DISCUSSION PAPER SERIES

DP14483

DIRECT STANDARD ERRORS FOR REGRESSIONS WITH SPATIALLY AUTOCORRELATED RESIDUALS

Morgan Kelly

ECONOMIC HISTORY

ISSN 0265-8003

DIRECT STANDARD ERRORS FOR REGRESSIONS WITH SPATIALLY AUTOCORRELATED RESIDUALS

Morgan Kelly

Discussion Paper DP14483
Published 12 March 2020
Submitted 12 March 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

DIRECT STANDARD ERRORS FOR REGRESSIONS WITH SPATIALLY AUTOCORRELATED RESIDUALS

Abstract

Abstract Regressions using data with known locations are increasingly used in empirical economics, and several standard error corrections are available to deal with the fact that their residuals tend to be spatially correlated. Unfortunately, different corrections commonly return significance levels that vary by several orders of magnitude, leaving the researcher uncertain as to which, if any, is valid. This paper proposes instead an extremely fast and simple procedure to derive standard errors directly from the spatial correlation structure of regression residuals. Importantly, because the estimated covariance matrix gives optimal weights to predict each residual as a linear combination of all residuals, the reliability of these standard errors is self-checking by construction. The approach extends immediately to instrumental variables, balanced and unbalanced panels, and a wide class of nonlinear models. A step by step guide to estimating these standard errors is given in the accompanying tutorials.

JEL Classification: N/A

Keywords: N/A

Morgan Kelly - morgan.kelly@ucd.ie
University College Dublin and CEPR

Acknowledgements
The tutorials are available at https://github.com/kellyspatial/direct-se

Direct Standard Errors for Regressions with Spatially Autocorrelated Residuals

Morgan Kelly*

Abstract

Regressions using data with known locations are increasingly used in empirical economics, and several standard error corrections are available to deal with the fact that their residuals tend to be spatially correlated. Unfortunately, different corrections commonly return significance levels that vary by several orders of magnitude, leaving the researcher uncertain as to which, if any, is valid. This paper proposes instead an extremely fast and simple procedure to derive standard errors directly from the spatial correlation structure of regression residuals. Importantly, because the estimated covariance matrix gives optimal weights to predict each residual as a linear combination of all residuals, the reliability of these standard errors is self-checking by construction. The approach extends immediately to instrumental variables, balanced and unbalanced panels, and a wide class of nonlinear models. A step by step guide to estimating these standard errors is given in the accompanying tutorials.
Keywords: Spatial regressions. Direct standard errors.

1 Introduction.

Cross-sectional and panel regressions that use data from neighbouring spatial locations are an increasing staple of empirical economics. Although the residuals from these regressions tend to be spatially correlated, the researcher has a choice of well established standard error corrections to deal with this. Most commonly, residuals are simply clustered into natural groups but, if desired, more explicitly spatial corrections are available, notably the HAC adjustment of Conley (1999) and the large cluster approach of Bester, Conley and Hansen (2011). However, at this stage things start to become complicated.

For concreteness, suppose that the data are from US counties. Should they be clustered by states or by the larger census divisions? Should the standard errors then be corrected using the default Stata adjustment or the Bell and McCaffrey (2002) procedure advocated by Imbens and Kolesár (2016)? If one chooses the HAC path, should the correlation cutoff

*University College Dublin and CEPR. The tutorials are available at `https://github.com/kellyspatial/direct-se`

be set at 100 km, or 500 km or 1,000 km? Or if the large cluster approach advocated by Barios et al. (2012) is used, should the data be sliced into two, three, four or more groups? In practice, most studies cluster at the lowest level available using *Stata* defaults, the implicit assumption being that other adjustments would return standard errors that are fairly similar.

Unfortunately, this is not the case in general: different corrections commonly lead to substantially different standard errors. To illustrate this, Table 1 gives the results of a regression using some variables from Autor, Dorn and Hanson (2013). Clustering at state level, the significance level of the main variable changes from 0.00004 using *Stata* defaults to 0.01 using a Bell-McCaffrey correction; and to 0.04 if the residuals are clustered by census division, again using Bell-McCaffery. If the Bester, Conley and Hansen (2011) correction is used, the significance level changes from 0.03 with four clusters to 0.21 with three.

Which adjustment then is the correct one? Monte Carlo simulations in the Appendix indicate that, when it comes to the probability of Type 1 errors, the Bester, Conley and Hansen (2011) approach with three clusters is fairly robust for varying patterns of spatial correlation. The performance of other corrections varies widely according to the underlying spatial pattern of the data, and they tend to reject noise regressions far too rarely. However, the very high *t* values needed for a Bester *et al* adjustment to be significant suggest that its power is low, and this turns out to be the case.

The very different significance levels of these standard adjustments, then, makes them of uncertain utility in practice. The current approach is to impose a weighting matrix or kernel where observations are assigned a weight, usually one or zero, depending on whether they belong to an assumed cluster or fall within an assumed cutoff range. This paper proposes instead an alternative approach of estimating standard errors directly from the spatial correlation structure of the regression residuals. Start with the vector of normally distributed regression residuals $U \sim \mathcal{N}(0, \Sigma)$ where Σ has generic entry Σ_{ij}. The approach is to impose some functional form on the covariance matrix, so $\Sigma_{ij} = f(h, \pi)$ where h is the distance between the locations of i and j, and π is a vector of parameters. Once we have chosen a functional form f we can estimate the parameters $\hat{\pi}$ straightforwardly by maximum likelihood. This gives an estimated covariance matrix $\hat{\Sigma}$ that allows us to derive regression standard errors in the usual way.

The immediate question is how can we know that these standard errors are any more trustworthy than existing ones? The answer is simple: the estimated covariance matrix $\hat{\Sigma}$ gives optimal forecasting weights to predict each residual as a linear combination of all residuals. By comparing the actual residuals with these predicted values we can judge the plausibility of the assumed functional form, and the reliability of the estimated standard errors.

The procedure extends to longitudinal regressions, where the covariance matrix is computed as a Kronecker product of temporal and spatial autocorrelations. There is one complication: fixed effects will absorb some degree of the spatio-temporal correlations in the residuals, leading to unreliable estimates of the covariance matrix. A simple two step

procedure can be used to deal with this. First, residuals are computed from a regression without fixed effects. These raw residuals are used to generate an estimated covariance matrix. Coefficients are then estimated from a regression with fixed effects, and the first stage covariance matrix is used to compute standard errors. For unbalanced data, the rows and columns of the covariance matrix corresponding to the missing observations are simply removed before estimating the standard errors.

Turning to nonlinear regressions where $y = g(X, \theta)$ and θ is a vector or parameters. Direct standard errors may be calculated in cases where X and θ only enter g in the form $X^{'}\theta$. This covers most regressions in common use including generalized linear models (logistic, Poisson, and gamma); negative binomials (along with hurdle and zero inflated extensions) for counts, survival models, and beta regressions for proportions. Direct standard errors must be applied cautiously in these cases however. The empirical densities of residuals from these regressions will typically not resemble Gaussian ones closely, so it is necessary to check that the predicted residuals are a reasonable match for the actual ones before deciding how much trust in the estimated spatial covariance matrix.

The rest of the paper is as follows. Section 2 shows how different choices of spatial correction can lead to widely different standard errors. This provides the motivation in Section 3 to estimate standard errors from a parametric covariance matrix and to show how these standard errors can be validated by comparing actual and predicted residuals. This procedure is extended to panel and nonlinear regressions in Sections 4 and 5. The technique is illustrated using a variety of examples from well known studies in Sections 6 to 8. An accompanying set of tutorials demonstrates the simple code used to apply these techniques.

2 The Impact of Spatial Robustness Adjustments

To illustrate the large changes in significance levels resulting from different ways to adjust for spatial correlation, I apply standard tests to some of the China Shock variables used by Autor, Dorn and Hanson (2013). The dependent variable is the change in manufacturing employment share and the equation is estimated with OLS using data for the 2000–2007 period, and unweighted observations for 722 Commuting Zones.

Table 1 shows the effect on significance levels of applying a variety of standard error corrections. Although it is increasingly realized that significance levels are a poor route to understanding regression results, because of the very low degrees of freedom involved in some corrections it is uninformative to report standard errors on their own, and confidence intervals would make the table too cluttered to be easily comprehensible.

The first column presents the commonest approach to such corrections, clustering at the state level and using the default *Stata* robustness correction (where the estimated covariance matrix is multiplied by $\frac{N-1}{N-K}\frac{S}{S-1}$ where N, K and S are the numbers of observations, explanatory variables, and clusters respectively). The second column uses the Bell

	48 State Clusters		8 Census Clusters		HAC Cutoff			Slice	
	Stata	BM	Stata	BM	100	500	1000	Quarters	Thirds
Chinese Imports	0.00004	0.00986	0.00151	0.03539	0.00038	0.00198	NaN	0.02772	0.21235
Mfg Employment	0.00001	0.00010	0.02370	0.06090	0.00000	0.00000	NaN	0.11791	0.11423
College Educated	0.11727	0.13339	0.22481	0.24420	0.05007	0.26479	0.00833	0.35178	0.32031
Foreign Born	0.58438	0.63015	0.54768	0.59802	0.61002	0.82716	0.83915	0.38566	0.39871
Female Employment	0.00227	0.00542	0.06251	0.18398	0.00146	0.24803	0.33224	0.16519	0.17604
Routine Tasks	0.00000	0.00000	0.00103	0.00757	0.00000	0.00252	0.02885	0.01442	0.00659
X	0.70191	0.73271	0.46146	0.54595	0.67651	0.89584	0.92044	0.89124	0.89477
Y	0.06181	0.11706	0.00244	0.03212	0.02009	0.50556	0.62746	0.32426	0.34296
XY	0.15824	0.23900	0.00114	0.04359	0.06841	0.59553	0.68382	0.30239	0.35847

Least squares regression of the impact of Chinese imports on manufacturing employment for 722 Commuting Zones for 2000-2007 using data from Autor et al (2013). For columns clustered by 48 states and 8 census divisions values are given for standard errors adjustments computed using *Stata* defaults and the Bell-McCaffrey correction. For HAC standard errors, significance levels are given for cutoffs at 100, 500, and 1000 kilometres. Slice denotes clustering by longitude where observations are divided into three or four groups of equal size, and Bell-McCaffrey corrections applied.

Table 1: Significance levels of China shock explanatory variables under a variety of spatial covariance adjustments.

and McCaffrey (2002) correction recommended by Imbens and Kolesár (2016). The next two columns use the same adjustments, but clustering at the level of Census Divisions.

The next set of columns apply the HAC correction of Conley (1999), with a rectangular kernel and cutoffs of 100, 500 and 1000 kilometres (the median distance between the centroids of commuting zones is 1300 km, 10 per cent are within 500 km of each other, and only 0.5 per cent within 100 km). The last set of columns implements the small number of large clusters approach of Bester, Conley and Hansen (2011), advocated by Barios et al. (2012). As the Appendix shows, this procedure has a reasonably correct size when three clusters are used. The data here are divided into three and four groups with the same number of observations, clustered by longitude.

What can be seen immediately is how these, probably equally plausible, corrections produce radically different significance levels. Taking the Chinese imports variable, clustering at the level of states with a *Stata* correction returns a significance level of 0.00004 that changes to 0.01 after applying a Bell-McCaffrey correction at state level, and to 0.04 at the level of census divisions.

Applying the Bester, Conley and Hansen (2011) procedure with three clusters increases significance to 0.23 whereas a HAC correction with a 500 km cutoff changes the significance level to 0.002. HAC corrections with long ranges can return covariance matrices with negative diagonal elements, as can be seen for the 1000 km cutoff.

3 Covariance Matrix Estimation and Validation.

Given the substantially different significance levels returned by existing covariance corrections, it is worth exploring a parametric approach of specifying the functional form of the covariance matrix and estimating its parameters by maximum likelihood.

The existing approach to standard error calculation is to estimate a covariance matrix

$$\hat{\Sigma} = \sum_{ij} K(i,j)\, x_i \hat{u}_i x_j^{'} \hat{u}_j \tag{1}$$

where $K(i,j)$ is a weighting kernel. For clustered models, this kernel is one for observations within each cluster and zero otherwise, while for HAC standard errors, in the usual case of a rectangular kernel, $K(i,j)$ consists of ones for observations within the cutoff range, and zero otherwise.

Our starting point is a vector of normally distributed regression residuals $U \sim \mathcal{N}(0, \Sigma)$ where Σ has generic entry Σ_{ij}. The approach is to assume some functional form for the covariance matrix, so $\Sigma_{ij} = f(\|i-j\|, \pi)$ where $\|i-j\|$ is the distance between points i and j and π is a vector of parameters. Once we have chosen a functional form f we can estimate the parameters $\hat{\pi}$ by maximum likelihood. From the estimated covariance matrix $\hat{\Sigma}$ we can derive regression standard errors in the usual way.

The natural objection to this procedure is how can we know that the assumed functional form is a valid one. It turns out that we can test this straightforwardly.

Suppose that we wish to predict V, the values of a variable at some unobserved sites, based on actual values U at some observed sites. Assuming that these observations are normally distributed

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{UU} & \Sigma_{UV} \\ \Sigma_{VU} & \Sigma_{VV} \end{pmatrix}\right) \tag{2}$$

then the conditional distribution of V given U (see for example Gneiting and Gutthorp 2010) is

$$(V|U) = \mathcal{N}\left(\Sigma_{UV}\ \Sigma_{UU}^{-1}\ U, \Sigma_{VV} - \Sigma_{UV}\ \Sigma_{UU}^{-1}\ \Sigma_{VU}\right). \tag{3}$$

The covariance function is assumed further to decompose into a systematic spatial part and idiosyncratic noise so Σ has generic entry

$$\Sigma_{ij} = \sigma^2 C\left(\|i-j\|, \pi\right) + \tau^2 \mathbf{1}_{ij} \tag{4}$$

where C is a correlation function, $\|i-j\|$ is the distance between i and j, and the indicator $\mathbf{1}_{ij} = 1$ when $i = j$ and 0 otherwise. In matrix terms, $\Sigma_{UU} = \sigma^2 C_{UU} + \tau^2 I$.

This means that if we have a holdout sample of residuals $V \neq U$, then their predicted values are $\hat{V} = \sigma^2 C_{UV}\ \left(\sigma^2 C_{UU} + \tau^2 I\right)^{-1} U$. If the assumed functional form of Σ is unreliable, the predicted values $\hat{V}$ will differ substantially from the observed ones.

An analogous procedure applies when making smoothed predictions of the original observations U except that the idiosyncratic noise τ^2 is omitted from the estimate: Cressie (1991, 127–129).[1] This leads to the predicted residuals

$$\hat{U} = \sigma^2 C_{UU}\ \left(\sigma^2 C_{UU} + \tau^2 I\right)^{-1} U. \tag{5}$$

By comparing the predicted value of the regression residuals with their actual values, we can gauge the validity of the assumed functional form of the correlation matrix Σ and, through this, the reliability of the standard errors derived from it.

3.1 The Functional Form of the Covariance Matrix.

Having come up with a straightforward way to evaluate the performance of a parametric covariance matrix, it is now necessary to choose a specific functional form. There are,

[1] In cases where idiosyncratic variance is included in the forecast, so we are interpolating rather than smoothing, the reliability of the covariance matrix may be evaluated instead very simply through leave one out cross validation. Define $Q = \Sigma^{-1}$ with (i,j) -th entry q_{ij}. Then the leave one out prediction of the i-th residual $\hat{u}_{-i} = -\sum_{j \neq i} q_{ij} u_j / q_{ii}$ (Dubrule, 1983). For the examples below, these predictions are highly correlated with those from (5) but estimates for large negative residuals are too low, and for large positive ones too high: fitted values "regress" away from the mean.

naturally, a large number of potential choices but the most common is based on the Whittle-Matérn function where the correlation between sites at distance h apart is

$$C(h) = \frac{2^{1-\kappa}}{\Gamma(\kappa)} \left(\frac{h}{\rho}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\rho}\right) \quad (\kappa > 0, \rho > 0) \tag{6}$$

where Γ is a gamma function and K_κ is a Bessel function of the second kind.[2] The parameter κ denotes the smoothness of the function, and ρ is a scale parameter giving the correlation range.[3] Specifically, when two sites are separated by a distance $h = \rho/\sqrt{8\kappa}$, the correlation between them is 0.14: this distance is commonly called the effective range and is the range that is reported for the regression examples below. The decay of correlation with distance is controlled by κ. For $\kappa = \frac{1}{2}$, correlation decays exponentially. Setting the scale $\rho = 1$, $C(h) = \exp(-h)$. When $\kappa = \frac{3}{2}$, $C(h) = (1+h)\exp(-h)$; and for $\kappa = \frac{5}{2}$, $C(h) = \left(1 + h + \frac{h^2}{3}\right)\exp(-h)$. As κ goes to infinity C falls off as a Gaussian curve.[4]

To recall, for a given vector of regression residuals $u \sim \mathcal{N}(0, \Sigma)$ the covariance matrix has the form $\Sigma = \sigma^2 C(\kappa, \rho) + \tau^2 I$ where C has functional form (6), $\|i - j\|$ is the distance between i and j, and the indicator $\mathbf{1}_{ij} = 1$ when $i = j$ and 0 otherwise. The four parameters giving the spatial structure of the residuals—the correlation range ρ, smoothness κ, variance σ^2 and noise τ^2 can be estimated by maximum likelihood. Once the parameters have been estimated, the covariance matrix $\hat{\Sigma}$ can immediately be calculated from equations (6) and (4) and the standard errors of the coefficients are the usual diagonal elements of $(X'X)^{-1}\left(X'\hat{\Sigma}X\right)(X'X)^{-1}$. This approach extends immediately to instrumental variables, using projected X values.[5]

As noted above, correlation between sites equals 0.14 when they are a distance $h = \rho/\sqrt{8\kappa}$ apart. This means that is not feasible to estimate smoothness κ and range ρ simultaneously, as the effects of increasing κ and reducing ρ largely cancel out. The usual approach is therefore to set κ over a range of values, typically from 0.5 to 2.5 in increments of 0.5, and choose the value that maximizes the likelihood, or alternatively the one that minimizes the sum of squared errors between the actual and predicted residuals. In practice it is usual to find exponential falloff of correlation with distance ($\kappa = 0.5$).

[2]Other correlation functions that are occasionally used include the double exponential; and the Cauchy which can capture the long tails encountered, for instance, in meteorology: Gneiting and Gutthorp (2010).

[3]Notation varies: in place of ρ some authors and statistical packages use $\rho' \equiv \rho/\sqrt{8\kappa}$.

[4]The Whittle-Matérn function is the solution to a stochastic differential equation that is the continuous analogue of an ARMA process. This is the basis of the Integrated Nested Laplace Approximation (INLA) approach of Lindgren, Rue and Lindstrom (2011) to solving the parameters of (6).

[5]We will ignore here the fact that $\hat{\Sigma}$ is computed using estimated coefficients and that the computed standard errors should reflect this extra uncertainty by, for instance, repeatedly sampling from the estimated distributions of these coefficients.

It can be seen that covariance breaks down into a systematic spatial component and idiosyncratic noise. The spatial structure parameter

$$s = \frac{\sigma^2}{\sigma^2 + \tau^2} \tag{7}$$

gives the share of spatial correlation in the variance of the residuals, a signal to noise ratio reflecting the amount of (Gaussian) spatial structure present. In cases where s is small, the predicted residuals (5) will match the actual residuals poorly.

So far we have supposed that the correlation between residuals is isotropic, or equally strong in all directions. In general, however, the falloff of correlation with distance will vary by direction: the isocorrelation contours will not be circles but ellipses at angle ϕ to the horizontal whose major axis is r times as long as their minor one. These parameters can be estimated, for instance by the nonparametric approach of Chorti and Hristopulos (2008) that I apply below, and used to stretch and rotate the original spatial coordinates of the observations prior to estimating the spatial covariance parameters in (4).

4 Panel Regressions

We can now consider a panel regression

$$y_{it} = \beta x_{it} + j_i + k_t + u_{it}$$

where j_i and k_t are individual and time fixed effects. Residual covariance is assumed to be separable between space and time: for each period the vector of residuals $u_{i\cdot} \sim \mathcal{N}(0, \Sigma)$ and for all sites the vector of residuals $u_{\cdot t} = \alpha u_{\cdot t-1} + \epsilon_t$.[6]

This gives the temporal autocorrelation matrix[7]

$$A = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^T \\ \alpha & 1 & \ddots & \ddots & \vdots \\ \alpha^2 & \ddots & \ddots & \ddots & \alpha^2 \\ \vdots & \ddots & \ddots & 1 & \alpha \\ \alpha^T & \cdots & \alpha^2 & \alpha & 1 \end{pmatrix}$$

which, combined with the spatial covariance matrix Σ leads to the longitudinal covariance matrix

$$\Sigma_L = A \otimes \Sigma. \tag{8}$$

[6] See Gneiting, Genton and Guttorp (2006) for a survey of non-separable covariances. These are important, for instance, in meteorology where the weather changes in upwind places before downwind ones.

[7] Allowance can be made for the fact that the temporal covariances α may vary between periods, so that the first row of A becomes $(1, \alpha_1, \alpha_1\alpha_2, \ldots)$ and so on.

	$p = 0.05$		Panel Residuals			Raw Residuals	
α	Cluster	Direct	$\hat{\alpha}$	$\hat{\sigma}^2$	$\hat{\sigma}^2_{\mathrm{FE}}$	$\hat{\alpha}$	$\hat{\sigma}^2$
0.2	0.190	0.076	-0.011	0.845	0.170	0.339	0.999
0.5	0.164	0.094	0.241	0.705	0.304	0.504	0.932
0.9	0.184	0.176	0.545	0.269	0.665	0.877	1.012

The first two columns give the proportion of noise regressions that are significant at 5 per cent when clustering on individuals, or when using direct standard errors computed from fixed effect residuals. These are reported for different degrees of temporal autocorrelation α. The block of columns under Panel Residuals give the estimated values of temporal autocorrelation, spatial residual variance (this should equal one) and fixed effect spatial variance based on fixed effect residuals. The fil block gives estimated values of temporal autocorrelation and spatial variance when these are estimated directly with "raw residuals" from a first stage regression with no fixed effects.

Table 2: Simulations of panel regressions showing how individual and time fixed effects absorb some of the spatio-temporal structure of regression residuals, which means that estimated covariance matrices must be based on the residuals of a first stage regression without fixed effects.

The spatial covariance matrix Σ is for all sites, and the temporal autocorrelation matrix A is for all time periods, so that Σ_L gives the spatio-temporal covariance for a balanced model, whether the original data are balanced or not. To compute the covariance matrix for an unbalanced panel, the rows and columns of Σ_L corresponding to missing observations are simply deleted.

When predicting residuals using (8), the procedure is simplified numerically because $\Sigma_L^{-1} = A^{-1} \otimes \Sigma^{-1}$. In the case where the panel is unbalanced however, the full covariance matrix Σ_L is calculated as before as the product $A \otimes \Sigma$ and the rows and columns corresponding to missing observations are then removed. In this case Σ_L must be inverted directly.[8]

4.1 Fixed Effects.

In computing the parameters for A and Σ empirically, one complication is that fixed effects for individuals and time periods will absorb some fraction of the spatial and temporal autocorrelation structure of the residuals. This will lead to underestimates of α and σ^2. The simulations in Table 2 illustrate this difficulty.

[8] In general spatial covariance matrices Σ are not sparse, increasing the difficulty of inverting them by a standard Cholesky decomposition. A common technique to generate a sparse approximation without compromising accuracy is tapering, where a cutoff distance is chosen so that each site has around 50 nearest neighbours, with longer correlations being set to zero: Furrer, Genton and Nychka (2006).

The observations are over 100 randomly distributed individual sites on a unit square over ten time periods. Correlation falls off exponentially ($\kappa = 0.5$) and the effective correlation range (2ρ) is 0.2. The spatial variance $\sigma^2 = 1$ and there is no idiosyncratic variance so $\tau^2 = 0$. The Table gives the results of 1000 simulations where one noise series with these spatio-temporal characteristics is regressed on another, for three different values of temporal autocorrelation $\alpha = 0.2, 0.5, 0.9$.

The first two columns give the percentage of these fixed effect noise regressions with fixed effects that are significant at 5 per cent, for two cases. The first is the standard approach of clustering by individuals, using a Bell-McCaffery adjustment. It can be seen that, as expected, considerably more than five per cent of regressions are significant: around one fifth in fact. The second column computes direct standard errors using the fixed effect regression residuals to estimate the average temporal autocorrelation (by averaging over the coefficients from 9 regressions of residuals on those from the previous period) and spatial autocorrelation parameters. These direct standard errors also perform poorly, with almost one fifth being significant when the temporal autocorrelation is 0.9.

The reason for this misbehaviour can be seen in the second block of columns, labelled Panel Residuals. This shows first that the estimated autocorrelation coefficient $\hat{\alpha}$ is too low. Similarly, the estimated variance $\hat{\sigma}^2$ is also severely underestimated: instead of 1 it falls successively from 0.84, to 0.7, and to 0.27 as the temporal autocorrelation increases from 0.2 to 0.9. At the same time, if we compute the spatial autocorrelation parameters of the individual fixed effects $\hat{\sigma}^2_{\mathrm{FE}}$, their variance rises from 0.17 to 0.3 to 0.27. In other words, we can see that these two estimated variances between them add up to the true value: the fixed effects are absorbing a large share of the spatial structure of the residuals.

The final block of columns, labelled Raw Residuals, uses the residuals computed without fixed effects to compute the spatial and temporal autocorrelations. It can be seen that the values returned here are correct. These values can then be used to compute the covariance matrix for the fixed effect regressions.

In summary, a two step procedure separating the estimation of the coefficients and covariances is needed. First, the regression is run with no fixed effects, and its residuals used to compute the spatial and temporal autocorrelation statistics. These parameters in turn are used to compute the spatio-temporal covariance matrix Σ_L in (8). Then, the regression is run with the fixed effects to get the true coefficients, and the first step covariance matrix is used to construct the standard errors.

5 Nonlinear Regressions

The analysis extends to a large class of nonlinear models.[9] There is an objective function $\Psi(y, X, \theta)$ that has an estimating function

$$\psi(y, X, \theta) = \frac{\partial \Psi(y, X, \theta)}{\partial \theta}. \tag{9}$$

Inference is based on a central limit theorem of the form $\sqrt{n}\left(\hat{\theta} - \theta\right) \to \mathcal{N}(0, S(\theta))$ where the covariance matrix $S(\theta)$ can be written in sandwich form as

$$S(\theta) = B(\theta) M(\theta) B(\theta). \tag{10}$$

The "bread" and "meat" of the sandwich are $B(\theta) = \left(E\left(-\psi'(y, X, \theta)\right)\right)^{-1}$ and $M(\theta) = \mathrm{Cov}(\psi(y, X, \theta))$ respectively.

To apply direct spatial standard errors it is necessary that $\Psi(y, X, \theta)$ takes the form $\Psi(y, \eta)$ where $\eta = X'\theta$. In that case where y depends X and θ only through a linear predictor the estimating function can be written

$$\psi\left(y, X'\theta\right) = \frac{\partial \Psi}{\partial \eta}\frac{\partial \eta}{\partial \theta} = wX \tag{11}$$

where the entries $w(y, \eta) = \partial\Psi/\partial\eta$ are sometimes called the working residuals.

The bread is calculated as $\hat{B} = \left(\frac{1}{n}\sum_i -\psi'\left(y, X, \hat{\theta}\right)\right)^{-1}$. The conventional approach, ignoring spatial correlation, is to estimate the meat as

$$\hat{M} = \frac{1}{n}\left(X'\hat{W}X\right) \tag{12}$$

where $\hat{W}$ is a diagonal matrix with elements $\hat{w_i}^2$.

Computing direct standard errors simply involves computing the spatial parameters of the working residuals $\hat{w}$ as before and using them in turn to compute a spatial covariance matrix $\hat{\Sigma}$. This covariance is then used to replace $\hat{W}$ when calculating $\hat{M}$ in (12).

The procedure applies directly to logistic regressions, Poisson and negative binomial counts (including hurdle and zero inflated extensions), survival models, and beta regressions of proportions. When the procedure is applicable, a complication arises from the fact that the empirical density of the working residuals will often not look very normal. It is therefore particularly important to ensure that the predicted residuals are a reasonable match for the actual ones before basing any inference on the estimated standard errors.

[9] The terminology here closely follows Zeileis (2006).

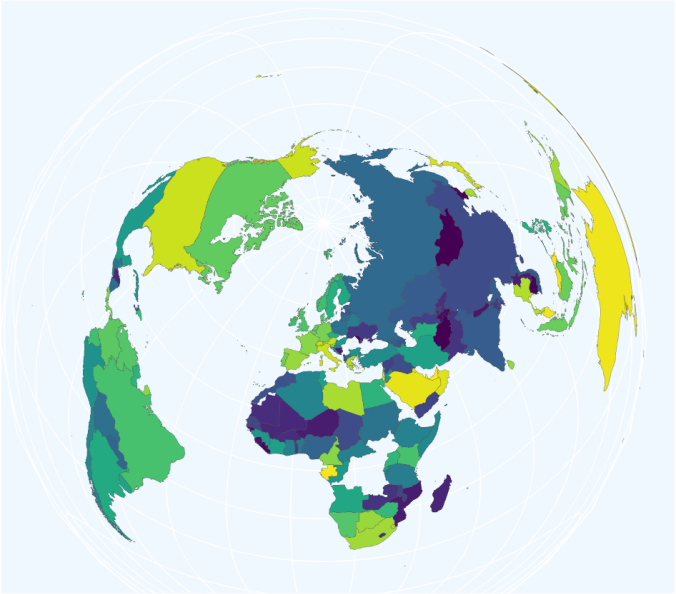

Figure 1: The residuals from Ashraf and Galor (2013) show correlation over long distances.

6 Illustrations: OLS regressions

The procedure to generate standard errors and to ascertain their reliability, then, is extremely straightforward. Take the estimated regression residuals, estimate their spatial structure parameters (adjusting for anisotropy if necessary) and then generate a covariance matrix with these parameters. We now take some examples to illustrate the process.

6.1 Ashraf and Galor (2013)

The first illustration comes from an equation in Ashraf and Galor (2013) which regresses GDP per capita on a measure of genetic diversity and diversity squared. Given the need to remove spatial trends prior to estimating the spatial structure of the regression residuals, I include two directional variables that are routinely used in such studies in an attempt to control for the impact of omitted variables: distance from the equator, and longitude.

Results are presented in Table 3, and the residuals are plotted in Figure 1: the strong autocorrelation over long distances is apparent. After the coefficients, the second column of the Table presents familiar robust standard errors. Several diagnostics are presented

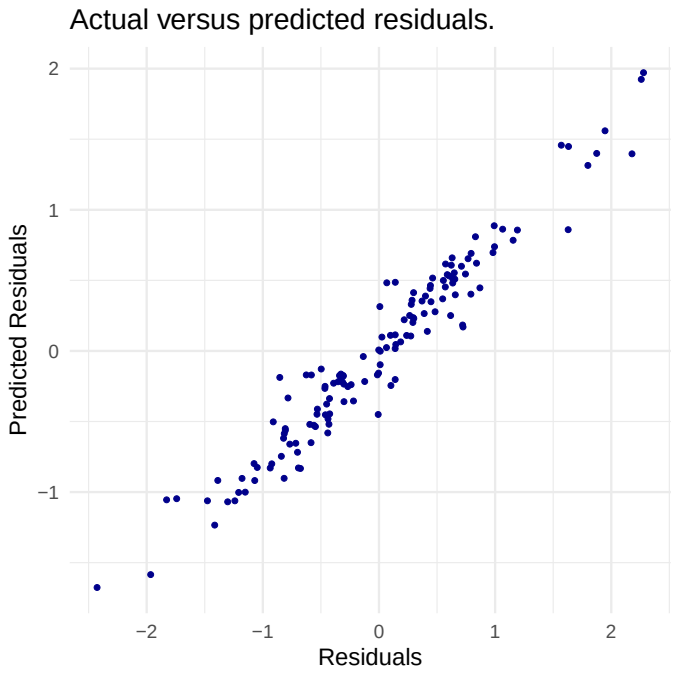

Figure 2: Actual residuals versus the values predicted by the estimated spatial covariance matrix for Ashraf and Galor (2013).

below. The first is the significance level of the standard test for spatial autocorrelation, Moran's I

$$I = \frac{N}{W} \frac{\sum_{i \neq j} w_{ij} u_i u_j}{\sum_i u_i^2} \tag{13}$$

where w_{ij} are weights and $W = \sum_{i,j} w_{ij}$. Here we give the 5 nearest neighbours, by great circle distance, a weight of one, and others zero. Changing this range rarely alters the results materially: the statistic is not intended as a yes-no, "significant-insignificant" test of autocorrelation, but to serve as an indicator of possible spatial structure in residuals. In most cases the significance level tends either above 0.2 or below 10^{-5}. In this instance, the significance level is of the order 10^{-23}. With cross-sectional regressions it is very often the case that the reported results are driven by a handful of extreme outliers. The Table therefore reports a standard influence measure, the largest Cook's distance of the observations. As a rule of thumb a value above 0.1 suggests an observation with substantial influence, which is not the case here.

The equation also reports the maximum likelihood estimates of the covariance parameters in (6) and (4). As noted above, changing the smoothness parameter κ does not affect the results markedly as range ρ moves in compensation. Here, changing smoothness from 0.5 to 4.0 caused the log likelihood to increase from -157.2 to -156.3 but caused the

	Coef	Robust SE	Direct SE
(Intercept)	-53.887	(28.763)	(62.064)
Diversity	182.286	(81.898)	(177.145)
Diversity Squared	-134.711	(58.028)	(126.006)
Longitude	-0.003	(0.001)	(0.003)
Abs Latitude	0.041	(0.004)	(0.012)
R^2	0.445		
Moran	0.000		
Cook	0.070		
Residual Fit	0.941		
Spatial Structure	0.782		

140 observations. Effective range is 1747 kilometres, and the anisotropy ratio and angle are 1.12 and 6 degrees. Moran is the p value of the I statistic. Cook is the maximum Cook's distance of the observations. Residual Fit gives the squared correlation between the regression residuals and their predicted value based on the estimated spatial covariance matrix. Structure is the share of spatial structure in residual variance given in equation 7.

Table 3: Summary statistics for a regression of income on genetic diversity from Ashraf and Galor (2013) giving robust and direct standard errors.

squared correlation between actual and fitted residuals to fall to 0.80. The estimated standard errors were effectively unchanged. The Table reports results for exponential falloff of correlation $\kappa = 0.5$. The reported range parameter is 2ρ, the effective range where correlation has fallen to 0.14. Spatial structure is the share of spatial signal in the variance of the residuals from (7): large values convey similar information to a significant Moran statistic, that there is substantial spatial structure present. The anisotropy ratio and angle give the stretching and rotation of the major correlation axis to give equal correlation in each direction, prior to estimating the covariance parameters.

Naturally, before we can put any faith in this estimated covariance matrix we need to know that it is accurate, by comparing the predicted value of the residuals from (3) with their actual values. This is done in Figure 2. It can be seen immediately that the two match closely, with predicted residuals regressing towards the origin as expected.

Given that we can place reasonable faith in our estimated covariance matrix $\hat{\Sigma}$, the third column of Table 3 reports the estimated direct standard errors: the square root of the diagonal terms of the sandwich matrix $(X'X)^{-1}\left(X'\hat{\Sigma}X\right)(X'X)^{-1}$. It can be seen that they are about twice as large as the original robust ones for the explanatory variables, and three times as large for the directional ones.

	Coef	Robust SE	Direct SE
(Intercept)	7.954	(1.045)	(1.101)
Mortality	-0.385	(0.204)	(0.208)
Longitude	0.000	(0.002)	(0.002)
Abs Latitude	0.005	(0.013)	(0.015)
R^2	0.080		
Moran	0.630		
Cook	0.104		
Residual Fit	0.197		
Spatial Structure	0.000		

60 observations. Effective range is 6360 kilometres, and the anisotropy ratio and angle are 1.36 and 39 degrees. Moran is the p value of the I statistic and Cook is the maximum Cook's distance of the observations. Residual Fit gives the squared correlation between the regression residuals and their predicted value based on the estimated spatial covariance matrix. Spatial structure is the share of spatial structure in residual variance given in equation 7.

Table 4: Acemoglu, Johnson and Robinson (2001): Impact of European mortality on appropriation risk.

6.2 Acemoglu, Johnson and Robinson (2001)

Now I illustrate a case where regression residuals have no spatial structure, making direct residuals superfluous. The data come from Acemoglu, Johnson and Robinson (2001) on the impact of European mortality on institutional quality measured by appropriation risk. I make two adjustments. First, regions with a mortality rate of over 250 are capped at 250: if this is not done, extreme outliers distort the residuals and lead to a spuriously large Moran statistics. Second, based on Cook's Distances, five observations (Australia, Canada, New Zealand, Singapore, and the United States) distort the regression estimates and are omitted from the reported results.

It is immediately evident that there is no spatial structure in the residuals: the p value of the Moran statistic is 0.67. Similarly, the spatial structure parameter (7) is 0.00004. As a result, the direct residuals are effectively identical to the unadjusted ones, and the predicted residuals are unrelated to the actual ones as Figure 3 indicates.

6.3 Autor, Dorn and Hanson (2013) and Chetty et al. (2014)

Next I examine two studies using data for US commuting zones. The first continues the Autor, Dorn and Hanson (2013) example in Table 1 of Section 2 which demonstrated how

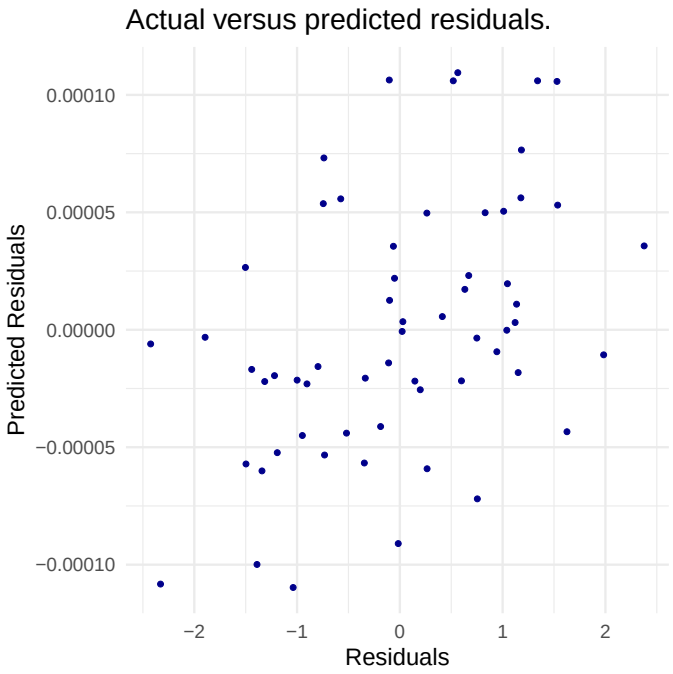

Figure 3: Actual versus predicted residuals for a case where there is no spatial correlation in the data: Acemoglu, Johnson and Robinson (2001).

different spatial corrections returned significance levels that varied by several orders of magnitude. The second uses data on the determinants of inter-generational mobility due to Chetty et al. (2014).

The Autor, Dorn and Hanson (2013) results are shown in Table 5. It can be seen that there is strong autocorrelation in residuals (the significance level of the Moran statistic is of the order 10^{-57}, and the spatial structure parameter is 0.66). There is no anisotropy in the residuals, and the effective range (where residual correlation falls to 0.14) is 129 km.

Importantly, the Residual Fit, the squared correlation between the residuals and their predicted values based on the estimated spatial covariance matrix, is 0.93. The table shows standard errors clustered by state using the default *Stata* standard error correction, the same that were used in the first column of Table 1, and the direct standard errors. What can be seen is that, for the spatial structure of these data, the original clustered standard errors are almost identical to the direct standard errors. Naturally, this behaviour does not occur in general, even with data for the same commuting zones, as the next example shows.

Table 6 gives the results of a regression of inter-generational mobility on a variety of social and inequality indicators from Chetty et al. (2014) for commuting zones, again clustered at state level. Here the relationship between clustered and direct standard errors is

	Coefficient	Clustered SEs	Direct SEs
(Intercept)	7.369	(5.471)	(4.956)
Chinese Imports	-0.121	(0.027)	(0.027)
Mfg Employment	-0.060	(0.012)	(0.010)
College Educated	0.031	(0.019)	(0.017)
Foreign Born	-0.009	(0.016)	(0.017)
Female Employment	0.089	(0.028)	(0.028)
Routine Tasks	-0.344	(0.047)	(0.043)
Longitude	-0.007	(0.019)	(0.021)
Latitude	-0.133	(0.069)	(0.070)
Longitude*Latitude	-4.240	(2.957)	(2.800)
R^2	0.564		
Moran	0.000		
Cook	0.116		
Residual Fit	0.923		
Spatial Structure	0.663		

Clustered SEs denote standard errors clustered by state using the default *Stata* correction. 722 observations. Effective range is 129 kilometres, and the anisotropy ratio and angle are 1.03 and 9 degrees. Moran is the p value of the I statistic. Cook is the maximum Cook's distance of the observations. Residual Fit gives the squared correlation between the regression residuals and their predicted value based on the estimated spatial covariance matrix. Structure is the share of spatial structure in residual variance given in equation 7.

Table 5: Autor, Dorn and Hanson (2013). Impact of China Shock.

more interesting, with the size of direct standard errors varying from 0.7 to 1.8 times the size of the original cluster ones.

7 Illustration: An Unbalanced Panel

I now give an example of direct standard errors for an unbalanced panel, based on Donaldson (2018). This paper examines the impact of railroad construction on Indian development from 1870 to 1929. I focus on the first column of Table 4 which regresses the log of real agricultural income of each district on a dummy variable for whether or not the district had a railroad, with fixed effects for district and year. Standard errors were clustered by district using the default *Stata* adjustment. Including longitude and latitude coordinates here did not alter the results materially and led to occasional problems of convergence

	Coef	Robust SE	Direct SE
(Intercept)	94.763	(31.430)	(38.697)
Gini Bottom 99	-0.056	(0.050)	(0.065)
Fraction Short Commute	0.245	(0.032)	(0.057)
High School Dropout Rate	-0.108	(0.047)	(0.035)
Social Capital Index	0.173	(0.041)	(0.050)
Fraction Single Mothers	-0.417	(0.060)	(0.051)
Longitude	-123.296	(41.443)	(50.959)
Longitude squared	44.179	(15.351)	(18.797)
Longitude cubed	0.000	(0.000)	(0.000)
R^2	0.811		
Moran	0.000		
Cook	0.075		
Residual Fit	0.956		
Spatial Structure	0.852		

580 observations. Effective range is 162 kilometres, and the anisotropy ratio and angle are 1.3 and 31 degrees. Moran is the p value of the I statistic. Cook is the maximum Cook's distance of the observations. Residual Fit gives the squared correlation between the regression residuals and their predicted value based on the estimated spatial covariance matrix. Structure is the share of spatial structure in residual variance given in equation 7.

Table 6: Chetty et al. (2014): Determinants of Inter-generational Mobility.

due to strong multicollearity with district dummies, so they are not used in the reported results.

It is necessary to decide on values for the temporal autocorrelation parameter α and the parameters underlying the spatial covariance matrix Σ. Using the raw (no fixed effect) regression residuals, the mean correlation between residuals across consecutive years was $\alpha = 0.94$.

For spatial correlation, the parameters of the covariance function we estimated for each year and the median used to calculate Σ. Applying an exponential covariance $\kappa = 0.5$, the median range 2ρ was 256 km, the spatial variance σ^2 was 0.3, and the spatial structure parameter $\sigma^2/(\sigma^2+\tau^2)$ was 0.6. The median anisotropy of residuals across years involved a ratio of 1.3 and an angle of 17 degrees: imposing a ratio of one and an angle of zero degrees (isotropy) reduced the standard error estimates by 1 per cent and these uncorrected results are reported here.

The data are heavily unbalanced: of a potential 11,520 entries there are 7,086 observations. This unbalancedness reflects in particular the sparseness of early data: before 1890

	1870–1929	1870–1889	1890–1929
Coefficient	0.1636	0.1107	0.0787
Clustered SE	(0.0501	(0.0711)	(0.0491
Direct SE	(0.0581)	(0.1458)	(0.0845)
R^2	0.9996	0.9999	0.9996
Cook	0.0428	0.0601	0.0456
N	7086	991	6095
Residual Fit	0.8355	0.8160	0.8410

Cook is the maximum Cook's distance of the observations. Residual Fit gives the squared correlation between the regression residuals and their predicted value based on the estimated autocovariance matrix. The covariance matrix is based on the averages across all years, with a temporal autocorrelation of 0.94, smoothness $\kappa = 0.5$, effective range $2\rho = 128$ km, variance $\sigma^2 = 0.3$ and structure $\sigma^2/(\sigma^2 + \tau^2) = 0.6$. No correction for anisotropy was applied.

Table 7: Donaldson (2018). Impact of railways on real agricultural income.

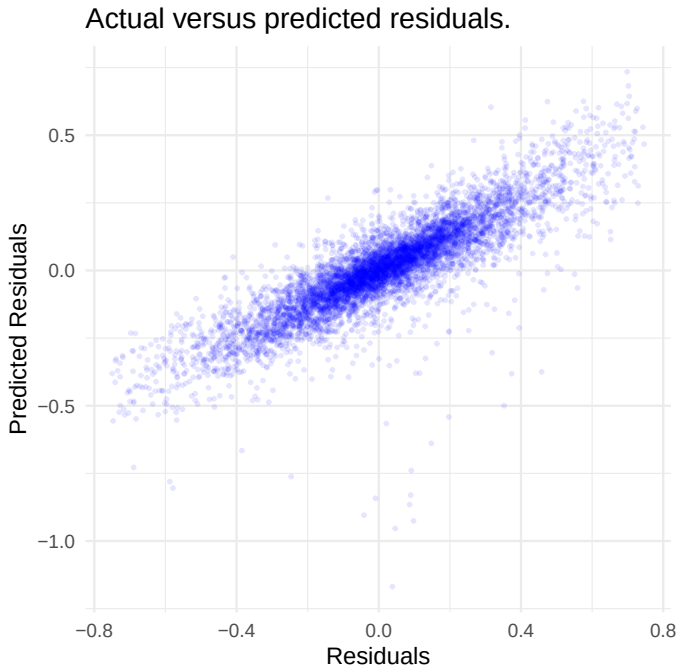

Figure 4: Actual and predicted residuals from the estimated spatial covariance matrix from Donaldson (2018).

	Coef	Robust SE	Direct SE
(Intercept)	-6.368	(8.837)	(11.313)
Tsetse SI	0.767	(0.374)	(0.606)
Tropics	3.252	(1.728)	(1.742)
Temperature	0.132	(0.357)	(0.425)
Humidity	0.037	(0.160)	(0.210)
Temperature*Humidity	-0.001	(0.006)	(0.009)
Long	-0.002	(0.041)	(0.061)
Abs Lat	0.145	(0.080)	(0.104)
Long*Abs Lat	-0.001	(0.002)	(0.003)
Moran	0.000		
Cook	0.067		
Residual Fit	0.824		
Spatial Structure	0.551		

423 observations. Effective range is 358 kilometres, and the anisotropy ratio and angle are 1.1 and -58 degrees. Moran is the p value of the I statistic. Cook is the maximum Cook's distance of the observations. Residual Fit gives the squared correlation between the regression residuals and their predicted value based on the estimated spatial covariance matrix. Structure is the share of spatial structure in residual variance given in equation 7.

Table 8: Logistic regression of the impact of Tsetse fly suitability on the presence of indigenous slavery from Alsan (2015).

there are 991 data points. Table 7 therefore reports results for each sub-period as well as for the entire sample; giving the estimated impact of having a railway, the original clustered standard error, and the direct standard error. Because the data are longitudinal, no Moran statistic is reported, although in each year the autocorrelation is substantial as shown by the estimated spatial structure parameter of 0.6.

The reliability of the assumed covariance matrix, measured by the squared correlation between actual and predicted residuals, ranges from 0.83–0.84. The direct standard errors in each case range from 1.16 to 2 times the original clustered ones.

8 Illustration: A Logistic Regression

The illustration here comes from Table 1 of Alsan (2015) and regresses the presence or absence of indigenous slavery on an index of Tsetse fly suitability and some climate controls. Whereas the original regression assumed a linear model, I will use a logistic function here as an example of a non-linear model. The second column of 8 gives standard errors clus-

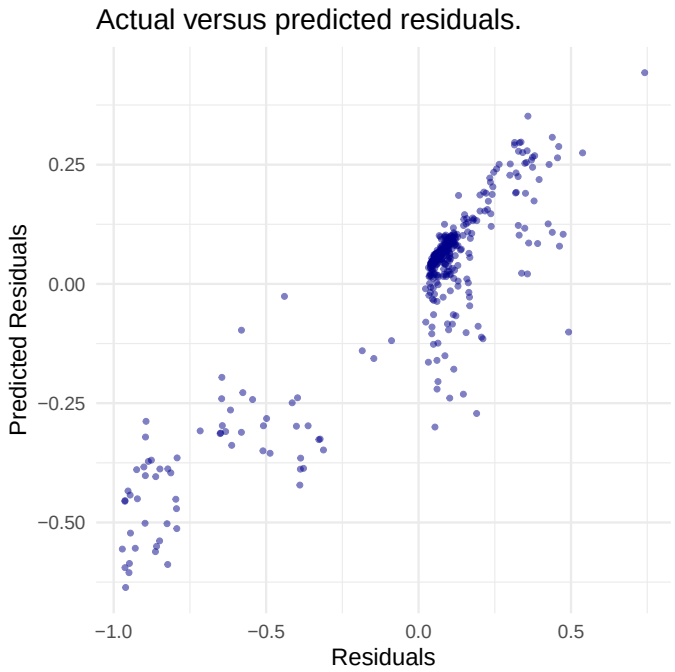

Figure 5: Actual and predicted residuals from the estimated spatial covariance matrix for a logistic regression based on Alsan (2015)

tered at the level of province using *Stata* defaults as in the original study. The residuals show strong spatial autocorrelation, and the residual fit and spatial structure statistics of 0.8 and 0.55 suggest that the estimated spatial covariance structure is not unreasonable. The direct standard error on the Tsetse variable is about two thirds larger than the clustered estimate.

9 Conclusions

With the growing popularity of regressions using spatial data it is worth bearing in mind Tobler's (1970) First Law of Geography: "everything is related to everything else, but near things are more related than distant things." Spatial data, in other words, tend to be autocorrelated. Although there exist a variety of standard error corrections to handle this, in practice as we have seen, their results can vary substantially. This paper instead proposed a simple technique to estimate standard errors directly from the spatial correlation structure of regression residuals and showed how their reliability can easily be tested by comparing the actual values of the residuals with the values predicted by the assumed covariance matrix.

Appendix

This Appendix looks at the size of various standard error corrections, based on regressing two spatial noise series on one another. Each noise series is generated as a Whittle-Matérn process with smoothness $\kappa = 1$, variance $\sigma^2 = 1$ and idiosyncratic variance $\tau^2 = 0$, at a variety of scales. We use three patterns of points. The first is 100 points spread at random over a square with sides of length 100. The second uses the pattern of the 722 commuting zones in the 48 contiguous US states from Autor, Dorn and Hanson (2013), and the third is 522 African tribal areas from Alsan (2015). The last two are rescaled to turn the maximum east-west and north-south distances to 100.

Table A1 reports the fraction of 1000 simulations where a *t* statistic significant at 5 per cent was returned. First, an unadjusted standard error is given. Next, standard errors are clustered and corrected using a Bell-McCaffrey adjustment. For the random data, the data are clustered into 16 squares, for commuting zones they are clustered by state, and tribes are clustered by district. Next are Bester, Conley and Hansen (2011) large clusters, where the data are divided into two, three and four vertical stripes, and a Bell-McCaffrey correction again applied. Finally there are Conley (1999) corrections with a rectangular kernel of varying cutoff lengths.

Several things are immediately apparent. The first is that, when it comes to random noise regressions, uncorrected and clustered standard errors are a good deal too small. The Conley HAC correction works fairly well when the data have a short scale, but in general its performance depends strongly on knowing the correct cutoff to use, and tends to disimprove at longer scales. The most reliable performance for random sites and commuting zones comes from the Bester *et al* correction when the data are split into three clusters, although it tends to accept more regressions with the African pattern of points.

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.

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

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

Autor, David H., David Dorn and Gordon Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103:2121–2168.

			Large Clusters			HAC Cutoff					
Range	None	Clustered	2	3	4	5	10	15	20	30	40
100 random sites.											
10	0.185	0.067	0.035	0.057	0.050	0.062	0.022	0.026	0.051	0.082	0.102
20	0.343	0.101	0.027	0.047	0.059	0.165	0.063	0.036	0.033	0.059	0.088
30	0.487	0.146	0.028	0.061	0.084	0.296	0.145	0.072	0.050	0.048	0.074
40	0.550	0.169	0.027	0.079	0.094	0.367	0.204	0.122	0.064	0.048	0.081
US Commuting Zones.											
10	0.571	0.269	0.030	0.049	0.056	0.042	0.017	0.034	0.076	0.116	0.146
20	0.751	0.404	0.031	0.064	0.077	0.181	0.063	0.050	0.077	0.124	0.189
30	0.813	0.499	0.035	0.074	0.092	0.336	0.141	0.099	0.112	0.160	0.204
40	0.843	0.500	0.026	0.067	0.108	0.375	0.186	0.125	0.115	0.171	0.206
African Tribal Areas.											
10	0.521	0.193	0.031	0.060	0.049	0.114	0.051	0.046	0.055	0.100	0.120
20	0.663	0.308	0.019	0.063	0.073	0.269	0.152	0.106	0.090	0.109	0.138
30	0.731	0.390	0.021	0.089	0.094	0.409	0.247	0.185	0.147	0.150	0.168
40	0.777	0.432	0.027	0.079	0.102	0.466	0.305	0.226	0.163	0.151	0.193

Table A1: Percentage of simulations where noise regressions are significant at 5 per cent under different standard error corrections. See the text for details.

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.

Bell, Robert M. and Daniel F. McCaffrey. 2002. "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples." *Survey Methodology* 28:169–182.

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.

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.

Chorti, A. and D. T. Hristopulos. 2008. "Non–parametric Identification of Anisotropic (Elliptic) Correlations in Spatially Distributed Data Sets." *IEEE Transactions on Signal Processing* 56:4738–4751.

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

Cressie, Noel A. C. 1991. *Statistics for Spatial Data*. New York: Wiley.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* 108:899–934.

Dubrule, Olivier. 1983. "Cross Validation of Kriging in a Unique Neighborhood." *Mathematical Geology* 15:687–699.

Furrer, Reinhard, Marc G Genton and Douglas Nychka. 2006. "Covariance Tapering for Interpolation of Large Spatial Datasets." *Journal of Computational and Graphical Statistics* 15(3):502–523.

Gneiting, Tilmann, Marc G. Genton and Peter Guttorp. 2006. Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry. In *Statistical Methods for Spatio-Temporal Systems*. London: Taylor and Francis.

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.

Imbens, Guido W. and Michal Kolesár. 2016. "Robust Standard Errors in Small Samples: Some Practical Advice." *Review of Economics and Statistics* 98:701–712.

Lindgren, Finn, Havard Rue and Johan Lindstrom. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach (with Discussion)." *Journal of the Royal Statistical Society: Series B* 73:423–498.

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

Zeileis, Achim. 2006. "Object-oriented Computation of Sandwich Estimators." *Journal of Statistical Software* 16(9):1–16.