49	0.41	0.66	0.57	0.46	0.39
4	0.41	0.74	0.67	0.58	0.51	0.71	0.63	0.53	0.46
5	0.50	0.79	0.73	0.66	0.60	0.77	0.69	0.61	0.54
Caicedo, The Mission									
3	0.46	0.62	0.50	0.39	0.29	0.71	0.63	0.53	0.46
4	0.50	0.65	0.55	0.43	0.33	0.74	0.66	0.57	0.50

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Continued on next page

Table A3: Europe and South America. (cont.)

Correlation Range[c]	Out-performs[d]	Explanatory Noise[a] $p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	Dependent Noise[b] $p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
5	0.52	0.68	0.57	0.44	0.35	0.76	0.69	0.59	0.53
Hornung. Huguenots									
3	0.18	0.55	0.44	0.30	0.20	0.47	0.32	0.18	0.09
4	0.28	0.61	0.51	0.39	0.30	0.53	0.39	0.24	0.12
5	0.33	0.68	0.58	0.46	0.35	0.58	0.43	0.24	0.11
Juhasz, Napoleonic Blockade									
3	0.03	0.35	0.17	0.03	0.00	0.56	0.44	0.31	0.23
4	0.07	0.44	0.25	0.07	0.01	0.64	0.54	0.42	0.33
5	0.11	0.49	0.32	0.11	0.01	0.69	0.60	0.50	0.41
Satyaneth et al, Bowling for Fascism									
3	0.18	0.33	0.19	0.09	0.03	0.37	0.22	0.09	0.03
4	0.23	0.40	0.25	0.11	0.05	0.44	0.28	0.14	0.06
5	0.29	0.46	0.30	0.16	0.08	0.48	0.34	0.19	0.09
Squicciarini-Voigtlaender, Age of Enlightenment									
3	0.00	0.17	0.07	0.01	0.00	0.21	0.08	0.03	0.01

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

Continued on next page

Table A3: Europe and South America. (cont.)

Correlation Range[c]	Explanatory Noise[a] Out-performs[d]	$p = 0.05$[e]	$p = 0.01$	$p = 0.001$	$p = 0.0001$	Dependent Noise[b] $p = 0.05$[f]	$p = 0.01$	$p = 0.001$	$p = 0.0001$
4	0.00	0.20	0.08	0.02	0.00	0.20	0.09	0.02	0.01
5	0.00	0.21	0.09	0.02	0.00	0.19	0.09	0.02	0.01
Voigtlaender-Voth. Persecution Perpetuated									
3	0.45	0.58	0.47	0.33	0.23	0.29	0.15	0.06	0.02
4	0.53	0.65	0.54	0.42	0.32	0.31	0.17	0.08	0.03
5	0.58	0.69	0.58	0.47	0.37	0.36	0.21	0.09	0.03
Dell, Peruvian Mita									
0.5	.	.	.	.	.	0.33	0.19	0.09	0.05
1	.	.	.	.	.	0.48	0.35	0.23	0.15
1.5	.	.	.	.	.	0.55	0.43	0.31	0.22

[a] Explanatory variable replaced by noise. [b] Dependent variable replaced by noise.
[c] East-west correlation range in degrees. North-south range is half this.
[d] Proportion of simulations where noise has higher explanatory power than original persistence variable.
[e] Proportion of simulations where noise is significant at level p. [f] Proportion of simulations where persistence variable explains noise with significance level p.

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.

Acemoglu, Daron, Tarek A. Hassan and James A. Robinson. 2011. "Social Structure and Development: A Legacy of the Holocaust in Russia." *Quarterly Journal of Economics* 126:895–946.

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.

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.

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.

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119:249–275.

Bivand, Roger S., Edzer J. Pebesma and Virgilio Gomez-Rubio. 2008. *Applied Spatial Data Analysis with R*. New York: Springer.

Brown, Patrick. 2015. "Model-Based Geostatistics the Easy Way." *Journal of Statistical Software* 63(12):1–24.
URL: *https://www.jstatsoft.org/v063/i12*

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.

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.

Gelman, Andrew and John Carlin. 2014. "Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors." *Perspectives in Psychological Science* 9.

Granger, C. W. J. and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2:111–120.

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

Juhász, Réka. 2018. "Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade." *American Economic Review* 108:3339–3376.

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

Michalopoulos, Stelios. 2012. "The Origins of Ethnolinguistic Diversity." *American Economic Review* 102:1508–1539.

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.

Murdock, George P. 1967. *Ethnographic Atlas*. Pittsburgh: University of Pittsburgh Press.

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

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.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* 37:187–204.

Putterman, Louis and David N. Weil. 2010. "Post 1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics* 125:1627–1682.

Satyanath, Shanker, Nico Voigtländer and Hans-Joachim Voth. 2017. "Bowling for Fascism: Social Capital and the Rise of the Nazi Party." *Journal of Political Economy* 125:478–526.

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

Squicciarini, Mara and Nico Voigtländer. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *Quarterly Journal of Economics* 30:1825–1883.

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.

Young, Alwyn. 2018. Consistency without Inference: Instrumental Variables in Practical Application. Working paper London School of Economics.