Modeling Spatial Patterns in Weather and Crop Yields

Seong Do Yun[1], Benjamin M. Gramig[2], Michael S. Delgado[2], and Raymond J.G.M. Florax[2,3]

[1]Yale School of Forestry and Environmental Studies, [2]Department of Agricultural Economics, Purdue University

[3]Department of Spatial Economics, Vrije Universiteit Amsterdam, The Netherlands

Introduction

- The rapidly growing availability and accessibility of spatially-explicit weather and climate data
- Relatively less attention to spatial correlation in weather variables and econometric methods
- Two approaches in the previous literature
  - Applied economics: Conley's Method
  - Spatial econometrics
- Reasoning for specification strategy in modeling spatial patterns is not stated explicitly
- Research Question:
  Which specification strategy of spatial correlation performs best in terms of prediction capability?

An Example Model: Crop Yield Response

- Schlenker and Roberts (2009)

$$y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}\,dh + \gamma_1 P_{it} + \gamma_2 P_{it}^2 + \gamma_3 t + \gamma_4 t^2 + \mu_i + \varepsilon_{it}$$

- Very popular non-Ricardian model in climate-yield literature
- Spatial correlation in the VC matrix
  - To take into account omitted weather variable bias
- This study argues that the main reason for spatial correlation in the VC matrix is aggregation bias rather than omitted weather variables
- Six comparable models are specified and prediction performance is compared

Model Specifications

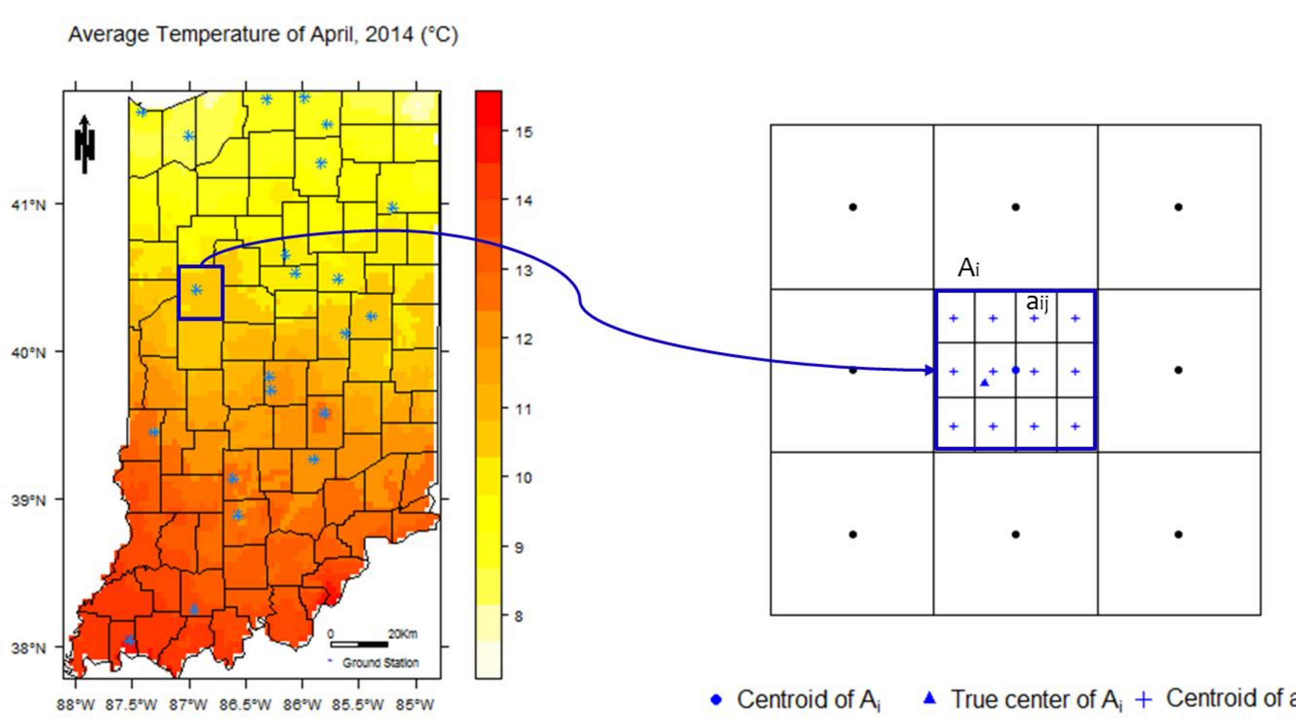
Specification Motivations and Models

Motivation	Fixed Effects Model	Random Effects Model with Soil Vars.
Aggregation Bias	FE (Schlenker and Roberts, 2009)	
Omitted Socio-economic Vars.	FE SLAG (ρWy)	
Omitted Weather Vars.		RE, KKP-RE (Auffhammer et al., 2013)
Geophysical Process	SLX ($WC\theta$)	

Models Compared

- Pooled Regression
  $y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}\,dh + \gamma_1 P_{it} + \gamma_2 P_{it}^2 + \gamma_3 t + \gamma_4 t^2 + \varepsilon_{it}$
- FE (Schlenker and Roberts, 2009): aggregation bias
  $y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}\,dh + \gamma_1 P_{it} + \gamma_2 P_{it}^2 + \gamma_3 t + \gamma_4 t^2 + \mu_i + \varepsilon_{it}$
- FE Spatial Lag (FE-SLAG): Omitted socio-economic vars.
  $y_{it} = \rho \mathbf{w}_i \mathbf{y}_t + \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}\,dh + \gamma_1 P_{it} + \gamma_2 P_{it}^2 + \gamma_3 t + \gamma_4 t^2 + \mu_i + \varepsilon_{it}$
- Spatial Lagged X (SLX): geophysical process of weather Vars.
  $y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}\,dh + \gamma_1 P_{it} + \gamma_2 P_{it}^2 + \gamma_3 t + \gamma_4 t^2 + \mathbf{w}_{it}\mathbf{C}_{jt}\boldsymbol{\theta} + \mu_i + \varepsilon_{it}$
- RE: Omitted Weather Vars.
  $y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}\,dh + \gamma_1 P_{it} + \gamma_2 P_{it}^2 + \gamma_3 t + \gamma_4 t^2 + \boldsymbol{\gamma}\mathbf{S}_i + \varepsilon_{it}$
- Kapoor, Kelejian, and Prucha RE (KKP-RE): RE + KKP error structure $\mathbf{u}_N = \rho(I_T \otimes W_N)\mathbf{u}_N + \varepsilon_N$

Aggregation Bias

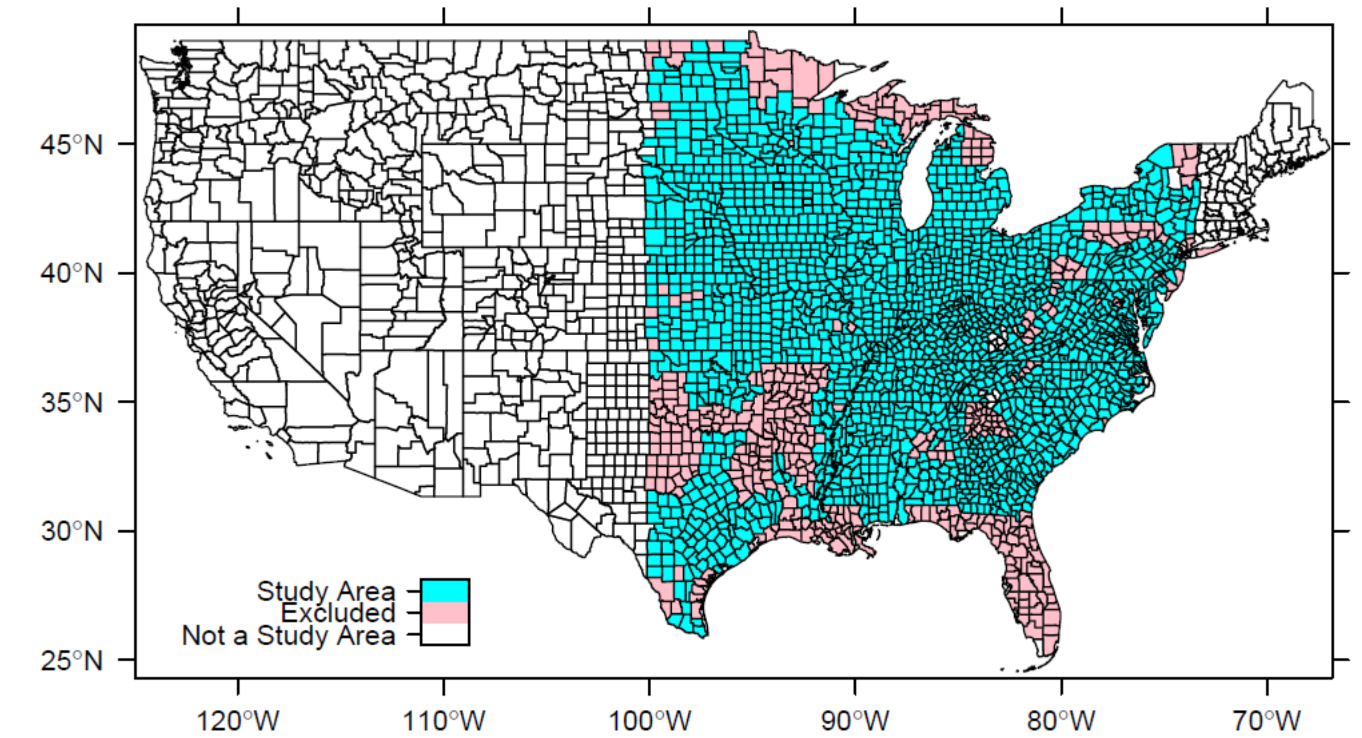


Temperature at j-th cell of i-th county (H_{ij})

$$H_{ij} = \underbrace{\overline{H}_i}_{\text{county average}} + \underbrace{\nu_{ij}}_{\text{cell-level variation}} = \sum_{j=1}^{n_i} w_{ij} H_{ij} + \nu_{ij}$$

- The error terms absorb all spatial variation
  - **Proposition 1**
    The larger the aggregation losses, the less spatial variation.
    $\lim_{n_i \to \infty} \sum_{j=1}^{n_i} |\nu_{ij}|$ does not define.
  - **Proposition 2**
    The larger aggregation has the less spatial correlation.
    $\rho_k \leq \rho_i$ when k is larger aggregation level than i
  - **Proposition 3**
    The areal center of weather variable is unknown.
    Distance based (fixed) weights matrix could be inappropriate.
    $\Rightarrow$ Conley's or SHAC estimator

Data

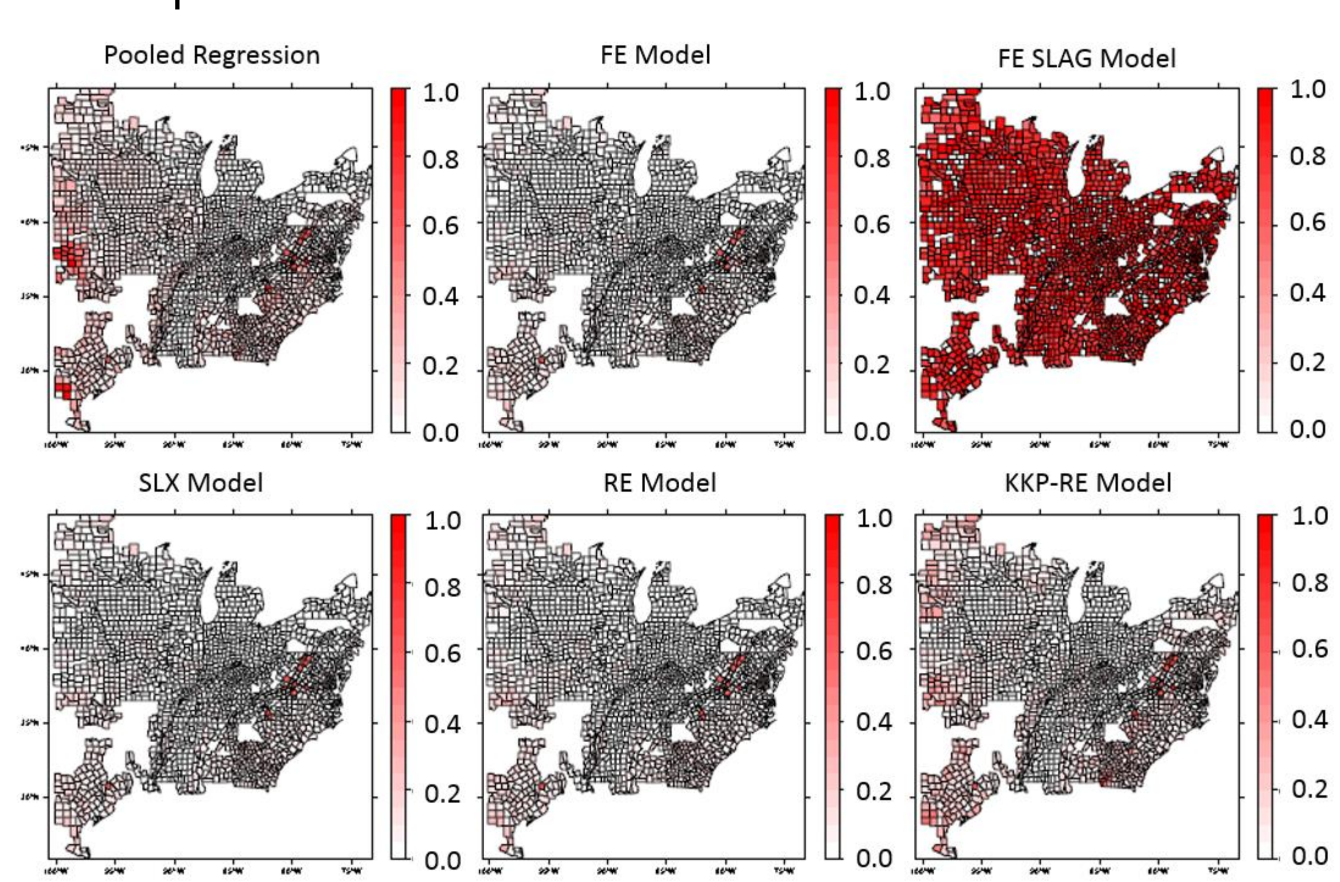


Study Areas: (1,964 counties X 33 years) Balanced Panel

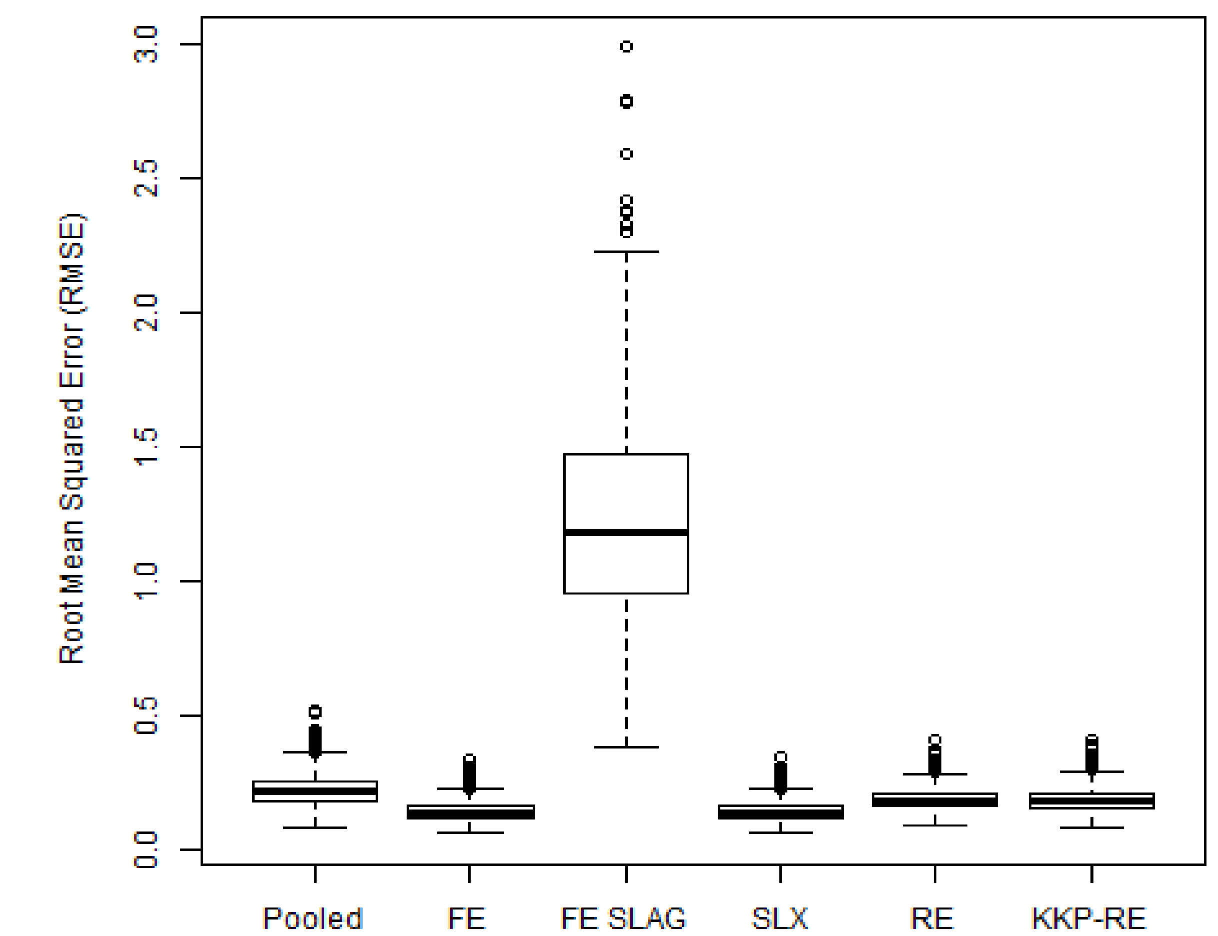
- Non-agricultural land masked out
  1992 LandSat (USGS, 30m X 30 m) applied
- Corn yields
  NASS county-level yields from 1981 – 2013
- Counties are based on 2013 (500K map) from the Census Bureau
- Temperature and total precipitation
  PRISM (Oregon State University, 4 Km X 4 Km)
- Soil Variables
  gSSURGO (NRCS, 10m X 10m)

Results

In Sample Prediction



Out of Sample Prediction (1,000 Simulations):
27 years for estimation / 6 years for prediction



- Best performance: SLX ~ FE and KKP-RE ~ RE
- FE-SLAG performed worst in our sample

Conclusion

- Aggregation bias is one of the major sources of spatial correlation in addition to omitted variables and geophysical processes
- Depending upon the causes of spatial correlation, specification of the regression equation can be different
- Soil characteristic should be included in RE specification
- SHAC estimator provide tighter CI band estimates

❖ Consider economic and geophysical reasoning more to guide model selection in prediction capability of crop yield response function

Acknowledgement

- Cluster computing resources for this research is supported by Agriculture and Food Research Initiative Competitive Grant no. 2011-68002-30220 from the USDA National Institute of Food and Agriculture (NIFA), entitled "Useful to Usable (U2U): Transforming Climate Variability and Change Information for Cereal Crop Producers." (http://www.agclimate4u.org/)
- The presentation travel to the 2015 Heartland Environmental & Resource Economics Workshop at Illinois is supported by the Knoblock Family Foundation and the Yale School of Forestry and Environmental Studies.