Space-Time Models for Retail Credit

Bob Stine
Department of Statistics
The Wharton School of the University of Pennsylvania

William Lang
Federal Reserve Bank of Philadelphia

August 2007

Questions to Consider

Question 1

- Is there spatial variation in credit and macro-economic conditions in the US?
  - Often hear numbers like "the" unemployment rate or level of disposable income...
  - How much variation around the overall numbers is present?
  - What is the spatial distribution of the variation?

Question 2

- Do local economic conditions affect the fit of models of retail credit risk?

  - Models routinely incorporate "bank data" that includes past default rates, utilization, ...
  - Do economic variables such as local employment levels add value beyond information in the bank data?

$$P(\text{default}|\text{bank},\text{macro}) = P(\text{default}|\text{bank})?$$

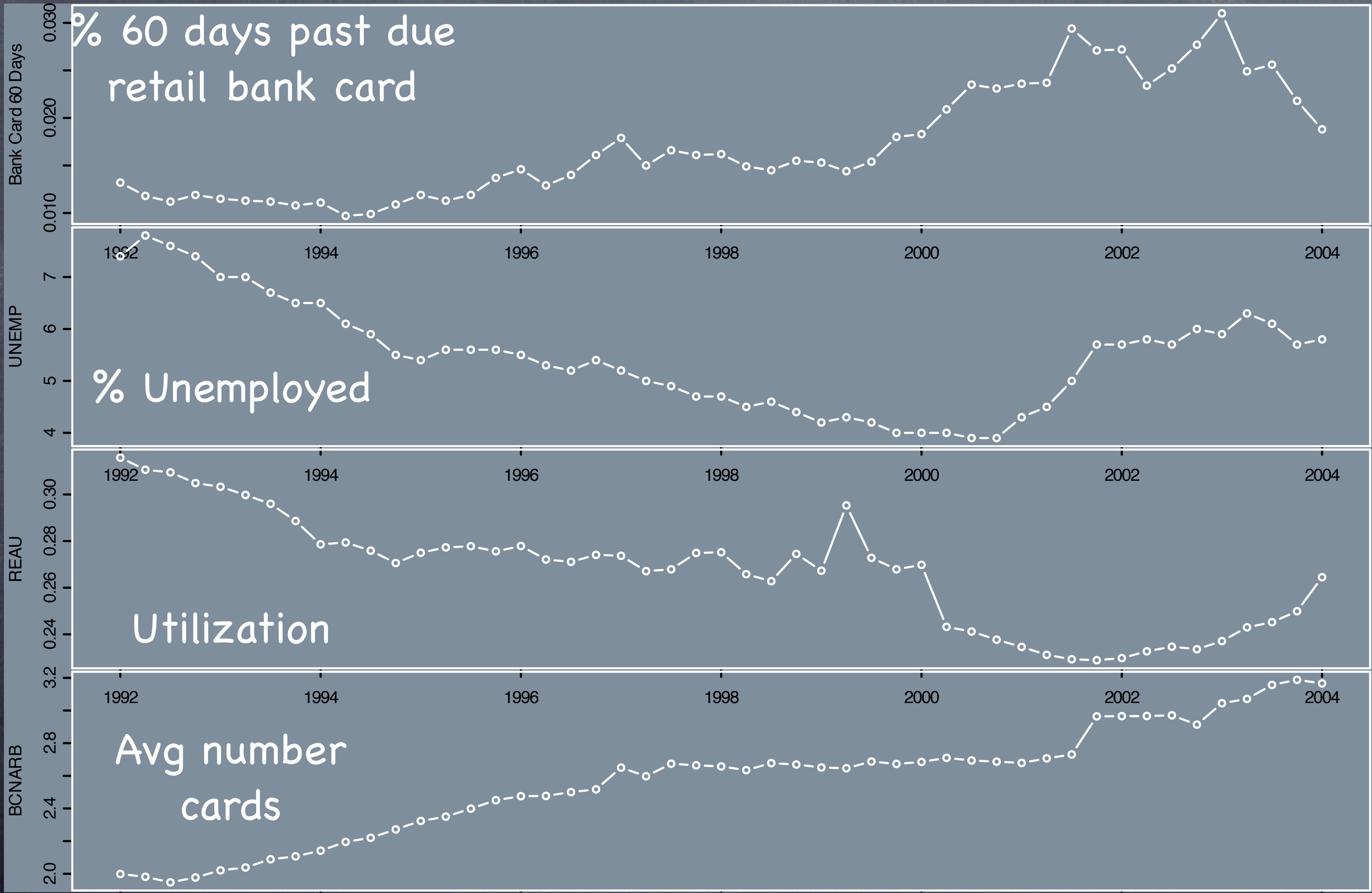
Question 3

- Does spatial variation in economic conditions produce a form of stress-testing?
  - Stress-test
    Does an overall model fit well when applied in times of economic distress?
  - Back-testing prescribed in Basel II
  - Tricky to prescribe conditions for stress test
  - Does spatial variation provide a natural framework for exploring model accuracy in periods of high economic stress?

County-Level Data

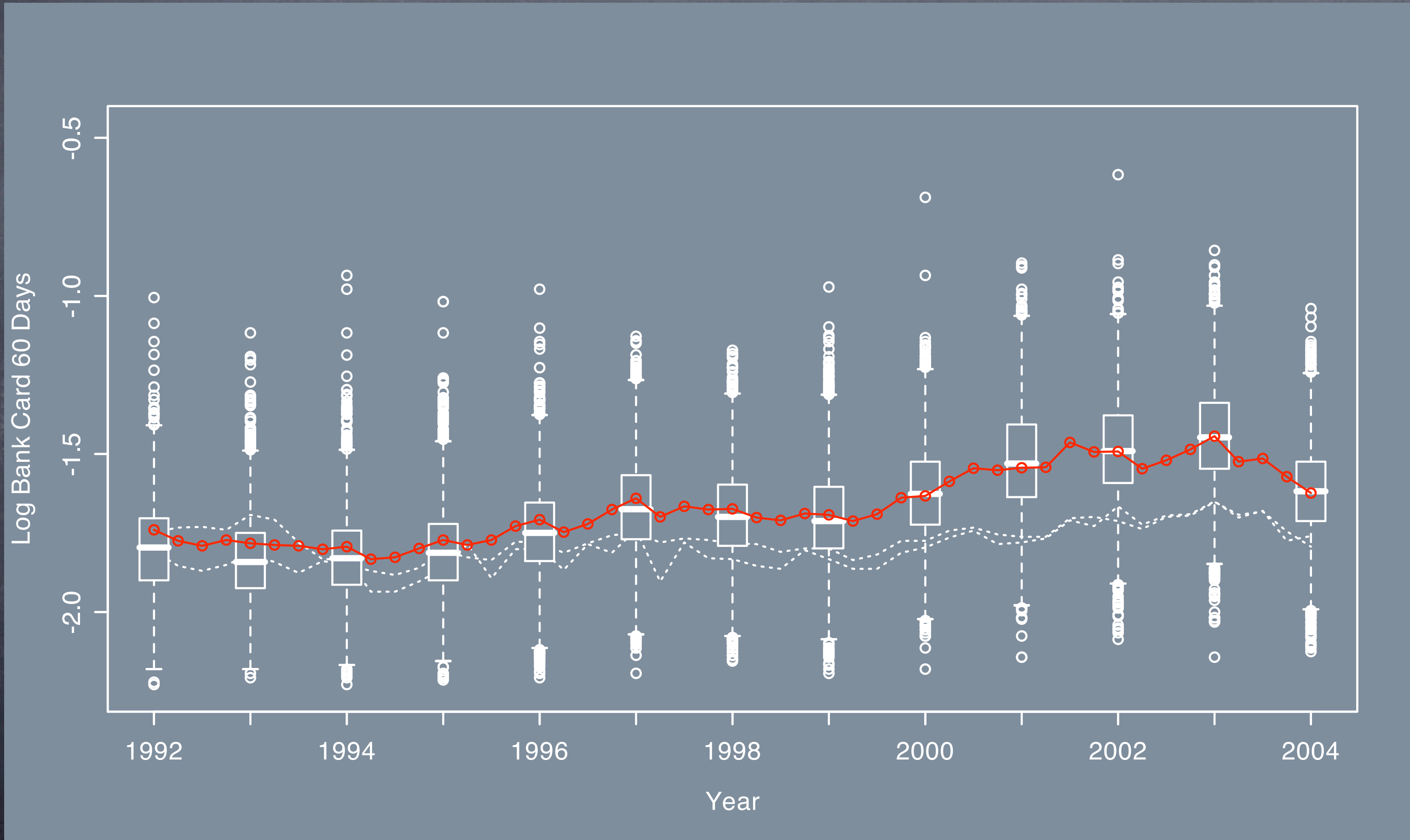
- Trend Data (TransUnion)
  - https://products.trendatatu.com/welcome.asp
  - Quarterly, 1992Q1 - 2004Q1 (49 quarters)
    - % 60 days past due among retail bank cards
    - % utilization among bank cards
    - average number of bank cards
- Macro-economic data
  - Monthly unemployment (Bureau of Labor Statistics, LAUS, Department of Labor)
  - Annual median income, % in poverty (SAIPE, US Census Bureau)
    - I'll use estimates derived from 2000 Census
    - Census now estimates annually as part of the American Community Survey (replacing old "long form")

National Trends

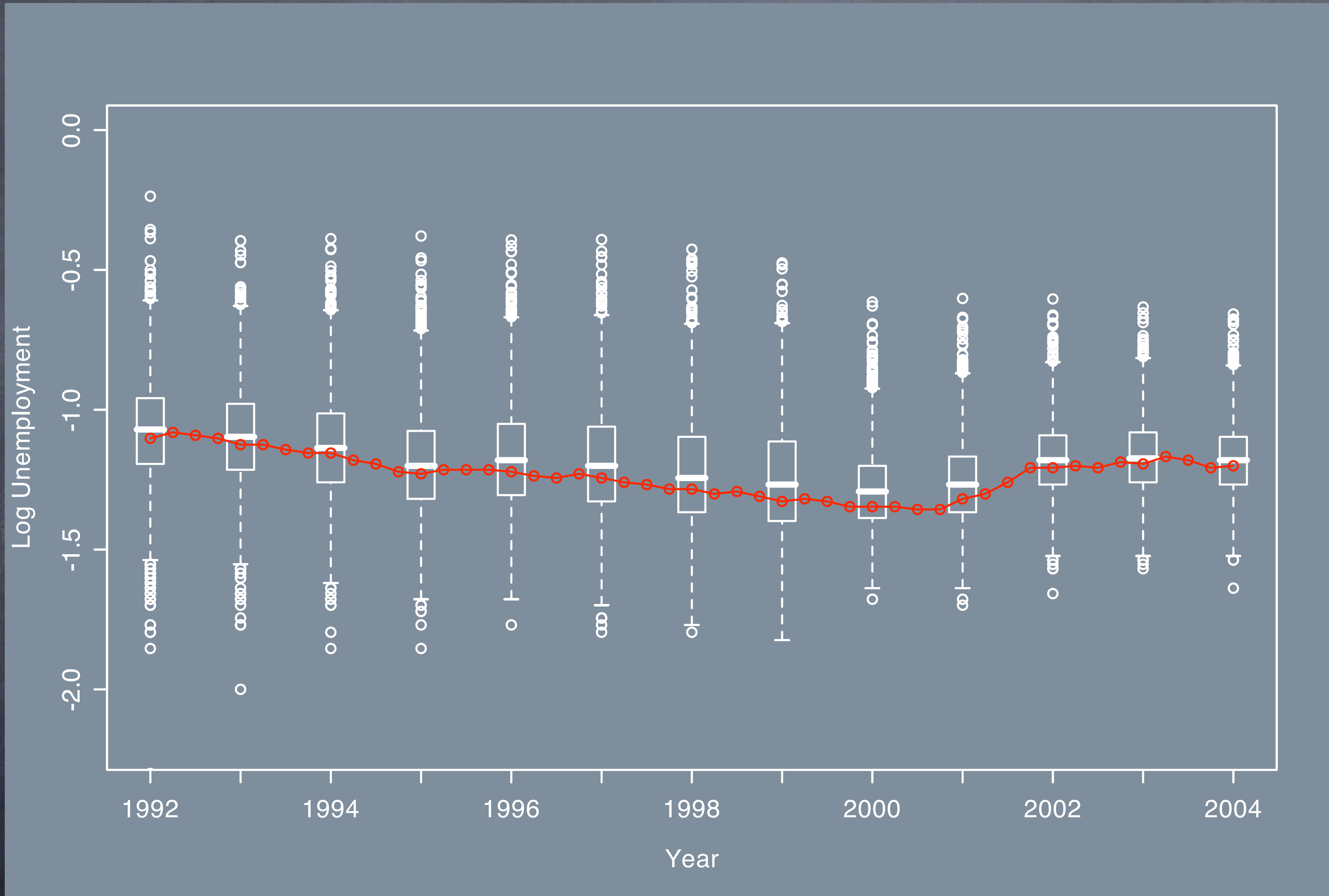


Regional Variation

Log % 60 Days Past Due



Variation in Unemployment

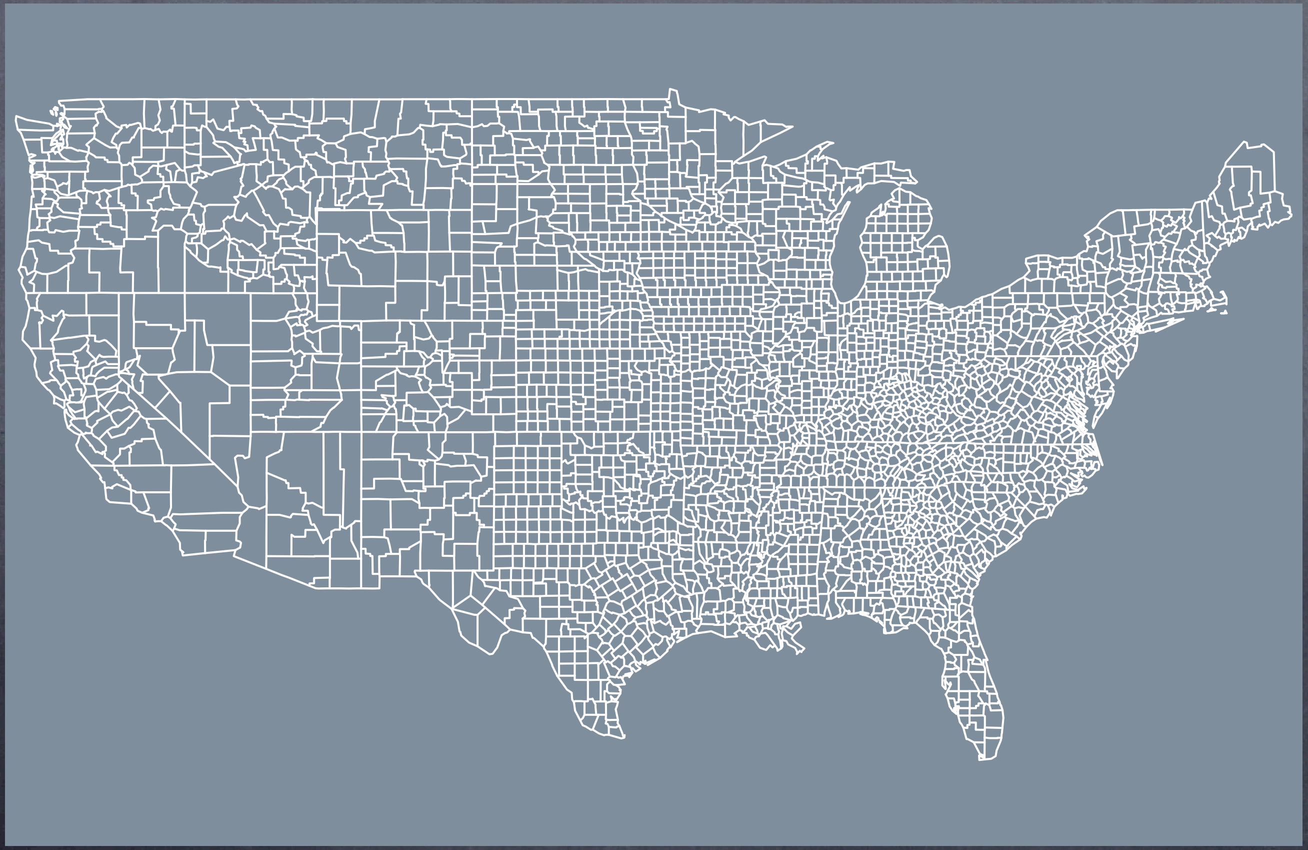


Quite large, with diminishing variation

Spatial View of Data

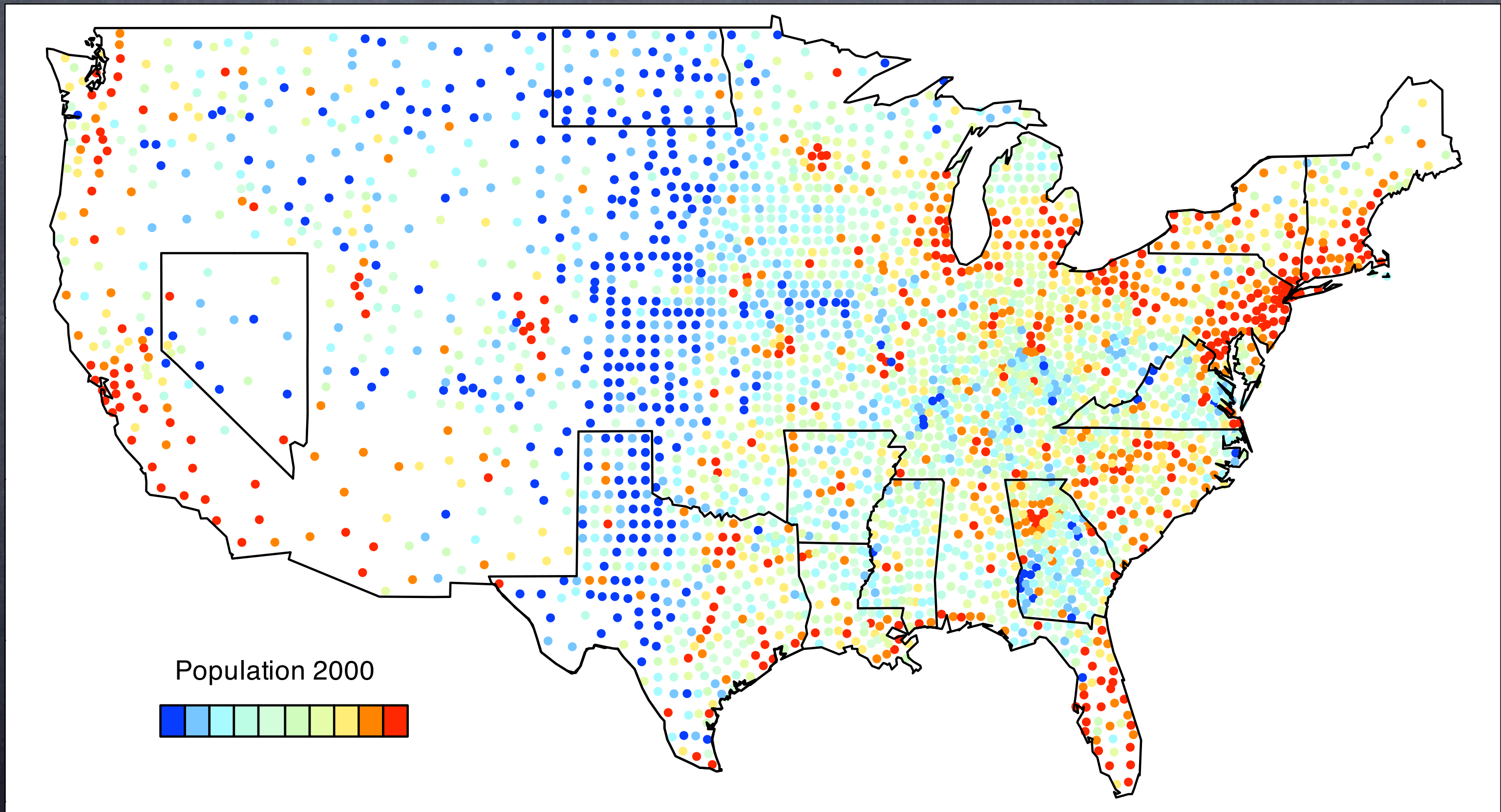
Regions

- County
  - Continental US has 3,000 counties
  - Diverse range of shapes and sizes



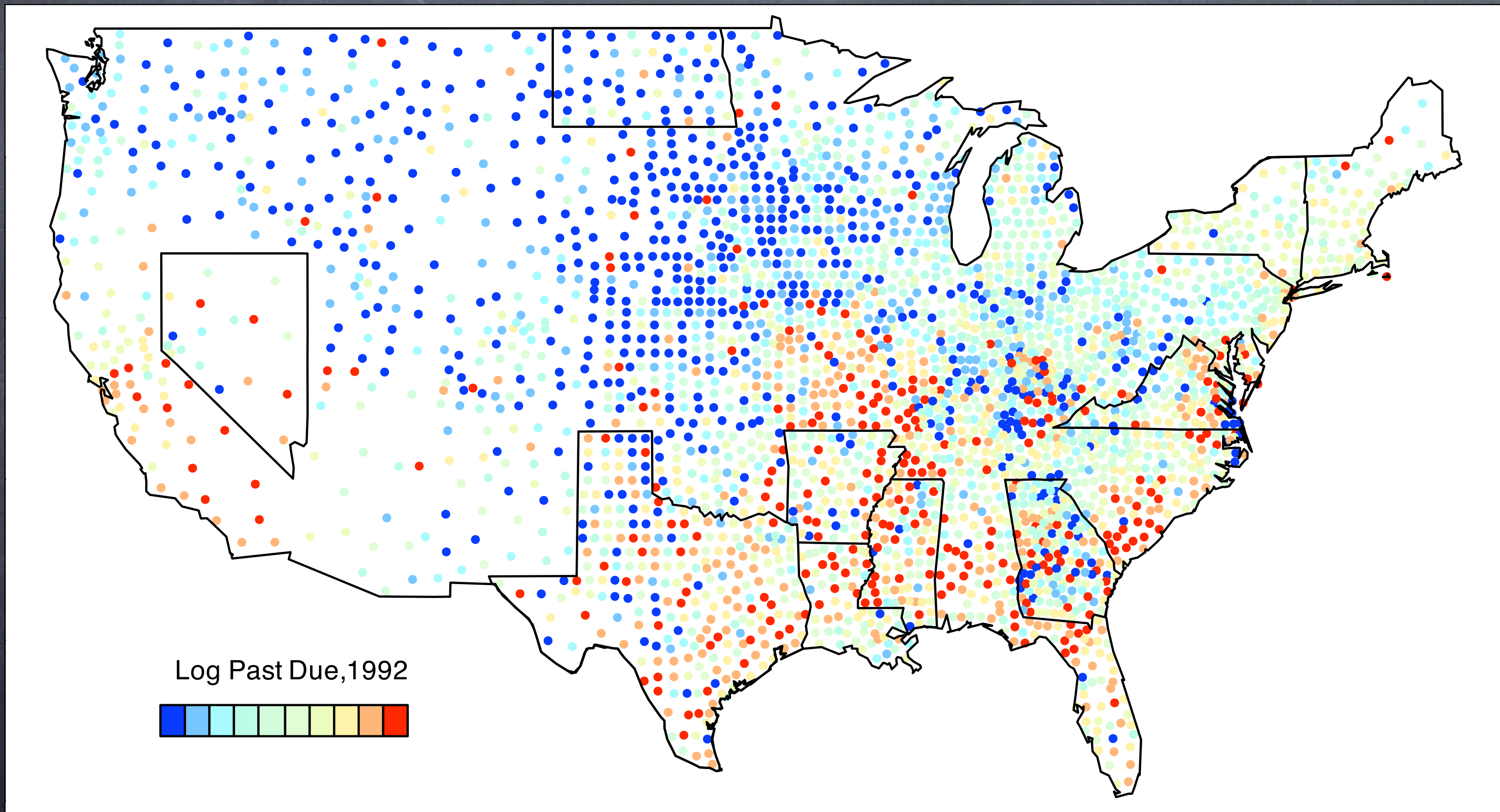
Spatial Clustering

- Concentrations of high population
  - Evident urban clusters
  - Confounding: urban and population



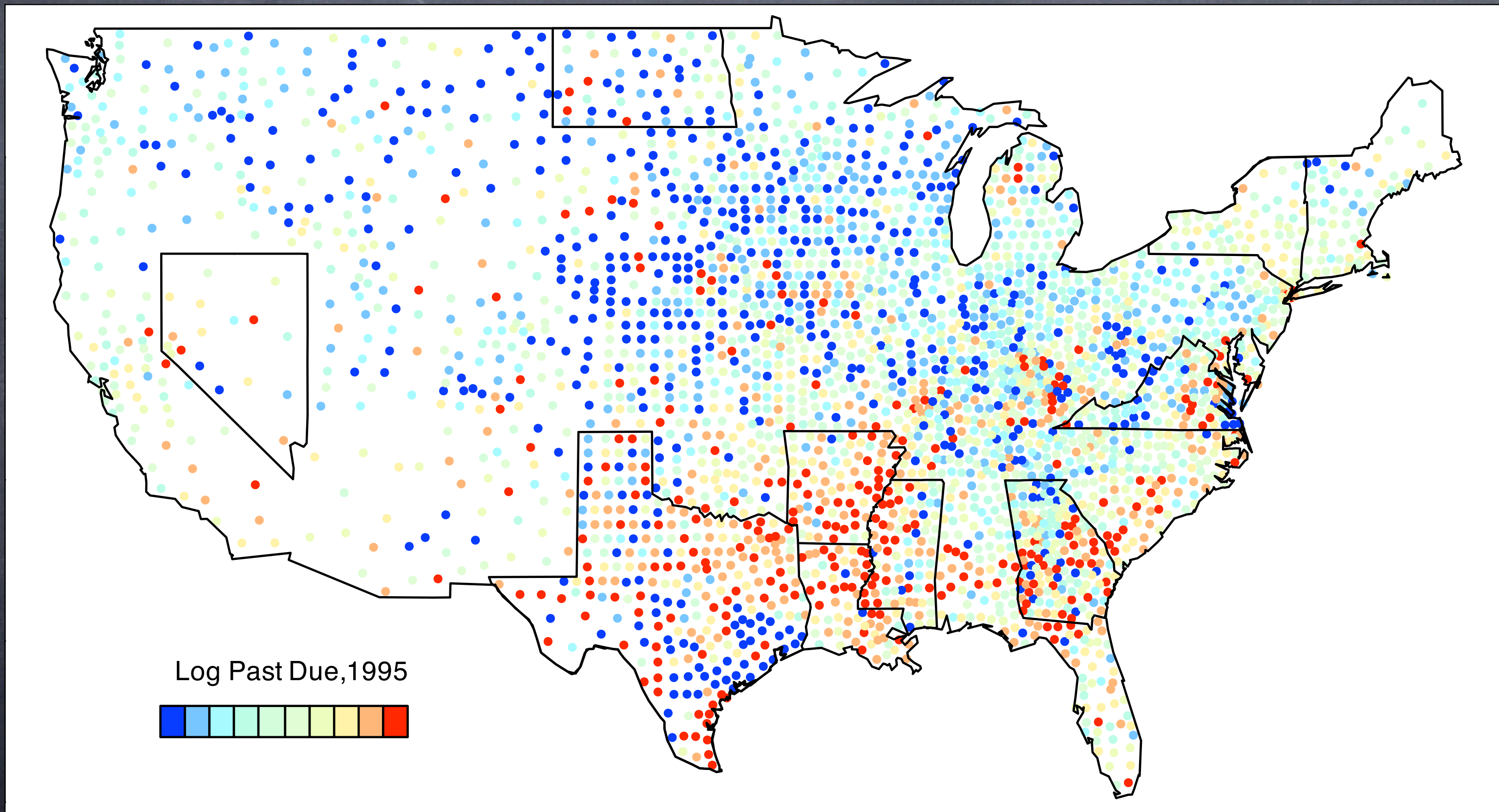
Spatial Patterns

Concentrations of high and low default



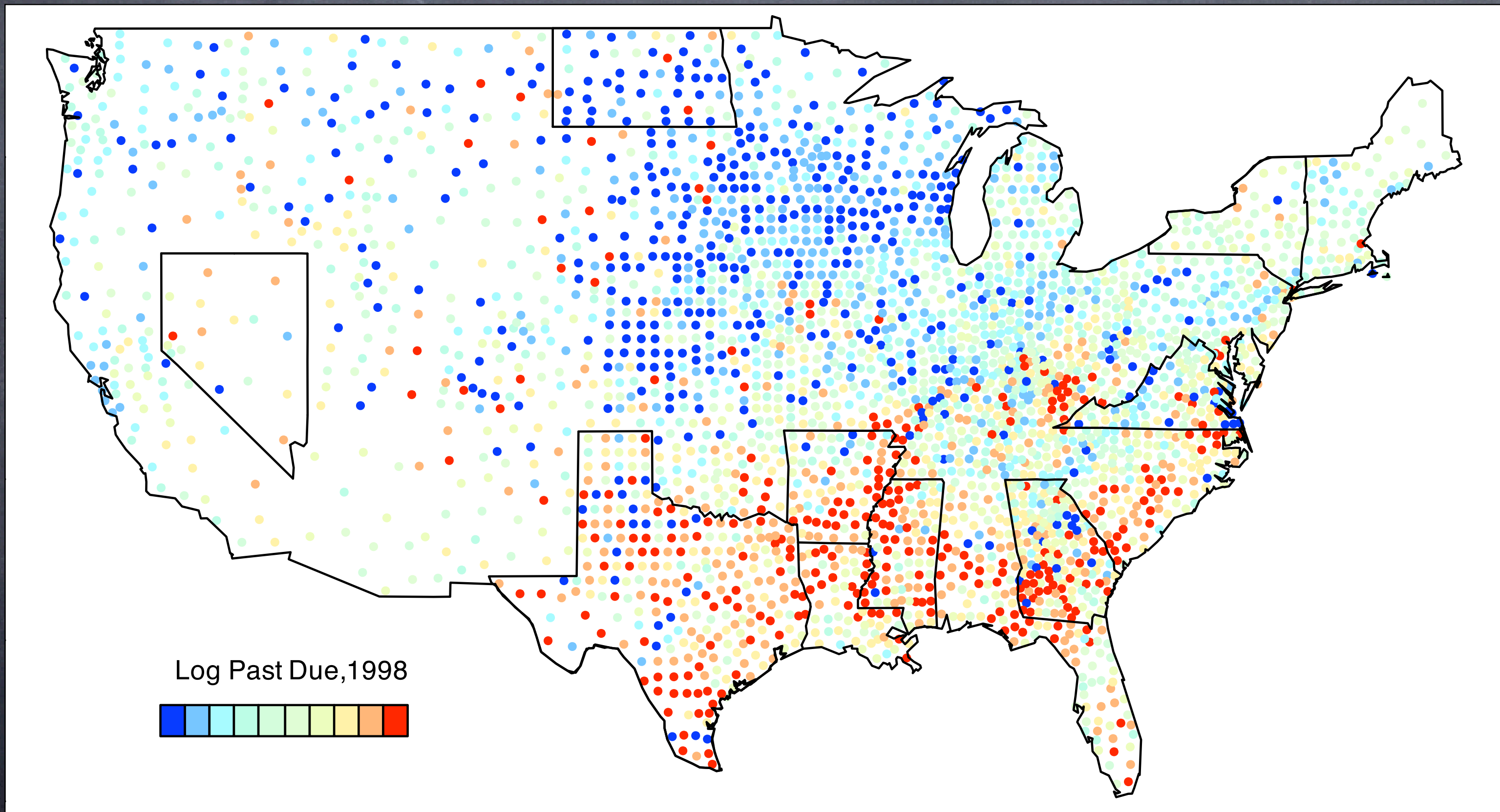
Spatial Patterns

Concentrations of high and low default



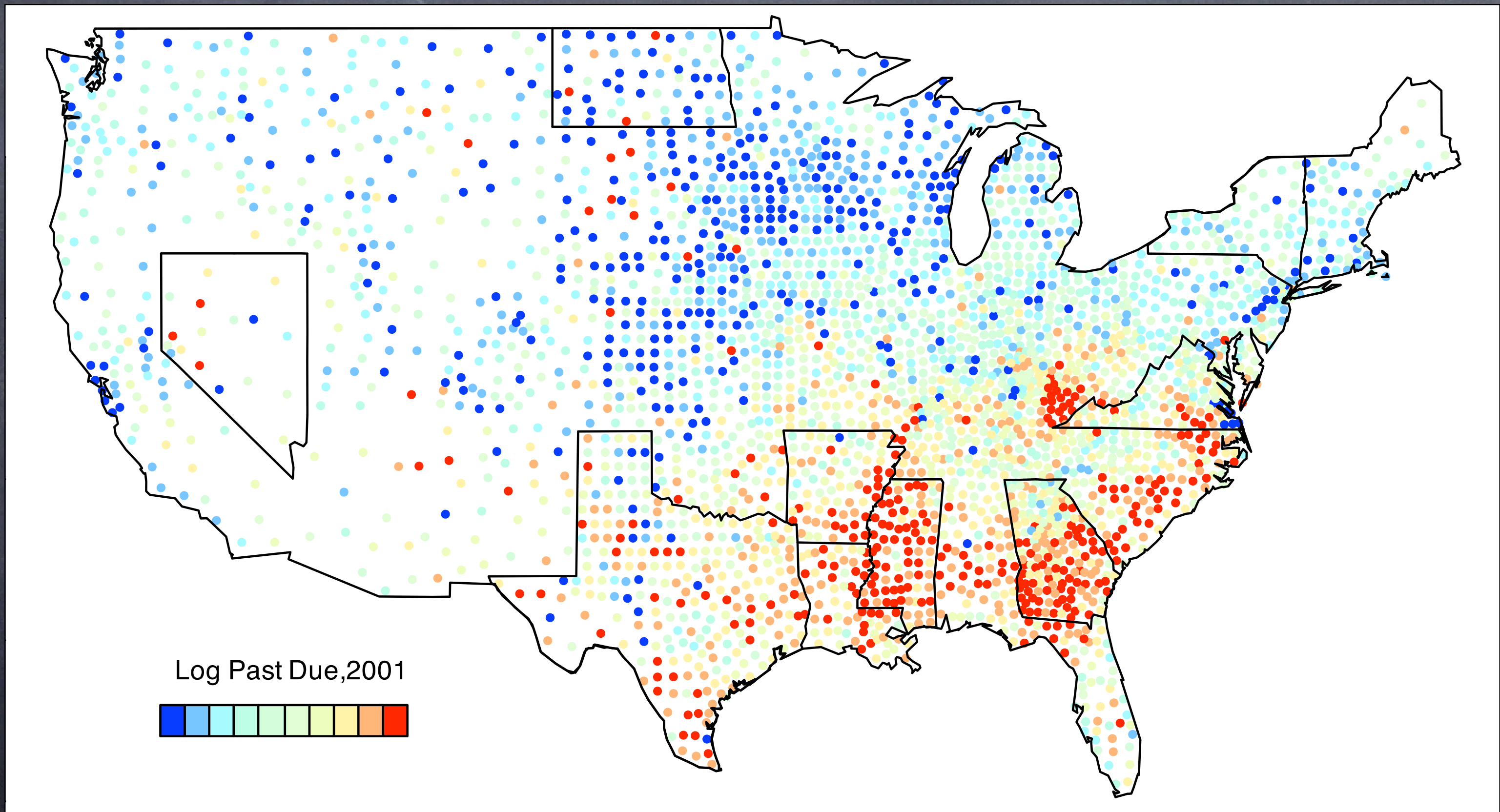
Spatial Patterns

Concentrations of high and low default



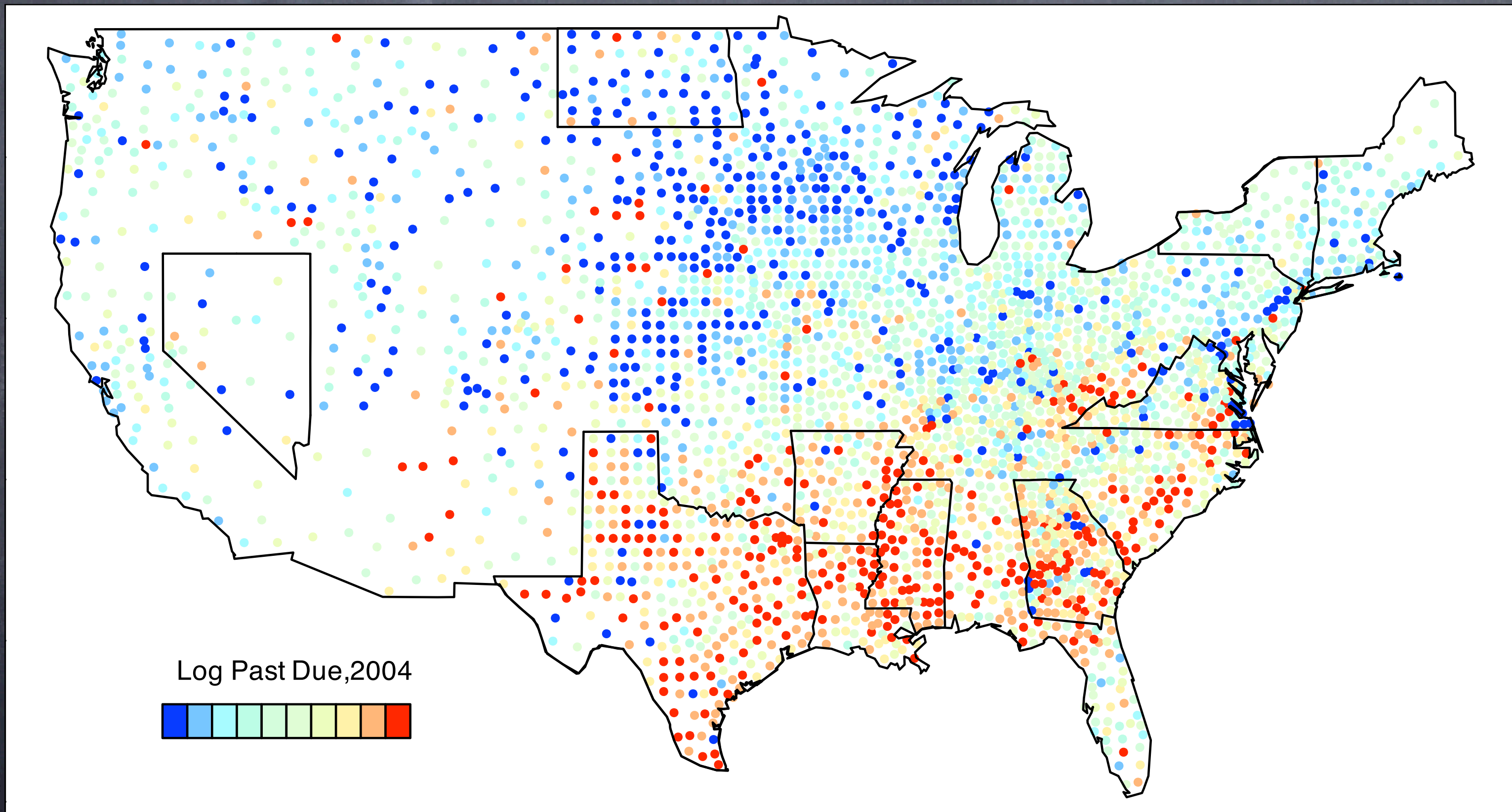
Spatial Patterns

Concentrations of high and low default



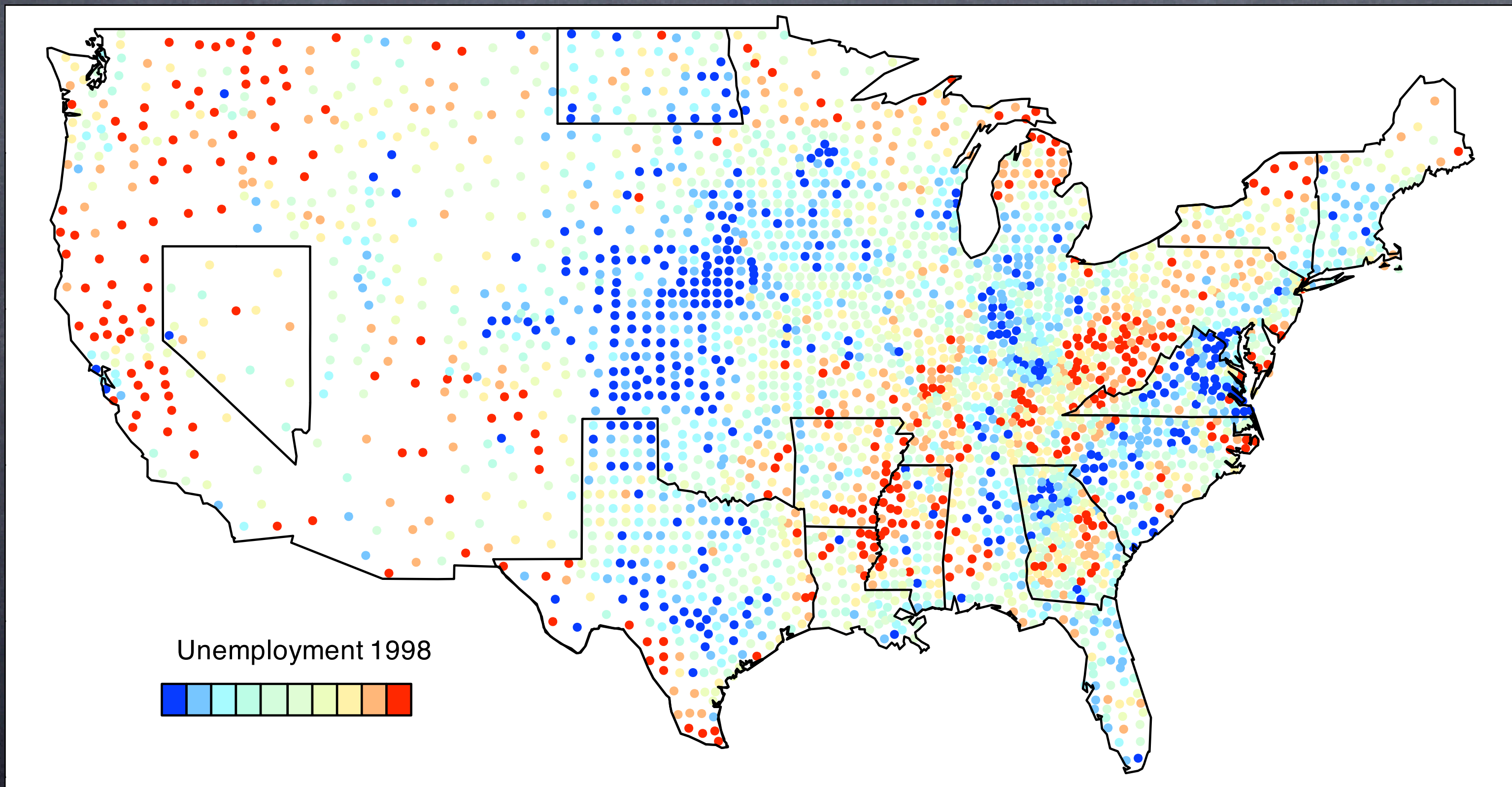
Spatial Patterns

Concentrations of high and low default



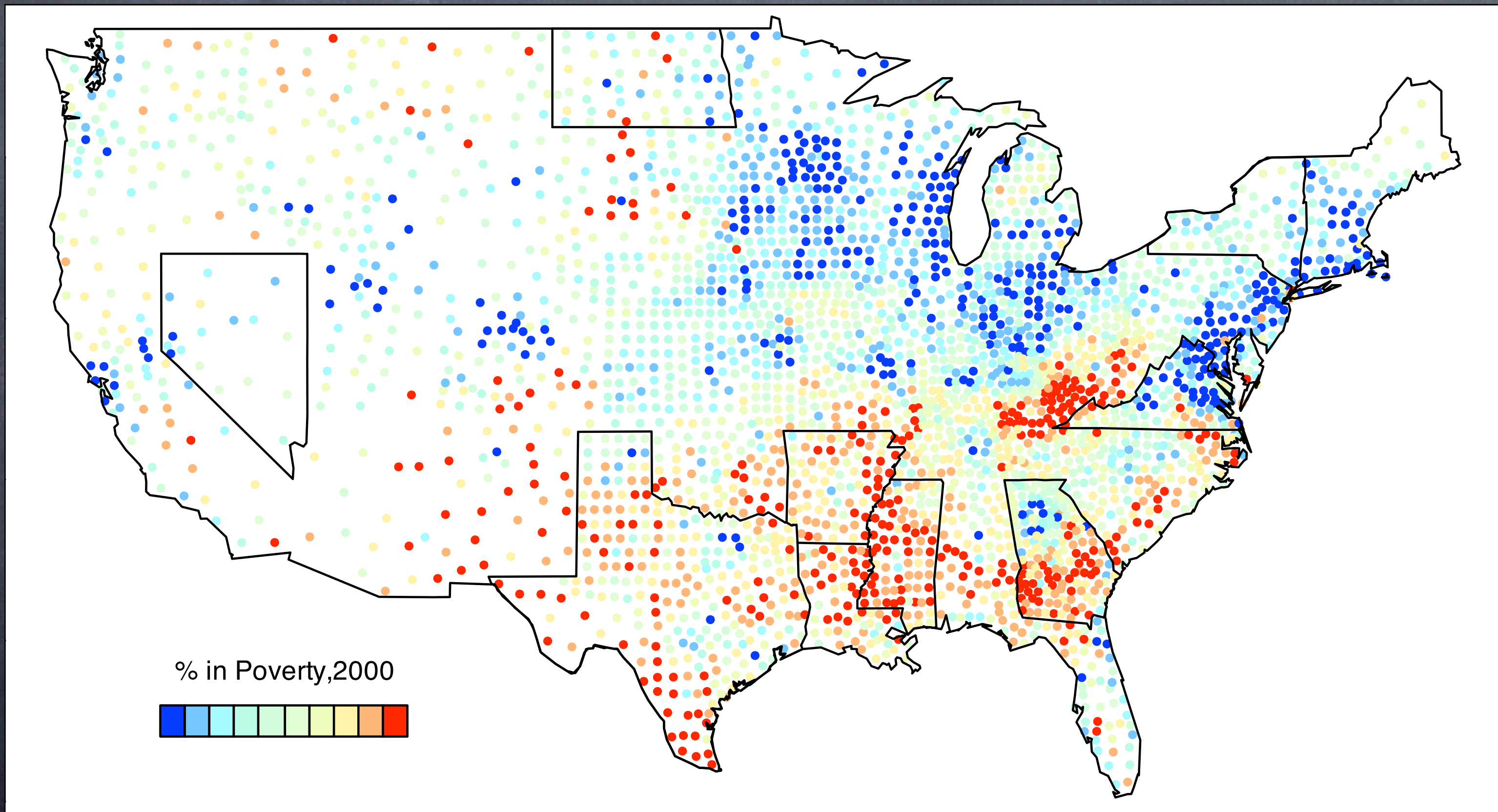
Unemployment

- Substantial regional concentrations in Mississippi valley, Appalachia, West



Spatial Patterns

Poverty also concentrated in southeastern US



Return to Questions

1. Is there adequate spatial variation to support modeling credit risk?

   Yes. Maybe too much!

2. Do local macro-economic variables add value beyond usual bank information?

3. Do models suffer under local economic stress?

Models

Models

- Predict percentage late payments
  - Log scale, one point in time
- Baseline model ignores covariates
  - $\text{Log}(\text{Late}_t) = b_{0,t} + b_{1,t}\ \text{Log}(\text{Late}_{t-1}) + e_t$
- More complex models include bank variables plus macroeconomic variables
  - Add lagged covariates of several types

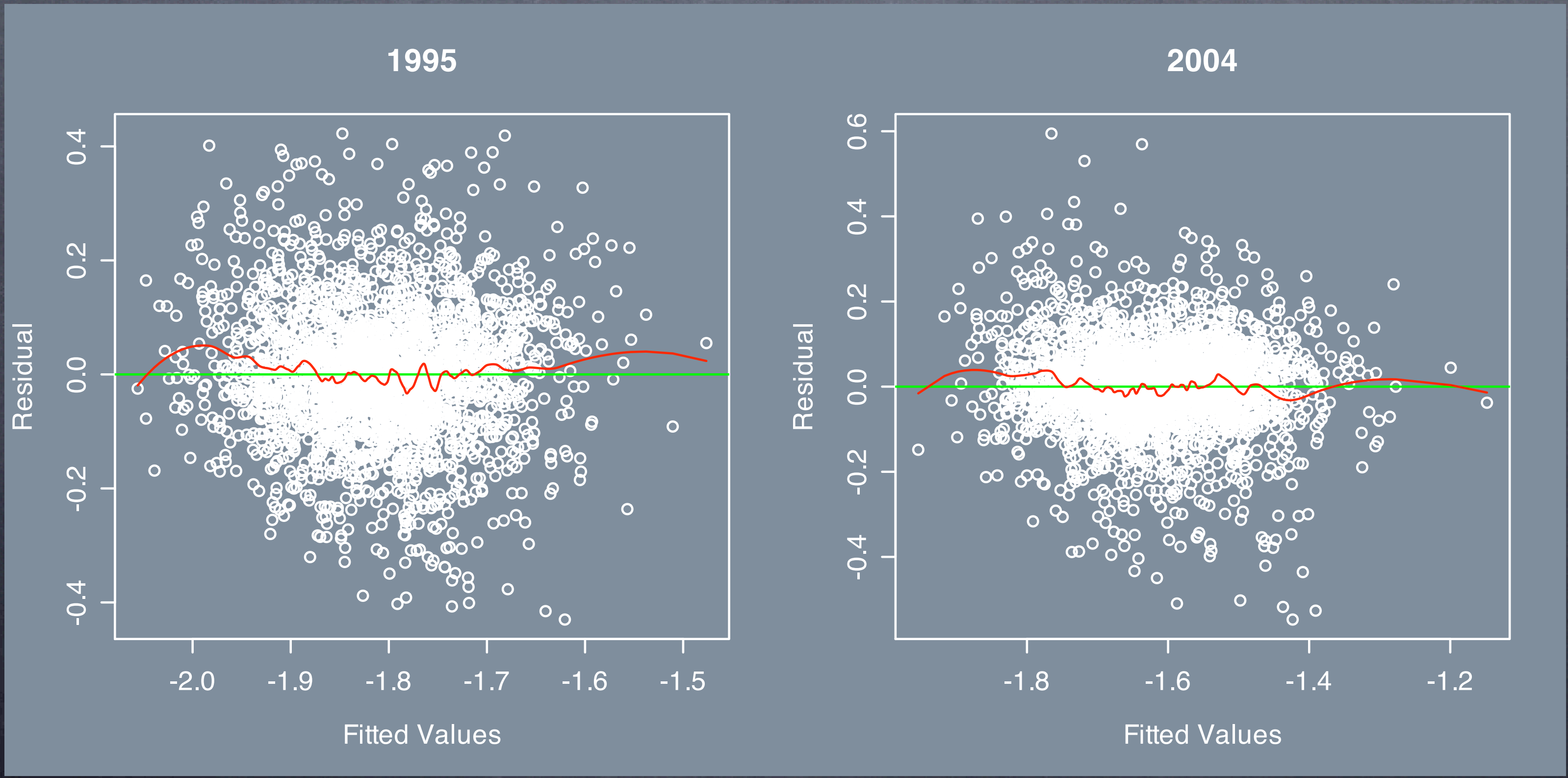
$$\text{Log}(\text{Late}_t) = b_{0,t} + b_{1,t}\ \text{Log}(\text{Late}_{t-1}) +$$

"bank" $\quad b_{2,t}\ \text{Log}(\text{Util}_{t-1,\ t-2}) + b_{3,t}\text{Log}(\text{Cards}_{t-1,t-2}) +$

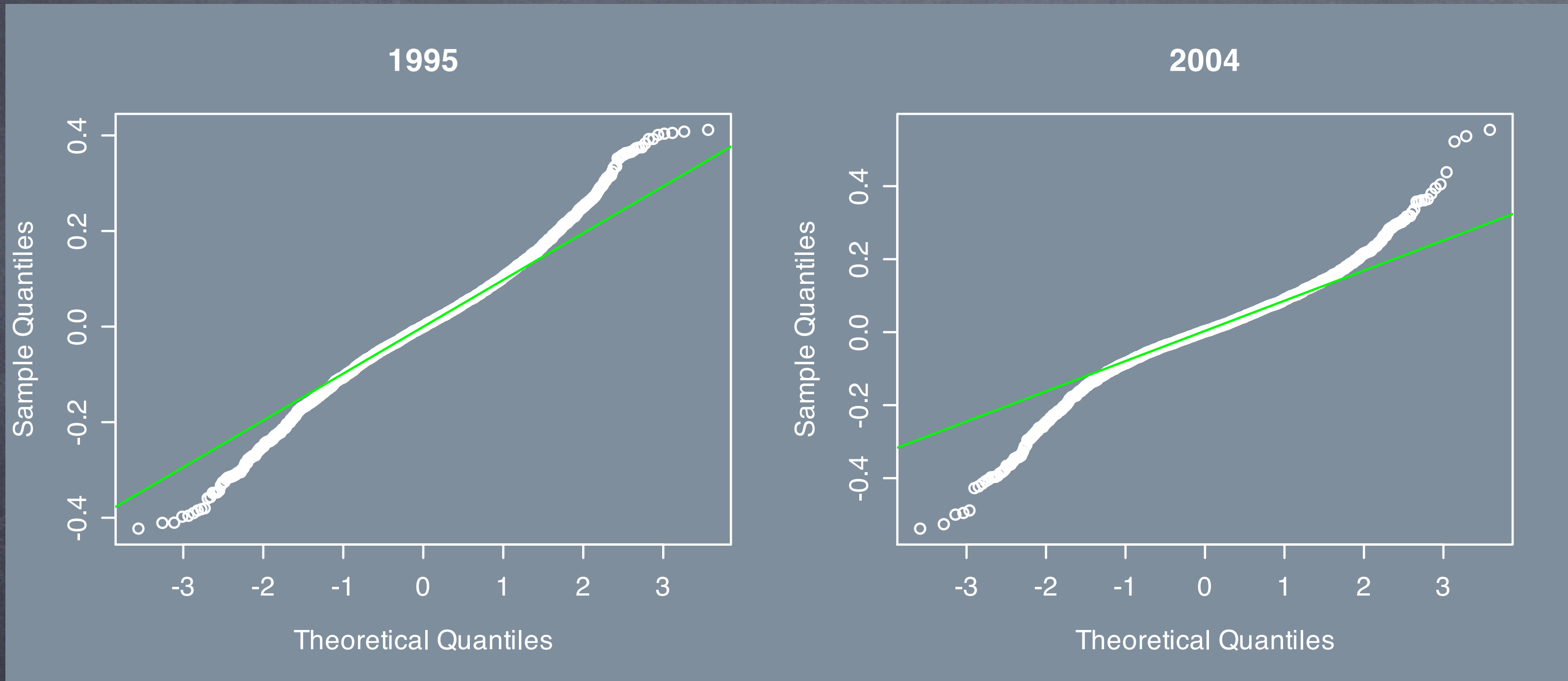
"macro" $\quad b_{4,t}\ \text{Log}(\text{Un}_{t-1,\ t-2,\ t-3,\ t-4}) + b_{5,t}\ \text{Log}(\text{Pov}) + e_t$

Residual Plots

- Models are well-calibrated, with fitted values linearly related to response.
- Larger residuals randomly scattered



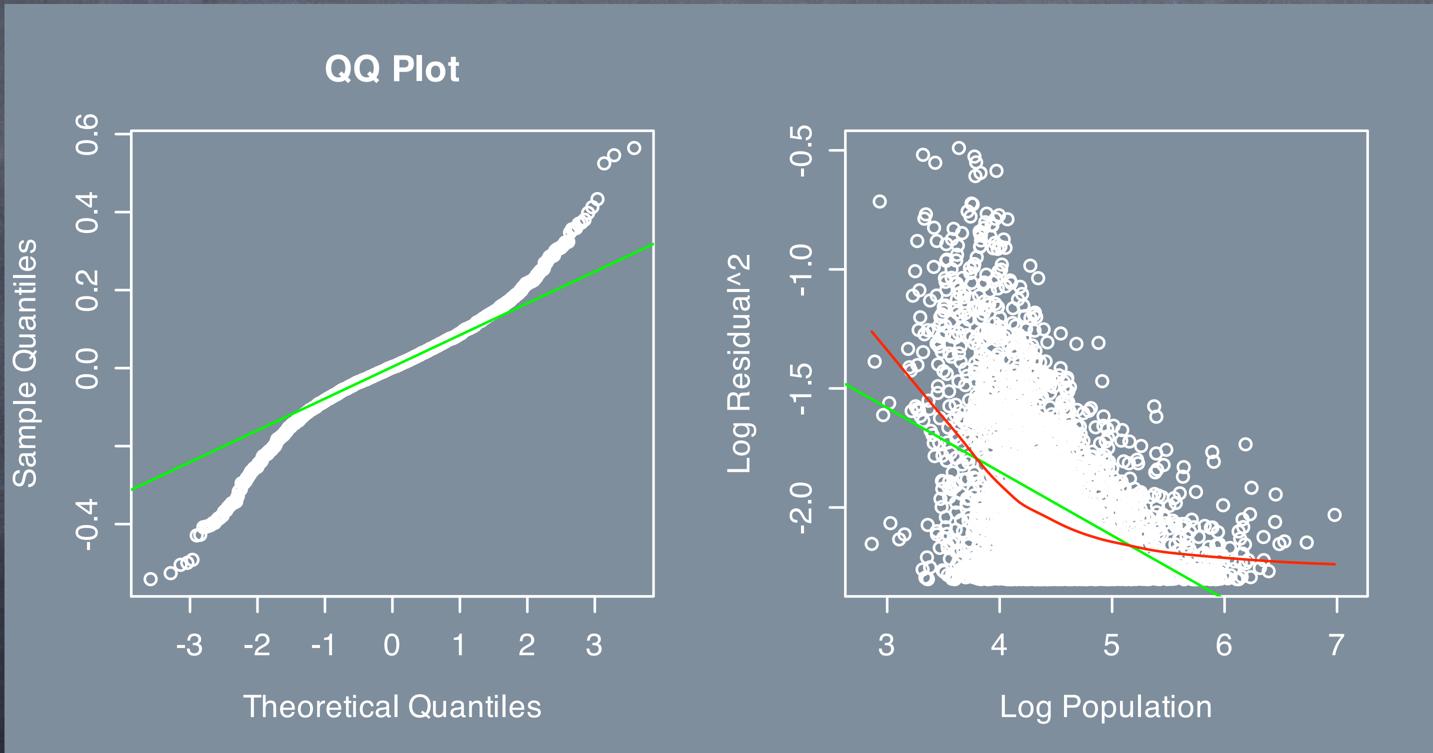
Uh-Oh!



Fat-tails: due to spatial heterogeneity?

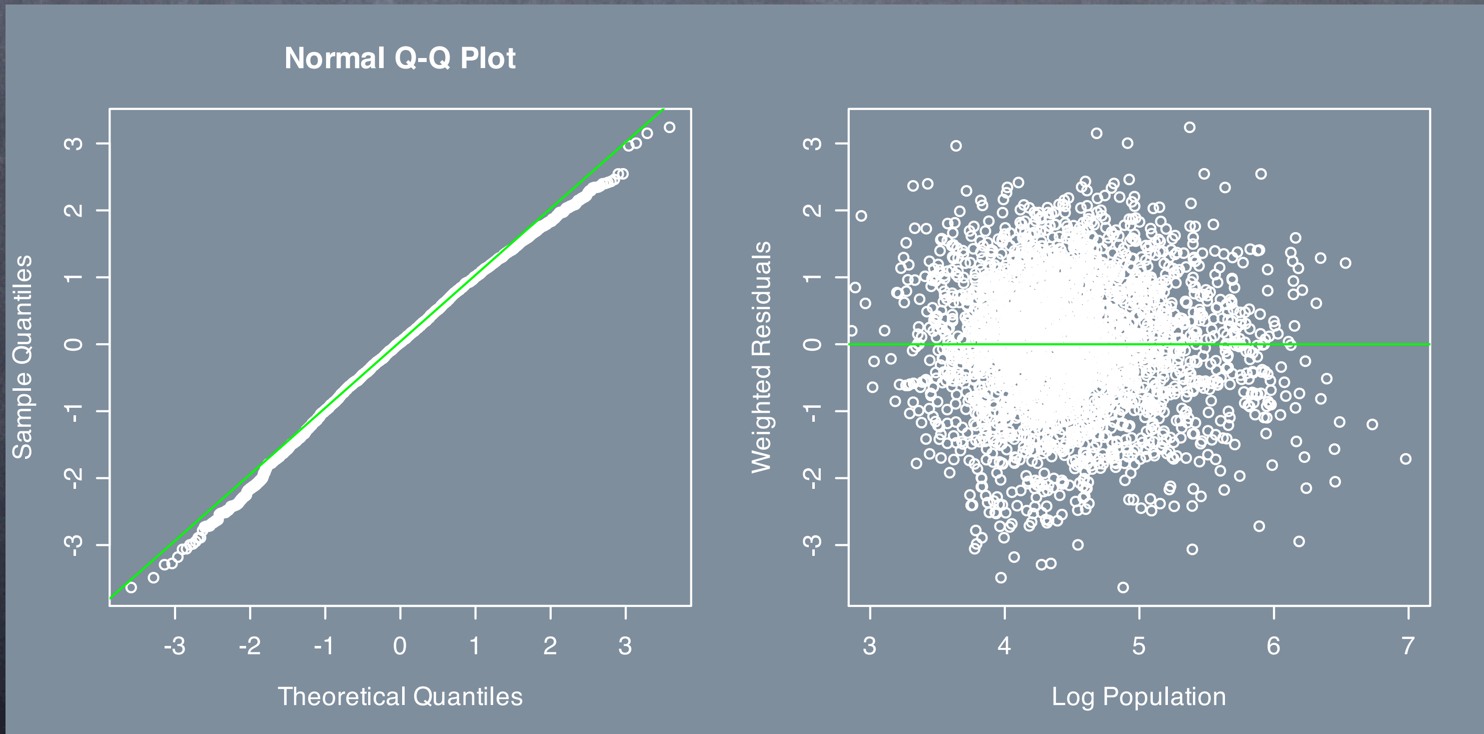
Explanation of Fat Tails

- Residual variance related to pop size
  - $Var(e_t)$ does not fall off with population as rapidly as usual calculation would suggest



Stabilize Variance

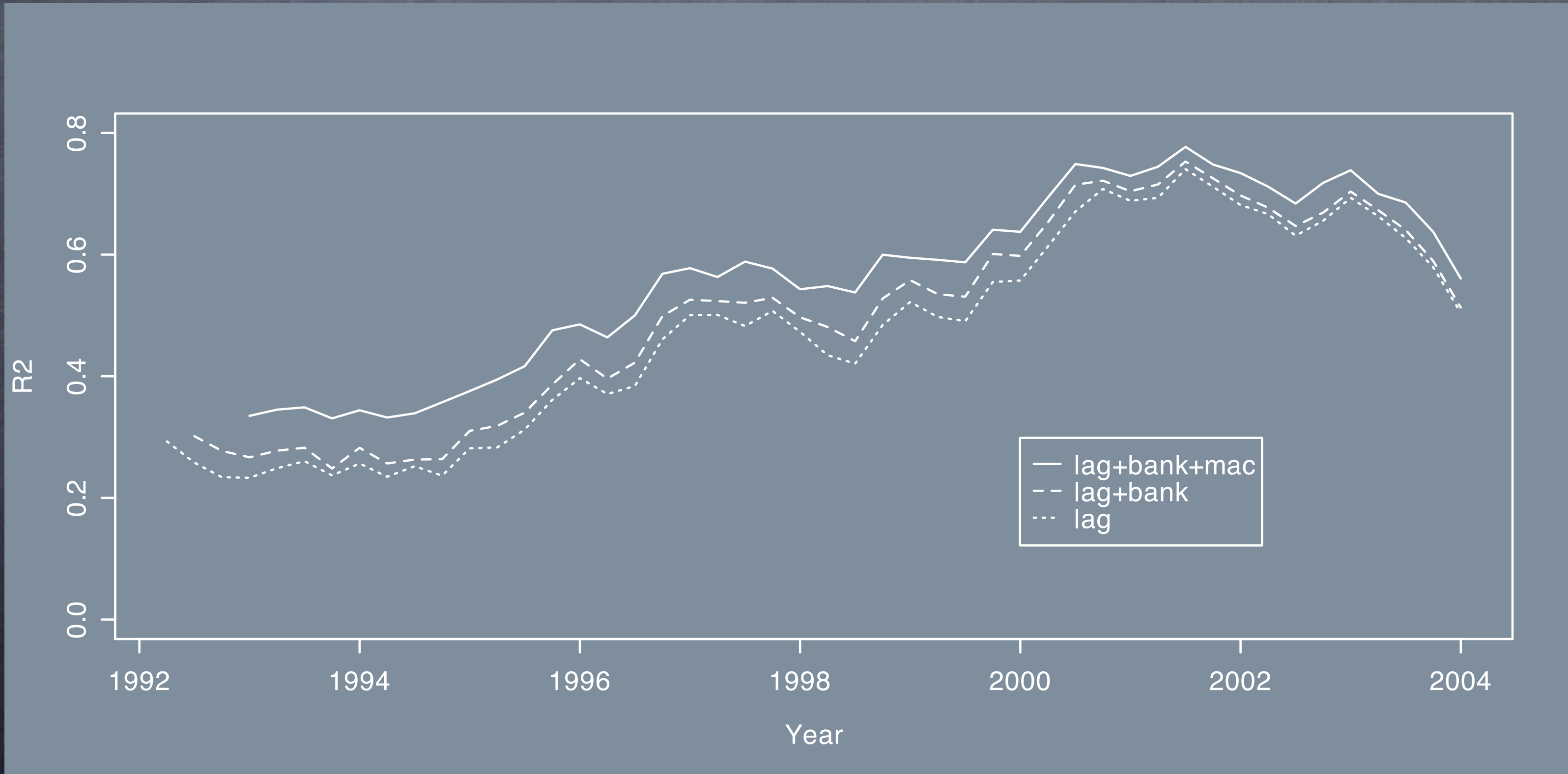
- After weighting by the estimated variance function, residuals are much nicer.
  - Much easier than spatial adjustments



Now that we have a reasonable model, take a look at its properties...

Goodness of Fit

- Fit improves over time
- Macro variables are statistically significant
- Gain worth the effort?



Coefficients over Time

- Estimated coefficient "drifts"
  - Size of effect of lagged endogenous grows
- Less drift when use macro variables

Borrowing Strength?

- I estimated the model M_t with data for a specific quarter, ignoring models in prior quarters M_{t-1}, M_{t-2}, ...
- Extension
  Smooth the models by capturing the dynamics of the drifting estimates and goodness of fit.
- Caveat
  Must capture seasonal effects like that in unemployment rather than smooth over. Smoothing forces similarity.

Questions

1. Is there adequate spatial variation to support modeling credit risk?

Yes. Maybe too much!

2. Do local macro-economic variables add value beyond usual bank information?

Yes. The gain is small (albeit significant) and these variables stabilize the model structure

3. Do models suffer under local economic stress?

Spatial Effects

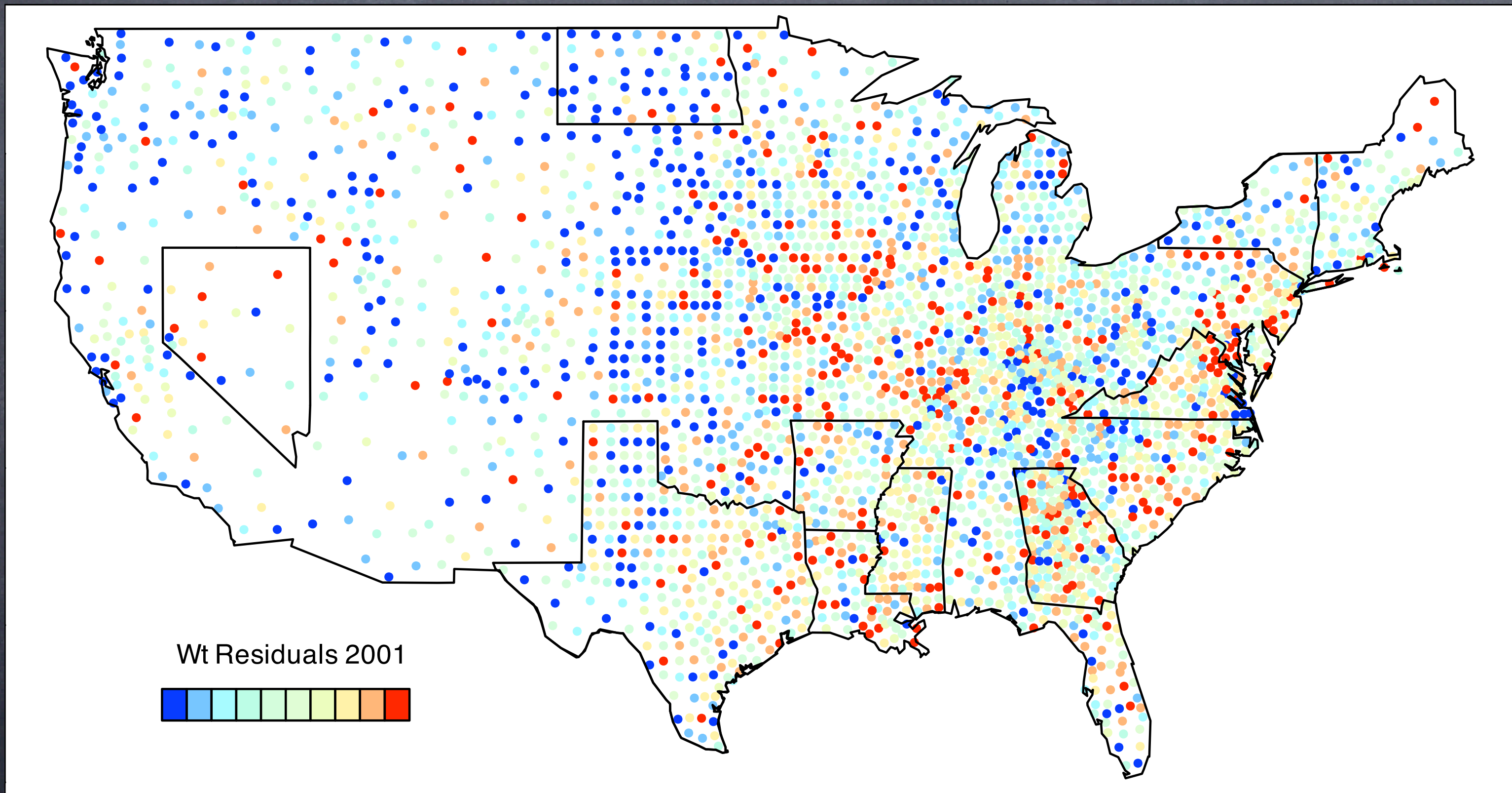
Plan

- Removed the heterogeneity due to population size, but still need to check for spatial dependence among residuals.

- Methods
  - Descriptive: variogram
  - Conceptual: Markov random field

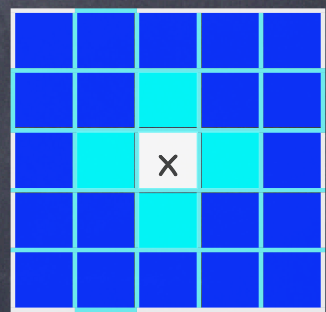
Spatial View of Residuals

Substantial clustering or natural variation?



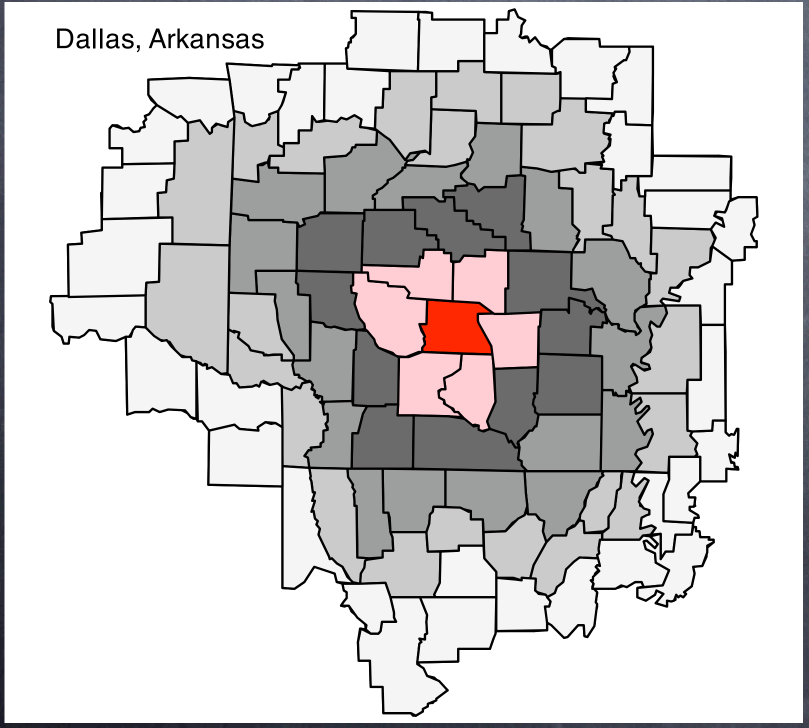
Spatial Models

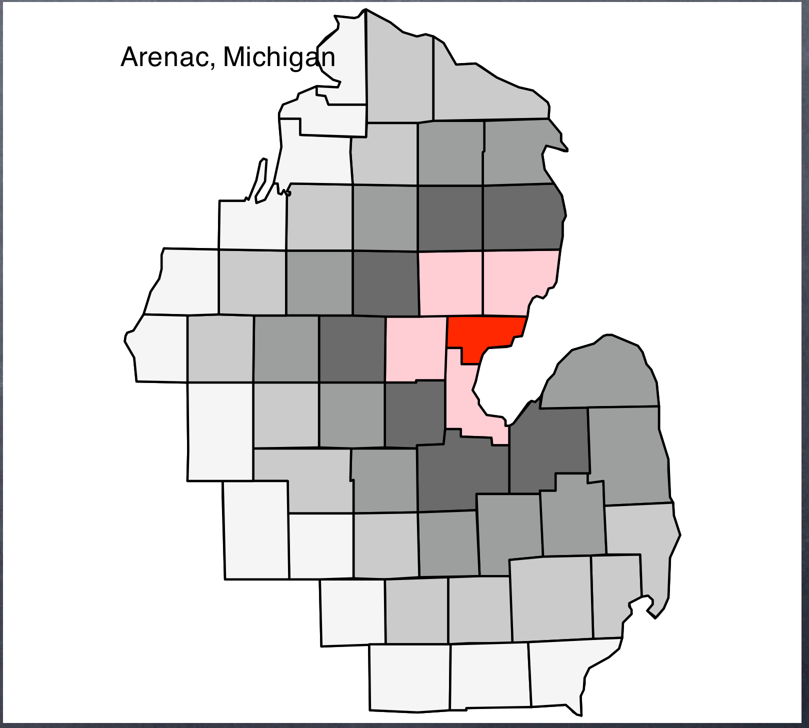
- Broad class
  - More models than those available in time series analysis because the dependence is not naturally "one-sided"
    - e.g., two types of first order autoregressions
    - 2-D plane only partially ordered
- Neighborhood
  - Which observations are "close"?
  - Easily defined in computer graphics because data live on well-defined grid
  - Less clear for map-based geographical units



Neighborhoods

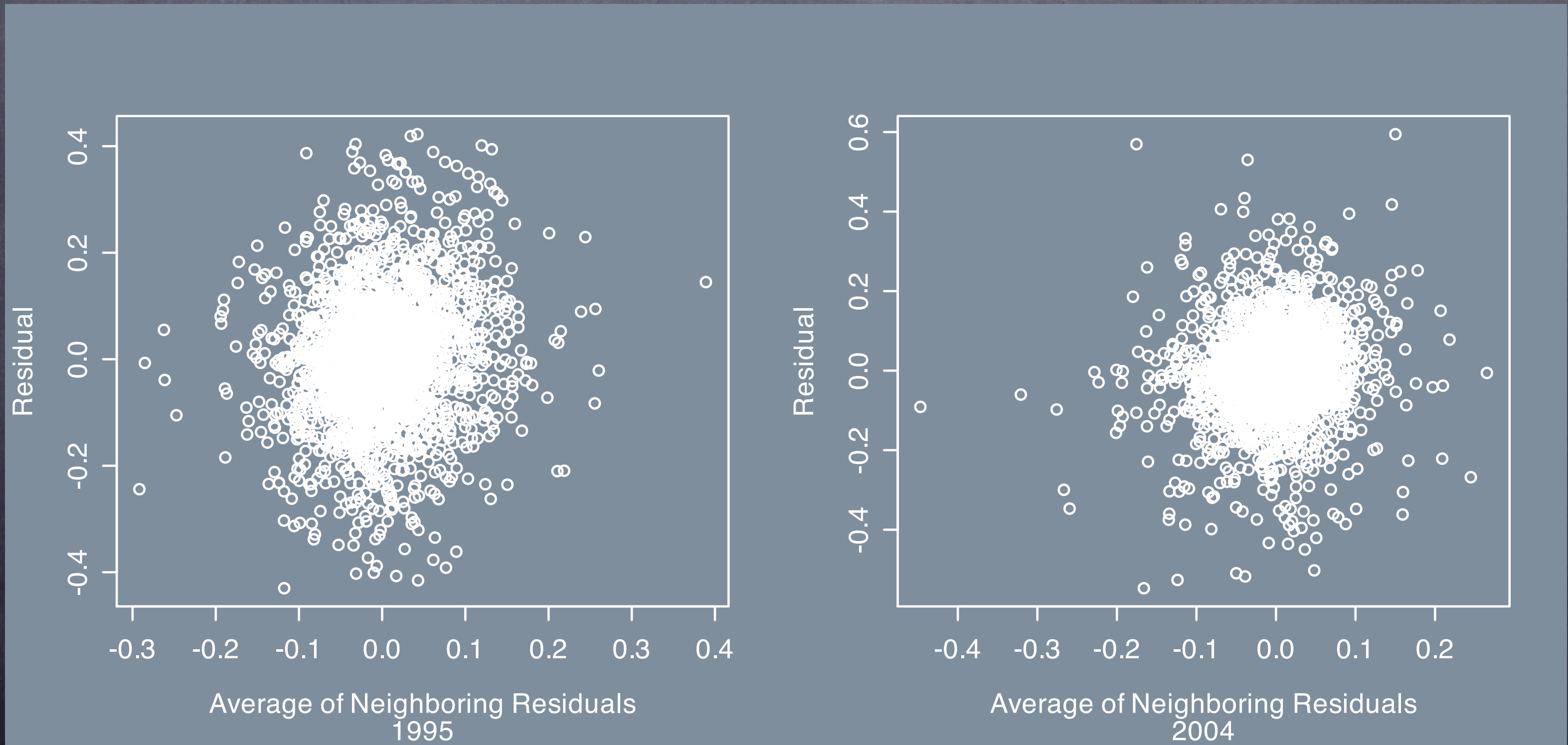
- Model dependence using adjacent counties
- Layers define neighborhoods
  - CAR model assumes P(Y|all) = P(Y|neighbors)
  - Conditional on neighbors, independent of others





Link to Neighbors

- Relationship of residuals to average residual over neighboring counties
- No evidence of dependence (r ≈ 0.15)

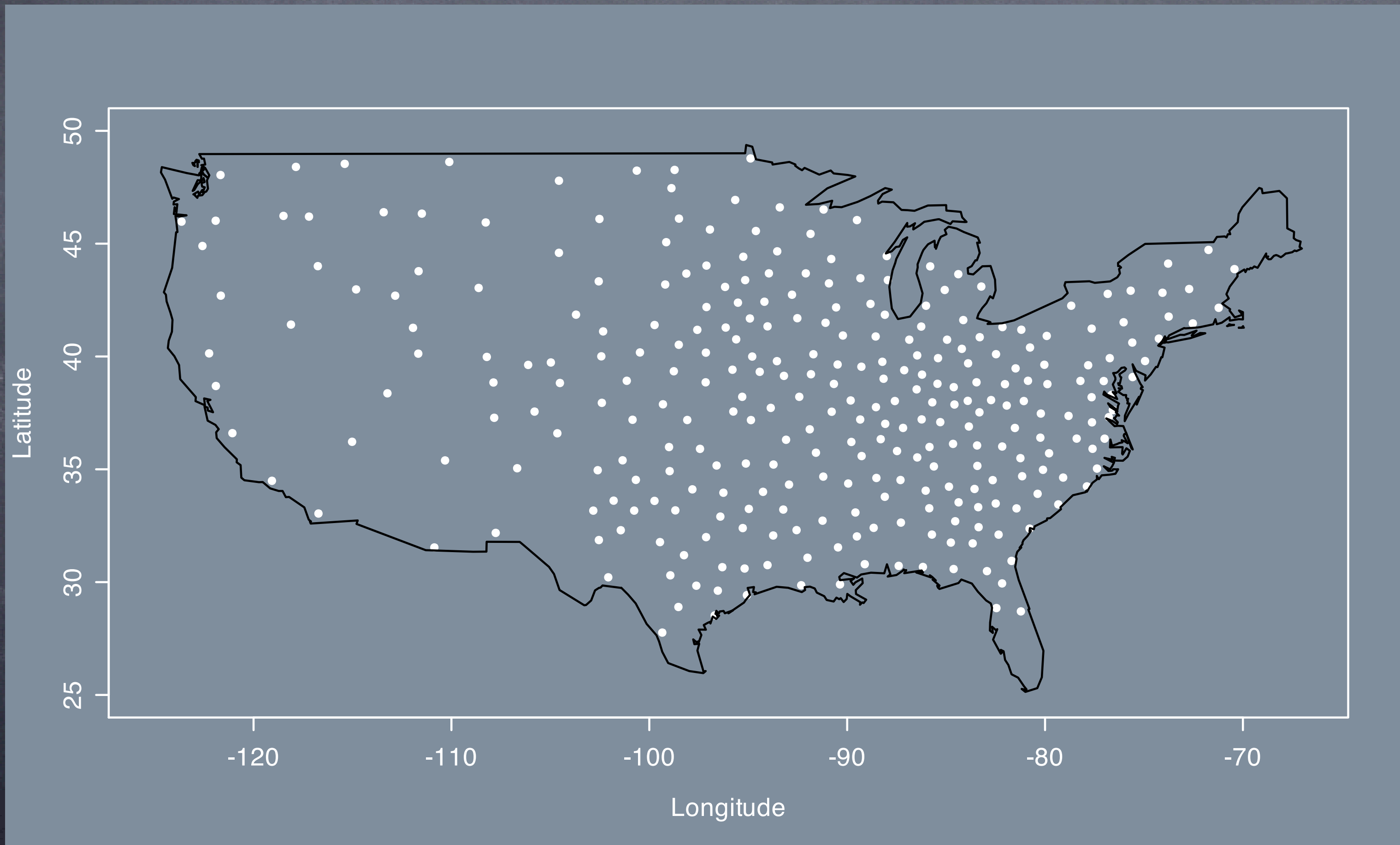


Fit Local Models

- Spatially local fits require "small" samples
  - Want small enough area so that do not lose ability to localize spatial properties
  - Cannot spread over time since we know these models change over time
- Sample 300 spatially separated points around the US
  - None is adjacent to another
  - Use 5 "layers" to define each neighborhood
  - Each neighborhood has 50 to 100+ counties

Local Models

Fit locally defined model at each position.

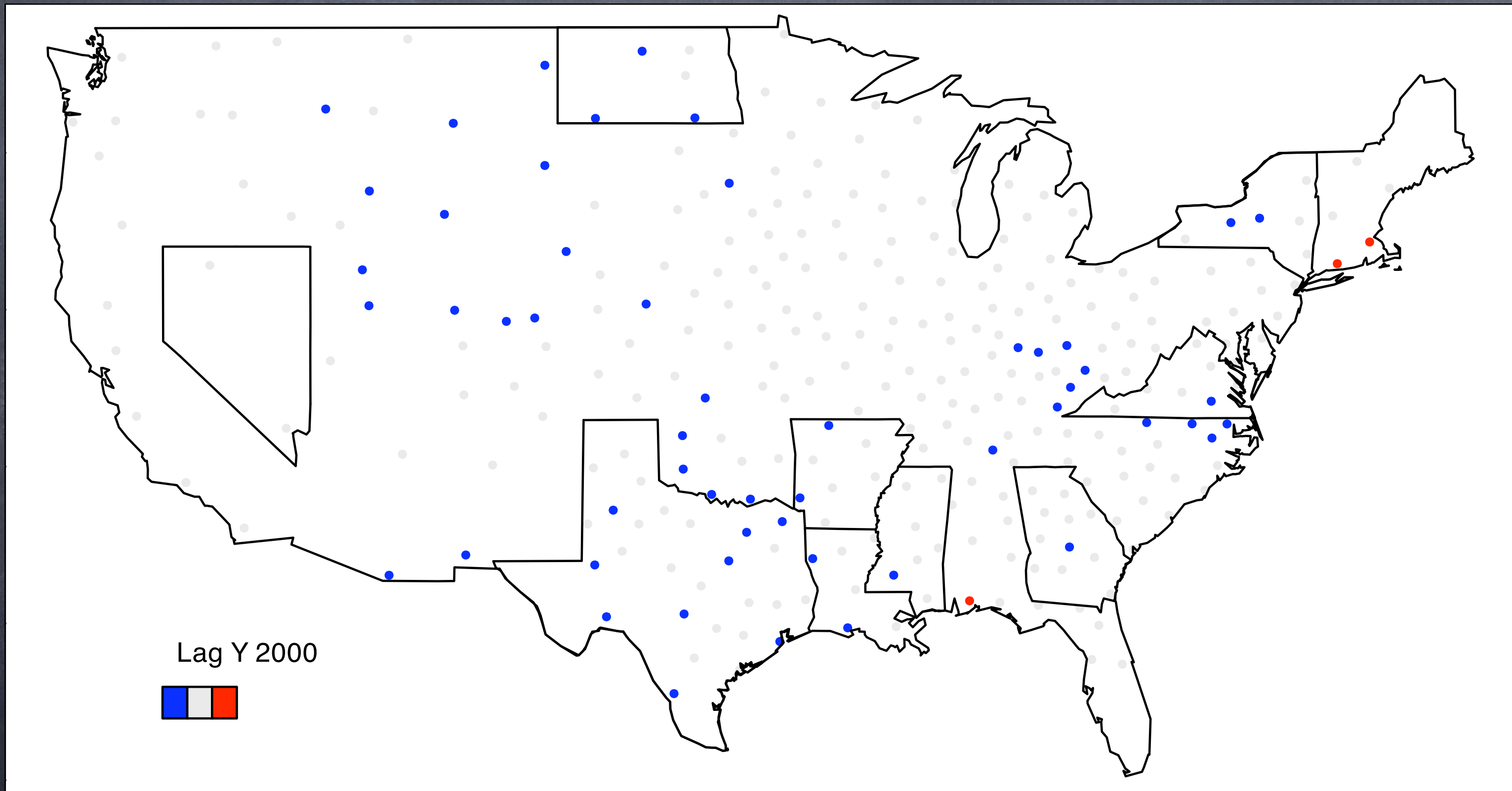


Finding Spatial Deviations

- Estimate deviations from overall estimates
  - Replace Y_t by e_t as the response
- Only interested in meaningful deviations from the overall fit
  - 300 fits offer many chances for accidentally estimating large deviations from overall WLS fit
  - Akin to over-fitting in variable selection
- Solution
  Use an approach that avoids the problem in variable selection: shrinkage
  - Ridge (ie, Bayes estimator under normal prior)
  - Adaptive estimator (Polyshrink)

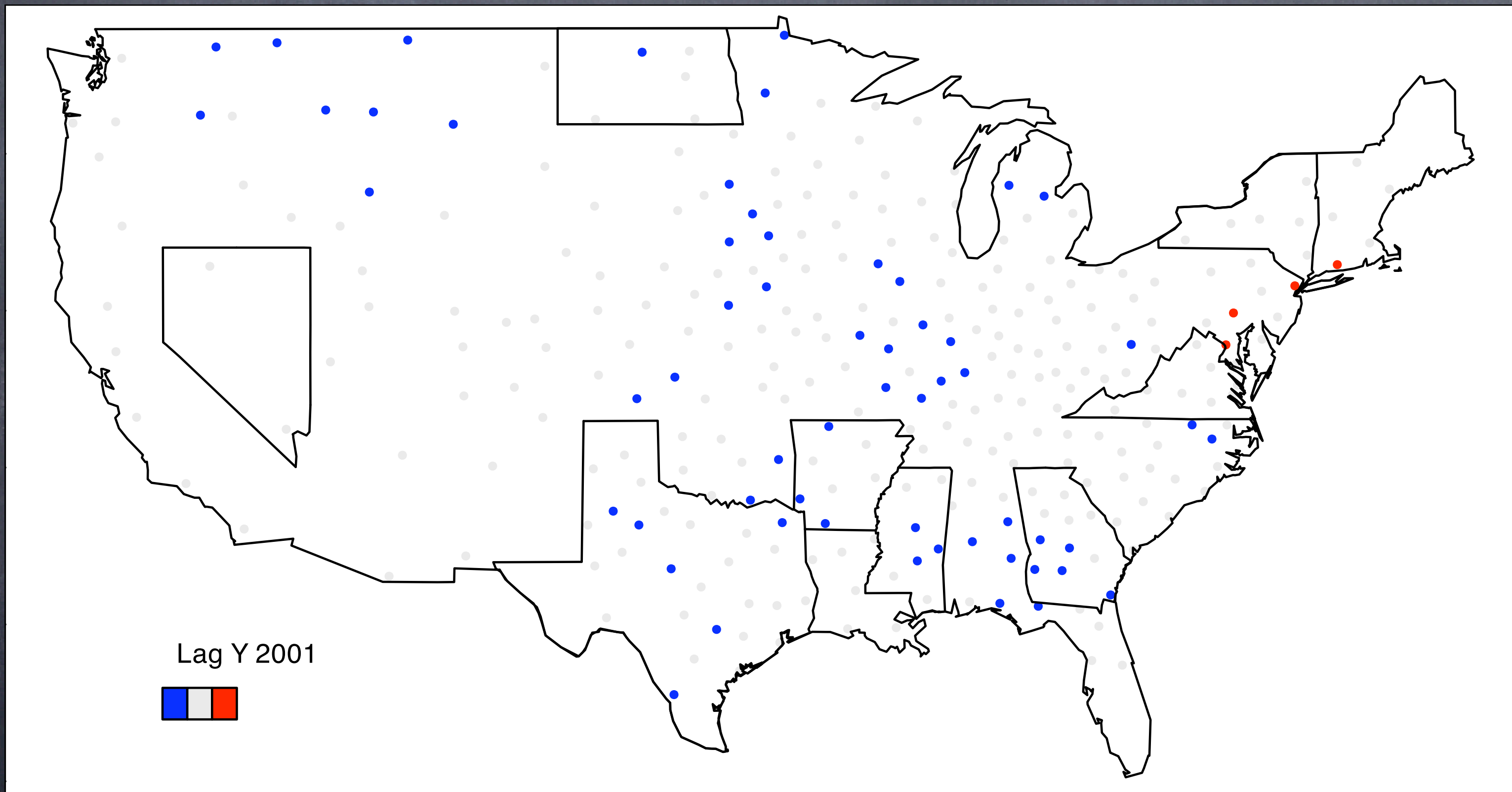
Deviations from Overall

- Polyshrink estimates of the coefficient of the lagged endogenous variable, 2000 Q1



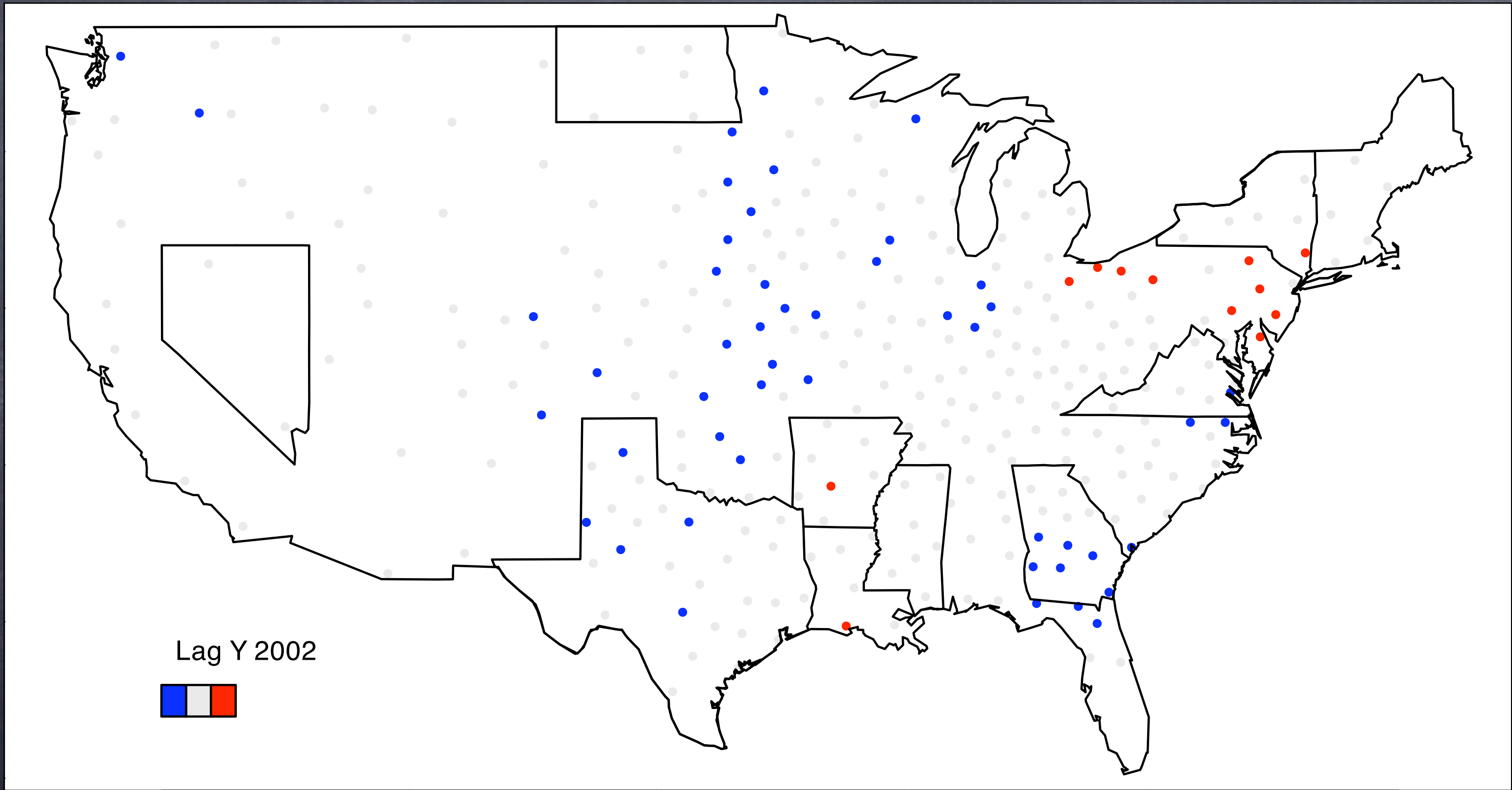
Deviations from Overall

- Polyshrink estimates of the coefficient of the lagged endogenous variable, 2001 Q1



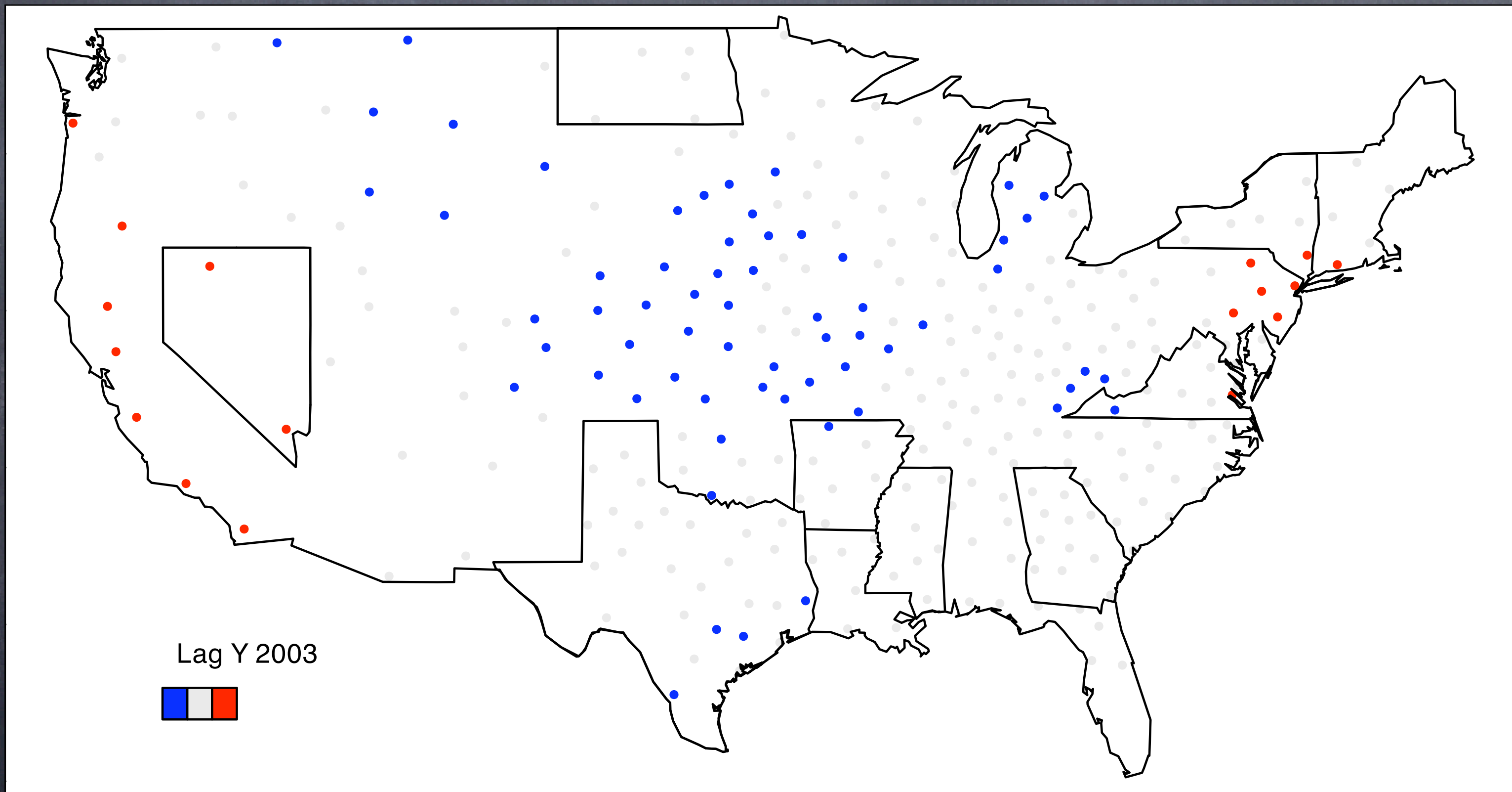
Deviations from Overall

- Polyshrink estimates of the coefficient of the lagged endogenous variable, 2002 Q1



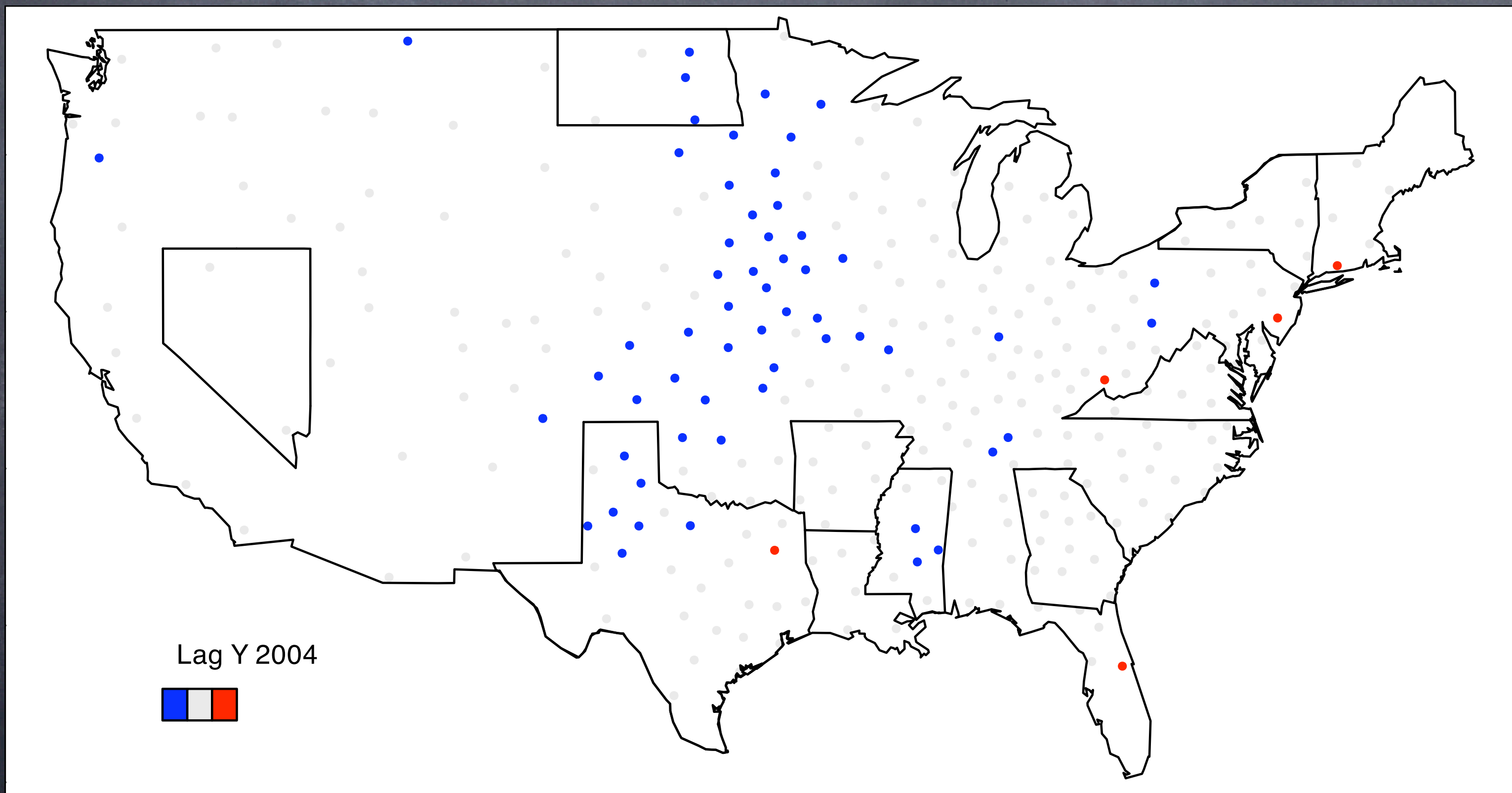
Deviations from Overall

- Polyshrink estimates of the coefficient of the lagged endogenous variable, 2003 Q1



Deviations from Overall

- Polyshrink estimates of the coefficient of the lagged endogenous variable, 2004 Q1



Questions

1. Is there adequate spatial variation to support modeling credit risk?

Yes. Maybe too much!

2. Do local macro-economic variables add value beyond usual bank information?

Yes. The gain is small (albeit significant) and these variables stabilize the model structure

3. Do models suffer under local economic stress?

Models vary spatially as well as over time. Explanations?

What next?

- Incremental
  - Multivariate response (mortgage, installment)
  - More population/demographic information
- Modeling
  - "Global" model that describes the evolution of parameters over time and spatial clusters
  - Combine with micro-level data
- Methods
  - Stress testing

    Comparison of accuracy in regions of changing parameters versus regions of stability

Thanks!

www-stat.wharton.upenn.edu/~stine