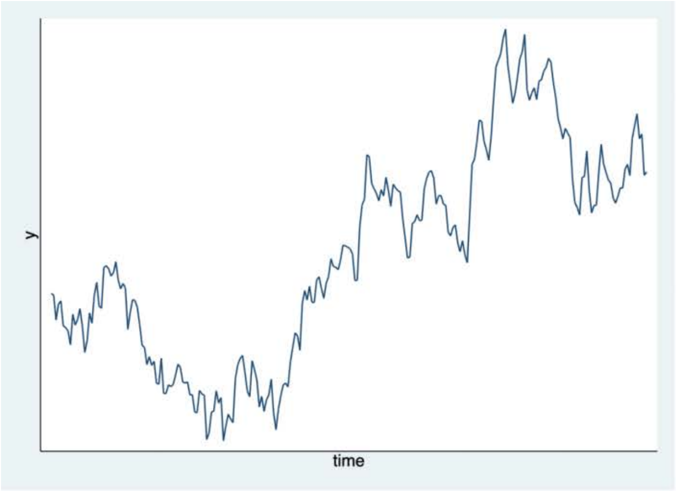
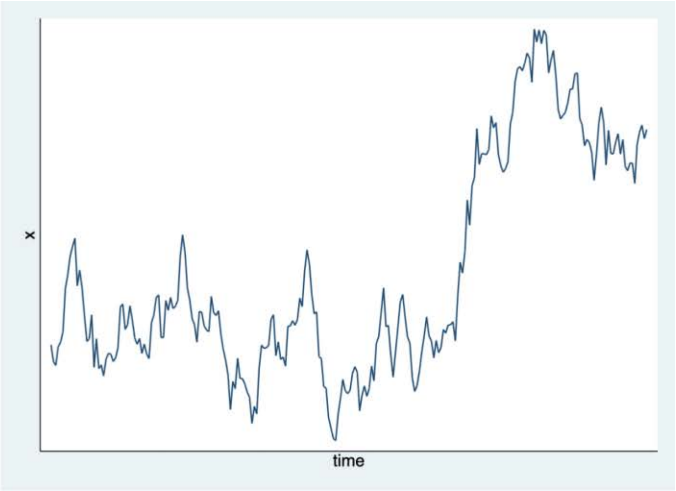
Spatial Unit Roots

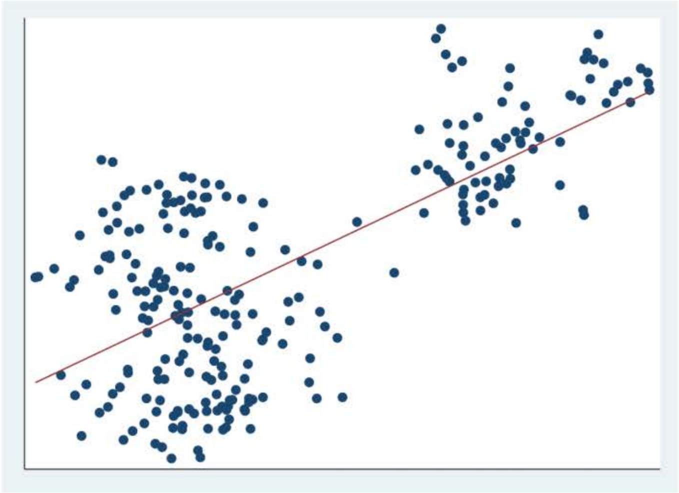
Ulrich K. Müller and Mark W. Watson
Princeton University

February 2023

Time Series Spurious Regression



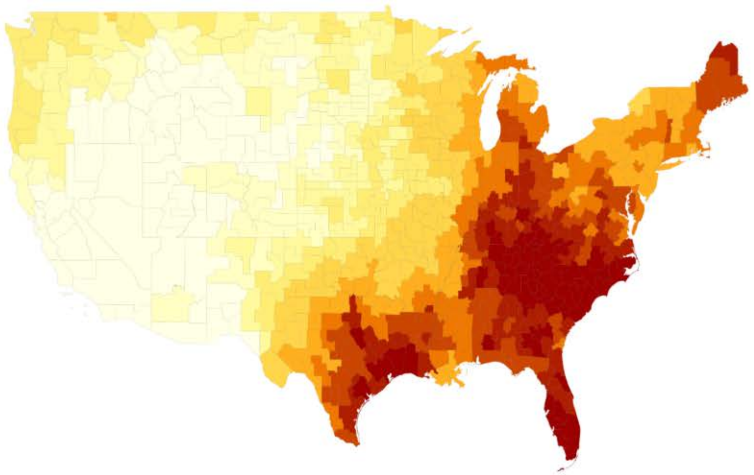


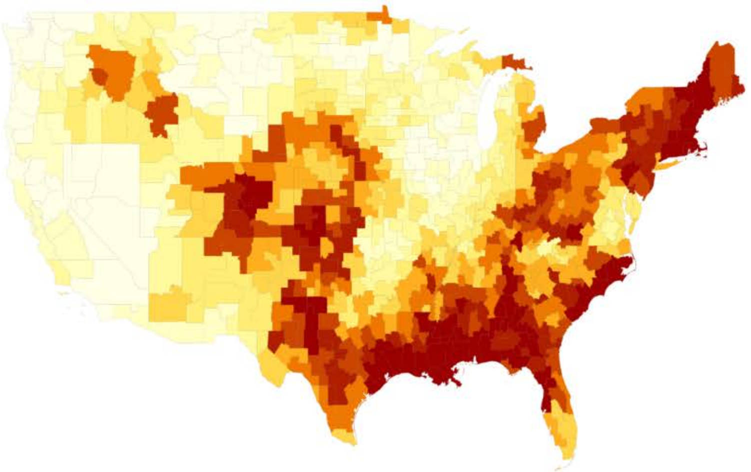


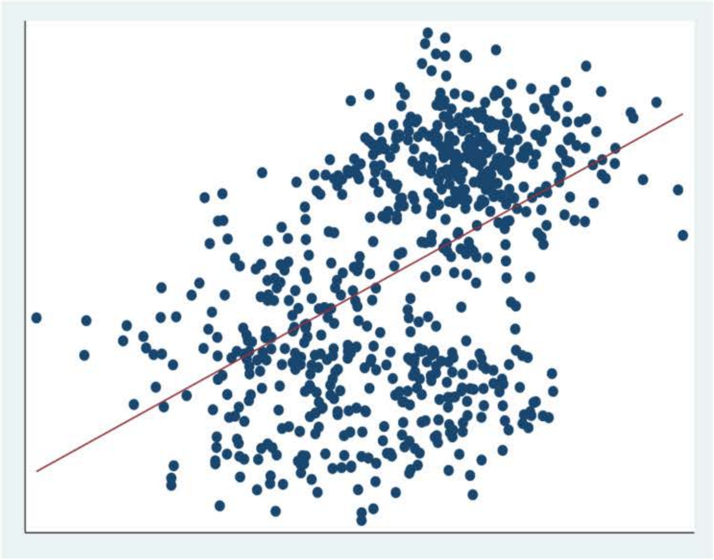
t-stat $= 8.31$

$R^2 = 0.48$

Spatial Spurious Regression



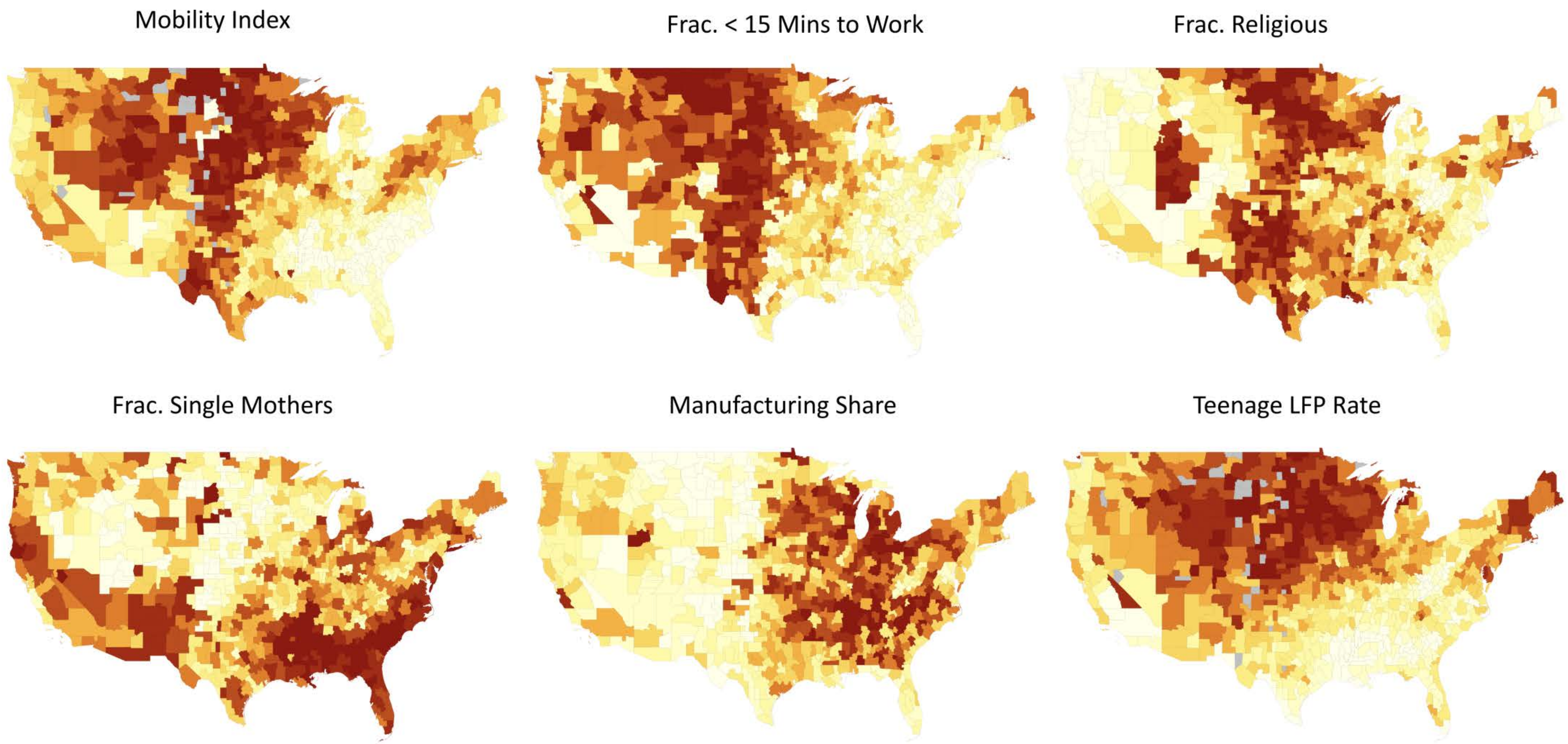




t-stat $= 5.95$

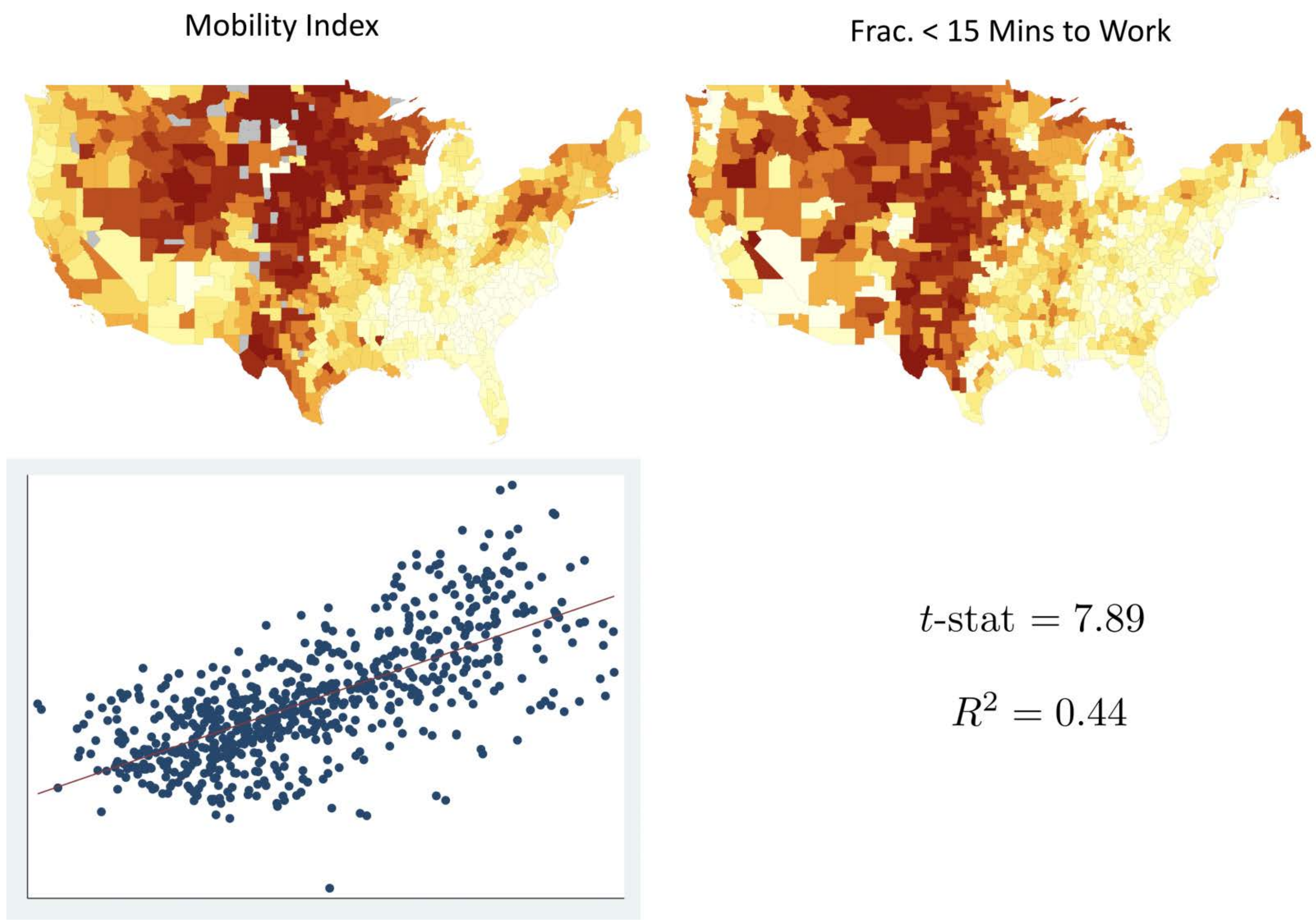
$R^2 = 0.23$

Data in Chetty et al. (2014)



Chetty, Hendren, Kline and Saez (2014): "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States", *QJE*

Regression in Chetty et al. (2014)



t-stat $= 7.89$

$R^2 = 0.44$

Regressions in Chetty et al. (2014)

	R^2	$\hat{\beta}$ [95% CI]
Frac. $<$ 15 Mins to Work	0.48	0.69 [0.55, 0.84]
Frac. Religious	0.28	0.53 [0.36, 0.70]
Frac. Single Mothers	0.59	-0.77 [-0.92,-0.63]
Manufacturing Share	0.09	-0.30 [-0.46,-0.13]
Teenage LFP Rate	0.44	0.66 [0.50, 0.82]

Notes: Pairwise regressions of Mobility Index on different socioeconomic variables, $n = 722$ commuting zones in contiguous U.S., standard errors clustered by state

Outline of Paper

1. Persistent Spatial Processes
   (a) Canonical Unit Root and I(1) Processes
   (b) Functional Central Limit Theorem
   (c) Local-to-Unity Processes
2. Spurious Regression
3. "Isotropic" Regression
4. Inference for Spatial Persistence: Unit Root and Stationarity Tests
5. "First Differencing" Spatial I(1) Processes

Related Literature

1. Literature on unit roots in spatial autoregressive model

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \boldsymbol{\varepsilon} \quad \text{with} \quad \rho = 1 \text{ or } \rho \to 1$$

   Fingleton (1999), Mur and Trivez (2003), Lee and Yu (2009, 2013), Rossi and Lieberman (2022)

2. Spatial HAR/HAC

   Conley (1999), Kelejian and Prucha (2007), Kim and Sun (2011), Bester, Conley, Hansen, and Vogelsang (2016), Sun and Kim (2012), Müller and Watson (2022a, 2022b), etc.

3. Strong Spatial Correlation in Economic Data

   Kelly (2019, 2020, 2022)

4. Large literature on continuous parameter stochastic processes

   Lévy (1948), Matérn (1986), Ivanov and Leonenko (1989), Adler (2010), etc.

Persistent Spatial Processes: Set-up

- Data y_l observed at locations $s_l \in \mathcal{S}_n \subset \mathbb{R}^d$ generated via

$$y_l = Y(s_l), \quad l = 1, \ldots, n$$

  with $Y(\cdot)$ a continuous parameter process on $\mathcal{S}_n$

- For fixed compact $\mathcal{S}^0 \subset \mathbb{R}^d$ and non-decreasing sequence λ_n,

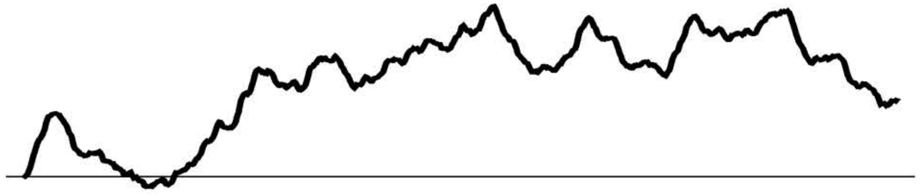
$$\mathcal{S}_n = \lambda_n \mathcal{S}^0 = \{s : \lambda_n^{-1} s \in \mathcal{S}^0\}$$

- $s_l = s_{n,l}$ nonstochastic, and $\{\lambda_n^{-1} s_l\}_{l=1}^n \subset \mathcal{S}^0$ has empirical distribution $G_n \Rightarrow G$ with G absolutely continuous on $\mathcal{S}^0$

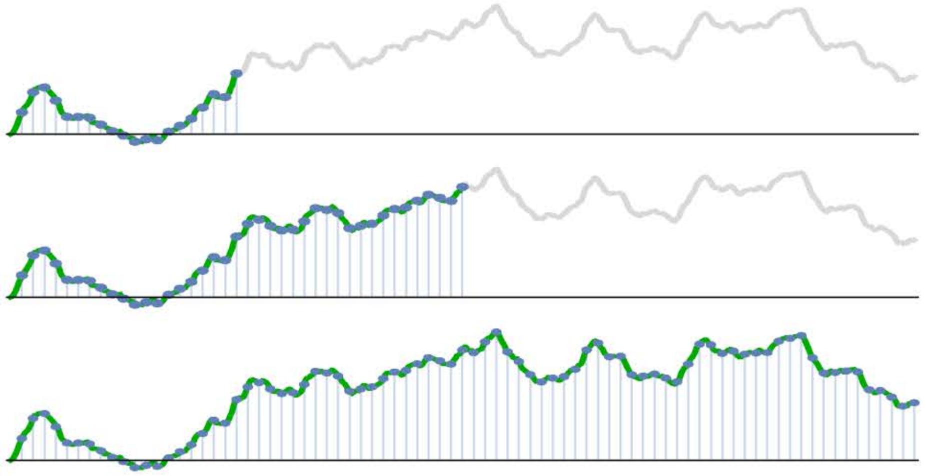
- Standard time series: $s_l = l$, $\lambda_n = n$, $\{\lambda_n^{-1} s_l\}_{l=1}^n = \{l/n\}_{l=1}^n \subset [0, 1]$ and $G_n \Rightarrow U_{[0,1]}$

Standard Time Series

Y



$Y(s_l)$



Spatial Generalization of Wiener Process

- Canonical persistent process for $d = 1$: Wiener process $W(s)$ with $\mathbb{E}[W(s)W(r)] = \min(s, r)$

- Two potential generalizations of W to $d > 1$:

1. Brownian sheet: For $s \geq 0$, $H(s) = \int_{\mathbb{R}^d} \mathbf{1}[0 \leq r \leq s] dW(r)$, so

$$\mathbb{E}[H(s)H(r)] = \prod_{i=1}^{d} \min(s_i, r_i)$$

2. Lévy (1948) Brownian motion: $L(\cdot)$ mean-zero Gaussian with

$$\mathbb{E}[L(s)L(r)] = \frac{1}{2}(|s| + |r| - |s - r|)$$

and $|a| = \sqrt{a'a}$

Sample Realizations for $d = 2$

Lévy Brownian Motion

Brownian Sheet

Properties of Lévy Brownian Motion

- Isotropy: From $\mathbb{E}[L(s)L(r)] = \frac{1}{2}(|s| + |r| - |s - r|)$, $L(Os) \sim L(s)$ for all rotation matrices O

- $L(s) \sim W(s)$ for $d = 1$: $\frac{1}{2}(|s| + |r| - |s - r|) = \min(s, r)$ for $s, r \geq 0$

- Evaluation along line yields Wiener process: $L(a + bs) - L(a) \sim W(s)$ for $s \in \mathbb{R}$ and $a, b \in \mathbb{R}^d$ with $|b| = 1$

- Self-similarity: $L(\lambda s) \sim \lambda^{1/2} L(s)$

$\Rightarrow$ Use $Y = L$ as "canonical unit root process"

In standard time series case with $s_l = l$, yields $y_l = Y(s_l) \sim W(l)$, i.e. Gaussian random walk

Two Representations of Lévy Brownian Motion

1. Karhunen-Loève representation: On $\mathcal{S}^0$

$$L(\cdot) = \sum_{j=1}^{\infty} \xi_j \varphi_j(\cdot), \quad \xi_j \sim id\mathcal{N}(0, \nu_j)$$

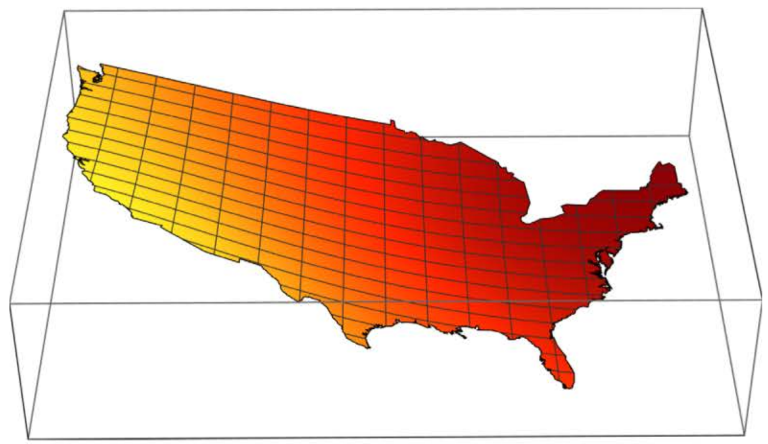
where (ν_j, φ_j) are eigenvalue/eigenfunction pairs of covariance kernel

$$\mathbb{E}[L(s)L(r)] = \frac{1}{2}\left(|s| + |r| - |s-r|\right) = \sum_{j=1}^{\infty} \nu_j \varphi_j(s)\varphi_j(r)$$

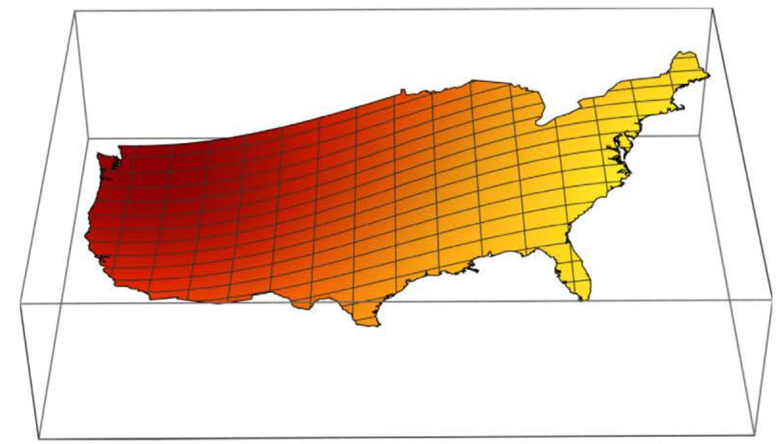
with $\varphi_j : \mathcal{S}^0 \mapsto \mathbb{R}$ satisfying $\int \varphi_i(s)\varphi_j(s)dG(s) = \mathbf{1}[i=j]$

Eigenfunctions for U.S. and Uniform G

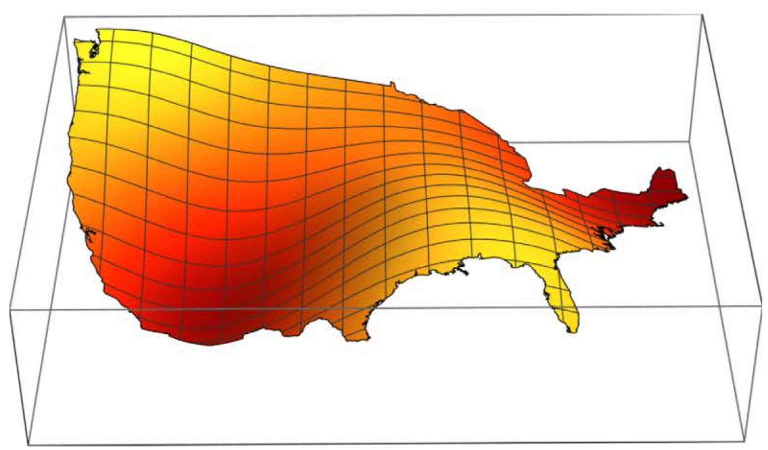
1^{st} eigenfunction



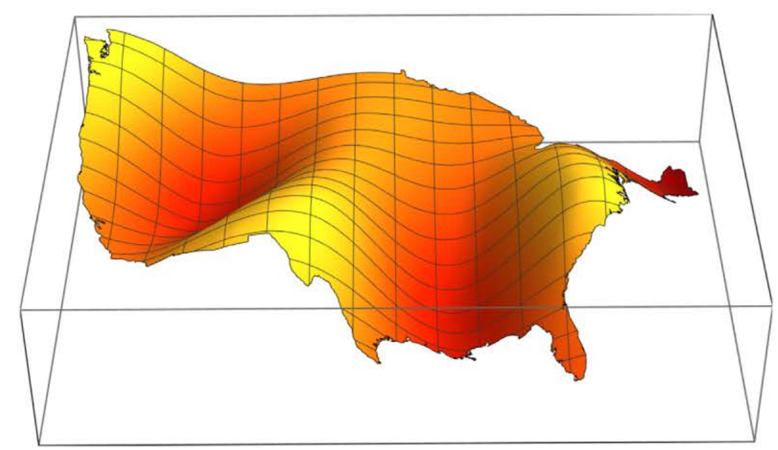
2^{nd} eigenfunction



5^{th} eigenfunction



10^{th} eigenfunction



Two Representations of Lévy Brownian Motion

2. Moving average representation: From Lindstrom (1993)

$$L(s) = \int_{\mathbb{R}^d} h(r,s)\,dW(r) = \begin{cases} \int_0^s dW(r) \text{ for } d = 1 \\ \kappa_d \int_{\mathbb{R}^d} (|s-r|^{(1-d)/2} - |r|^{(1-d)/2})\,dW(r) \text{ for } d > 1 \end{cases}$$

Spatial I(1) Processes

- Motivation: Richer class of strongly persistent processes with flexible high frequency properties

- Replace dW in $L(s) = \int_{\mathbb{R}^d} h(r,s)dW(r)$ by weakly dependent process B on $\mathbb{R}^d$, so

$$
\begin{aligned}
Y(s) &= \int_{\mathbb{R}^d} h(r,s)B(r)dr \\
&= \begin{cases} \int_0^s B(r)dr \text{ for } d = 1 \\ \kappa_d \int_{\mathbb{R}^d} (|s-r|^{(1-d)/2} - |r|^{(1-d)/2}) B(r)dr \text{ for } d > 1 \end{cases}
\end{aligned}
$$

- Time series case: $y_l = Y(l) = \int_0^l B(r)dr = \sum_{t=1}^l u_t$ with $u_t = \int_{t-1}^t B(r)dr \sim I(0)$

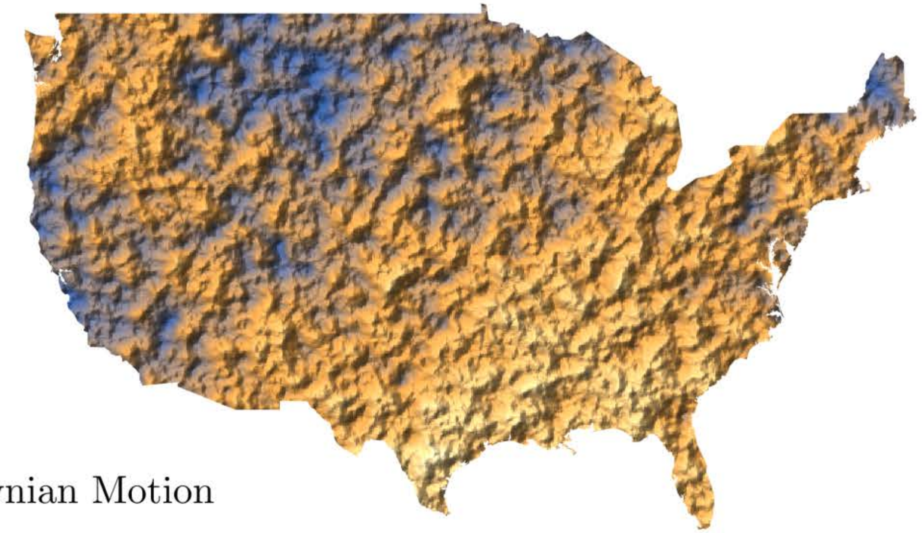
- For $d > 1$, $\int_{\mathbb{R}^d} |h(r,s)|dr$ does not exist, but $\int_{\mathbb{R}^d} h(r,s)^2 dr < \infty$

  Lemma: Under regularity conditions on $B(\cdot)$, $Y(\cdot)$ is a.s. continuous on $\mathcal{S}_n \subset \mathbb{R}^d$

Spatial I(1) Processes

Y with $B = B_1$

Y with $B = B_2$



Lévy Brownian Motion

Functional Central Limit Theorem

- Recall that $y_l = Y(s_l)$ with $s_l \in \mathcal{S}_n = \lambda_n \mathcal{S}^0$. On $\mathcal{S}^0$, define $Y_n^0(s) = \lambda_n^{-1/2} Y(\lambda_n^{-1} s)$

  (Cf. normalization for time series FCLT: $n^{-1/2} \sum_{t=1}^{\lfloor sn \rfloor} u_t \Rightarrow \omega W(s)$)

- **Theorem:** Let B be stationary with $\mathbb{E}[B(s)B(r)] = \sigma_B(s-r)$. Under suitable weak dependence conditions for B

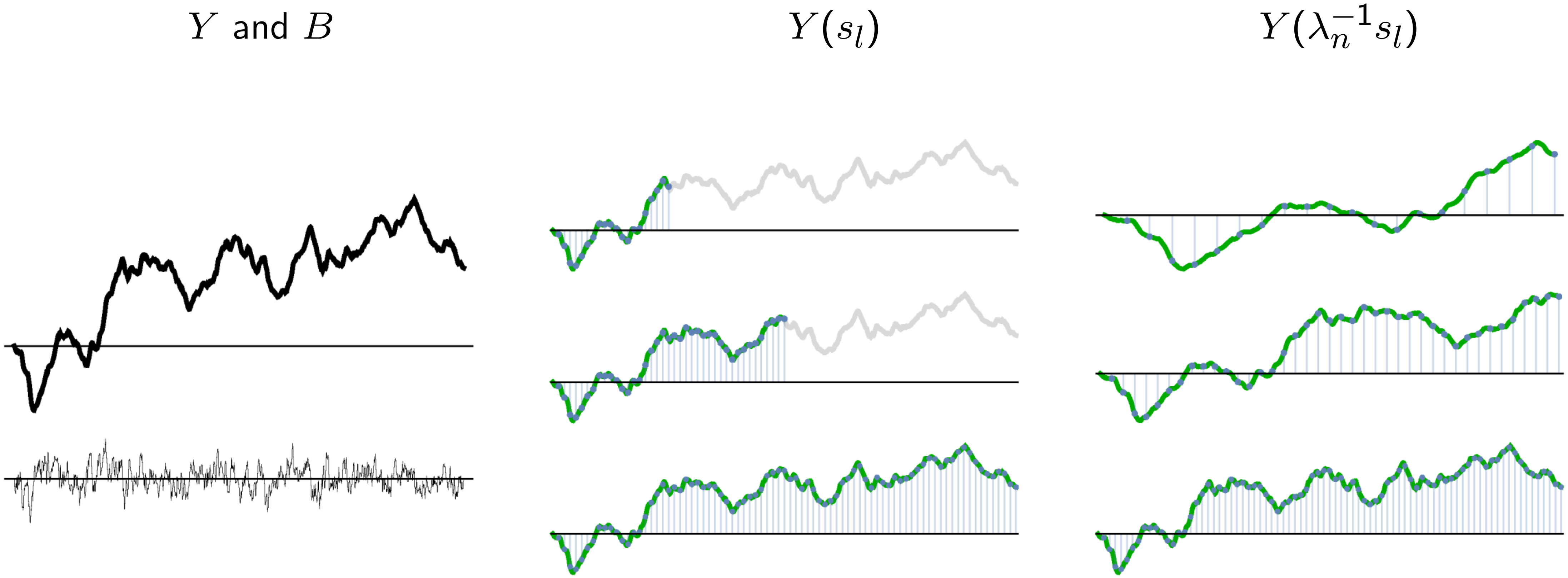
$$Y_n^0(\cdot) \Rightarrow \omega L(\cdot) \qquad \omega^2 = \int_{\mathbb{R}^d} \sigma_B(r) dr$$

  as $\lambda_n \to \infty$ as a process on $\mathcal{S}^0$.

- $\lambda_n \to \infty$: increasing domain (over Y) asymptotics as $n \to \infty$

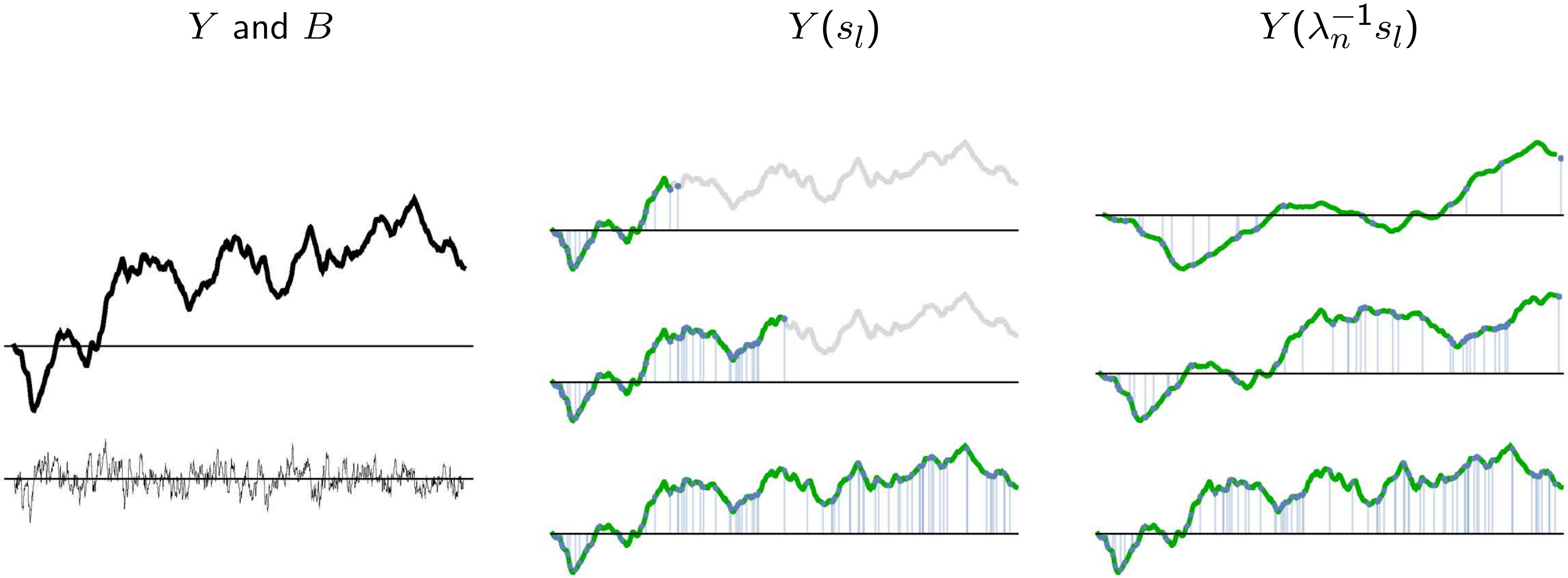
- Typical distance of neighboring locations is $O(n^{-1/d})$. Difference $Y_n^0(s + n^{-1/d} r) - Y_n^0(s)$ *does not* generically become Gaussian (cf. $\sum_{t=1}^{s+1} u_t - \sum_{t=1}^{s} u_t = u_{s+1}$)

FCLT Illustration: Standard Time Series



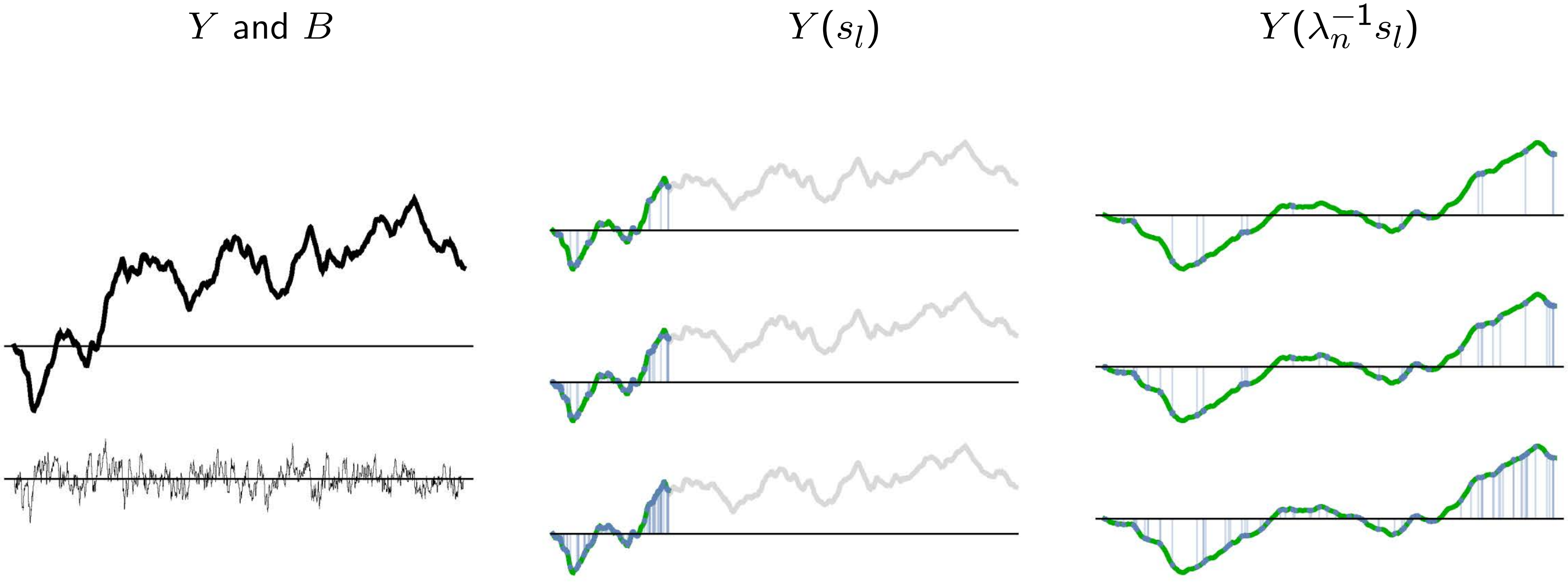
$\Rightarrow$ Increasing domain, FCLT holds

FCLT Illustration: $d = 1$ with Irregular Sampling



$\Rightarrow$ Increasing domain, FCLT holds

$d = 1$ with Pure Infill Sampling ($\lambda_n =$ const.)



$\Rightarrow$ Fixed domain, FCLT does not hold

FCLT Illustration: Increasing Domain Asymptotics

Generalization to Local-to-Unity Processes

- Consider process $J_c(\cdot)$ on $\mathbb{R}^d$ with covariance kernel $\mathbb{E}[J_c(s)J_c(r)] = e^{-c|s-r|}/(2c)$
  - weak mean reversion as measured by $c > 0$
  - $J_c(a + bs)$ scalar Ornstein-Uhlenbeck process for any $a, b \in \mathbb{R}^d$ with $|b| = 1$
  - $J_c(\cdot) - J_c(0) \Rightarrow L(\cdot)$ as $c \to 0$

- Moving average representation: $J_c(s) = \int_{\mathbb{R}^d} h_c(r, s)dW(r)$

- General spatial local-to-unity process: Let $Y_c(s) = \int_{\mathbb{R}^d} h_c(r, s)B(r)dr$ and set

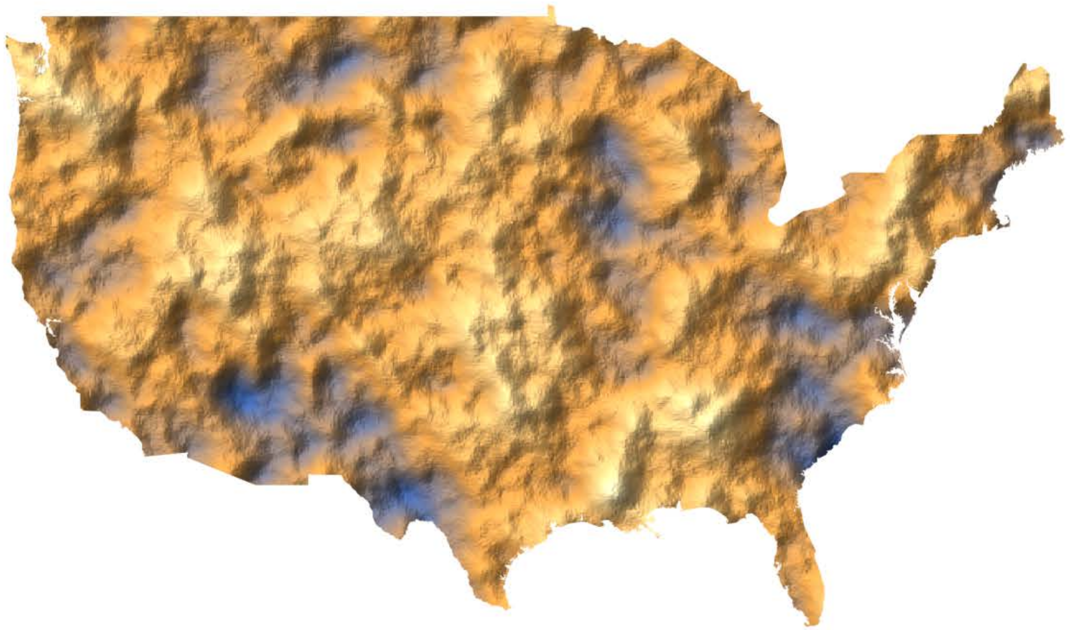
$$Y_n^0(s) = \lambda_n^{-1/2} Y_{c/\lambda_n}(\lambda_n s), \quad s \in \mathcal{S}^0$$

- Under same assumptions on B as for I(1) process: $Y_n^0(\cdot) \Rightarrow \omega J_c(\cdot)$ on $\mathcal{S}^0$

Realization of J_c and LTU Process

J_c

LTU



Spatial Spurious Regression

- Consider regression

$$y_l = \alpha + x_l'\beta + u_l$$

where $(y_l, x_l) = (Y(s_l), X(s_l)) \in \mathbb{R}^{p+1}$ and $(Y(\cdot), X(\cdot)) \sim$ independent spatial I(1)

- **Theorem**:
  - OLS coefficients, R^2 etc. behave asymptotically like continuous parameter regressions involving $p+1$ independent Lévy Brownian motions, weighted by G

    Exactly same limit as in Phillips (1986) if $d = 1$ and G uniform
  - F^{HAC}-statistics diverge under null hypothesis for any kernel HAC estimator with bandwidth that induces consistent variance estimates under weak dependence

    $\Rightarrow$ also if $(y_l, x_l) \sim$ independent local-to-unity processes

Isotropic Regression

- In time series, learn about persistence by regressing Δy_l on y_{l-1}

- Isotropic difference: $y_l^* = \sum_{\ell \neq l} \kappa(|s_\ell - s_l|)(y_\ell - y_l)$, isotropic regression: regress y_l^* on y_l

- Special case in time series: $y_l^* = \frac{1}{2}(y_{l+1} - y_l) + \frac{1}{2}(y_{l-1} - y_l) = \frac{1}{2}(y_{l+1} + y_{l-1}) - y_l$

  Pantula, Gonzalez-Farias and Fuller (1994): With $\Delta y_l \sim iid(0,1)$,

$$n^{-1} \sum_{t=2}^{n-1} y_t^* y_t \xrightarrow{p} -1/2, \quad \text{so} \quad \frac{\sum_{t=2}^{n-1} y_t^* y_t}{n^{-1} \sum_{t=2}^{n-1} y_t^2} \Rightarrow \frac{-1/2}{\int_0^1 W(r)^2 dr}$$

- We establish general "constant limiting numerator" result for spatial isotropic regression coefficient under $b \to 0$ limits after "fixed-b kernel κ" $n \to \infty$ limits

Unit Root and Stationarity Tests

- Let $\mathbf{\Sigma}_L$ be $n \times n$ covariance matrix of $L(s_l)$, $l = 1, \ldots n$, $\mathbf{M} = \mathbf{I}_n - \mathbf{e}\mathbf{e}'/n$, $\mathbf{e} = (1, \ldots, 1)'$ and consider spectral decomposition

$$\mathbf{M}\mathbf{\Sigma}_L\mathbf{M} = \mathbf{R}\mathbf{D}\mathbf{R}'$$

- By construction, eigenvectors $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_{n-1})$ satisfy $\mathbf{r}_i'\mathbf{e} = 0$ and $\mathbf{r}_i'\mathbf{r}_j = \mathbf{1}[i = j]$

- Let $\mathbf{R}_q = (\mathbf{r}_1, \ldots, \mathbf{r}_q)$ associated with largest q eigenvalues. Treat $q \times 1$ vector

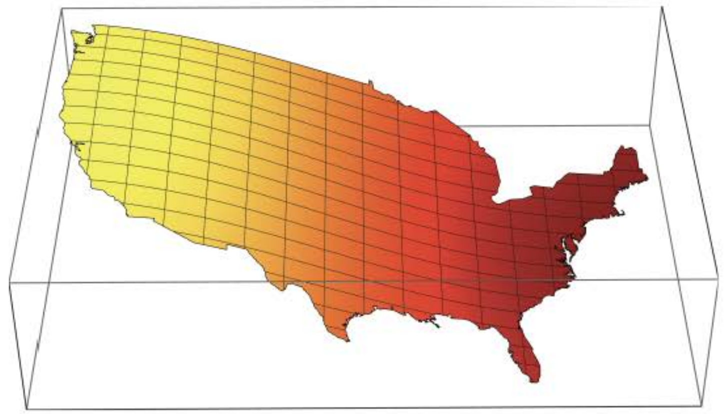
$$\mathbf{Z} = \mathbf{R}_q'\mathbf{y} = \mathbf{R}_q'(\mathbf{y} - \mu\mathbf{e})$$

  as effective observations for inference (cf. Müller and Watson (2008))

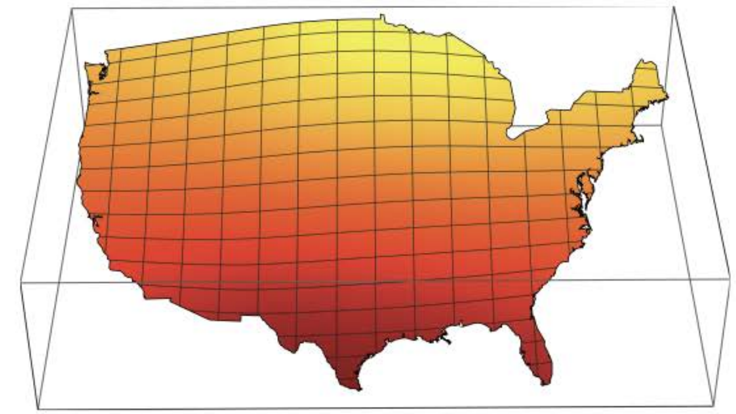
- **Lemma**: $\mathbf{R}_q$ converges in suitable sense to eigenfunctions of demeaned covariance kernel of L

  $\Rightarrow$ Adapts previous result by Rosasco, Belkin and Vito (2010) and Müller and Watson (2022)

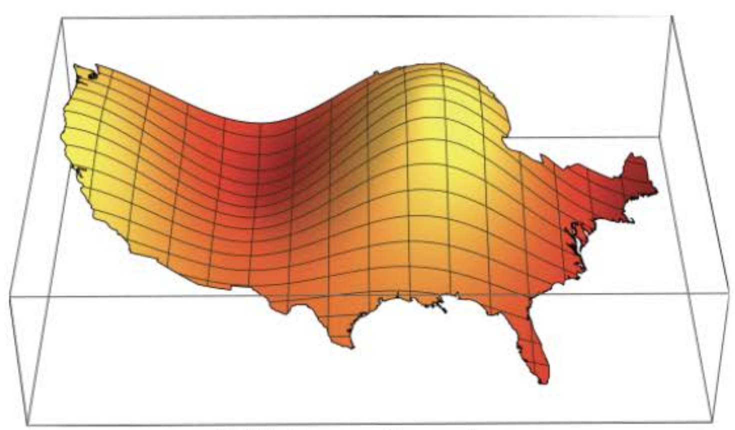
Eigenfunctions for U.S. under Uniform Distribution



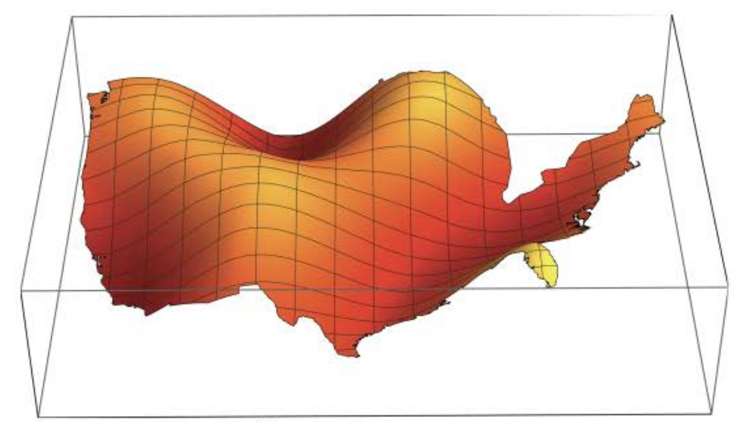
1^{st} Eigenfunction



2^{nd} Eigenfunction



5^{th} Eigenfunction



10^{th} Eigenfunction

Tests for Degree of Spatial Dependence

- In canonical I(1) and I(0) model, $\mathbf{y} \sim \mathcal{N}(\mu\mathbf{e}, \mathbf{\Sigma}_L)$ and $\mathbf{y} \sim \mathcal{N}(\mu\mathbf{e}, \mathbf{I}_n)$, so $\mathbf{Z} = \mathbf{R}_q'\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$ with $\mathbf{\Omega} = \mathbf{D}_q$ and $\mathbf{\Omega} = \mathbf{I}_q$

- In general I(1) or I(0) model, (F)CLT and eigenvector convergence yield approximate normality of $\mathbf{Z}$ under fixed q asymptotics

- I(1) and I(0) tests amount to tests about covariance matrix of q-dimensional normal: standard problem, see paper

  $\Rightarrow$ Can apply I(0) test to $y_l - \beta_0' x_l$ to test null of $(1, -\beta_0)$ cointegrating vector, cf. Wright (2000)

- Extends naturally to inference about c in spatial LTU model

  $\Rightarrow$ For ease of interpretation, map c to half-life, same units as s_l

Spatial Persistence of Chetty et al. (2014) Data

Variable	p-value for Test of		Half-life
	$I(1)$ Null	$I(0)$ Null	95% CI
Mobility Index	0.39	<0.01	[0.10,∞]
Frac. $<$ 15 Mins to Work	0.58	<0.01	[0.14,∞]
Frac. Religious	0.27	0.04	[0.07,∞]
Frac. Single Mothers	0.18	<0.01	[0.05,∞]
Manufacturing Share	0.21	<0.01	[0.06,∞]
Teenage LFP Rate	0.51	<0.01	[0.12,∞]

Notes: Based on $q = 15$. Half-lives are multiples of largest distance between commuting zones (ca. 2800 miles)

Residual Based Tests

- $y_l = x_l'\beta + u_l$, seek inference about persistence of u_l

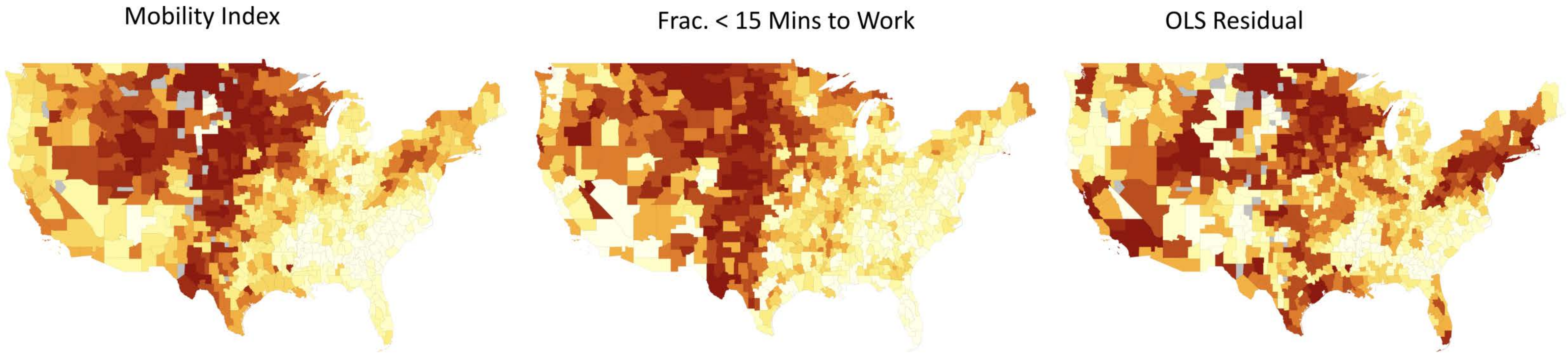
  $\Rightarrow$ Special case: Engle-Granger (1987)-type test of null hypothesis of no cointegration, $u_l \sim I(1)$

- Proceed as before, except now use q eigenvectors $\mathbf{R}_X$ of $\mathbf{M}_X \mathbf{\Sigma}_L \mathbf{M}_X$ with $\mathbf{M}_X = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$, so that $\mathbf{R}_X \mathbf{y} = \mathbf{R}_X \mathbf{u}$

- Asymptotic validity if x_l is small-sample or "asymptotically" independent from u_l

  $\Rightarrow$ **Result**: Always holds if (y_l, x_l) is jointly I(1), i.e. under null hypothesis of no cointegration

Chetty et al. (2014) Residual Example



Engle-Granger Test for Chetty et al. (2014) Regressions

Variable	R^2	p-value no coint.
Frac. $<$ 15 Mins to Work	0.48	0.14
Frac. Religious	0.28	0.26
Frac. Single Mothers	0.59	0.11
Manufacturing Share	0.09	0.37
Teenage LFP Rate	0.44	0.29

Notes: Based on $q = 15$. p-value of residual based test of null hypothesis of no cointegration between Mobility Index and different variables

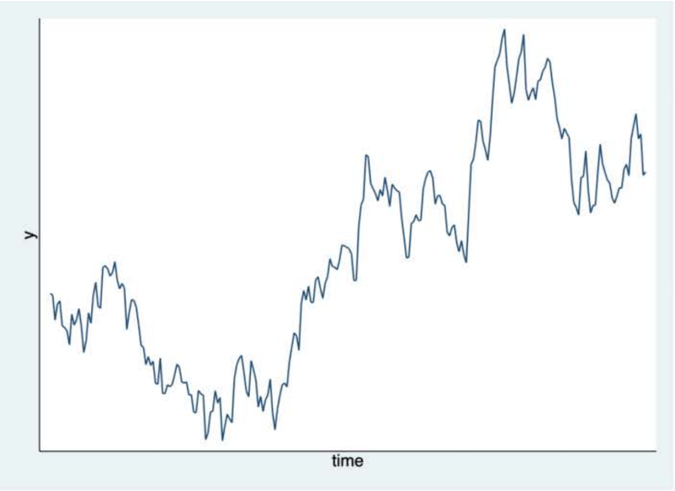
"First Differencing" Spatial I(1) Processes

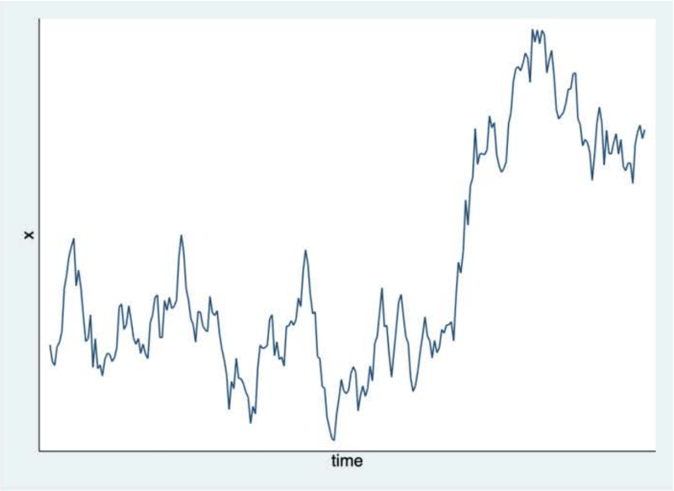
- For standard time series, avoid spurious regression by running regression in first differences

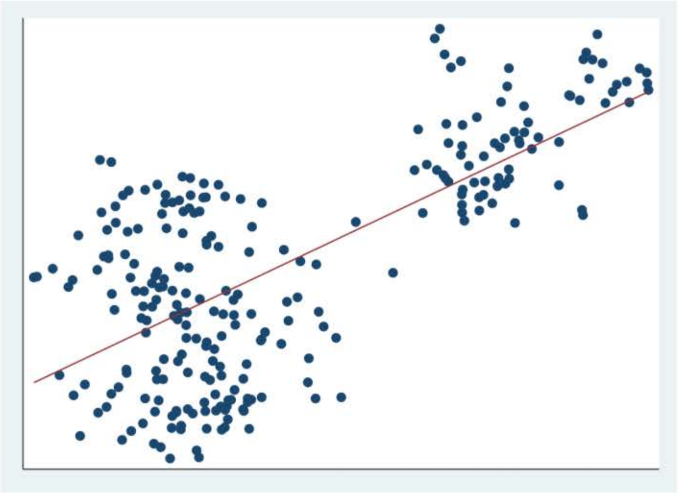
$$\Delta y_l = \Delta x_l' \beta + u_l$$

- Analogue for spatial I(1) data?

Time Series Spurious Regression



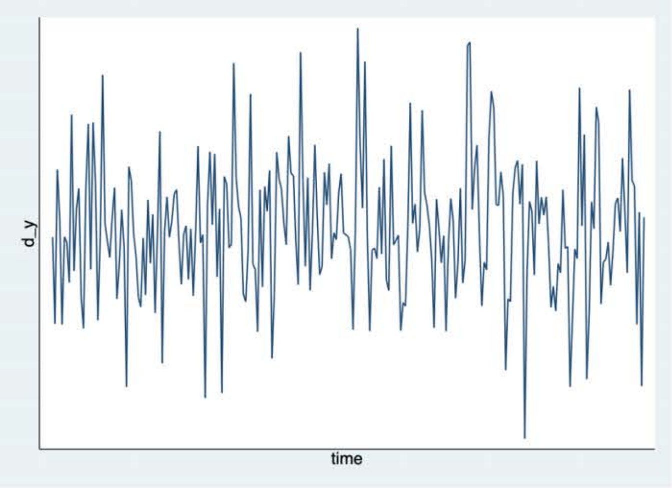


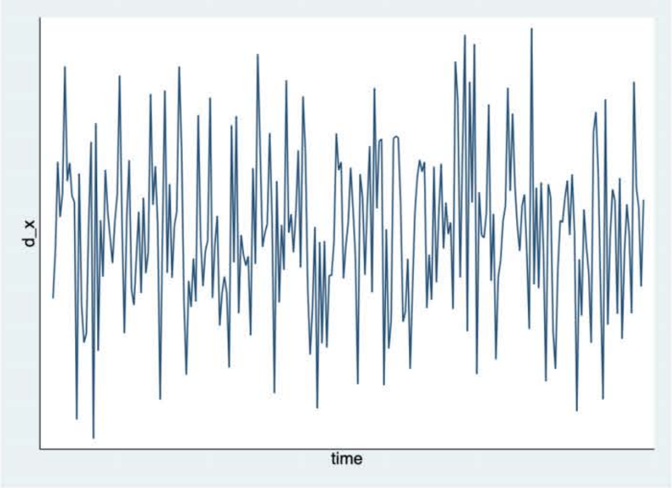


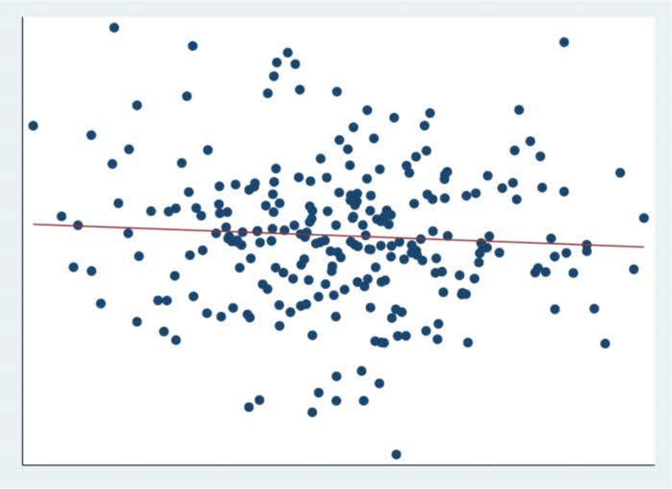
t-stat $= 8.31$

$R^2 = 0.48$

Time Series First Difference Regression







$$t\text{-stat} = -1.09$$

$$R^2 = 0.00$$

Spatial Data Transformations

- Isotropic differences $y_l^* = \sum_{\ell \neq l} \kappa(|s_l - s_\ell|)(y_l - y_\ell)$ and HAR adjustment

- Cluster fixed effects and cluster HAR adjustment

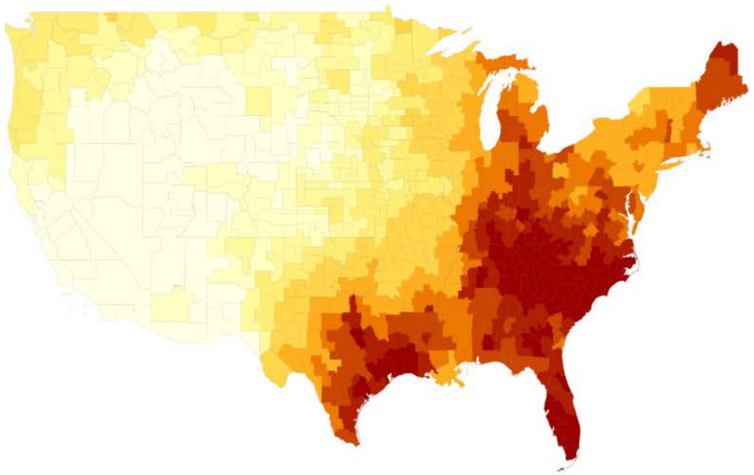
- Lévy Brownian Motion (LBM) GLS $\mathbf{y}^* = (\mathbf{M}\boldsymbol{\Sigma}_L\mathbf{M})^{-1/2}\mathbf{y}$ and HAR adjustment

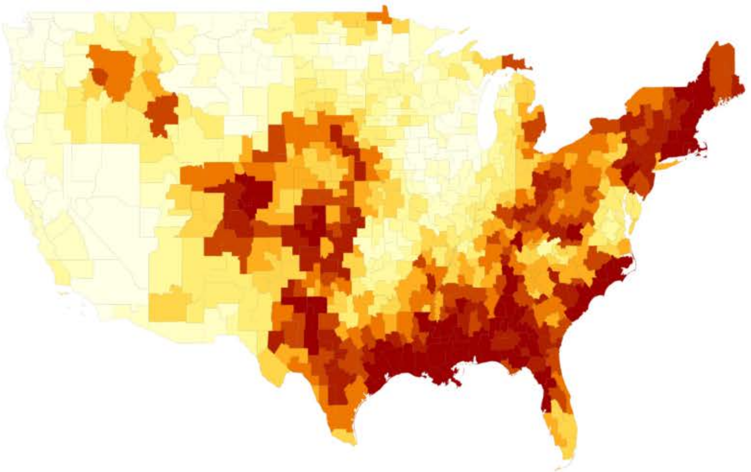
- High-pass filter $\mathbf{y}^* = \mathbf{y} - \mathbf{R}_q\mathbf{R}_q'\mathbf{y}$ and HAR adjustment

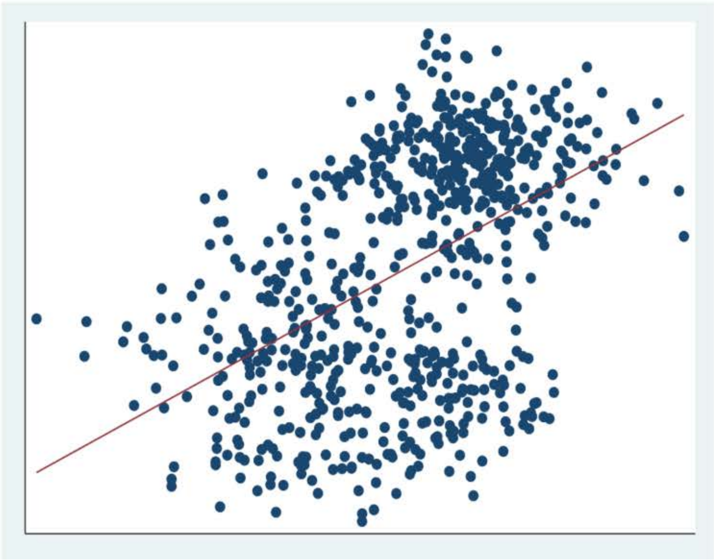
- Low-frequency transformation $\mathbf{y}^* = \mathbf{R}_q'\mathbf{y}$ and use implication for I(1) model

Use C-SCPC of Müller and Watson (2022b) for HAR adjustments

Spatial Spurious Regression



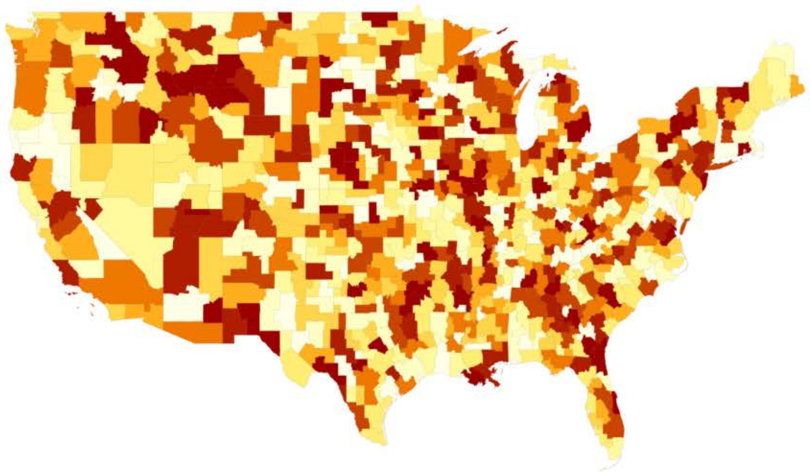


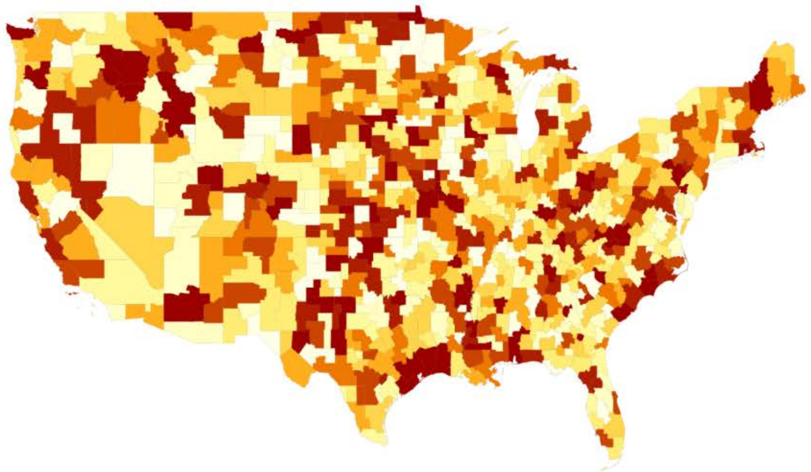


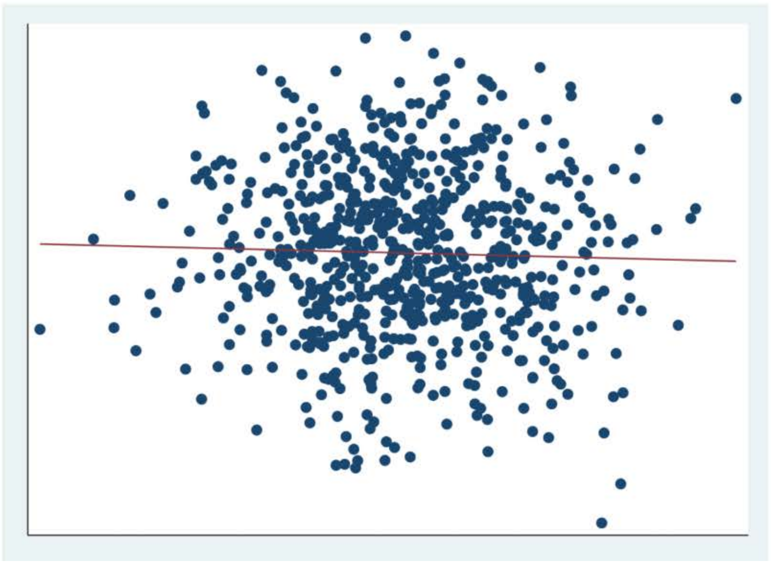
t-stat $= 5.95$

$R^2 = 0.23$

Spatial LBM-GLS Regression







t-stat $= -0.89$

$R^2 = 0.00$

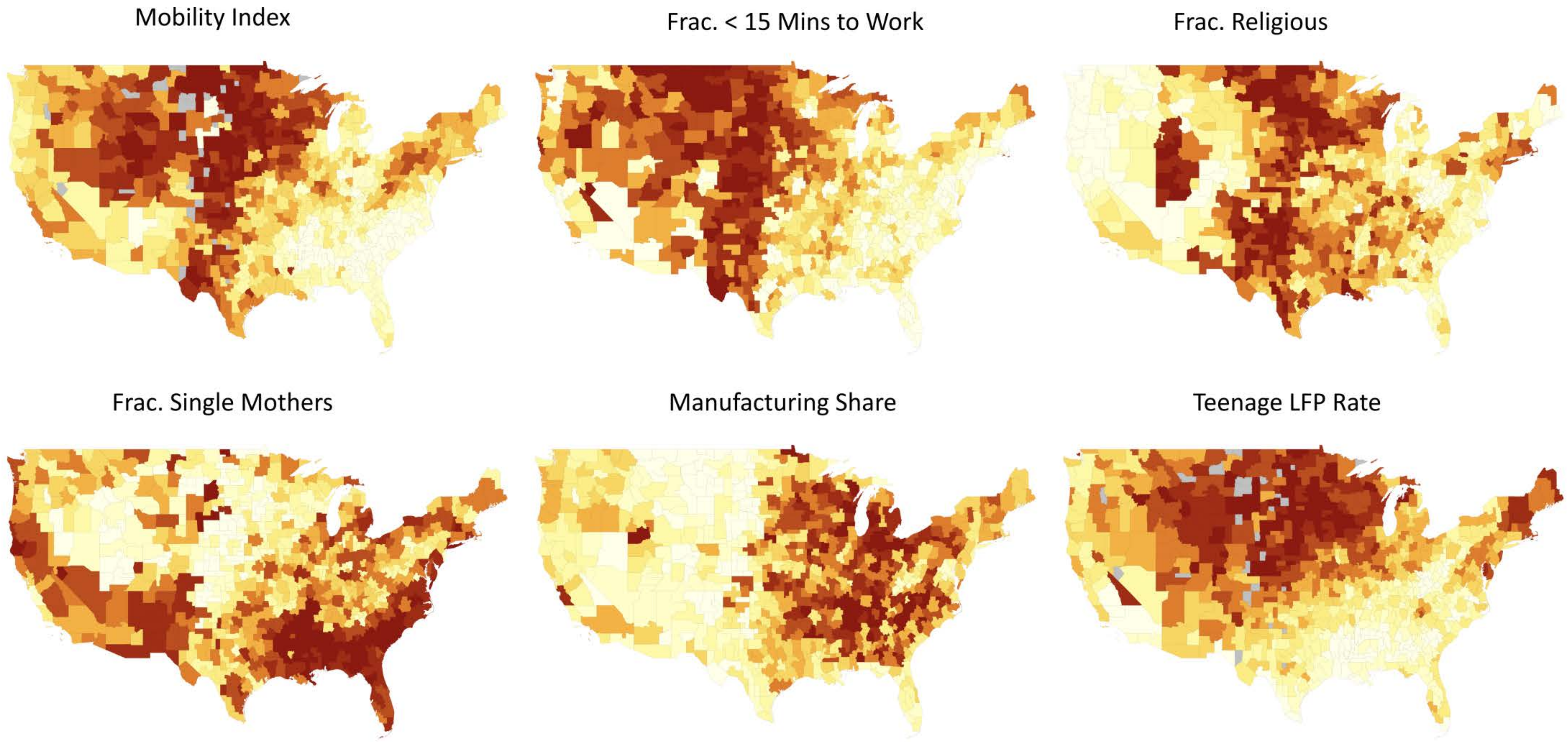
Monte Carlo Properties of Transformations

- Draw 250 locations uniformly within each of the 48 contiguous U.S. states

- Generate data from Lévy BM, I(1), J_c, LTU(c), Brownian Sheet

- Summary:
  - Level regression with HAR or clustered standard errors, cluster fixed effects and high-pass filter don't control size well
  - LF transformation, isotropic differences and LBM-GLS have good size control
  - LBM-GLS yields shortest confidence intervals

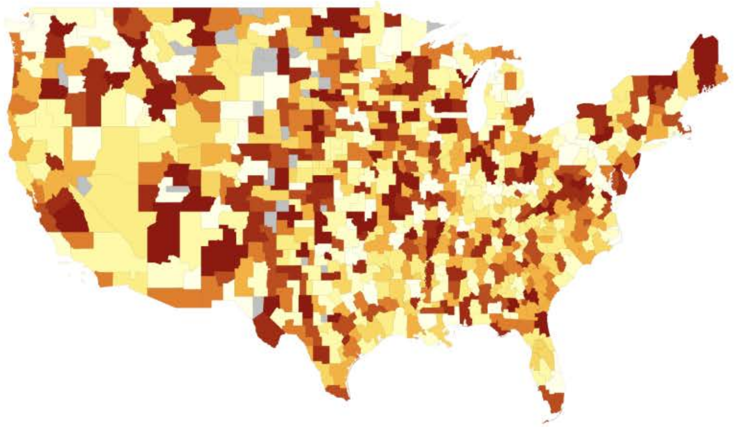
Data in Chetty et al. (2014)



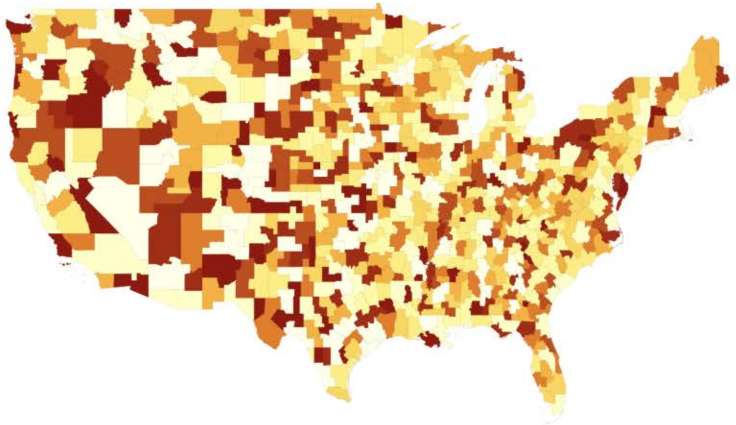
LBM-GLS "Differences" of Data in Chetty et al. (2014)

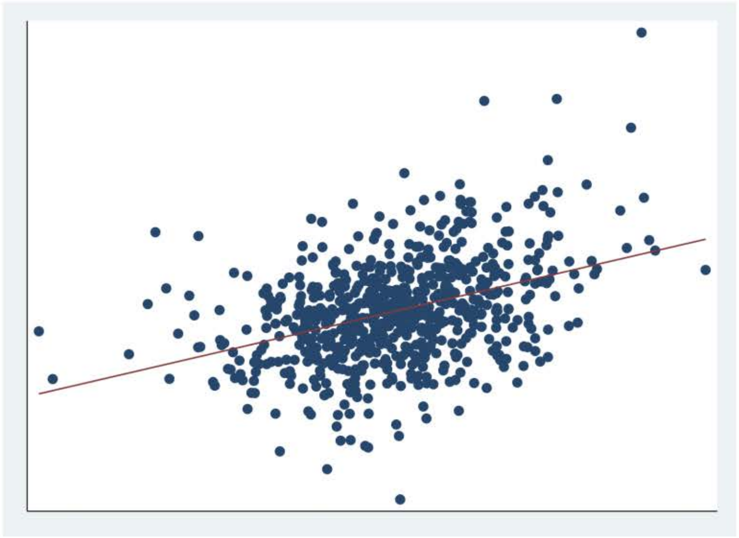
Spatial LBM-GLS Regression

Mobility Index



Frac. < 15 Mins to Work





t-stat$= 6.04$

$R^2 = 0.16$

Chetty et al. (2014) LBM-GLS Regression

Variable	Levels		LBM-GLS	
	R^2	$\hat{\beta}$ [95% CI] Cluster	R^2	$\hat{\beta}$ [95% CI] C-SCPC
Frac. $<$ 15 Mins to Work	0.48	0.69 [0.55, 0.84]	0.16	0.37 [0.24, 0.50]
Frac. Religious	0.28	0.53 [0.36, 0.70]	0.14	0.32 [0.13, 0.51]
Frac. Single Mothers	0.59	-0.77 [-0.92,-0.63]	0.51	-0.61 [-0.68,-0.53]
Manufacturing Share	0.09	-0.30 [-0.46,-0.13]	0.01	0.06 [-0.01, 0.13]
Teenage LFP Rate	0.44	0.66 [0.50, 0.82]	0.04	0.25 [0.11, 0.40]

Notes: Pairwise regressions of Mobility Index on different socioeconomic variables, $n = 722$ commuting zones in contiguous U.S.

Conclusions

- Strong persistence as much a problem for spatially correlated data as for time series
  - Paper provides formal analysis
  - Kelly (2019, 2020, 2022) suggests empirical relevance
- Tests for degree of spatial persistence, analogous to time series case
- LBM-GLS transformation most promising analogue of "first differencing"

  More research required to fully understand its (asymptotic) properties