Spurious weather effects

Jo Thori Lind†

Friday 10th February, 2017

Abstract

Rainfall is exogenous to human actions and hence popular as an exogenous source of variation. But it is also spatially correlated. I show that this can generate spurious relationships between rainfall and other spatially correlated outcomes both theoretically and using simulation. As an example, rainfall on almost any day of the year has seemingly high predictive power of electoral turnout as well as median incomes in Norwegian municipalities. In Monte Carlo analyses, I find that standard tests can reject true null hypotheses in as much as 99% of cases, and standard approaches to estimating consistent standard errors do not solve the problem. Instead, I suggest controlling for spatial and spatio-temporal trends using multi-dimensional polynomials to solve the problem.

Keywords: Rainfall, electoral turnout, spurious relationship, spatial correlation

JEL Codes: C13, C14, C21, D72

Word count: 8211 (including 5 half page graphs)

*I am grateful for useful comments from Monique de Haan, Andreas Kotsadam, Kyle Meng, Per Pettersson-Lidbom, and participants at the (EC)² conference, the Oslo Empirical Political Economics workshop, and seminar participants at the University of Oslo. I also got invaluable help with the meteorological data from Ole Einar Tveito. While carrying out this research I have been associated with the center Equality, Social Organization, and Performance (ESOP) at the Department of Economics at the University of Oslo. ESOP is supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 179552.

†Department of Economics, University of Oslo, PB 1095 Blindern, 0317 Oslo, Norway. Email: j.t.lind@econ.uio.no. Tel. (+47) 22 84 40 27.
1 Introduction

Both human behavior and economic outcomes, such as individual effort and agricultural productivity, depend on weather conditions. Moreover, weather phenomena can generally be seen as exogenous to human behavior. The effect of rainfall on the demand for umbrellas, known from numerous introductions to decisions under uncertainty, is only one example. In empirical economic research, truly exogenous variables are sought after as sources of exogenous variation to provide causal inference. At least going back to Koopmans (1949), economists have been using rainfall as an exogenous shock. In recent years, the strategy has become increasingly popular. A search on Google Scholar for “rainfall” and “exogenous” yields 3410 papers written in 2000 increasing to 12 700 in 2015.\(^1\)

Few suspect that human actions affect the weather in the short run,\(^2\) and the weather has a potential impact on a number of outcomes. But by its very nature, rainfall is spatially correlated. If it is raining in one location, the likelihood of rain in nearby areas is high. Autocorrelated explanatory variables are not usually considered problematic, but I show that this induces a danger of spurious correlations when outcomes of interest also contain spatially correlated variation.

The key problem is that in both cross sectional and panel data, it is common to observe spatial patterns in most outcomes of interest. When these are regressed on rainfall, the spatial patterns in the two variables are almost always going to concur in one way or another. Although there may be no real relationship between the two, conventional tests will indicate a relationship. In panel data, where spatial trends can be controlled by fixed effects, the same problem arises if there are spatially dependent time trends in outcomes of interest. Moreover, this is primarily an effect on the estimated parameters, so estimating standard errors that take into account spatial clustering at geographic entities (Moulton, 1986) or following e.g. Conley’s (1999) approach, does not solve the problem.\(^3\) Specifically, his proof of consistency of OLS with spatially dependent data relies on spatially stationary data, which is typically not the case with rainfall data.

To illustrate the magnitude of the problem, consider the relationship between electoral turnout and rainfall. There may be good reasons to expect a relationship between turnout and rainfall on election day. But rainfall on other days, with a possible exception of a few

\(^1\)The exact search term is “(rainfall OR rain OR weather) exogenous”. The search for “exogenous” alone reaches a peak at 136 000 hits in 2006, but has fallen to 49 600 in 2015.

\(^2\)As pointed out by Miller (2015), however, although weather phenomena are exogenous, they may in many cases be predicted in advance.

\(^3\)An econometric model that properly models the spatially correlated pattern of rainfall as well as other variables of interest might be able to yield correct results. However, no such model is readily available to practitioners and in the fairly extensive literature reviewed below, no such attempts are made.
days prior to the election, should not have any impact. Using data from Norwegian municipal elections and daily rainfall data in the window from 600 days before to 600 days after the election, I find that rainfall on almost every day has an impact on electoral turnout. As the regressions include both municipality and year dummies, explanations such as probability of rain can not explain the findings.

In these analyses we should only expect to find significant results due to the expected Type I errors determined by the level of significance. That is not the case. Rather, a 5% significance test reject the hypothesis of no effect of precipitation in 70.8% of the cases. The estimated t-values are shown in Panel (a) of Figure 1. Although the distribution is symmetric around zero, the variance is much higher than the expected level of unity. Moreover, the distribution is not normal as the tails are lighter than the normal kurtosis. Clustering standard errors at the regional, rather than the local level, improves the situation somewhat. But still a 5% test rejects in 46.4% of cases, and this level of clustering is much stronger than what is used in most applied studies.

If this phenomenon was unique to electoral turnout, it would be a narrow problem. However, rainfall exhibits a similar pattern with other variables. As an example, I regressed median gross income by municipality and year on rainfall on every day of the year, one day a time. The distribution of the t-values are shown in Panel (b) of Figure 1. A 5% test of the (probably) true null hypothesis of no relationship rejects on 62.5% of days. One could expect to find a relationship between income and rainfall in countries with a large agricultural sector (Paxson, 1992), but in that case only between rainfall over longer periods of time. Moreover, Norway does not have an important agricultural sector any more.

In both of the cases considered above, there are spatio-temporal tends in the dependent variable: turnout in the eastern part of the country has decreased relative to national averages whereas it has increased in relative terms in the western part. The underlying explanation is probably differential development of macro-regional common factors. In the case of income, the oil industry has had the strongest positive effect on incomes in the South-Western part.

---

4The estimation uses data from ten elections between 1971 and 2007 using a two way fixed effects specification. See Section 6.1 for further details.

5One explanation for this feature could be that the distribution of precipitation has heavy tails or other irregularities in the data. However, results remain the same if one looks at dummies for precipitation above different thresholds, normalize by municipality means or variances, normalize the turnout variable and so on.

6The problem is not unique to rainfall. First, various terrain characteristics such as ruggedness and gradients, the use of which was popularized by e.g. Duflo and Pande (2007), clearly exhibit spatial correlation. In the empirical analysis of violent conflict, for instance, researchers have found strong spatial and spatio-temporal patterns (see e.g. Buhaug and Gleditsch, 2008; Weidmann and Ward, 2010). There is also a vast literature on the spatial and spatio-temporal nature of housing and property prices (see e.g. Holly et al., 2011; Brady, 2014). More examples could easily be listed. In all these cases, there is a danger of spurious relationships stemming from joint spatial patterns.
Notes: The graph shows the coefficient from two way fixed effects regression of electoral turnout (Panel a) and median gross income (Panel b) on daily precipitation. Standard errors are clustered at the municipal level. In Panel a, precipitation for 600 days before to 600 days after election day employed, but data from +/- 10 days are excluded. In Panel b, data for the 365 days of the year are used.

of the country whereas the North, dependent on small scale fisheries, has seen economic decline. In some ways this mimics the intricate spatial patterns detected in earnings data by Barrios et al. (2012), but here also with a temporal component.

In a stylized model with an omitted spatially correlated variable, I show theoretically that OLS estimates are inconsistent with almost surely diverging point estimates and t-statistics. Using Monte Carlo simulations, I further show that the test of the irrelevance of the irrelevant variable is vastly over-rejected. This holds both with large sample sizes, with several independent clusters of spatially dependent variables, using cross-sectional and panel data, as well as several specifications of the omitted trend including spatial AR processes. Also, the simulations indicate that neither ordinary clustered standard errors nor Conley (1999) standard errors solve the problem.

My suggested solution to the problem is to control for the trend using orthogonal polynomials.\textsuperscript{7} In the cross sectional spatial case, such a trend would be a polynomial in geographical coordinates. In the case of panel data, we need a time trend whose slope varies geographically, so the slope of the trend is modeled by a similar polynomial in geographical

\textsuperscript{7}A number of papers using rainfall as an exogenous source of variation, e.g. Brückner and Ciccone (2011) and Fujiwara et al. (2016), also include spatially varying trends. However, it is not yet well understood why this is necessary.
coordinates. Although any polynomial can in theory be used, sequences of orthogonal polynomials have good numerical stability. In the current study, I focus on tensor products of Legendre polynomials, which seem to perform well.

The use of meteorological data in empirical analyses has skyrocketed in recent years. Some of these take worries of spurious correlations into account by running placebo studies, but this is not yet widespread. Among the first applications of meteorological data where studies using annual and seasonal weather conditions to study agricultural output and hence serve as an instrument for income – see Dell et al. (2014) for a survey of this literature. More recently, short term weather conditions have also caught researchers’ attention. One strand of literature is based on the relationship between weather conditions and people’s mood. An early contribution along these lines in economics is Saunders’ (1993) finding that US daily stock prices are affected by weather conditions in New York City, the city where they were traded.

Starting with Gomez et al. (2007) and Hansford and Gomez (2010), there is also by now a fairly large literature on the relationship between election day weather and turnout. Beyond the US, the question has been studied in Japan, Holland, Spain, Italy, Sweden, and Norway (Horiuchi and Saito, 2009; Eisinga et al., 2012b,a; Artés, 2014; Sforza, 2013; Lo Prete and Revelli, 2014; Persson et al., 2014; Lind, 2014). In many studies, it is found that rain on election day reduces turnout, but in Sweden there seems to be essentially no relationship between the two and in Norway the relationship is positive. Daily weather conditions have also been found to have an impact on participation in civil rights riots in the 1960s (Collins and Margo, 2007), Tea Party rallies (Madestam et al., 2013), and May day demonstrations (Kurrid-Klitgaard, 2013).

There is also clear evidence that the weather on a specific day affects the labor market: Male works have been found to work more on rainy days (Connolly, 2008) and labor productivity seems to be higher (Lee et al., 2014). Graff Zivin et al. (2015) find that NLSY survey respondents’ math performance depends on the temperature on the day of observations. Connolly (2013) shows that answers to well-being surveys are affected by the weather on the interview day. Guven and Hoxha (2014) build on this research and use sunshine as an instrument for happiness to find the effect of happiness on willingness to take risk.

In other studies of the effect of daily weather conditions, Simonsohn (2010) finds that the probability of enrollment into colleges is related to cloud cover of the day of visiting the college. Busse et al. (2014) find that the car purchase decisions are affected by daily weather

---

8There is a large literature on this topic in psychology. See Cunningham (1979) for a seminal contribution and e.g. Denissen et al. (2008) and Keller et al. (2005) for more recent contributions.

9See e.g. Frühwirth and Sögner (2015) for an updated overview of the relationship between financial markets and the weather.
conditions: it is more common to buy convertibles on warm days without rain and 4x4s on cold days with rain or snow. Carr and Doleac (2014) use variation in rainfall in the afternoon on incapacitating potential offenders to derive the causal effect of potential offenders on gun violence. Sen and Yildirim (2015) use rain on a given day as an instrument for the number of readers an online newspaper article gets, based on the idea that potential readers spend more time indoors on rainy days and hence have more time to read online newspapers.

There are a few papers scrutinizing the validity of rainfall as an exogenous source of variation. Sarsons (2015) look at the relationship between rainfall and conflict in India, where rainfall typically is believed to be an instrument for income. However, she finds strong effects of rainfall both in rain-fed and dam-fed regions, possibly invalidating this identification.

Methodologically, the present paper is also closely related to the work of Bertrand et al. (2004). They show that differences in differences estimation tends to find effects effects of placebo “reforms” on female wages. In one way, the present paper is a converse to their study as they focus on outcomes with spatial patterns (wages) whereas I focus on explanatory variables with spatial patterns. The study by Barrios et al. (2012) also indicate that failing to accounting for spatial dependency may severely bias the results, but they also focus on standard errors.

The paper is also related to literature on spurious regressions in time series. In some ways it relates to the presence of spurious regression in regressions with non-stationary variables (Granger and Newbold, 1974; Phillips, 1986). Also, my suggested solution to estimate spatial or spatio-temporal trends relates to the literature on estimating time trends (Sims et al., 1990). As this concerns units in space, it also relates to the massive literature on spatial statistics and the more modest literature on spatial econometrics. The tiny literature on spatial co-integration yields results resembling those I find. In a number of Monte Carlo analyses, Fingleton (1999) finds distributions of t-values resembling those in Figure 1 for unrelated spatially non-stationary processes.

The literature on spatio-temporal statistics has a strong focus on space-time autoregressive moving average (STARMA) type models (Cliff et al., 1975; Pfeifer and Deutsch, 1980), characterized by linear dependence lagged in both space and time. The literature on estimating models with spatially dependent error terms is particularly relevant (Kelejian and Prucha, 1999; Chudik et al., 2011; Pesaran and Tosetti, 2011). Such models can also be extended to regression frameworks with spatial autoregressive distributed lags models (Elhorst, 2001). Although these models may be suited to handle the problem at hand, their

---

10 Cressie (1993) and Ripley (2004) provide introductions to parts of the literature.
11 See e.g LeSage and Pace (2009) for an introduction.
Notes: The graph shows the distribution of the $t$-values when regressing municipal turnout on daily precipitation for 600 days before and after election day. The 10 days before and after the actual election day are omitted. Panel (a) shows results from regressing levels on levels. Panel (b) shows the regression of turnout on a dummy for more than 25 mm rain while Panel (c) employs a dummy for any rain. Panel (d) shows results from a regression where the rank of turnout is measured on the rank of rain, i.e. both variables are uniform on the unit interval.

main problem is that they are difficult to identify and estimate by themselves. When we also want to add panel data features, clustered standard errors, instrumental variables or discontinuity designs, they become intractable and not useful for practical applications. Hence I have chosen to rely on a simpler approach.

My suggested solution is to allow for a spatially varying time trend. This relates to the literature on varying coefficients (Hastie and Tibshirani, 1993) and particularly spatially varying coefficients (Gelfand et al., 2003). Specifically, Hoover et al. (1998) and Huang et al. (2002) estimate varying coefficients models where they model the coefficients by regularized basis functions as I suggest (albeit using B-splines rather than polynomial bases). However, they consider coefficients varying in time, not in space. To the best of my knowledge, the only spatial application of the methodology is Zhu et al. (2014) who study MRI images.

Finally, there is a quite substantial literature on spatio-temporal modeling of weather phenomena (Stern and Coe, 1984; Brown et al., 2001; Velarde et al., 2004), but this literature generally has completely different objectives than the current paper.

2 The nature of the problem

As already illustrated, regressions with rainfall as the independent variable seems to yield too high $t$-values. Before exploring the explanation based on spatio-temporal trends further,

\footnote{See also Matsui et al. (2011, 2014) for some recent development.}
it is worthwhile dispensing with the alternative explanation of outliers in precipitation. It is well known that rainfall data has a heavy right tail, which could affect the regression analyses. To show that this cannot be the sole explanation, Figure 2 shows the distribution of the t-values in a number of specifications that reduces the leverage of outliers. Panel (a) is the specification shown in the introduction, where the level of turnout is regressed on the level of rain in millimeters. Panels (b) and (c) replace the measure of precipitation with dummies for substantial rain, defined as above 2.5 mm, and any rain at all. Finally, in Panel (d) both rainfall and turnout are measured using their ranks so they both have a uniform distribution on the unit interval. In all four cases, the distribution of the test statistic is far from the standard normal or t-distributions we would expect. The four measured t-values are indeed heavily correlated.\footnote{See also the matrix plot in Appendix Figure A-3.} This should indicate that mere outliers cannot explain the findings.

Rather, it seems that the nature of the problem is omitted variable exhibiting spatio-
Table 1: Monte Carlo simulations based on actual rainfall

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Skewness</th>
<th>Share rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) No clustering</td>
<td>0.38</td>
<td>9.13</td>
<td>0.43</td>
<td>0.84</td>
</tr>
<tr>
<td>2) Clustered on region</td>
<td>0.25</td>
<td>3.41</td>
<td>1.20</td>
<td>0.59</td>
</tr>
<tr>
<td>3) Conley standard errors</td>
<td>0.18</td>
<td>3.15</td>
<td>0.62</td>
<td>0.59</td>
</tr>
<tr>
<td>4) Random municipality</td>
<td>-0.04</td>
<td>1.00</td>
<td>0.17</td>
<td>0.05</td>
</tr>
<tr>
<td>5) Panel</td>
<td>-0.62</td>
<td>7.96</td>
<td>0.12</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Notes: The Table shows the results of Monte Carlo analyses simulating data as in (2) with $\epsilon_i \sim N(0,1)$ and 1000 replications. All regressions are cross sectional analyses except the last row which is based on a two way fixed effects panel data estimator.

temporal trends and potentially spatial non-stationarity\(^{14}\) in the rainfall data in combination with spatio-temporal trends in the turnout data. To explore the latter trends, I run a regression of the type

$$\text{Turnout}_{it} = \alpha_i + \tau_t + \delta_i t + \epsilon_{it}$$ (1)

Here $\delta_i$ is an estimate of the municipality specific trend. I have plotted the values of $\delta_i$ in Figure 3. Panel (a) shows the geographical distribution of temporal trends. It is clear that there is a strong negative trend in the eastern part of the country and a positive trend in parts of the west and the center. Panel (b) shows a Moran plot where the municipality specific coefficient $\delta_i$ is plotted against the average $\delta_i$ in the adjacent municipalities. Again it is clear that there is a spatial pattern. Formally, Moran’s I statistic is $I = 0.456$ and the Moran test for no spatial dependency rejects with a p-value of $2.2 \times 10^{-16}$. We conclude that when controlling for two way fixed effects, turnout has been declining in the eastern part of the country and increasing in the western part. As shown in Sections 3 and 4, this can explain the t-values shown in Figure 2.

To show more forcefully that spatially dependent outcome variables may create spurious correlation patterns with weather data, I run some simulations. The independent variable is the observed rainfall in Norwegian municipalities on a random day between January 1st 1968 and November 30th 2012.\(^{15}\) First, I consider a cross sectional setting where I simulate dependent variables

$$z_i = x_i + y_i + \epsilon_i$$ (2)

\(^{14}\)Spatial non-stationarity is commonly found in meteorological data even after demeaning, see e.g. Fuglstad et al. (2015) and Ingebrigtsen et al. (2015). However, to the best of my knowledge no general procedure for formally testing for stationarity exists.

\(^{15}\)See Appendix B for details on the meteorological data.
where $x_i$ and $y_i$ is the longitude and latitude of the municipal centre of municipality $i$ and $\epsilon_i$ is a standard normally distributed residual. Regressing $z_i$ on observed rainfall yields the results shown in Row 1 of Table 1. The full distributions are shown in Appendix Figure A-2. As in the example above, we see massive over rejection of false null hypotheses: Here the test at the 5 % level rejects in 84 % of cases.

To account for the spatial pattern, I also attempt to cluster standard errors at the 19 regions of Norway as well as using Conley (1999) standard errors. Both approaches seems to have about the same effects of increasing the standard errors. However, as can be seen from Rows 2 and 3 of Table 1, we still have rejection of the false null in 59 % of cases using a 5 % test. Moreover, we notice that this tends to yield a heavily skewed distribution of the t-values. First, this indicates that at least a substantial part of the problems lies in the estimated regression coefficient $\hat{\beta}$, not in the computation of the standard errors. In some ways, this is unsurprising as biased regression coefficients are well known results from omitted variables problems. Second, it tells us that the problem can not be solved by correcting the standard errors alone.

I also ran simulations as above, but where municipalities were assigned rainfall data from a randomly selected municipality. This maintains the distribution of the independent variable but removes the spatial pattern. As we can see from Row 4 of Table 1, we now get almost a perfect replication of the standard normal distribution. This indicates that the cause of the problem is the spatial correlations and not the skewed distribution of the rainfall data.

A typical way to control for such spatial patterns is to include municipal fixed effects, as was done in the analysis of the turnout data above. As long as the dependent variable only exhibits spatial correlation, this solves the problem. If, however, the variable has a spatially dependent trend, fixed effects can't solve the problem. To illustrate this, I draw panels starting on a random day, with observations every four years and totally 10 observations per municipality. Then I construct a dependent variable as

$$y_{it} = (x_i + y_i) t + \epsilon_{it}$$

Regressing $y_{it}$ on observed rainfall yields the results shown in Row 5 of Table 1. Again, the false null is over rejected, this time in 81 % of cases for the 5 % test.
3 A theoretical exposition of the problem

To show how severe the problem of spurious regressions really is, Monte Carlo simulations are probably the best tool. But to get a better grasp of the nature of the problem at hand, I first develop a very simple econometric model that is able to generate the phenomenon and may approximate the problem in real world situations. Specifically, we want to show that an irrelevant spatially correlated explanatory variable may seem relevant in the presence of spatially correlated omitted variables.

Consider a case where space is reduced to one dimension, so observations can be represented as points \( i = 1, \ldots, N \) on a line. This implies that as \( N \) grows, the size of the line grows. The case with simply more densely distributed points is considered below. We generate an explanatory variable \( r_i \) as the sum of a number of \( K \) spatially correlated shocks. Specifically, each shock has a location \( p_k \in [1, N] \) and a value \( \nu_k \) drawn from some continuous distribution. In the case of precipitation, we may think of each shock as a weather system with intensity \( \nu_k \) and center at \( p_k \). At position \( i \), the total effect of shocks is

\[
    r_i = \sum_k \frac{\nu_k}{1+d(i,p_k)}
\]

where \( d \) is a distance function which satisfies \( d(i, i) = 0 \) and \( d(i, j) > 0 \) when \( i \neq j \). Notice that this is essentially a radial basis function network, which is commonly used to approximate functions (Buhmann, 2003). Hence this model should approximate a wide range of spatial patterns found in real life. In the theoretical exposition, I focus on the metric \( d(i, j) = |i - j| \).

The outcome variable \( y_i \) may depend on \( r_i \) - this is the key empirical question. In addition it depends on a variable with a spatial trend which is omitted from the analysis, generating an omitted variables problem with a specific structure. To emulate this, the true model is

\[
y_i = \alpha + \beta r_i + \tau i + \epsilon_i.
\]

The spatial trend is modelled as \( \tau i \) real number \( \tau \). We want to test the hypothesis that \( \beta = 0 \), and the issue is the effect of neglecting the trend \( \tau i \). The core of the problem is that the regression analysis may mistake the trend \( \tau i \) for the signal \( r_i \).

The case with \( K = 1 \), i.e only one shock \( p \), is illustrated in Figure 4. As is apparent from the figure, whenever the “position” of the shock is \( p \neq \frac{N}{2} \), there is scope for the shock to pick up parts of the trend. I show formally that this is indeed so below. Moreover, as \( N \) grows, the problem does not diminish but rather get more acute.

Specifically, when the data are generated as in equation (3), but where we fail to control

\[16\] For simplicity I condition on given values of \( p_k \), but little would change if these were drawn from some continuous distribution on \([0, N]\).
for the trend $\tau_i$ in the analysis, the OLS estimator yields

$$
\hat{\beta} = \beta + \frac{1}{N} \sum_{i=1}^{N} (r_i - \bar{r})\epsilon_i + \tau\frac{1}{N} \sum_{i=1}^{N} (r_i - \bar{r})i 
$$

(4)

When $\epsilon_i \sim NID(0, \sigma^2)$ and the number of shocks $K$ grows at the same rate as $N$, the first fraction has converges to zero and has an asymptotic normal distribution. This is handled by ordinary estimation and hypothesis testing procedures.\textsuperscript{17} The second term, which stems from the omitted variable, is more problematic. In finite samples it it non-zero unless $p = \frac{N}{2}$. Moreover, the problem is exacerbated with growing sample sizes, as the following result demonstrates:

**Proposition 1.** When $N \to +\infty$ and $\frac{K}{N} \to \kappa < +\infty$, we have $Pr\left( |\hat{\beta} - \beta| \to +\infty \right) = 1$.

The full proof is provided in Appendices A.1 and A.2. Denote by $\bar{w}$ the average $\bar{w} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{1}{1+|p_k-i|}$. In Appendix A.1 I show that the expression $\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \left( \frac{1}{1+|p_k-i|} - \bar{w} \right)i$ converges to a logarithmic function and hence diverges as $N \to \infty$. The proof is based on showing that the expression can be sandwiched between two harmonic sequences which both have logarithmic growth. Moreover, in Appendix A.2 I show that if $\frac{K}{N} \to \kappa \in \mathbb{R}^+$, we have that as $N \to \infty$, the denominator $\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \left( \frac{1}{1+|p_k-i|} - \bar{w} \right)^2 \to Q \in \mathbb{R}^+$. This is based on showing that $\sum_{i=1}^{N} \sum_{k=1}^{K} \left( \frac{1}{1+|p_k-i|} - \bar{w} \right)^2$ is closely related to the sum of recip-

\textsuperscript{17}As shown in Appendix A.2, the growth of $K$ is necessary to avoid the denominator converging to 0.
rocals of squares of natural numbers. The sum of this sequence is known to converge to $\frac{\pi^2}{6}$. As a consequence, the sum at hand also converges to a constant. Consequently, the expression goes to 0 at rate $O\left(\frac{1}{N}\right)$. If on the other hand we allow $K$ to grow linearly as $N$ grows, the denominator converges to a constant. Still, the estimator $\hat{\beta}$ explodes and hence is inconsistent.

In applied research much emphasis is on statistical significance, i.e. the t-values. As the denominator in (4) converges to a constant, the standard error of $\hat{\beta}$ also converge to a constant. As $\hat{\beta} - \beta$ diverges, this implies that the t-values also diverge. This could hence explain the unusually high t-values observed above.

One objection to this analysis could be that as $N$ increases, the size of space, and hence the range of $r_i$, increases. An alternative model could be to restrict space to say $[0, 1]$ and increase the density of observations as $N$ increases. Then the trend should be modeled as $\frac{r_i}{N}$, and the distance metric for the shocks replaced by $\left| \frac{i}{N} - \frac{p_k}{N} \right|$. However, it is easily seen that in computing OLS estimates, this does not change the final expression and hence the regression estimate still diverges as $N$ increases.

The above results depends on the the specification of $r_i$ being proportional to the inverse of the distance. If the weighting decays more quickly, a may be realistic in many applications, the proof of divergence has to be changed. However, it seems that the main insight would go through with other specifications of the weighting too.

4 Monte Carlo evidence

4.1 The linear case

Table 2 shows a Monte Carlo analysis of the model from Section 3 for sample sizes between 10 and 10000 and number of shocks varying from 1 to 20000. The simulations are based on a model where the true $\beta = 0$ so the fraction of t-tests rejecting the null should correspond to the level of the test, here 5 %. First, we recognize the diverging t-values: The larger the sample gets, the more likely the t-test is to reject. The test at the 5 % level rejects in about half of the cases for small samples and in more than 80 % of cases in larger samples. Rejections rates and values of $|t|$ are slightly smaller for larger numbers of shocks, but this is not enough to take levels down to reasonable magnitudes.

As explained above, results should remain in a model where the size of space (i.e. the

---

18 Typically one can show that although the numerator in (4) converges to zero, the denominator converges at a higher rate so the fraction diverges.

19 These numbers could of course be reduced by increasing the noise, i.e. increasing the variance of $\epsilon_i$, but this does not reduce the importance of the problem.
Table 2: A Monte Carlo analysis of the simple model

<table>
<thead>
<tr>
<th>( K )</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>100</th>
<th>N</th>
<th>2N</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.59</td>
<td>0.57</td>
<td>0.51</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>2.6</td>
<td>2.4</td>
<td>2.4</td>
<td>2.3</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.77</td>
<td>0.67</td>
<td>0.64</td>
<td>0.66</td>
<td>0.62</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>3.5</td>
<td>3.4</td>
<td>3.4</td>
<td>3.3</td>
<td>3.3</td>
<td>3.5</td>
</tr>
<tr>
<td>100</td>
<td>0.79</td>
<td>0.74</td>
<td>0.69</td>
<td>0.68</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>4.0</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.86</td>
<td>0.82</td>
<td>0.82</td>
<td>0.80</td>
<td>0.78</td>
<td>0.75</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>6.1</td>
<td>6.0</td>
<td>5.7</td>
<td>5.6</td>
<td>5.4</td>
<td>5.9</td>
</tr>
<tr>
<td>10000</td>
<td>0.90</td>
<td>0.88</td>
<td>0.86</td>
<td>0.87</td>
<td>0.83</td>
<td>0.86</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>8.8</td>
<td>8.5</td>
<td>8.2</td>
<td>8.3</td>
<td>7.9</td>
<td>7.7</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Notes: The table shows the fraction of cases where a t-test of \( \beta = 0 \) is rejected at the 5% level (first line) and the average of the absolute value of the associated t-value (second line). The true model is \( \beta = 0, \tau = 1, \epsilon_i \sim N(0,1) \), and for each \( k, \nu_k \sim N(0,1) \) and the position \( p_k \sim U(0,N) \). Each model is replicated 1000 times.

length of the line) doesn’t increase when \( N \) increases. Monte Carlo analyses of a similar model constrained to lie on the interval \([0,1]\) yields similar results as above, see Appendix Table A-1.

4.2 Spatial models

In many real world applications, the assumption of a linear world is too restrictive.\(^{20}\) A more realistic assumption is a spatial data structure where it is meaningful to talk about the distance between two observations, and where units tend to be correlated with nearby units. Denoting observation \( i \)'s geographical position \((x_i, y_i)\), we can use the Euclidean distance function \( d(i,p) = \sqrt{(x_i - x_p)^2 + (y_i - y_p)^2} \). A linear geographical trend can be defined as \( \tau_i = \tau_x x_i + \tau_y y_i \) for constants \( \tau_x \) and \( \tau_y \). Such trends, sometimes with more sophisticated specifications, are widespread in geographical data and their study goes at least back to Krumbein (1959; 1963) and Tobler (1969).

Without going into the formalism, it is easily seen that this model is essentially equivalent to the model studied in Section 3, and hence that the same problems arise. Monte Carlo

\(^{20}\)An exception is time series data, but the current modeling of shocks does not seem particularly relevant to that case.
analyses reported in Appendix Table A-2 also show that the problem is indeed still present and if anything stronger than in the basic model.

4.3 Panel data models

In many applications including most of those mentioned in the introduction, we have access to a panel of observations. This allows for controlling for unit fixed effects, which would rule out the problem of the spatial trend $\tau_i$. Time trends are also unproblematic as they are routinely handled by year dummies. But if time trends depend on geography, that is we have spatio-temporal trends, the problem studied above reappears. Consider the case where

$$z_{it} = \alpha_i + \beta r_{it} + \tau_i t + \epsilon_{it}$$

with the trend $\tau_i = \tau_x x_i + \tau_y y_i$ for constants $\tau_x$ and $\tau_y$. If we assume a balanced panel so we can differentiate expression (5), we get

$$\Delta z_{it} = \beta \Delta r_{it} + \tau_i + \Delta \epsilon_{it}$$

which essentially is specification (3). De-meaning of course yields similar results. The only major difference is that we look at differenced shocks (or deviations from means). However, these have the exact same properties of spatial correlation as the undifferenced shock, so the issues studied in Section 3 still remain.

To see the effect of omitted spatio-temporal trends, Table 3 shows the results from some Monte Carlo simulations of model (5) for different panel lengths, sample sizes, and number of shocks. The conclusions are generally as above – the null hypothesis of no relationship which should have been rejected in 5% of cases is rejected far too often and t-values are typically high. Moreover, the problem is exacerbated by increasing sample sizes. There are some indications that increased panel lengths reduces the problem. As time periods are independent of each other, increasing $T$ increases the (random) variation in $\Delta r_t$ which helps uncover its independence to $\Delta z_{it}$.

4.4 General spatial dependence

The mechanism generating the spurious rejections of hypothesis tests above is that the precipitation shocks in sum are non-zero and with opposite sign in different corners of space. With a spatial trend, the two combines to form spurious correlation. The assumption of a deterministic spatial (or spatio-temporal) trend may be too strong in certain applications. However, the problem may persist in more general models of spatial dependence.
Table 3: A Monte Carlo analysis of the panel data model

<table>
<thead>
<tr>
<th>T</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>100</th>
<th>N</th>
<th>2N</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>100</td>
<td>N</td>
<td>2N</td>
</tr>
<tr>
<td>9</td>
<td>0.90</td>
<td>0.83</td>
<td>0.82</td>
<td>0.81</td>
<td>0.82</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>4.51</td>
<td>4.18</td>
<td>4.23</td>
<td>4.16</td>
<td>4.22</td>
<td>4.11</td>
<td>4.18</td>
</tr>
<tr>
<td>25</td>
<td>0.93</td>
<td>0.92</td>
<td>0.88</td>
<td>0.89</td>
<td>0.88</td>
<td>0.87</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>7.23</td>
<td>6.90</td>
<td>6.71</td>
<td>6.54</td>
<td>6.44</td>
<td>6.57</td>
<td>6.52</td>
</tr>
<tr>
<td>49</td>
<td>0.94</td>
<td>0.91</td>
<td>0.89</td>
<td>0.90</td>
<td>0.90</td>
<td>0.89</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>9.75</td>
<td>9.16</td>
<td>8.78</td>
<td>8.80</td>
<td>8.94</td>
<td>8.87</td>
<td>8.66</td>
</tr>
<tr>
<td>100</td>
<td>0.94</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>13.29</td>
<td>11.87</td>
<td>11.84</td>
<td>11.50</td>
<td>11.45</td>
<td>11.86</td>
<td>14.32</td>
</tr>
<tr>
<td>400</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>22.06</td>
<td>20.94</td>
<td>19.66</td>
<td>19.16</td>
<td>19.73</td>
<td>19.39</td>
<td>19.69</td>
</tr>
<tr>
<td>1024</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>31.18</td>
<td>28.84</td>
<td>28.25</td>
<td>28.79</td>
<td>28.70</td>
<td>27.38</td>
<td>27.07</td>
</tr>
<tr>
<td>10000</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>76.86</td>
<td>70.88</td>
<td>65.82</td>
<td>68.79</td>
<td>68.76</td>
<td>67.55</td>
<td>66.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T</th>
<th>10</th>
<th>20</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>100</th>
<th>N</th>
<th>2N</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>100</td>
<td>N</td>
<td>2N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.49</td>
<td>0.48</td>
<td>0.51</td>
<td>0.52</td>
<td>0.51</td>
<td>0.49</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.40</td>
<td>2.28</td>
<td>2.42</td>
<td>2.43</td>
<td>2.44</td>
<td>2.37</td>
<td>2.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.69</td>
<td>0.67</td>
<td>0.66</td>
<td>0.69</td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.87</td>
<td>3.89</td>
<td>3.91</td>
<td>3.66</td>
<td>3.77</td>
<td>3.75</td>
<td>3.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>0.76</td>
<td>0.73</td>
<td>0.76</td>
<td>0.74</td>
<td>0.73</td>
<td>0.78</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.09</td>
<td>5.01</td>
<td>5.13</td>
<td>5.09</td>
<td>5.04</td>
<td>5.41</td>
<td>5.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.82</td>
<td>0.82</td>
<td>0.85</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.09</td>
<td>6.78</td>
<td>7.70</td>
<td>7.30</td>
<td>7.16</td>
<td>7.43</td>
<td>7.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.92</td>
<td>0.90</td>
<td>0.90</td>
<td>0.92</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.32</td>
<td>12.66</td>
<td>13.23</td>
<td>13.55</td>
<td>13.30</td>
<td>13.23</td>
<td>12.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20.93</td>
<td>20.36</td>
<td>20.47</td>
<td>19.70</td>
<td>19.85</td>
<td>18.82</td>
<td>20.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.43</td>
<td>54.44</td>
<td>55.01</td>
<td>52.13</td>
<td>53.25</td>
<td>52.16</td>
<td>53.80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The Table shows Monte Carlo simulations of model (5) for different panel lengths $T$, sample sizes $N$, and number of shocks $K$. The table shows the fraction of cases where a $t$-test of $\beta = 0$ is rejected at the 5% level (first line) and the average of the absolute value of the associated $t$-value (second line). The true model is $\beta = 0$, $\tau = 1$, $\epsilon_i \sim N(0, 1)$, and for each $k$, $\nu_k \sim N(0, 1)$ and the position $p_k \sim U([0, N]^2)$. Each model is replicated 1000 times.
Consider a model in one-dimensional space of the form

\[ y_i = \alpha + \beta r_i + u_i \]
\[ u_i = \rho u_{i-1} + \epsilon_i, \]

i.e. a simple first order auto-regressive model.\(^{21}\) Unbiased test of the hypothesis \( \beta = 0 \) depends on \( E \sum_i r_i u_i = 0 \). With the current specification, we have \( u_i = \sum_{j=0}^{i-1} \rho^j \epsilon_i \). Say for simplicity that the location \( p \) is at an integer. Then

\[ \sum_i r_i u_i = \sum_{j=0}^{N-p} \frac{\nu}{1+j} u_{j+p} + \sum_{j=1}^{p-1} \frac{\nu}{1+j} u_{p-j}. \]

Using the autoregressive structure of the residuals, this can be rewritten

\[ \sum_i r_i u_i = \sum_{i=1}^{N} \epsilon_i \sum_{j=0}^{N-1} \frac{\nu}{1+|p - i + j|} \rho^j. \]

With high spatial dependence (i.e. \( \rho \) close to unity), the loading on a few of the innovations \( \epsilon_i \) close to \( p \) is going to be large, and these innovations determine the whole estimated \( \beta \). Even as \( N \) gets large, this effect persists.

For \( K > 1 \), we compare the loading on \( \epsilon_i \) around several clusters. Still, some of the clusters dominate and hence the bias remains the same if not worse. This is illustrated in Table 4, where I report results from a Monte Carlo analysis of the above model. We notice that when the spatial dependency is high (\( \rho \) close to unity), we get spurious relationships in the vast majority of cases. Increasing the sample size or the number of shocks does not seem to improve the situation.

5 Detecting and solving the problem

In the case of the weather, the problem of spurious correlations can usually be detected by examining the weather at counterfactual dates as in Figure 1. With other independent variables, other placebos may be feasible. If rejection rates differ markedly from the expected rates, some spatial or spatio-temporal dependency may be the explanation although of course other explanations obviously also exist. The next step should be to try to get some impression of the spatial dependency. One way to do this is to simply plot maps of spatial values or

\(^{21}\)To simplify, correlation is only to the left. The model would be essentially unchanged if \( u_i \) was correlated with both \( u_{i-1} \) and \( u_{i+1} \).
Table 4: A Monte Carlo analysis of a model with autoregressive spatial correlation

<table>
<thead>
<tr>
<th>ρ</th>
<th>.5</th>
<th></th>
<th>.75</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 5 10 100 N 2N</td>
<td></td>
<td>1 2 5 10 100 N 2N</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>10</td>
<td>0.19 0.19 0.17 0.18 0.17 0.19 0.18</td>
<td>0.25 0.26 0.24 0.22 0.25 0.24 0.22</td>
<td>1.2 1.2 1.1 1.2 1.2 1.2 1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5 1.4 1.4 1.3 1.4 1.4 1.3</td>
<td>1.8 1.7 1.7 1.7 1.7 1.7 1.6</td>
<td>1.3 1.3 1.2 1.3 1.2 1.3 1.3</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.24 0.22 0.21 0.22 0.22 0.22 0.22</td>
<td>0.36 0.35 0.35 0.36 0.36 0.35 0.31</td>
<td>1.3 1.3 1.2 1.3 1.2 1.3 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.8 1.7 1.7 1.7 1.7 1.7 1.6</td>
<td>1.8 1.8 1.8 1.8 1.8 1.8 1.8</td>
<td>1.8 1.8 1.8 1.8 1.8 1.8 1.8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.24 0.22 0.23 0.19 0.24 0.22 0.21</td>
<td>0.38 0.42 0.39 0.38 0.40 0.34 0.38</td>
<td>1.3 1.3 1.2 1.3 1.2 1.3 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.8 1.8 1.8 1.8 1.8 1.8 1.8</td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.25 0.20 0.23 0.23 0.26 0.19 0.26</td>
<td>0.44 0.39 0.39 0.39 0.39 0.39 0.42</td>
<td>1.3 1.3 1.2 1.3 1.2 1.3 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>0.24 0.22 0.25 0.22 0.22 0.24 0.22</td>
<td>0.41 0.41 0.41 0.40 0.40 0.40 0.39</td>
<td>1.3 1.3 1.2 1.3 1.2 1.3 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.9 1.9 1.9 1.9 1.9 1.9 1.9</td>
<td>1.9 1.9 1.9 1.9 1.9 1.9 1.9</td>
<td>1.9 1.9 1.9 1.9 1.9 1.9 1.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ρ</th>
<th>.95</th>
<th></th>
<th>1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 5 10 100 N 2N</td>
<td></td>
<td>1 2 5 10 100 N 2N</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>10</td>
<td>0.40 0.33 0.33 0.29 0.31 0.34 0.29</td>
<td>0.39 0.39 0.34 0.34 0.34 0.34 0.34</td>
<td>1.8 1.7 1.6 1.6 1.6 1.6 1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.9 1.8 1.8 1.7 1.6 1.6 1.6</td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
<td>1.9 1.8 1.8 1.8 1.8 1.8 1.8</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.57 0.50 0.49 0.51 0.47 0.49 0.49</td>
<td>0.59 0.55 0.52 0.55 0.54 0.52 0.53</td>
<td>2.6 2.4 2.3 2.4 2.3 2.3 2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.8 2.7 2.5 2.6 2.6 2.6 2.6</td>
<td>2.8 2.7 2.5 2.6 2.6 2.6 2.6</td>
<td>2.8 2.7 2.5 2.6 2.6 2.6 2.6</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.56 0.58 0.56 0.56 0.53 0.53 0.52</td>
<td>0.68 0.64 0.60 0.59 0.59 0.61 0.58</td>
<td>2.7 2.8 2.7 2.8 2.6 2.5 2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.4 3.2 3.1 3.1 3.1 3.1 3</td>
<td>3.4 3.2 3.1 3.1 3.1 3.1 3</td>
<td>3.4 3.2 3.1 3.1 3.1 3.1 3</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.60 0.61 0.63 0.61 0.64 0.63 0.59</td>
<td>0.81 0.79 0.77 0.75 0.75 0.74 0.75</td>
<td>3.1 3.3 3.2 3.1 3.3 3.2 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.5 5.1 5.1 5 5 5 5</td>
<td>5.5 5.1 5.1 5 5 5 5</td>
<td>5.5 5.1 5.1 5 5 5 5</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>0.65 0.64 0.64 0.65 0.64 0.61 0.63</td>
<td>0.86 0.86 0.80 0.85 0.82 0.83 0.83</td>
<td>3.2 3.3 3.3 3.4 3.4 3.2 3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.8 7.5 7.3 7.1 7.3 7.2 7.4</td>
<td>7.8 7.5 7.3 7.1 7.3 7.2 7.4</td>
<td>7.8 7.5 7.3 7.1 7.3 7.2 7.4</td>
</tr>
</tbody>
</table>

Notes: The Table shows Monte Carlo simulations of model (6) for different autoregressive coefficients ρ, sample sizes N, and number of shocks K. The table shows the fraction of cases where a t-test of β = 0 is rejected at the 5 % level (first line) and the average of the absolute value of the associated t-value (second line). The true model is β = 0, τ = 1, ε_i ~ N(0,1), and for each k, ν_k ~ N(0,1) and the position p_k ~ U([0,N]^2). Each model is replicated 1000 times.
estimated spatial trends. In some cases it may also be useful to use testing procedures such as Moran’s I statistic.

If a spatial pattern is found, two possible solutions can be pursued. The ideal solution is to find the source of the dependency and expand the model specification to take this into account. If, for instance, geographically varying trends are due to geographical differences in demographic patterns (say young people moving toward large cities), one could potentially solve the problem by adding demographic controls. However, it may not always be easy to find a simple explanation and there may not be a single explanation for the geographical trend. In such cases, it may be a better option to attempt to control for the geo-spatial trend. In the time series literature, this is usually done by simply including the date as a variable, sometimes with a few polynomial terms. In the case of geographical data, this may be too limiting.

In the cross sectional case, we want to control for some unknown function $T(x, y)$. As the shape of $T$ is unknown, a flexible estimator in two-dimensional space is called for. Kernel based and other standard non-parametric estimators are computationally intensive, and as their rate of convergence is typically below $\sqrt{n}$, inference of the other variables in the regression can’t always be made using standard techniques. Consequently, a simpler form may be advisable.

In the case of a panel, we need to estimate a function $T(x, y, t)$. As this is a function of three variables, a fully flexible non-parametric approach gets even more demanding. At least for short panels, it seems reasonable that the trend may be kept linear, so we can rewrite $T(x, y, t) = U(x, y)t$ for some function $U$. One solution that seems to work well for the electoral turnout data considered below is one where $U$ is specified as a tensor product of Legendre polynomials.\(^{22}\) The choice of orthogonal polynomials is to reduce problems of multicollinearity and improve numerical stability. One justification for choosing Legendre polynomials is their orthogonality property with regard to an $L^2$ inner product given a uniform spatial distribution of units. Although the distribution is not exactly uniform, this approach is likely to give better behavior than most other orthogonal polynomial bases that provide orthogonality given various bell shaped distributions. Still, it seems that the choice of polynomial base has little effect on the final outcomes.

Given dimensionalities $K$ and $L$, we can specify

$$T(x, y, t) = t \sum_{k=0}^{K} \sum_{\ell=0}^{L} \theta_{k\ell} P_k(x) P_\ell(y)$$  \(^{(7)}\)

---

\(^{22}\)See e.g. Judd (1998, Ch. 6) for an overview of Legendre polynomials and other polynomial basis with applications in economics and Totik (2005) for the mathematical background.
where $P_i(\cdot)$ is the $i$'th order Legendre polynomial. The $(K + 1)(L + 1)$ parameters $\theta_{k\ell}$ can be estimated together with the other parameters in an ordinary regression model.

The choice of the dimensions $K$ and $L$ has to be chosen to make the polynomial (7) provide a reasonable fit of the data. If $K$ and $L$ are chosen too high, there is both a danger of over fitting (Hastie et al., 2008, Ch. 7) and loosing so much variation that it becomes impossible to identify the effect of the variable of interest. Hence we want would like a good fit with a low dimensional polynomial. To make a good trade off, I recommend to consider choosing $K$ and $L$ by maximizing a linear penalty function

$$R^2 - \xi(K + 1)(L + 1)$$

where $R^2$ is the fit of the model and $\xi$ a penalty on the number of parameters to estimate. This is closely related to maximizing the AIC and BIC criteria, but varying the degrees of freedom penalty. Varying the parameter $\xi$, we can trace out the class of potentially good polynomial compositions. It is also important to undertake counter factual estimations as in Figure 1 to check that the polynomial at hand actually solves the problem. If the fit is good enough, most of the placebo variables should have little effect on the outcome. Another approach could also be to choose $K$ and $L$ high, but constrain the $\theta_{k\ell}$ by employing ridge regression, LASSO, or other versions of constrained estimation (Belloni et al., 2014; Hastie et al., 2008, Ch. 3).

6 Turnout in Norwegian elections

6.1 Controlling for the spatio-temporal trend

As argued in Section 5, one way to handle the problem of spation-temporal trends is to control them out in the estimation. I approximate the trend with the tensor product of Legendre polynomials. The first step needed is to make a choice of how many polynomial terms to include in each of the two dimensions. Figure 5 shows the the model’s fit (net of a baseline model without spatio-temporal controls) for each combination of between 0 and 10 terms in each dimensions. Combinations of polynomial orders $K$ and $L$ that are maxima of the penalized model (8) for some value of $\xi$, i.e. those which are elements of the convex hull of the points, are shown in red. There is a strong increase in fit going up to about 15 terms, then the effect of additional terms seems to flatten out. To avoid over fitting the data and preserve some degrees of freedom, my preferred model specifies spatio-temporal trends using

\begin{align*}
\text{These polynomials are usually defined recursively with } P_0(x) &= 1, \\
P_1(x) &= x, \\
\text{and for } i \geq 2, P_i(x) &= [(2i - 1)xP_{i-1}(x) - (i - 1)P_{i-1}(x)]/i \\
\text{where the variable } x \text{ is normalized to be in the interval } [-1, 1].
\end{align*}
Figure 5: The number of terms in the nonparametric trend model

(a) Model fit and number of longitude and latitude terms
(b) Model fit and total number of terms

Notes: Panel (a) shows model fit as a function of the number of terms in the longitudinal and latitudinal polynomials, whereas Panel (b) shows fit as a function of the total number of terms included in the tensor product. Approximation is with tensor products of Legendre polynomials of varying degrees. In Panel (b), combinations that belong to the convex hull are shown with solid orange dots and other combinations with hollow green dots.

Adding more terms not only have a minor impact on the model’s fit, it turns out that the exact specification of the spatio-temporal has little importance once we reach a minimum level of complexity. Figure 6 shows the distribution of t-values for eight specification with increasing complexity of the tensor product of Legendre polynomials and regional dummies and with linear and quadratic time trends. Adding more terms not only have a minor impact on the model’s fit, it turns out that the exact specification of the spatio-temporal has little importance once we reach a minimum level of complexity. Figure 6 shows the distribution of t-values for eight specification with increasing complexity of the tensor product of Legendre polynomials and regional dummies and with linear and quadratic time trends. The distributions are almost perfectly overlapping for each of the eight models. Indeed, the correlation between the most and the least complex models are between .85 and .9.

Also, it seems that using region specific trends has a comparable effect in improving estimation results to spatio-temporal trends. As argued above there are many cases where it is implausible that the trend has a spatial discontinuity at regional borders. Still, it seems that this possible mis-specification has little impact in practice.

Moreover, we notice that the distribution of t-values is much more well behaved than the extreme values found in Figure 2. Although the distribution is somewhat fatter than the theoretical Student’s t distribution, the distribution is much more sensible to work with.

Table 5 shows the actual estimation results regressing turnout on election day weather.

\[ \text{Table 5: Actual estimation results regressing turnout on election day weather.} \]

\[ \text{The distribution of the estimated coefficients can be found in Appendix Figure A-4.} \]
Notes: The graph shows the distribution of the t-values when regressing municipal turnout on daily precipitation for 600 days before and after election day. The 10 days before and after the actual election day are omitted. Panel (a) shows results from regressing levels on levels. Panel (b) shows the regression of turnout on a dummy for more than 25 mm rain while Panel (c) employs a dummy for any rain. Panel (d) shows results from a regression where the rank of turnout is measured on the rank of rain, i.e. both variables are uniform on the unit interval.
Spatio-temporal trends are controlled for using tensor products of Legendre polynomials with $1 \times 6$, $3 \times 10$, and $7 \times 8$ terms as well as regional trends. Linear temporal trends are shown in solid lines and quadratic linear trends in dashed lines.
For reference, Panel A shows the estimation results using a two-way fixed effects specification without controlling for spatio-temporal trends. We notice that estimation results are very sensitive to the exact specification of the independent variable, and most results are insignificant. As argued above, this specification is probably not trustworthy.

Panel B of 5 shows the estimation results from the preferred specifications. The general pattern is that rain seems to increase turnout in Norway – see Lind (2014) for a discussion of the rationale behind this. Column (1) shows the plain regression of turnout on precipitation in cm. The effect of 1 cm increase in precipitation is about .3 percentage point increase in turnout. Columns (2) and (3) turns the attention to dummies for positive rain and substantial rain, defined as above 2.5 mm. Comparing elections with and without rain, turnout is about .5 to .7 percentage points higher in the former. Columns (5) and (6) tests for the presence of a change in the parameters estimates over time. The effects seem to be fairly stable. Finally, Columns (7) and (8) test for non-linearities in the relationship. There is a weak tendency for extreme amounts of precipitation to reduce turnout, but the overall pattern is still close to linearity.

7 Conclusion

In this paper, I have shown that when outcomes of interest are regressed on weather data, there is a danger of spuriously detecting relationships. To illustrate the occurrence of the problem, I have shown nonsensical relationships such as a relationship between electoral turnout and rainfall 100 days before the election. In such cases, the relationship can be rejected by common sense. But for more relevant questions, such as whether rain on the day of the election affect turnout, the problems of spatial correlation remain the same. To give a satisfactory answer in the potentially interesting cases, we need a proper understanding of the phenomena generating the spurious relationships.

The reason for these relationships, I argue, is that spatial patterns in weather conditions are likely to align up with spatial or spatio-temporal patterns in the outcomes of interest. I have shown that this does indeed occur in a simple model of spatially dependent data as well as in an extensive range of Monte Carlo analyses. Moreover, the analyses reveal that standard techniques, such as clustering on spatial entities (Moulton, 1986) or using Conley’s (1999) approach to computing standard errors does solve the problem.

Rather, I suggest introducing controls for spatial or spatio-temporal trends in regressions to solve the problem. This is a simple remedy that can easily be combined with other techniques, such as instrumental variables of regression discontinuity designs. In a sample of Norwegian municipal elections, I show that this reduces the problem of spurious relationships.
Table 5: The effect of precipitation on turnout

**Panel A: Estimation without spatio-temporal trends**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain (in cm)</td>
<td>-0.000339</td>
<td>-0.000436</td>
<td>0.00134**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.50)</td>
<td>(-0.51)</td>
<td>2.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rain positive</td>
<td></td>
<td></td>
<td></td>
<td>0.00546***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.55)</td>
<td></td>
</tr>
<tr>
<td>Rain above 2.5 mm</td>
<td></td>
<td>-0.000143</td>
<td>0.000396</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.12)</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rain × Year</td>
<td></td>
<td></td>
<td>0.00185***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(6.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean dep. var</td>
<td>0.681</td>
<td>0.681</td>
<td>0.681</td>
<td>0.681</td>
<td>0.681</td>
</tr>
<tr>
<td>Obs</td>
<td>4417</td>
<td>4417</td>
<td>4417</td>
<td>4417</td>
<td>4417</td>
</tr>
<tr>
<td>R²</td>
<td>0.612</td>
<td>0.613</td>
<td>0.612</td>
<td>0.612</td>
<td>0.624</td>
</tr>
</tbody>
</table>

**Panel B: Estimation with spatio-temporal trends**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain (in cm)</td>
<td>0.00299***</td>
<td></td>
<td>0.00244***</td>
<td>0.00283***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.34)</td>
<td></td>
<td>(3.58)</td>
<td>(5.29)</td>
<td></td>
</tr>
<tr>
<td>Rain positive</td>
<td></td>
<td>0.00742***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rain above 2.5 mm</td>
<td></td>
<td></td>
<td>0.00511***</td>
<td>0.00215*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.94)</td>
<td>(1.69)</td>
<td></td>
</tr>
<tr>
<td>Rain × Year</td>
<td></td>
<td></td>
<td></td>
<td>-0.000356</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-1.56)</td>
<td></td>
</tr>
<tr>
<td>Mean dep. var</td>
<td>0.681</td>
<td>0.681</td>
<td>0.681</td>
<td>0.681</td>
<td>0.681</td>
</tr>
<tr>
<td>Obs</td>
<td>4417</td>
<td>4417</td>
<td>4417</td>
<td>4417</td>
<td>4417</td>
</tr>
<tr>
<td>R²</td>
<td>0.698</td>
<td>0.697</td>
<td>0.697</td>
<td>0.698</td>
<td>0.698</td>
</tr>
</tbody>
</table>

*Notes: Outcome variable is municipal electoral turnout. All specifications include municipal and year fixed effects. All specifications in Panel B also include the tensor product of Legendre polynomials with $1 \times 6$ terms to control for spatio-temporal trends. Standard errors are clustered at the municipality level (using the 2010 municipal structure). t-values in parentheses, and *, **, and *** denotes significant at the 10%, 5%, and 1% levels.*
in statistical tests to close to the theoretical properties.

The question of more sophisticated approaches to controlling for spatial and spatio-temporal trends, possibly borrowing from the literature on spatial statistics and econometrics is left for future research. There are probably possibilities to do better, but it is unclear that such approaches are sufficiently simple to implement that they actually matter for the applied researcher.

In studies of the effect of short term weather changes, as studied in this paper, weather data are typically available for a large number of periods, of which only a few matter. Then there is ample supply of placebo data. Such data should regularly be used to test the validity of empirical approaches used. In studies of the effect of long term weather effects, such as the effect on agricultural production, surplus data are harder to find. Still it may be possible to run placebo studies by temporally moving the whole or parts of the rainfall pattern.

When placebo data can be constructed, one may also ask whether these they could be used to construct a more correct null distribution of the parameter of interest, somewhat along the lines of bootstrapping techniques. Saunders (1993) implements a version of this estimator, but does not go into its statistical properties and potential advantages compared to ordinary inference.
References


A Proofs

A.1 Proof of divergence of the numerator in (4)

Proof. Let $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceil and floor operators, and define the relative position of the location of a shock as $\zeta = \frac{p_i}{N}$. Finally define $\lambda = \lceil \zeta N \rceil - \zeta N$ (so $1 - \lambda = \zeta N - \lceil \zeta N \rceil$). To study the behavior of $\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \left( \frac{1}{1 + |p_k - i|} - \bar{w} \right) i$, we need the behavior of $\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{1}{1 + |p_k - i|}$ and $\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{i}{1 + |p_k - i|}$.

1) The behavior of $\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{1}{1 + |p_k - i|}$:

For simplicity of notation, we disregard the subscript $k$. We split the absolute value in the denominator into the terms with $i$ below and above $p_k = \zeta N$. This yields

$$\sum_{i=1}^{N} \frac{1}{|\zeta N - i| + 1} = \frac{1}{[\zeta N] - \zeta N + 1} + \frac{1}{[\zeta N] - \zeta N + 2} + \ldots + \frac{1}{[\zeta N] - \zeta N + (N - [\zeta N])}$$

$$+ \frac{1}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{1}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{\zeta N - \lceil \zeta N \rceil + \lceil \zeta N \rceil}$$

$$= \frac{1}{1 + \lambda} + \frac{1}{2 + \lambda} + \ldots + \frac{1}{(N - \lceil \zeta N \rceil) + \lambda}$$

$$+ \frac{1}{1 + (1 - \lambda)} + \frac{1}{2 + (1 - \lambda)} + \ldots + \frac{1}{\lceil \zeta N \rceil + (1 - \lambda)}$$

Define

$$S_{N-\lceil \zeta N \rceil}^1 = \frac{1}{1 + \lambda} + \frac{1}{2 + \lambda} + \ldots + \frac{1}{(N - \lceil \zeta N \rceil) + \lambda}$$

and

$$S_{\lceil \zeta N \rceil}^2 = \frac{1}{1 + (1 - \lambda)} + \frac{1}{2 + (1 - \lambda)} + \ldots + \frac{1}{\lceil \zeta N \rceil + (1 - \lambda)}$$

The series $S_n^1$ is a generalized harmonic series of length $n$. If we define $S_n^0 = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$ as the standard harmonic series of length $n$, we see that $S_n^0 = S_{N-\lceil \zeta N \rceil}^1 - 1 < S_{N-\lceil \zeta N \rceil}^1 < S_{N-\lceil \zeta N \rceil}^1$. For large $n$ we know that $S_n^0 \to \gamma + \ln n$ where $\gamma$ is the Euler–Mascheroni constant ($\gamma \approx .577$). Hence $\gamma + \ln \frac{N-\lceil \zeta N \rceil + 1}{e} < S_{N-\lceil \zeta N \rceil}^1 < \gamma + \ln (N - \lceil \zeta N \rceil)$. From a similar reasoning, $\gamma + \ln \frac{\lceil \zeta N \rceil + 1}{e} < S_{\lceil \zeta N \rceil}^2 < \gamma + \ln (\lceil \zeta N \rceil)$. It follows that

$$2\gamma + \ln \frac{N - \lceil \zeta N \rceil + 1}{e} + \ln \frac{\lceil \zeta N \rceil + 1}{e} < \sum_{i=1}^{\lceil \zeta N \rceil - 1} \frac{1}{|\zeta N - i| + 1} < 2\gamma + \ln (N - \lceil \zeta N \rceil) + \ln (\lceil \zeta N \rceil)$$

That is, for any $x \in \mathbb{R}^+$, $\lfloor x \rfloor = \min \{ y \in \mathbb{N} : x \leq y \}$ and $\lceil x \rceil = \max \{ y \in \mathbb{N} : x \geq y \}$.
It follows that \( \frac{1}{N} \sum_{|p|>|i|} \frac{1}{|\zeta N - i| + 1} \to 0 \) as \( N \to +\infty \). As this hold for any \( p_k \), it also holds for the sum so \( \bar{w} \to 0 \) as \( N \to +\infty \). Moreover, it also holds for the sum weighted by \( \nu_k \) so \( \bar{r} \to 0 \) as \( N \to +\infty \).

2) The behavior of \( \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{i}{1+|p_k-i|} \):

We proceed by splitting in the same manner, yielding

\[
\sum \frac{i}{|p_i| + 1} = \frac{\lceil \zeta N \rceil}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{\lceil \zeta N \rceil + 1}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{N}{\zeta N - \lceil \zeta N \rceil + [\zeta N] - \lceil \zeta N \rceil + 1} + \frac{\lceil \zeta N \rceil}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{\lceil \zeta N \rceil - 1}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{\zeta N - \lceil \zeta N \rceil + [\zeta N]}
\]

We have

\[
\frac{\lceil \zeta N \rceil + 1}{\lceil \zeta N \rceil - \zeta N + 1} + \frac{\lceil \zeta N \rceil + 1}{\lceil \zeta N \rceil - \zeta N + 2} + \ldots + \frac{N}{\zeta N - \lceil \zeta N \rceil + (N - [\zeta N])}
= ([\zeta N] - 1) \left( \frac{1}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{1}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{N - \lceil \zeta N \rceil} \right)
\]

\[
+ \frac{1}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{2}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{\zeta N - \lceil \zeta N \rceil + [\zeta N]}
\]

\[
= ([\zeta N] - 1) S_N^1 + T_{N-[\zeta N]}^1
\]

where the serie

\[
T_n^1 = \frac{1}{1+\lambda} + \frac{2}{2+\lambda} + \ldots + \frac{n}{n+\lambda}
\]

We know that

\[
\frac{n}{1+\lambda} < T_n^1 < n
\]

Similarly,

\[
\frac{\lceil \zeta N \rceil}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{\lceil \zeta N \rceil - 1}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{\zeta N - \lceil \zeta N \rceil + [\zeta N]}
= ([\zeta N] + 1) \left( \frac{1}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{1}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{\zeta N - \lceil \zeta N \rceil + [\zeta N]} \right)
\]

\[
- \left( \frac{1}{\zeta N - \lceil \zeta N \rceil + 1} + \frac{2}{\zeta N - \lceil \zeta N \rceil + 2} + \ldots + \frac{1}{\zeta N - \lceil \zeta N \rceil + [\zeta N]} \right)
\]

\[
= ([\zeta N] + 1) S_N^2 - T_{[\zeta N]}^2
\]

where the series

\[
T_n^2 = \frac{1}{1+(1-\lambda)} + \frac{2}{2+(1-\lambda)} + \ldots + \frac{n}{n+(1-\lambda)}
\]
We know that
\[ \frac{n}{2 - \lambda} < T_n^2 < n \]

It follows that
\[ \frac{1}{N} \sum \frac{i}{|p_i| + 1} = \frac{([\zeta N] - 1) S_{N-1}^1 + ([\zeta N] + 1) S_{\zeta N}^2 + T_{N-[\zeta N]}^1 - T_{\zeta N}^2}{N} \]

Hence
\[ \frac{1}{N} \sum \left( \frac{1}{|p_i| + 1} - \bar{W} \right) i = \left( \frac{[\zeta N] - 1}{N} - \frac{N + 1}{2N} \right) S_{N-1}^1 + \left( \frac{[\zeta N] + 1}{N} - \frac{N + 1}{2N} \right) S_{\zeta N}^2 + \frac{T_{N-[\zeta N]}^1 - T_{\zeta N}^2}{N} \]

When \( N \to +\infty \), we see that the two first parentheses converge to \( \zeta - \frac{1}{2} \) and the last fraction to a constant \( \Xi \in \left( -\frac{\lambda}{1+\lambda}, \frac{1-\lambda}{2-\lambda} \right) \). Hence the expression converges to a log function, so \( w_k = \frac{1}{N} \sum \left( \frac{1}{|p_i| + 1} - \bar{W} \right) i \to +\infty \) as \( N \to +\infty \).

3) Behavior of the numerator:

Then the full numerator can be written as \( \sum_{k=1}^{K} w_k \nu_k \). Define the set \( K_+ = \{ k : \nu_k > 0 \} \) and \( K_- = \{ k : \nu_k < 0 \} \). Now as \( N \to +\infty \), the numerator converges to
\[ \frac{1}{N} \sum_{k \in K_+} \nu_k w_k + \frac{1}{N} \sum_{k \in K_-} \nu_k w_k \]

At least of the sums diverge. If only one diverge, the numerator converges to either \( +\infty \) (only the first diverge) or \( -\infty \) (only the second diverge). Furthermore, if both diverge we have \( \Pr \left( \left| \sum_{k=1}^{K} w_k \nu_k \right| \to +\infty \right) = 1 \). To see this, notice that for a given realization of \( \{\nu_2, \ldots, \nu_K\} \) (and the \( p_k \)s if they are taken as random), there is only one value of \( \nu_1 \) that assures convergence. The probability of this realization is zero as the \( \nu_k \)s have continuous distributions. Hence with probability 1 the numerator of (4) diverges.
A.2 Proof of convergence of the denominator in (4)

Proof. We want to study the behavior of \( \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \sum_{k=1}^{K} \frac{1}{|p_{k-i+1}|} - \bar{w} \right)^{2}} =: \sqrt{\frac{1}{N^{2}} A_{N}} \). We have \( A_{N} = \sum_{i=1}^{N} \left( \sum_{k=1}^{K} \frac{1}{|p_{k-i+1}|} \right)^{2} - N \bar{w}^{2} \). We know from the proof in Appendix A.1 that \( \bar{w} \) converges to a log function so \( \frac{\bar{w}^{2}}{N} \to 0 \) as \( N \to \infty \).

We first study the behavior of \( \sum_{i=1}^{N} \left( \frac{1}{|p_{k-i+1}|} \right)^{2} \) for any choice of \( k \). We want to show that as \( N \to \infty \), it converges to a finite positive value \( Q_{k\infty} \). As for the proof in A.1, define \( \lambda = \lceil \zeta N \rceil - \zeta N \) (so \( 1 - \lambda = \zeta N - \lfloor \zeta N \rfloor \)). Then we have

\[
\sum \left( \frac{1}{\zeta N - i + 1} \right)^{2} = \left( \frac{1}{\lceil \zeta N \rceil - \zeta N + 1} \right)^{2} + \left( \frac{1}{\lceil \zeta N \rceil - \zeta N + 2} \right)^{2} + \ldots \\
+ \left( \frac{1}{\zeta N - \lceil \zeta N \rceil + 1} \right)^{2} + \left( \frac{1}{\zeta N - \lceil \zeta N \rceil + 2} \right)^{2} + \ldots \\
+ \left( \frac{1}{\zeta N - \lceil \zeta N \rceil + \lfloor \zeta N \rfloor} \right)^{2}.
\]

Define the series

\[
Q_{N-\lceil \zeta N \rceil}^{1} = \left( \frac{1}{1 + \lambda} \right)^{2} + \left( \frac{1}{2 + \lambda} \right)^{2} + \ldots + \left( \frac{1}{(N - \lceil \zeta N \rceil) + \lambda} \right)^{2}
\]

and

\[
Q_{\lceil \zeta N \rceil}^{2} = \left( \frac{1}{1 + (1 - \lambda)} \right)^{2} + \left( \frac{1}{2 + (1 - \lambda)} \right)^{2} + \ldots + \left( \frac{1}{\lceil \zeta N \rceil + (1 - \lambda)} \right)^{2}
\]

and define the sum of the the reciprocals of the squares of natural numbers \( Q_{n}^{0} = \sum_{i=1}^{n} \left( \frac{1}{i} \right)^{2} \). Then we see that \( Q_{1+N-\lceil \zeta N \rceil}^{0} - 1 \leq Q_{N-\lceil \zeta N \rceil}^{1} \leq Q_{N-\lceil \zeta N \rceil}^{0} \) and \( Q_{1+\lceil \zeta N \rceil}^{0} - 1 \leq Q_{\lceil \zeta N \rceil}^{2} \leq Q_{\lceil \zeta N \rceil}^{0} \). Hence for given \( \zeta \) we get \( \lim_{N \to \infty} \sum \left( \frac{1}{\lfloor \zeta N - i \rfloor + 1} \right)^{2} = Q_{k\infty} \). Moreover, as \( \lim_{n \to \infty} Q_{n}^{0} = \pi^{2}/6 \), we have \( \pi^{2}/3 - 2 \leq Q_{k\infty} \leq \pi^{2}/3 \).

Consider next the cross terms \( \sum_{i=1}^{N} \frac{1}{|p_{k-i+1}|} \frac{1}{|p_{k'-i+1}|} =: B_{kk'N} \) for any choices of \( k \neq k' \). As \( \frac{1}{|p_{k-i+1}|} > 0 \) for any \( p_{k} \in [1, N] \), we have \( B_{kk'N} > 0 \). Moreover, it follows from the
Cauchy-Schwarz inequality that $B_{kk'N} \leq \sqrt{Q_{k\infty}Q_{k'\infty}} \leq \frac{\pi^2}{3}$.

Hence as $N \to \infty$, we have $A_N \to \sum_{k=1}^{K} \sum_{k'=1}^{K} \sqrt{Q_{k\infty}Q_{k'\infty}}$, so $0 < A_n \leq K^2 \frac{\pi^2}{6}$. Hence if $\frac{K}{N} \to \kappa \in \mathbb{R}^+$, the denominator in (4) converges to a non-zero constant whereas if $\frac{K}{N} \to 0$, the denominator in (4) converges to 0. \qed
B The data

The meteorological data used in the paper are created by the Norwegian Meteorological Institute (met.no). The data are based on daily observations of precipitation at all 421 measurement stations in Norway, and based on spatial interpolation using a residual kriging approach Tveito and Førland (1999). First, each observation is regressed on a number of geographic properties to separate between a deterministic and a stochastic part. The residuals are then interpolated using kriging and combined with deterministic parts to obtain a grid of $1 \times 1$ km cells for Norway. As one would expect, average rainfall is larger along the west coast and in parts of the north.

To get municipal averages, I combine the data with GIS data on municipal boundaries to construct data on average precipitation by municipality for each election year. Municipal boundaries have changed over time, and GIS data on past municipal borders are essentially non-existent. To solve this I map municipalities that no longer exist into their current municipality and use weather data from the present day municipality.

Average precipitation values on election days are shown in Panel (a) of Appendix Figure A-1. Panel (b) of Appendix Figure A-1 show the average election day precipitation and turnout for the period 1971-2007. There are no clear geographical trends in average turnout.

Data on electoral turnout taken from the recent collection of Norwegian municipal data made available by Fiva et al. (2012), originating from Statistics Norway and the Norwegian Social Science Data Services. Data from the ten municipal elections between 1971 and 2007 were used. The data on median gross household incomes was taken from Statistics Norway’s StatBank. The sample covers all municipal medians for the years 1993-2012.
Notes: The graph shows average precipitation on election day, averaged over the elections 1971-2007. Dark colors indicate high levels of precipitation.
C Additional Monte Carlo results

Table A-1: A Monte Carlo analysis of the linear model with constant size of space

<table>
<thead>
<tr>
<th>K</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>100</th>
<th>N</th>
<th>2N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>0.15</td>
<td>0.13</td>
<td>0.15</td>
<td>0.16</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.37</td>
<td>0.37</td>
<td>0.35</td>
<td>0.33</td>
<td>0.34</td>
<td>0.36</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.57</td>
<td>0.57</td>
<td>0.55</td>
<td>0.55</td>
<td>0.53</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.90</td>
<td>0.89</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21</td>
<td>22</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Notes: The table shows the fraction of cases where a t-test of $\beta = 0$ is rejected at the 5% level (first line) and the average of the absolute value of the associated t-value (second line). The true model is $y_i = \alpha + \beta r_i + \tau \frac{i}{N} + \epsilon_i$, with $\beta = 0$, $\tau = 1$, $\epsilon_i \sim N(0,1)$, and for each $k$, $\nu_k \sim N(0,1)$ and the position $p_k \sim U(0,1)$. Each model is replicated 1000 times.
Table A-2: A Monte Carlo analysis of the spatial model

<table>
<thead>
<tr>
<th>K</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>100</th>
<th>N</th>
<th>2N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>0.37</td>
<td>0.36</td>
<td>0.31</td>
<td>0.32</td>
<td>0.30</td>
<td>0.29</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.76</td>
<td>1.71</td>
<td>1.59</td>
<td>1.59</td>
<td>1.56</td>
<td>1.53</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.57</td>
<td>0.58</td>
<td>0.55</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.84</td>
<td>2.78</td>
<td>2.57</td>
<td>2.52</td>
<td>2.44</td>
<td>2.49</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
<td>0.65</td>
<td>0.63</td>
<td>0.66</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.8</td>
<td>3.84</td>
<td>3.59</td>
<td>3.47</td>
<td>3.28</td>
<td>3.41</td>
<td>3.35</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.73</td>
<td>0.76</td>
<td>0.73</td>
<td>0.74</td>
<td>0.72</td>
<td>0.71</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.21</td>
<td>5.12</td>
<td>4.79</td>
<td>4.87</td>
<td>4.61</td>
<td>4.6</td>
<td>4.64</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.85</td>
<td>0.85</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.15</td>
<td>8.98</td>
<td>8.5</td>
<td>8.6</td>
<td>8.12</td>
<td>8.56</td>
<td>7.79</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>0.88</td>
<td>0.90</td>
<td>0.90</td>
<td>0.91</td>
<td>0.89</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.1</td>
<td>12.9</td>
<td>12.7</td>
<td>12.7</td>
<td>12.3</td>
<td>11.7</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>31.5</td>
<td>31.6</td>
<td>31.5</td>
<td>30.3</td>
<td>29.8</td>
<td>31</td>
<td>31.4</td>
</tr>
</tbody>
</table>

Notes: The table shows the fraction of cases where a t-test of $\beta = 0$ is rejected at the 5% level (first line) and the average of the absolute value of the associated t-value (second line). The true model is $z_i = \alpha + \beta \sum r_i + \tau (x_i + y_i) + \epsilon_i$ with $\beta = 0$, $\tau = 1$, $\epsilon_i \sim N(0,1)$, and for each $k$, $\nu_k \sim N(0,1)$ and the position $p_k \sim U([0, N] \times [0, N])$. Each model is replicated 1000 times.
D Additional graphs

Figure A-2: Distribution of simulated t-values

Notes: The graphs show the distribution of t-values obtained from the Monte Carlo analyses of spatial trends on actual rainfall. Panel A shows the distribution of the t-values without clustering in the cross sectional setting (solid green lines) and with a panel (dashed orange). Panel B shows the values in the cross section clustering at the regional level (solid green) and using Conley standard errors (dashed orange). The cut offs for the Conley standard errors were chosen as the standard deviation of the coordinates. Panel C shows the distribution of the placebo with randomly chosen municipalities (solid green) and the standard normal distribution (dotted orange).
Figure A-3: Association between the t-values in the different specifications

Notes: The graph shows the association between the t-values when regressing municipal turnout on daily precipitation for 600 days before and after election day using four different specifications. The 10 days before and after the actual election day are omitted.
Figure A-4: Distribution of estimated coefficients controlling for spatio-temporal trends

Notes: The graph shows the distribution of the estimated coefficients when regressing municipal turnout on daily precipitation for 600 days before and after election day. The 10 days before and after the actual election day are omitted. Panel (a) shows results from regressing levels on levels. Panel (b) shows the regression of turnout on a dummy for more than 25 mm rain while Panel (c) employs a dummy for any rain. Panel (d) shows results from a regression where the rank of turnout is measured on the rank of rain, i.e. both variables are uniform on the unit interval.

Spatio-temporal trends are controlled for using tensor products of Legendre polynomials with $1 \times 6$, $3 \times 10$, and $7 \times 8$ terms as well as regional trends. Linear temporal trends are shown in solid lines and quadratic linear trends in dashed lines.