Tying current observations to their preceding data generating process by utilizing interregional migration dynamics

Michael Tiefelsdorf & Xiaojun Pu

School of Economic, Political and Policy Sciences
The University of Texas at Dallas
Email: tiefelsdorf@utd.edu

RSAI Meeting, San Antonio, TX
November 10th, 2018

Introduction I: Change of perspective

- Change of spatial econometric modelling perspective population focused studies:
  - Instead of assuming somewhat arbitrary ***adjacency*** or ***distance decay*** based spatial dependencies,
  - a ***migration*** matrix is used to capture explicitly an underlying demographic process, which ties regions together.
- This allows adopting a discrete space-time perspective by linking a data generating process at ***t-1*** to the process realizations at ***t***.

Introduction II: Study Implementation

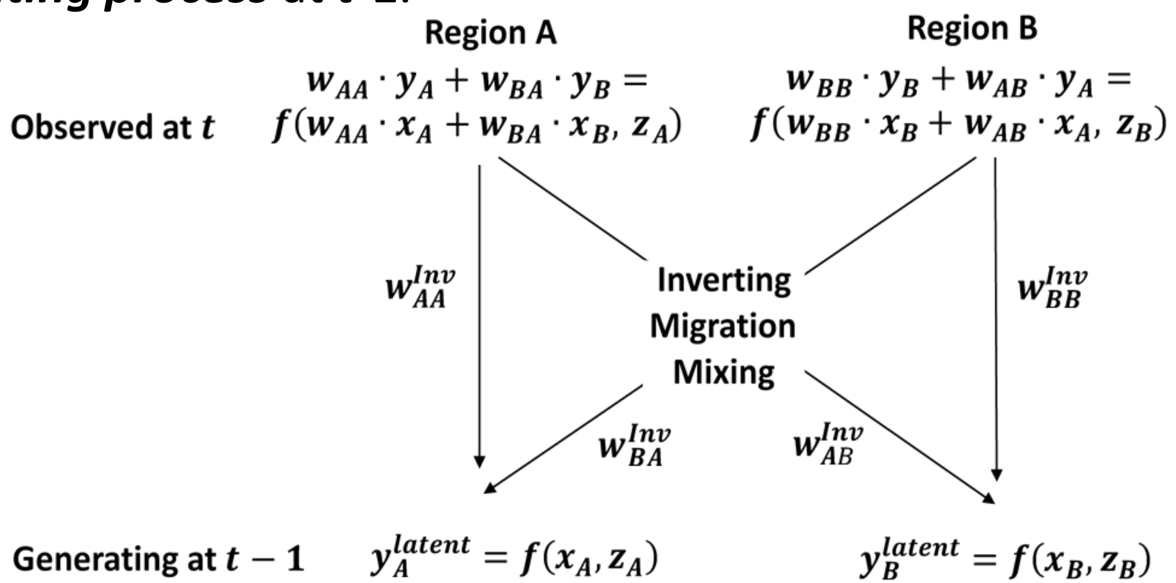
- Model implementation:
  - Underlying ***demographic*** structure of the data generating process.
  - Operationalization of as a ***stochastic process*** accounting for temporally misaligned data.
  - Investigation of ***model misspecifications***.
- Motivation for the misspecification investigation:
  - We found in our empirical analyses with the proposed method that the estimated ***autocorrelation coefficient*** can take values ***outside its feasible range***.
  - An empirical researcher can ***never be perfectly certain*** that she/he has selected the ***proper model specification***. Therefore, a ***sensitivity analysis*** with regards to model misspecifications is advisable.

Introduction III: Key Assumptions

- Key assumptions:
  - A ***discrete*** demographic process with a fixed starting time $t-1$ and end point t.
    Rather than a ***dynamic process*** in continuous time.
  - ***Homogeneity*** of demographic cohorts and strata.
  - ***Balance*** of birth and death rates.
  - The regional ***system*** is ***closed***.

Introduction IV: Underlying Model Structure

- Only ***process realizations*** at ***t*** can be observed.
- At ***t*** the ***observed*** endogenous regional rates/counts and exogenous ***population related*** factors are migration induced ***mixtures***.
- ***Region specific*** exogenous factors are stable because they are migration invariant.
- Reversing the mixing migration effect, allows identifying the ***preceding data generating process*** at ***t*-1**.



Theory I: Underlying Demographic Migration Process

- Let the $n \times n$ migration matrix among n regions be

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

- Its entries m_{ij} are the number of migrants between the i^{th} origin (row) the ***starting time point t*** $\boldsymbol{-1}$ and the j^{th} destination (column) at the ***end time point t*** with

$$m_{ij} = \begin{cases} i \neq j & \text{the number of migrants from origin } i \text{ to destination } j \\ j = i & \text{the number of residents staying in the origin } i \end{cases}$$

- The population $\mathbf{p}_{t-1}$ ***at the origin*** at time point $t-1$ can be derived from the ***row-sums*** of $\mathbf{M}$:

$$\mathbf{p}_{t-1} = \mathbf{M} \cdot \mathbf{1} \text{ where } \mathbf{1} = [1,1,\dots,1]^T$$

- The population $\mathbf{p}_t$ ***that arrived at the destination*** at time point t is the ***column sums*** of $\mathbf{M}$:

$$\mathbf{p}_t = \mathbf{M}^T \cdot \mathbf{1}$$

Theory II: The Transition Matrix

- The row-sum standardized ***transition matrix***

$$\mathbf{W} = \begin{bmatrix} m_{11}/p_{1,t-1} & m_{12}/p_{1,t-1} & \cdots & m_{1n}/p_{1,t-1} \\ m_{21}/p_{2,t-1} & m_{22}/p_{2,t-1} & \cdots & m_{2n}/p_{2,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1}/p_{n,t-1} & m_{n2}/p_{n,t-1} & \cdots & m_{nn}/p_{n,t-1} \end{bmatrix} = \mathrm{diag}(\mathbf{p}_{t-1})^{-1} \cdot \mathbf{M}$$

  projects the population count at $\mathbf{p}_{t-1}$ at $t-1$ to the population $\mathbf{p}_t$ at t

$$\mathbf{p}_t = \mathbf{W}^T \cdot \mathbf{p}_{t-1}$$

  under the assumption that the regional system is closed and that the number of births and deaths are in balance.
- Therefore, if only the ***current*** population counts at $\mathbf{p}_t$ are known the latent ***previous*** population $\mathbf{p}_{t-1}$ can be ***derived by***

$$\mathbf{p}_{t-1} = [\mathbf{W}^T]^{-1} \cdot \mathbf{p}_t$$

- The same transformation applies to the ***numerator*** and ***denominator*** of regional ***population related rates***.

Theory III: The Diagonal of a Migration Matrix

- The ***number of stayers*** on the diagonal and the ***number of movers*** on the off-diagonal cells depend on the ***duration*** of the accounting period from $t-1$ to t .
- The ***longer*** the accounting period, the more stayers leave their origins for other destinations.
- ***Shifting*** a proportion of stayers to the off-diagonals cells:
  - Decomposition of the migration matrix: $\mathbf{M} = \mathbf{M}_D + \boldsymbol{M}_{\bar{D}}$
  - Shifting a proportion of stayers $\mathbf{M}_D$ to the movers matrix $\boldsymbol{M}_{\bar{D}}$:

$$\mathbf{M}_{[\pi]} = \boldsymbol{M}_{\bar{D}} + \pi \cdot \mathbf{M}_D + (1-\pi) \cdot \mathbf{M}_D \cdot \mathbf{W}_{\bar{D}} \text{ with}$$

$$\mathbf{W}_{\bar{D}} = \text{diag}[\mathbf{M}_{\bar{D}} \cdot \mathbf{1}]^{-1} \cdot \mathbf{M}_{\bar{D}} \text{ and } 0 \leq \pi \leq 1$$

- Overall, the ***spatial relationships strengthen***.
- Caveat: ***Return migration*** is assumed to be not present during the timeframe in the regional system.

Theory IV: Stochastic Specification as Leroux Model

- In an empirical setting the ***exact demographic model*** needs to be ***relaxed*** due to:
  - Potential ***misspecifications*** of the transition matrix $\mathbf{W}$
  - ***Aging*** of the population from $\mathbf{p}_{t-1}$ to $\mathbf{p}_t$.
  - The migration ***accounting period*** that does not perfectly ***line up*** with timing of the other empirical data.
- The ***flexible Leroux (1999) specification*** of the transition matrix will abate some of these issues:

$$\boldsymbol{L}_\rho^T = (1-\rho)\cdot\mathbf{I} + \rho\cdot\mathbf{W}^T$$

- The autocorrelation parameter ρ of $\mathbf{L}_\rho$ measures the ***empirical strength*** of the migration exchange process:
  - For $\rho = 1$ the migration transition matrix $\mathbf{W}^T$ is fully effective
  - For $\rho = 0$ interregional migration effects are irrelevant.

Theory V: Doubly Stochastic Matrix

- The row-sums of $\mathbf{W}^T$ and $\mathbf{I}$ in the Leroux model mismatch.
- This imbalance can lead to biased model estimates.
- Therefore, a doubly stochastic matrix $\mathbf{D}$ of the migration matrix $\mathbf{M}$ is used.
  - It is iteratively standardized so that all row and column sums are one (Slater 2008; Slater 2009).
- Properties of the doubly stochastic matrix $\mathbf{D}$:
  - Controls for marginal size effect and focus on the relative spatial interaction effect (Slater 2008);
  - The odds ratios $\frac{m_{ij}/m_{ik}}{m_{jl}/m_{kl}} = \frac{m_{ij}m_{kl}}{m_{ik}m_{jl}}$ remains invariant (Slater 2008; Slater 2009)
- The final specification of Leroux spatial transition matrix becomes

$$\mathbf{L}_\rho^T = (1-\rho)\cdot\mathbf{I} + \rho\cdot\mathbf{D}^T$$

Theory VI: The Underlying Data Generating Process at $t-1$

- Key model components:
  - Let $\mathbf{y}_{t-1}$ denote a vector of n ***regional rates*** based on the population $\mathbf{p}_{t-1}$.
  - Assumption: The observed rates $\mathbf{y}_{t-1}$ are ***stochastically independent***.
  - $\mathbf{X}_{t-1}^{B}$ is a $n \times p$ matrix of p exogenous variables ***linked to*** the ***population*** $\mathbf{p}_{t-1}$.
  - $\mathbf{X}^{E}$ is a $n \times q$ matrix of q exogenous variables unrelated to population and therefore ***migration invariant***.
- Data generating process:

$$g[E(\mathbf{y}_{t-1})] = \mathbf{1} \cdot \beta_0 + \mathbf{X}_{t-1}^{B} \cdot \boldsymbol{\beta}^{B} + \mathbf{X}^{E} \cdot \boldsymbol{\beta}^{E}$$

where $g(\ )$ is the logistic link function.

Theory VII: Observed Process at t under the Influence of Migration

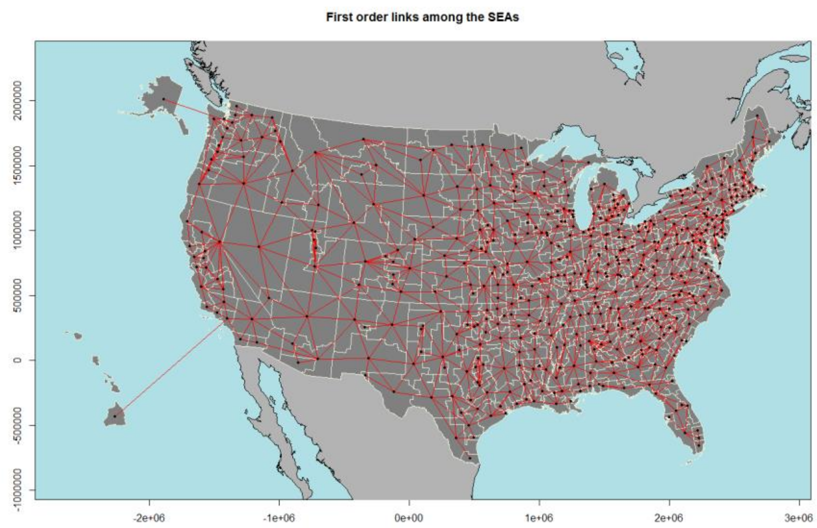
- Only $\mathbf{y}_t = \mathbf{L}_\rho^T \cdot \mathbf{y}_{t-1}$ and $\mathbf{X}_t^B = \mathbf{L}_\rho^T \cdot \mathbf{X}_{t-1}^B$ ***can be observed at*** t but not the the underlying data generating process at $t-1$. Thus
  - model estimates $\{\beta_0, \boldsymbol{\beta}^B, \boldsymbol{\beta}^E, \rho\}$ may become ***biased*** and
  - observations become ***autocorrelated*** with a covariance structure $Cov(\mathbf{y}_t) = \sigma^2 \cdot \mathbf{L}_\rho \cdot \mathbf{L}_\rho^T$ that is structurally equivalent to that of a ***moving average spatial process.***
- The model can be transformed back to its stochastically independent state at $t-1$ with $\left[\mathbf{L}_\rho^T\right]^{-1}$:

$$g\left[E\left(\left[\mathbf{L}_\rho^T\right]^{-1} \cdot \mathbf{y}_t\right)\right] = \left[\mathbf{L}_\rho^T\right]^{-1} \cdot \mathbf{1} \cdot \beta_0 + \left[\mathbf{L}_\rho^T\right]^{-1} \cdot \mathbf{X}_t^B \cdot \boldsymbol{\beta}^B + \mathbf{X}^E \cdot \boldsymbol{\beta}^E$$

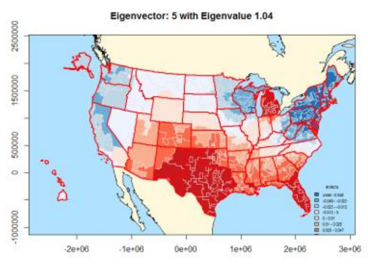
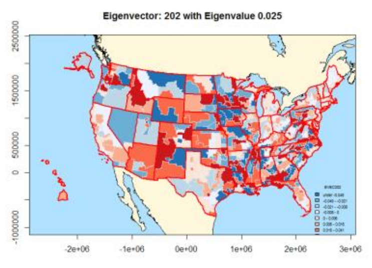
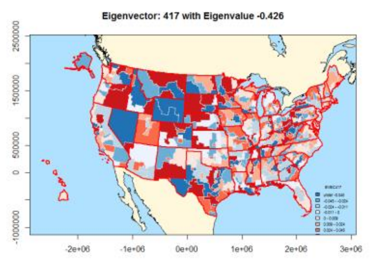
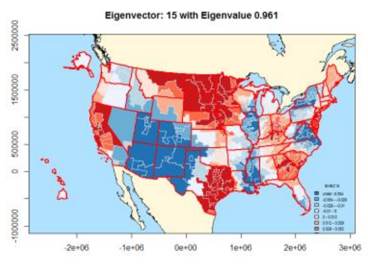
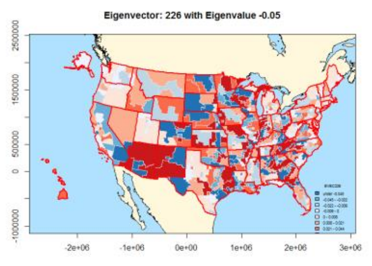
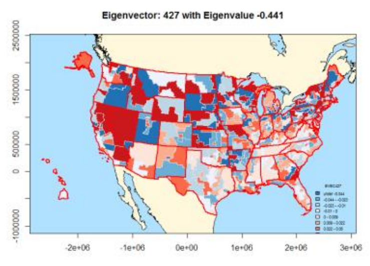
- The migration invariant $\mathbf{X}^E$ corresponds with spatial lag specification.
- Consequently, the proposed model becomes a hybrid moving average/spatial lag model.

Study Design I: Spatial Setting of the Data Generating Process

- The spatial layout: 508 State Economic Areas (SEA)
- The 508×508 migration matrix $\mathbf{M}$ (1965 to 1970) with estimated number of stayer on the ***diagonal***.
- Exogenous variables $\mathbf{X}_{t-1}^{B}$ and $\mathbf{X}^{E}$ are selected eigenvectors based on
  $$\mathbf{I} - \frac{\mathbf{1} \cdot \mathbf{1}^{T}}{n} \cdot \frac{1}{2} \cdot \left(\mathbf{A}_{[W]} + \mathbf{A}_{[W]}^{T}\right) \cdot \mathbf{I} - \frac{\mathbf{1} \cdot \mathbf{1}^{T}}{n}$$
  with $\mathbf{A}_{[W]}$ being the row-sum standardized spatial adjacency matrix.
- Properties of the eigenvectors:
  - They are ***uncorrelated*** among each other.
  - Each eigenvector exhibits a ***distinctive spatial pattern*** with a given spatial autocorrelation level.



Study Design II: Pattern of Selected Eigenvectors

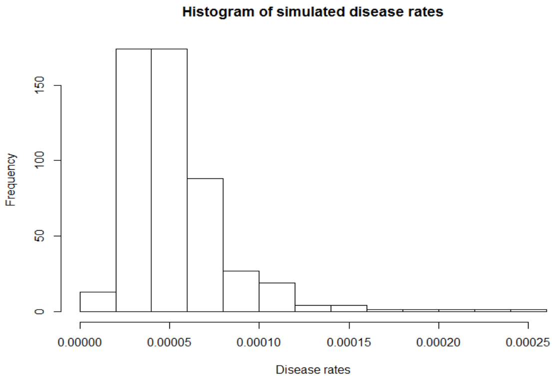
<table>
<tr><td>B1 (eigenvector 5: I=1.040)</td><td>B2 (eigenvector 202: I =0.025)</td><td>B3 (eigenvector 417: I =-0.426)</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>E1 (eigenvector 15: I =0.961)</td><td>E2 (eigenvector 226: I =-0.050)</td><td>E3 (eigenvector 427: I =-0.441)</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

- Organization of the table:
  - Columns: [1] Strong ***positive*** spatial autocorrelation, [2] spatial ***independence*** and [3] strong ***negative*** spatial autocorrelation
  - Row 1: Patterns $B1_{t-1}, B2_{t-1}$ and $B3_{t-1}$ represent ***population related*** exogenous variables affected by migration
  - Row 2: Patterns E1, E2 and E3 are ***migration invariant*** exogenous variables.

Study Design III: Specification at $t-1$ and t

- The data generating process at $t-1$:
  - All variables are scaled so that their coefficients are 1.

$$g[E(\boldsymbol{y}_{t-1})] = -10 + 1 \cdot B1_{t-1} + 1 \cdot B2_{t-1} + 1 \cdot B3_{t-1} + 1 \cdot E3 + 1 \cdot E2 + 1 \cdot E1$$



- The observed data at t:
  - The endogenous variable: $\mathbf{y}_t = \mathbf{L}_{\rho=\rho_0}^T \cdot \mathbf{y}_{t-1}$
  - The exogenous population based variables: $\mathbf{X}_t^B = \mathbf{L}_{\rho=\rho_0}^T \cdot \mathbf{X}_{t-1}^B$
  - The migration invariant variables $\mathbf{X}^E$ remain the same.
  - Unless otherwise stated the autocorrelation level is $\rho_0 = 1$.

Experiment I with $\mathbf{M}_{[\pi=1.0]}$: Misspecified autocorrelation structure

The values in bracket display α-errors of $H_0: \beta = 1$ against $H_1: \beta \neq 1$ and $H_0: \rho = 1$ against $H_1: \rho \neq 1$.

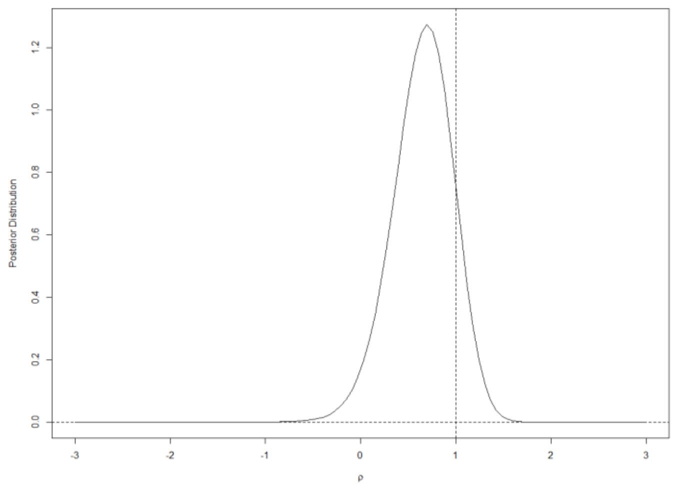
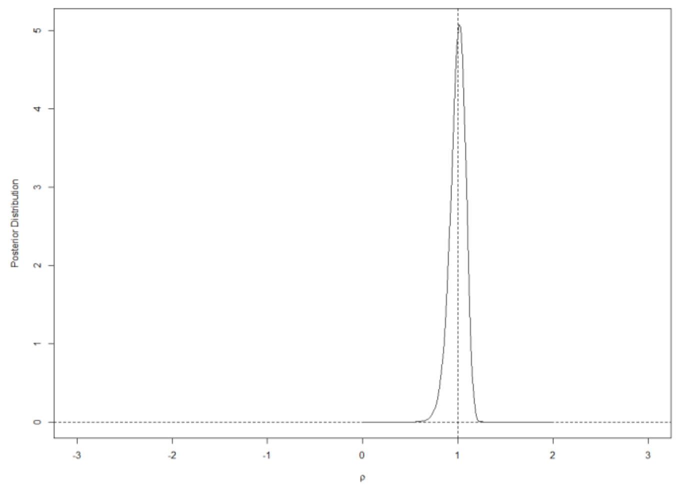
Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	*DIC*
Target parameters	1	1	1	1	1	1	1	
Properly specified Leroux model based on 25 simulations	0.99 (0.86)	1.00 (0.97)	1.03 (0.66)	0.98 (0.68)	0.94 (0.28)	0.91 (0.06)	0.64 (0.26)	2651.47
Properly specified Leroux model	1.00 (0.92)	0.99 (0.88)	1.02 (0.78)	0.98 (0.72)	0.93 (0.21)	0.91 (0.07)	0.65 (0.28)	***2184.62***
Leroux specification with $\rho_0 = 0$	1.00 (0.95)	1.00 (0.95)	1.00 (1.00)	1.00 (0.96)	1.00 (0.97)	1.00 (0.97)	-0.00 (0.99)	2157.48
Aspatial model (ignoring L_ρ)	0.99 (0.89)	0.99 (0.88)	0.99 (0.90)	0.89 (0.02)	0.83 (0.00)	0.80 (0.00)		2185.81
SAR model with spatial adjacency matrix in the doubly stochastic standardization	1.00 (0.91)	0.99 (0.88)	0.99 (0.90)	0.89 (0.02)	0.83 (0.00)	0.80 (0.00)	-0.01 (0.00)	2188.00
SMA model with spatial adjacency matrix in the doubly stochastic standardization	1.00 (0.91)	0.99 (0.89)	0.99 (0.89)	0.90 (0.04)	0.83 (0.00)	0.80 (0.00)	-0.01 (0.00)	2188.03
SMA model with the migration flow matrix (without diagonal) in the doubly stochastic standardization	0.99 (0.75)	0.99 (0.90)	1.00 (0.97)	0.83 (0.00)	0.84 (0.00)	0.83 (0.00)	0.20 (0.00)	2187.25

Expriment II: Increasing redistribution the non-movers

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	DIC
Target parameters	1	1	1	1	1	1	1	
Leroux model with $\mathbf{M}_{[\pi=1]}$	1.00 (0.92)	0.99 (0.88)	1.02 (0.78)	0.98 (0.72)	0.93 (0.21)	0.91 (0.07)	0.65 (0.28)	2184.62
Shifted matrix $\mathbf{M}_{[\pi=0.9]}$	0.99 (0.83)	0.99 (0.84)	1.02 (0.78)	0.99 (0.88)	0.94 (0.32)	0.89 (0.04)	0.78 (0.32)	2201.50
Shifted matrix $\mathbf{M}_{[\pi=0.75]}$	0.99 (0.88)	0.99 (0.89)	1.03 (0.72)	1.08 (0.91)	0.96 (0.56)	0.88 (0.08)	0.91 (0.53)	2218.61
Shifted matrix $\mathbf{M}_{[\pi=0.5]}$	0.99 (0.91)	0.99 (0.95)	1.03 (0.80)	1.02 (0.79)	1.02 (0.88)	0.99 (0.44)	1.00 (0.96)	2232.22
Shifted matrix $\mathbf{M}_{[\pi=0.25]}$	1.00 (1.00)	1.06 (0.80)	1.01 (0.97)	1.04 (0.66)	1.08 (0.73)	1.02 (0.95)	1.00 (1.00)	2235.75
Shifted matrix $\mathbf{M}_{[\pi=0.0]}$	1.02 (0.84)	1.10 (0.83)	0.96 (0.91)	1.07 (0.53)	0.90 (0.80)	1.06 (0.85)	1.00 (1.00)	2233.51

- The interregional dependencies increase with the autocorrelation level approached its expected value.

Experiment II: Posterior Distributions of ρ for $\mathbf{M}_{[\pi=1.0]}$ and $\mathbf{M}_{[\pi=0.5]}$

(a) The posterior distribution of ρ of unadjusted migration matrix	(b) The posterior distribution of ρ of 50% adjusted migration matrix
	

- The standard error of ρ's distribution shrinks for $\pi < 1$.

Experiment III with $\mathrm{M}_{[\pi=0.5]}$: Misspecification of the autocorrelation structure

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	DIC
Target parameters	1	1	1	1	1	1	1	
Properly specified Leroux model	0.99 (0.91)	0.99 (0.95)	1.03 (0.80)	1.02 (0.79)	1.02 (0.88)	0.91 (0.44)	1.00 (0.96)	**2232.22**
Aspatial model (ignoring L_ρ)	1.01 (0.87)	1.00 (1.00)	1.00 (1.00)	0.60 (0.00)	0.41 (0.00)	0.31 (0.00)		2264.00
SAR model with spatial adjacency matrix in the doubly stochastic matrix	1.01 (0.89)	1.00 (0.99)	1.00 (1.00)	0.56 (0.00)	0.41 (0.00)	0.32 (0.00)	0.07 (0.00)	2265.71
MA model with spatial adjacency matrix in the doubly stochastic matrix	1.01 (0.90)	1.00 (0.99)	1.00 (1.00)	0.57 (0.00)	0.42 (0.00)	0.33 (0.00)	0.11 (0.00)	2265.80
MA model with the migration flow matrix (without diagonal) in the doubly stochastic matrix	0.98 (0.67)	1.01 (0.96)	1.02 (0.89)	0.41 (0.00)	0.43 (0.00)	0.32 (0.00)	1.33 (0.48)	2254.68

- The coefficients of migration invariant variables are downwards biased.
- $\hat{\rho}$ of the MA migration model exceeds its theoretical upper bound 1.

Experiment IV with $\mathbf{M}_{[\pi=0.5]}$: Misspecification of the variable assignment to either the migration variant and invariant group

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	*DIC*
Target parameters	1	1	1	1	1	1	1	
Properly specified Leroux model	0.99 (0.91)	0.99 (0.95)	1.03 (0.80)	1.02 (0.79)	1.02 (0.88)	0.91 (0.44)	1.00 (0.96)	**2232.22**
Treat B1 as environmental factor	1.25 (0.00)	1.00 (0.98)	1.03 (0.84)	0.98 (0.73)	0.97 (0.76)	0.90 (0.36)	0.99 (0.91)	2237.29
Treat B2 as environmental factor	1.00 (0.93)	1.78 (0.00)	1.03 (0.82)	0.99 (0.94)	0.99 (0.93)	0.85 (0.18)	0.98 (0.88)	2235.38
Treat B3 as environmental factor	0.99 (0.89)	0.99 (0.93)	1.87 (0.00)	0.99 (0.84)	0.93 (0.54)	0.81 (0.07)	0.97 (0.58)	2238.06
Treat E1 as behavioral factor	1.03 (0.59)	1.00 (0.97)	1.02 (0.89)	0.61 (0.00)	1.03 (0.79)	1.00 (1.00)	1.00 (1.00)	2246.23
Treat E2 as behavioral factor	1.01 (0.94)	0.99 (0.91)	1.02 (0.90)	1.00 (0.97)	0.40 (0.00)	0.89 (0.32)	0.98 (0.88)	2240.72
Treat E3 as behavioral factor	0.97 (0.60)	1.00 (0.97)	1.03 (0.82)	0.99 (0.94)	0.97 (0.77)	0.28 (0.00)	0.94 (0.62)	2243.48

- The coefficients of the wrongly assigned migration variant variables increase significantly. The coefficients of the wrongly assigned migration invariant variables decrease.
- The DIC values for models with wrongly assigned migration variant variables increases.

Experiment V with $\mathrm{M}_{[\pi=1.0]}$: Missing relevant exogenous variables

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	DIC
Target parameters	1	1	1	1	1	1	1	
Properly specified Leroux model	1.00 (0.92)	0.99 (0.88)	1.02 (0.78)	0.98 (0.72)	0.93 (0.21)	0.91 (0.07)	0.65 (0.28)	**2184.62**
Missing B1		0.99 (0.85)	1.00 (0.99)	1.12 (0.13)	1.14 (0.09)	1.02 (0.78)	1.57 (0.05)	2542.43
Missing B2	1.00 (0.99)		1.04 (0.56)	0.95 (0.35)	1.01 (0.92)	0.89 (0.02)	0.54 (0.17)	2415.68
Missing B3	0.97 (0.44)	1.03 (0.70)		0.88 (0.01)	0.76 (0.00)	0.78 (0.00)	-0.69 (0.00)	2426.93
Missing E1	1.02 (0.77)	0.96 (0.56)	1.07 (0.33)		0.85 (0.01)	0.82 (0.00)	-0.01 (0.15)	2505.16
Missing E2	1.02 (0.58)	1.09 (0.17)	1.04 (0.54)	1.00 (0.97)		0.88 (0.02)	0.68 (0.39)	2453.60
Missing E3	0.99 (0.84)	1.00 (0.98)	1.06 (0.39)	0.96 (0.54)	0.89 (0.09)		0.55 (0.35)	2470.35

- For missing migration variant variables the autocorrelation level ρ changes with the autocorrelation level of the missing variable.

Experiment V with $\mathbf{M}_{[\pi=0.5]}$: Missing relevant exogenous variables

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	DIC
Target parameters	1	1	1	1	1	1	1	
Properly specified Leroux model	0.99 (0.91)	0.99 (0.95)	1.03 (0.80)	1.02 (0.79)	1.02 (0.88)	0.91 (0.44)	1.00 (0.96)	**2232.22**
Missing B1		0.96 (0.73)	1.01 (0.97)	1.15 (0.14)	1.29 (0.08)	1.10 (0.56)	1.16 (0.15)	2440.88
Missing B2	0.99 (0.80)		1.03 (0.84)	1.02 (0.84)	1.06 (0.62)	0.91 (0.46)	1.01 (0.95)	2298.59
Missing B3	0.98 (0.69)	1.00 (0.98)		1.01 (0.93)	0.98 (0.83)	0.90 (0.35)	0.98 (0.89)	2289.32
Missing E1	1.05 (0.45)	0.99 (0.86)	1.07 (0.60)		1.01 (0.91)	0.91 (0.45)	1.00 (0.97)	2412.78
Missing E2	1.02 (0.73)	1.03 (0.79)	1.04 (0.79)	1.04 (0.61)		0.95 (0.70)	1.01 (0.92)	2303.61
Missing E3	1.00 (0.99)	1.01 (0.91)	1.08 (0.54)	1.06 (0.43)	1.09 (0.48)		1.04 (0.73)	2288.69

- The spatial dependency structure is more pronounced which makes the model less sensitive to missing relevant autocorrelated exogenous variables.

Conclusions I

- Overall, the proposed migration effect model is sound and robust for properly specified models.
  $\Rightarrow$ It allows ***identifying*** the underlying data generating process ***at*** $t-1$ from ***observed data at*** t.
- Misspecifications of the estimated model can lead to biases:
  - Estimates of coefficients for ***migration invariant variables*** are more ***susceptible*** to model misspecifications.
  - The autocorrelation coefficient ρ can become ***biased*** or assume ***unrealistic values***.
- The ***increased strength*** of the spatial dependencies when using $\mathbf{M}_{[\pi=0.5]}$ has a substantial effect the estimates.

Conclusions II

- In empirical models it is ***not always clear*** whether an exogenous variable is ***migration variant*** or ***invariant***.
  - At substantial computational cost by allowing each exogenous variable having its own autocorrelation level it could be determined whether a variable is more akin to being migration variant or invariant.
- The proposed model structure also allows having Poisson or negative binomial distributed ***regional counts*** with a proper offset term.
- Without the use of the ***simulation-free*** integrate nested Laplace approximation and ***explicit parallel computations*** the calibration of the Bayesian models would not have been feasible.
- In addition to population focused analyses the proposed model specification can also be adapted for interregional dependencies such as ***international trade flow*** or regional ***input/output tables***.

Thank you!

Experiment with $M_{[\pi=1.0]}$: Wrong assignment of risk factors to either the environmental or bio-behavioral category $M_{[\pi=1]}$

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	ρ	DIC
Target parameters	1	1	1	1	1	1	1	
Properly specified Leroux model	1.00 (0.92)	0.99 (0.88)	1.02 (0.78)	0.98 (0.72)	0.93 (0.21)	0.91 (0.07)	0.65 (0.28)	**2184.62**
Treat B1 as environmental factor	1.06 (0.20)	0.99 (0.90)	1.02 (0.80)	0.98 (0.67)	0.93 (0.21)	0.92 (0.11)	0.68 (0.30)	2184.11
Treat B2 as environmental factor	1.00 (0.97)	1.08 (0.26)	1.02 (0.74)	0.98 (0.64)	0.92 (0.17)	0.90 (0.04)	0.60 (0.17)	2184.58
Treat B3 as environmental factor	1.00 (0.92)	0.99 (0.90)	1.08 (0.25)	0.95 (0.27)	0.89 (0.03)	0.86 (0.00)	0.37 (0.01)	2186.13
Treat E1 as behavioral factor	1.00 (0.98)	0.99 (0.83)	1.01 (0.86)	0.91 (0.04)	0.98 (0.78)	0.97 (0.56)	0.96 (0.94)	2184.31
Treat E2 as behavioral factor	0.99 (0.90)	0.99 (0.88)	1.01 (0.88)	0.96 (0.43)	0.83 (0.00)	0.89 (0.03)	0.52 (0.20)	2186.14
Treat E3 as behavioral factor	0.99 (0.88)	0.99 (0.92)	1.01 (0.84)	0.99 (0.82)	0.95 (0.37)	0.80 (0.00)	0.73 (0.51)	2185.23

- The spatial autocorrelation level drops substantially if the negatively autocorrelated bio-behavioral factor is mis-assigned to the environmental category.
- A mis-classification of the environmental factors biases their regression coefficients downwards

Experiment: Over-specified model with irrelevant variable

Expectation: An added irrelevant variable does not influence the other estimates because it is uncorrelated with the remaining relevant variables.

Model specifications	β_{B1}	β_{B2}	β_{B3}	β_{E1}	β_{E2}	β_{E3}	β_4	ρ	DIC
Target parameters	1	1	1	1	1	1	**0**	1	
Aspatial model	0.99 (0.90)	0.99 (0.88)	0.99 (0.89)	0.89 (0.02)	0.83 (0.00)	0.80 (0.00)	-0.01 (0.83)		2187.76
Original matrix $M_{[\pi=1]}$. Added as behavioral factor	1.00 (0.97)	0.99 (0.88)	1.02 (0.77)	0.99 (0.83)	0.95 (0.37)	0.92 (0.13)	-0.04 (0.34)	0.75 (0.45)	2185.85
Original matrix $M_{[\pi=1]}$. Added as environmental factor	1.00 (0.97)	0.99 (0.88)	1.02 (0.76)	0.99 (0.82)	0.95 (0.36)	0.92 (0.13)	-0.04 (0.36)	0.75 (0.44)	2185.90
Shifted matrix $M_{[\pi=0.5]}$. Added as behavioral factor	1.00 (0.93)	0.99 (0.96)	1.04 (0.78)	1.02 (0.78)	1.03 (0.79)	0.92 (0.51)	-0.02 (0.51)	1.00 (0.98)	2233.89
Shifted matrix $M_{[\pi=0.5]}$. Added as environmental factor	0.99 (0.92)	0.99 (0.96)	1.04 (0.79)	1.02 (0.79)	1.03 (0.84)	0.92 (0.48)	-0.02 (0.67)	1.00 (0.99)	2234.08

Reference

- *Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International regional science review, 26(2), pp.153–166.*
- *Bivand, R.S., Gomes-Rubio, V. & Rue, H., 2015. Spatial Data Analysis with R - INLA with Some Extensions. Journal of Statistical Software, 63(20), pp.1–31.*
- *Blangiardo, M. & Cameletti, M., 2015. Spatial and spatio-temporal bayesian models with R-INLA, John Wiley & Sons.*
- *Kliewer, E. V, 1992. Influence of Migrants on Regional Variations of Stomach and Colon Cancer Mortality in the Western United States. International Journal of Epidemiology, 21(3), p.442.*
- *Leroux, B., Lei, X. & Breslow, N., 1999. Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence. In M. Halloran & D. Berry, eds. Statistical Models in Epidemiology, the Environment and Clinical Trials. New York: Springer-Verlag, pp. 135–178.*
- *Rogerson, P.A. & Han, D., 2002. The effects of migration on the detection of geographic differences in disease risk. Social Science and Medicine, 55(10), pp.1817–1828.*
- *Rue, H., Martino, S. & Chopin, N., 2009. Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 71(2), pp.319–392.*
- *Tiefelsdorf, M., 2007. Controlling for migration effects in ecological disease mapping of prostate cancer. Stochastic Environmental Research and Risk Assessment, 21(5), pp.615–624.*
- *Tiefelsdorf, M., Griffith, D.A. & Boots, B., 1999. A variance-stabilizing coding scheme for spatial link matrices. Environment and Planning A, 31(1), pp.165–180.*
- *United States Census Bureau, 1972. 1970 Census - Subject Reports: Migration Between State Economic Areas. Available at: https://www.census.gov/library/publications/1972/dec/pc-2-2e.html [Accessed March 25, 2017].*
- Berglund, Svante, and Anders Karlstrom. 1999. "Identifying Local Spatial Association in Flow Data." *Journal of Geographical Systems* 1: 219–36.
- Bivand, R S, V Gomes-Rubio, and H Rue. 2015. "Spatial Data Analysis with R - INLA with Some Extensions." *Journal of Statistical Software* 63 (20): 1–31. doi:http://dx.doi.org/10.18637/jss.v063.i20.
- Black, William R. 1992. "Network Autocorrelation in Transport Network and Flow Systems." *Geographical Analysis* 24 (3): 207–22. doi:10.1111/j.1538-4632.1992.tb00262.x.
- Chun, Yongwan. 2008. "Modeling Network Autocorrelation within Migration Flows by Eigenvector Spatial Filtering." *Journal of Geographical Systems* 10 (4): 317–44. doi:10.1007/s10109-008-0068-2.
- Chun, Yongwan, and Daniel A. Griffith. 2011. "Modeling Network Autocorrelation in Space-Time Migration Flow Data: An Eigenvector Spatial Filtering Approach." *Annals of the Association of American Geographers* 101 (3): 523–36. doi:10.1080/00045608.2011.561070.
- Dijkstra, Edsger W. 1959. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1 (1). Springer: 269–71.
- Fan, C. Cindy. 2005. "Modeling Interprovincial Migration in China, 1985-2000." *Eurasian Geography and Economics* 46 (3): 165–84. doi:10.2747/1538-7216.46.3.165.
- Fischer, Manfred M., and Daniel A. Griffith. 2008. "Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data in the European Union." *Journal of Regional Science* 48 (5): 969–89. doi:10.1111/j.1467-9787.2008.00572.x.
- Fotheringham, A Stewart. 1983. "A New Set of Spatial-Interaction Models: The Theory of Competing Destinations." *Environment & Planning A* 15 (1): 15–36.
- Getis, Authur, and Daniel A. Griffith. 2002. "Comparative Spatial Filtering in Regression Analysis." *Geographical Analysis* 34 (2): 130–40. doi:10.1353/geo.2002.0009.
- Griffith, Daniel A. 2003. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*. Springer Science & Business Media.
- Griffith, Daniel A. 1996. "Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data." *Canadian Geographer* 40 (4): 351–67. doi:10.1111/j.1541-0064.1996.tb00462.x.
- Hilbe, Joseph M. 2011. *Negative Binomial Regression*. Cambridge University Press.
- Stouffer, Samuel A. 1940. "Intervening Opportunities: A Theory Relating Mobility and Distance." *American Sociological Review* 5 (6): 845–67.
- ———. 1960. "Intervening Opportunities and Competing Migrants." *Journal of Regional Science* 2 (1): 1–26. doi:10.1111/j.1467-9787.1960.tb00832.x.

Reference

- Tiefelsdorf, M. 2003. “Misspecifications in Interaction Model Distance Decay Relations-A Spatial Structure Effect.” Journal of Geographical Systems 5: 25–50. doi:10.1007/s101090300102.
- Tiefelsdorf, M., D. A. Griffith, and B. Boots. 1999. “A Variance-Stabilizing Coding Scheme for Spatial Link Matrices.” Environment and Planning A 31 (1): 165–80. doi:10.1068/a310165.
- Tiefelsdorf, Michael. 2007. “Controlling for Migration Effects in Ecological Disease Mapping of Prostate Cancer.” Stochastic Environmental Research and Risk Assessment 21 (5): 615–24. doi:10.1007/s00477-007-0148-8.
- Tiefelsdorf, Michael, and Daniel A. Griffith. 2007. “Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach.” Environment and Planning A 39 (5): 1193–1221. doi:10.1068/a37378.
- Waller, Lance A., and Carol A. Gotway. 2004. Applied Spatial Statistics for Public Health Data. John Wiley & Sons.
- Anderson, C., Lee, D., & Dean, N. (2014). Identifying clusters in Bayesian disease mapping. Biostatistics, 15(3), 457–469. http://doi.org/10.1093/biostatistics/kxu005
- Bentham, G. (1988). Migration and morbidity: Implications for geographical studies of disease. Social Science and Medicine, 26(1), 49–54. http://doi.org/10.1016/0277-9536(88)90044-5
- Buell, P. (1973). Changing incidence of breast cancer in japanese-american women. Journal of the National Cancer Institute, 51(5), 1479–1483. http://doi.org/10.1093/jnci/51.5.1479
- Cohen, B. L., & Colditz, G. A. (1994). Tests of the Linear-No Threshold Theory for Lung Cancer Induced by Exposure to Radon. Environmental Research, 64, 65–89.
- Colonna, M. (2004). Bootstrap investigation of the stability of disease mapping of Bayesian cancer relative risk estimations. European Journal of Epidemiology, 19(8), 761–768. http://doi.org/10.1023/B:EJEP.0000036570.33081.fa
- Elebead, F. M., Hamid, A., Hilmi, H. S. M., & Galal, H. (2012). Mapping cancer disease using geographical information system (gis) in gezira state-Sudan. Journal of Community Health, 37(4), 830–839. http://doi.org/10.1007/s10900-011-9517-9
- Elford, J., Phillips, A. N., Thomson, A. G., & Shaper, A. G. (1989). Migration and geographic variations in ischaemic heart disease in Great Britain. The Lancet, (18th February 1989), 343–346.
- Federation of Tax Administrators. (2014). Tax, 49, 12–25. Retrieved from http://www.taxadmin.org/assets/docs/Tobacco/papers/tax_burden_2014.pdf [Accessed January 1, 2017].
- Gelman A & Price NP. (2006). Radon Project. Available at: http:// www.stat.columbia.edu/radon/ [Accessed July 1, 2006].
- Hampton, K. H., Serre, M. L., Gesink, D. C., Pilcher, C. D., & Miller, W. C. (2011). Adjusting for sampling variability in sparse data: geostatistical approaches to disease mapping. International Journal of Health Geographics, 10(1), 54. http://doi.org/10.1186/1476-072X-10-54
- Jarup, L. (2004). Health and environment information systems for exposure and disease mapping, and risk assessment. Environmental Health Perspectives, 112(9), 995–997. http://doi.org/10.1289/ehp.6736
- Lagazio, C., Biggeri, A., & Dreassi, E. (2003). Age-period-cohort models and disease mapping. Environmetrics, 14(5), 475–490. http://doi.org/10.1002/env.600
- Lawson, A. B., & Clark, A. (2002). Spatial mixture relative risk models applied to disease mapping. Statistics in Medicine, 21(3), 359–370. http://doi.org/10.1002/sim.1022
- National Cancer Institute. (2017a). Cancer Mortality Maps. Retrieved February 1, 2017, from https://ratecalc.cancer.gov/
- National Cancer Institute. (2017b). Geographic Data. Retrieved February 1, 2017, from https://ratecalc.cancer.gov/geo.jsp
- Rogerson, P. A., & Han, D. (2002). The effects of migration on the detection of geographic differences in disease risk. Social Science and Medicine, 55(10), 1817–1828. http://doi.org/10.1016/S0277-9536(01)00316-1
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 71(2), 319–392. http://doi.org/10.1111/j.1467-9868.2008.00700.x
- Tiefelsdorf, M. (2007). Controlling for migration effects in ecological disease mapping of prostate cancer. Stochastic Environmental Research and Risk Assessment, 21(5), 615–624. http://doi.org/10.1007/s00477-007-0148-8
- Tong, S. (2000). Migration Bias in Ecologic Studies Author (s): Shilu Tong Published by : Springer Stable URL : http://www.jstor.org/stable/3581728 Migration bias in ecologic studies, 16(4), 365–369.
- Torabi, M., & Rosychuk, R. J. (2011). Spatio-temporal modelling using B-spline for disease mapping: analysis of childhood cancer trends. Journal of Applied Statistics, 38(9), 1769–1781. http://doi.org/10.1080/02664763.2010.529877
- United States Census Bureau. (1975). Migration Between State Economic Areas. 1970 Census of Population., Subject Re.
- Verheij, R. A., Mheen, H. D. Van De, Bakker, D. H. De, Groenewegen, P. P., Mackenbach, J. P., Mheen, V. De, … Dike, H. (2016). Urban-Rural Variations in Health in the Netherlands : Does Selective Migration Play a Part ? All use subject to JSTOR Terms and Conditions Urban-rural Netherlands : in health in the variations does selective migration play a part ?, 52(8), 487–493.
- Wheeler, D. C., Ward, M. H., & Waller, L. a. (2012). Spatial-Temporal Analysis of Cancer Risk in Epidemiologic Studies with Residential Histories. Annals of the Association of American Geographers, 102(5), 1049–1057. http://doi.org/10.1080/00045608.2012.671131