Spatial Autocorrelation and Interactions between Surface Temperature Trends and Socioeconomic Changes

Ross McKitrick
Department of Economics
University of Guelph

December, 2007

Spatial Autocorrelation and Interactions between Surface Temperature Trends and Socioeconomic Changes

ABSTRACT

McKitrick and Michaels (2007) tested for independence between the spatial pattern of trends in surface climate data and the spatial pattern of socioeconomic indicators that serve as proxies for measurement inhomogeneities and anthropogenic surface processes. They found the relationship to be statistically significant, and in counterfactual simulation concluded that the extraneous signals explain about half the post-1980 warming trend in surface data. This paper examines the robustness of these conclusions to treatment for possible spatial autocorrelation in the model residuals. Under a variety of weighting schemes, a robust LM test for no spatial autocorrelation is not rejected. Applying a correction for spatial autocorrelation anyway does not change the original conclusions.

Spatial Autocorrelation and Interactions between Surface Temperature Trends and Socioeconomic Changes

Introduction

[McKitrick and Michaels (2007), MM07] observed that if adjusted surface climate data are free of biases due to inhomogeneities and anthropogenic surface processes, then the spatial pattern of gridded temperature trends should not be significantly correlated with socioeconomic variables that determine the evolution of these extraneous factors. They estimated

$$\theta_i = \beta_0 + \beta_1 TROP_i + \beta_2 PRESS_i + \beta_3 DRY_i + \beta_4 DSLP_i + \beta_5 WATER_i + \beta_6 ABSLAT_i$$
$$+ \beta_7 p_i + \beta_8 m_i + \beta_9 y_i + \beta_{10} c_i + \beta_{11} e_i + \beta_{12} g_i + \beta_{13} x_i + u_i \qquad (1)$$

where θ_i is the 1979-2002 trend in gridded surface climate data, $TROP_i$ is the time trend of Microwave Sounding Unit (MSU)-derived temperatures in the lower troposphere in the same grid cell as θ_i over the same time interval, $PRESS_i$ is the mean sea level air pressure in grid cell *i*, DRY_i is a dummy variable denoting when a grid cell is characterized by predominantly dry conditions (which is indicated by the mean dewpoint being below 0 ^{o}C). $DSLP_i$ is $DRY_i \times PRESS_i$, $WATER_i$ is a dummy variable indicating the grid cell contains a major coastline, $ABSLAT_i$ denotes the absolute latitude of the grid cell, p_i is local population change from 1979 to 2002, m_i is per capita income change from 1979 to 2002, y_i is total Gross Domestic Product

(GDP) change from 1979 to 2002, c_i is coal consumption change from 1979 to 2002, g_i is GDP density (national Gross Domestic Product per square kilometer) as of 1979, e_i is the average level of educational attainment, and x_i is the number of missing months in the observed temperature series and u_i is the regression residual. Further details, including data sources and definitions, are in *[MM07]*.

The results indicated that extraneous factors (*p* through *x*) have significant explanatory power on surface climate trends ($P = 7.1\times10^{-14}$), even after controlling for latitude, the lower troposphere temperature trends etc. After testing for and ruling out various forms of misspecification and spurious correlation they concluded that the gridded surface data are contaminated with extraneous signals. The coefficients from (1) imply that the contamination adds up to a net warming bias that can account for half the mean post-1980 global warming trend over land.

While *MM07* applied corrections for error clustering and heteroskasticity, subsequent criticism (R. Benestad http://www.realclimate.org/index.php/archives/2007/12/are-temperature-trends-affected-by-economic-activity-ii/; R. Pielke Sr., pers. comm..) raised the possibility that spatial autocorrelation of the climate trend field might be present and if so, failure to correct it would lead to exaggerated significance.

Spatial autocorrelation of the dependent variable is not a problem if the model on the right hand side explains it and leaves an uncorrelated residual. A test for residual spatial dependence can be implemented as follows. The regression model (1) can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\beta + u \quad (2)$$

where **y** is the linear trend in the temperature series for each of 440 surface grid cells, **X** is the matrix of climatic and socioeconomic covariates, β is the vector of least-squares slope coefficients and u is the residual vector. Spatial autocorrelation in the residual vector can be treated using

$$u = \lambda \mathbf{W} u + e \quad (3)$$

where λ is the autocorrelation coefficient, **W** is a symmetric $n \times n$ matrix of weights that measure the influence of each location on the other, and e is a vector of homoskedastic Gaussian disturbances, *[Pisati (2001)]*.

A test of $H_0 : \lambda = 0$ measures whether the error term in (1) is spatially dependent. *[Anselin et al. (1996)]* provides a discussion of the distributional properties of common tests of H_0. Standard adaptations of Wald and Lagrange Multiplier (LM) formulae yield tests that are severely biased towards over-rejection of the null. The problem is that if the alternative model allows for possible spatial dependence of the **y** variables, i.e.

$$\mathbf{y} = \phi \mathbf{Z} \mathbf{y} + \mathbf{X}\beta + e \quad (4),$$

where **Z** is a matrix of spatial weights for **y** and may not be identical to **W**, then conventional tests of $\lambda = 0$ assuming an alternative model of the form $\mathbf{y} = \mathbf{X}\beta + e$ is a misspecification. *[Anselin et al. (1996)]* derive a $\chi^2(1)$ Lagrange Multiplier (LM) test of $\lambda = 0$ robust to possibly nonzero ϕ in (4), which has substantially superior performance in Monte Carlo evaluations compared to the non-robust LM test.

Hypothesis tests, and any subsequent parameter estimations, are conditional on the assumed form of the spatial weights matrix **W** in (3). We examine three forms herein. Denote the great circle distance between the grid cell centers from which observation *i* and observation *j* are drawn as g_{ij}. Weighting matrix 1 (W1) is computed such that each element is $1/g_{ij}$ and the rows are standardized to sum to one. Weighting matrix 2 (W2) is computed such that each element is $1/\sqrt{g_{ij}}$ and the rows are standardized to sum to one. Weighting matrix 3 (W3) is computed such that each element is $1/g_{ij}^2$ and the rows are standardized to sum to one. Results were similar without row-standardization but convergence problems arose as the likelihood function had non-concave segments, so these results are omitted. Matrix W1 assumes the influence of adjacent cells diminishes at a hyperbolic rate. Matrix W2 assumes the inter-cell influence declines more slowly with distance while W3 assumes it declines more rapidly with distance.

2. Results

Table 1 shows the robust LM test values for weighting matrices W1—W3. In none of the three cases is there evidence of significant spatial autocorrelation in the residuals of (1). The inverse-square weighting rule, which allows for the slowest decline in the influence of adjacent grid cells as the distance increases, shows the largest test score, though it is still insignificant.

Analysis of the model dependent variables (**y**) did indicate spatial dependence. This implies that the right hand side variables in (1) explain the spatial dependence sufficiently to leave an error term that is not itself spatially autocorrelated.

For the sake of completeness, the regression model was re-run allowing for non-zero λ, also applying *[White's (1980)]* correction for heteroskedasticity, using a maximum likelihood routine developed by *[Pisati (2001)]*. The results are in Table 2. The coefficients are stable, as are the inferences, except for m_i, y_i and c_i under model W3, which fall to weak significance, and the joint test of anthropogenic surface processes also falls to weak significance (P = 0.067). But in all three cases the joint test of inhomogeneities and anthropogenic surface processes remains significant (P <0.01). The stability of the parameters indicates that the calculation of adjusted surface temperature trends still shows a large drop in the unweighted mean, from 0.30 C per decade to 0.17—0.18 C per decade. Weighting the grid cells by the cosine of latitude yields a drop from 0.27 C per decade to 0.12—0.14 C per decade (not shown).

3. Conclusions

MM07 reported evidence that gridded surface climate data are contaminated with extraneous signals due to inhomogeneities and anthropogenic surface processes, which may account for half the measured warming trend after 1980. This paper presents a test of the *MM07* model for exaggerated significance due to spatial autocorrelation in the residuals. Across numerous weighting specifications a robust LM statistic fails to reject the null hypothesis that no spatial autocorrelation is present, indicating that the estimations and inferences reported in *MM07* are not affected by spatial dependence of the surface temperature field. Even if the model is extended to treat spatial autocorrelation—at a risk of overspecification—the original results remain intact, especially the finding of significant contamination implying an overall warming bias that is large relative to the trend in the gridded data itself.

ACKNOWLEDGMENTS

The author sincerely thanks Werner Antweiler and Glen Waddell for assistance in the econometric programming.

References

McKitrick, R.R. and P.J. Michaels (2007), Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data, J. Geophys. Res., 112, D24S09, doi:10.1029/2007JD008465.

Anselin, Luc, Anil K. Bera, Raymond Florax and Mann J. Yoon (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26: 77—104.

Pisati, Maurizio (2001) Tools for spatial data analysis. *Stata Technical Bulletin* STB-60, March 2001, 21—37.

Weighting Matrix	Weighting Formula	Robust LM Score (*P* value)
W1	Inverse-linear	0.032 (0.858)
W2	Inverse-square root	2.564 (0.109)
W3	Inverse-squared	0.094 (0.759)

Table 1. Hypothesis tests for spatial autocorrelation in model (1) of surface temperature trends and inhomogeneity-anthropogenic surface process biases. The null hypothesis is no spatial dependence in the model residuals.

Variable	Original	W1 Inverse Linear	W2 Inverse sqrt	W3 Inverse sq
trop	**0.8631**	**0.8231**	**0.8598**	**0.7515**
	(8.62)	(11.46)	(13.16)	(7.41)
slp	0.0044	0.0048	0.0045	0.0053
	(1.02)	(1.50)	(1.44)	(1.45)
dry	0.5704	0.8337	0.5813	1.4746
	(0.10)	(0.21)	(0.15)	(0.33)
dslp	-0.0005	-0.0007	-0.0005	-0.0014
	(-0.09)	(-0.18)	(-0.13)	(-0.31)
water	-0.0289	-0.0306	-0.0290	-0.0358
	(-1.37)	(-1.50)	(-1.44)	(-1.70)
abslat	0.0006	0.0008	0.0006	0.0014
	(0.51)	(0.85)	(0.70)	(1.23)
g	**0.0432**	**0.0446**	**0.0434**	**0.0438**
	(3.36)	(3.12)	(3.08)	(2.92)
e	**-0.0027**	**-0.0025**	**-0.0027**	**-0.0021**
	(-5.14)	(-4.00)	(-5.21)	(-2.54)
x	0.0041	0.0033	0.0041	0.0019
	(1.66)	(1.04)	(1.25)	(0.63)
p	**0.3839**	**0.3641**	**0.3822**	**0.3162**
	(2.72)	(2.91)	(3.13)	(2.35)
m	**0.4093**	**0.3589**	**0.4054**	0.2594
	(2.39)	(2.56)	(2.92)	(1.68)
y	**-0.3047**	**-0.2669**	**-0.3018**	-0.1961
	(-2.22)	(-2.41)	(-2.76)	(-1.64)
c	**0.0062**	**0.0056**	**0.0061**	0.0042
	(3.45)	(2.91)	(3.32)	(1.86)
Constant	-4.2081	-4.6341	-4.2481	-5.1241
	(-0.96)	-1.42	-1.35	-1.39
N	440	440	440	440
R^2	0.53			
Log likelihood	139.22	141.45	139.25	147.84
P(I)	**0.000**	**0.000**	**0.000**	**0.005**
P(S)	**0.000**	**0.002**	**0.001**	0.067
P(all)	**0.000**	**0.000**	**0.000**	**0.007**
Adj surf	0.17	0.17	0.17	0.18

Table 2. Coefficient estimates for equation (1). First column: as reported in *MM07*. Columns 2—4, allowing for spatial autocorrelation using weighting schemes described in text. Coefficient t-statistics underneath in parentheses, based on standard errors that account for heteroskedasticity. **Bold** denotes significant at 95%. Variable codes: *g*: 1979 GDP density; *e*: educational attainment; *x*: count of missing months; *p*: % change in population; *m*: %income growth; *y*: % growth in GDP; *c*: % growth in coal consumption. ll = loglikelihood value. P(I) = prob value of

test that all inhomogeneity factors (g—x) are jointly zero; P(S) = prob value of test that all surface process coefficients (p—c) are jointly zero; P(all) = prob value of test that g—c are jointly zero. Adj. surf: unweighted mean surface temperature trend across 440 grid cells after removing extraneous effects following methodology in *MM07*.