

Modeling migration

Ernesto F. L. Amaral

September 29, 2020
Migration (SOCL 647)



www.ernestoamaral.com

Outline

- Mathematical models to smooth migration rates
 - Rogers, Castro 1981
- Improved measures of age profiles of migration
 - Bernard, Bell, Charles-Edwards 2014
- Migration flows from population stocks of infants
 - Rogers, Jordan 2004
- Log-linear models for migration
 - Raymer, Rogers 2007
- Gravity models
 - Stillwell 2005
- Analysis of spatial association
 - Anselin 1995
- Spatial analysis
 - LeSage, Pace 2008
- Agent-based models and simulations
 - Klabunde, Willekens 2016



Smooth migration rates

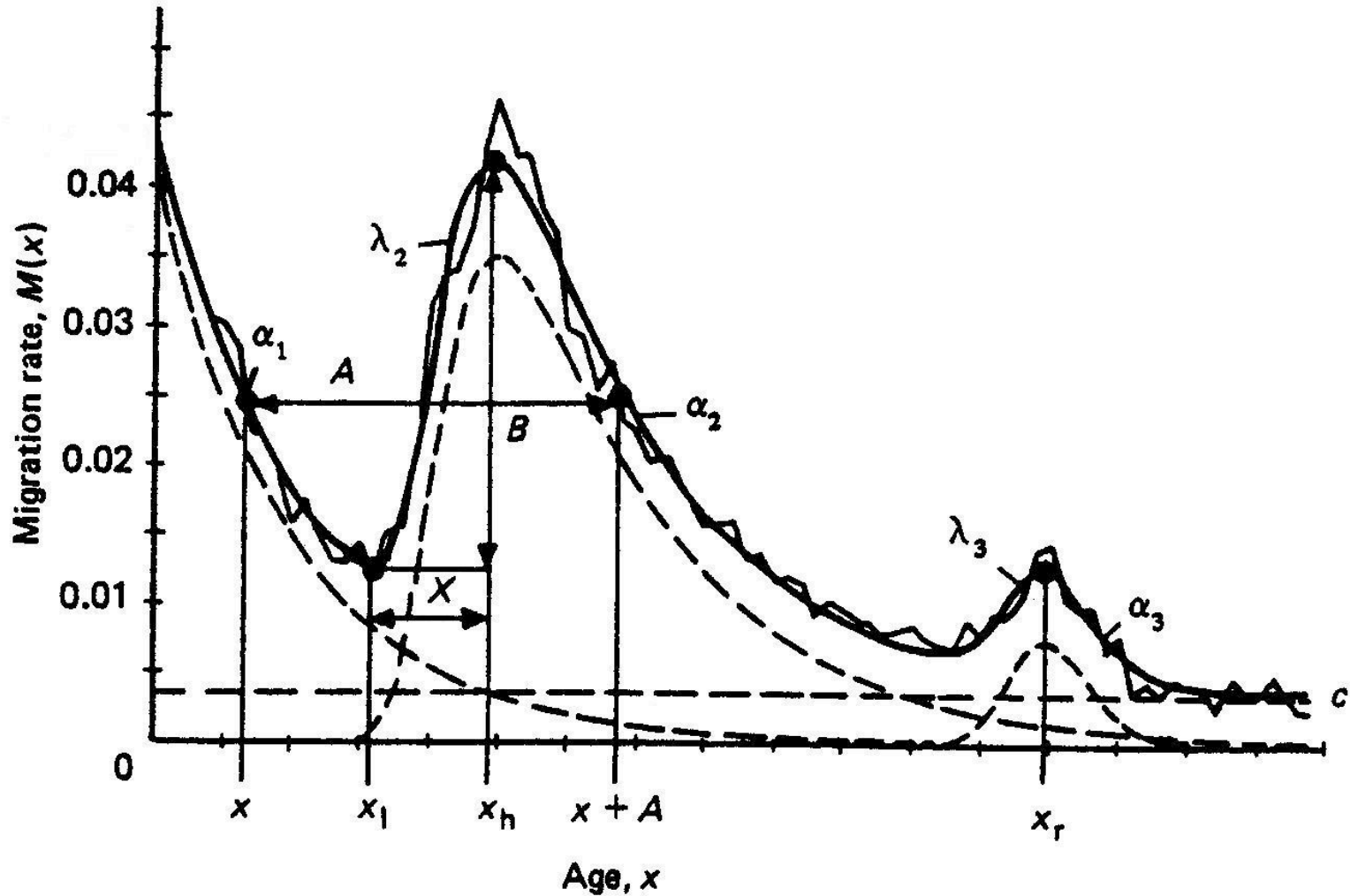
- After the estimation of migration rates by age group, mathematical models can be implemented on the results (Rogers, Castro 1981; Rogers, Jordan 2004)
- Regularities found in migration rates by age help develop hypothetical migration models that can be used in population studies with limited or inadequate data

Modeling migration schedules

- Mathematical models can be applied to estimated migration rates, in order to (Rogers, Castro 1981; Rogers, Jordan 2004)
 - Smooth the curves of migration rates
 - Originate parameters to assist in understanding levels and patterns of population flows among areas
- The mathematical proposition establishes that
 - Migration is highly influenced by economics, thus curves designate different moments of an individual's entrance into the labor market
 - The migration schedule is composed of four components related to the labor market



Model migration schedule



Source: Rogers, Castro 1981, p.6.



Four components of migration schedule

- **Pre-labor curve** is a negative exponential curve from 0 to 19 years-of-age (α_1 as the descendent indicator; a_1 as the level indicator)
- **Labor-age curve** has a parabolic shape (μ_2 as the mean age indicator; λ_2 as the ascendant indicator; α_2 as the descendent indicator; a_2 as the level indicator)
- **Post-labor curve** is a small parabola signifying the individuals around 65 years-of-age (μ_3 as the mean age indicator; λ_3 as the ascendant indicator; α_3 as the descendent indicator; a_3 as the level indicator)
- **A constant** is the last parameter of the model schedule (c), which adjusts the level of migration rates to the mathematic expression

Source: Rogers, Castro 1981.



Basic model migration schedule

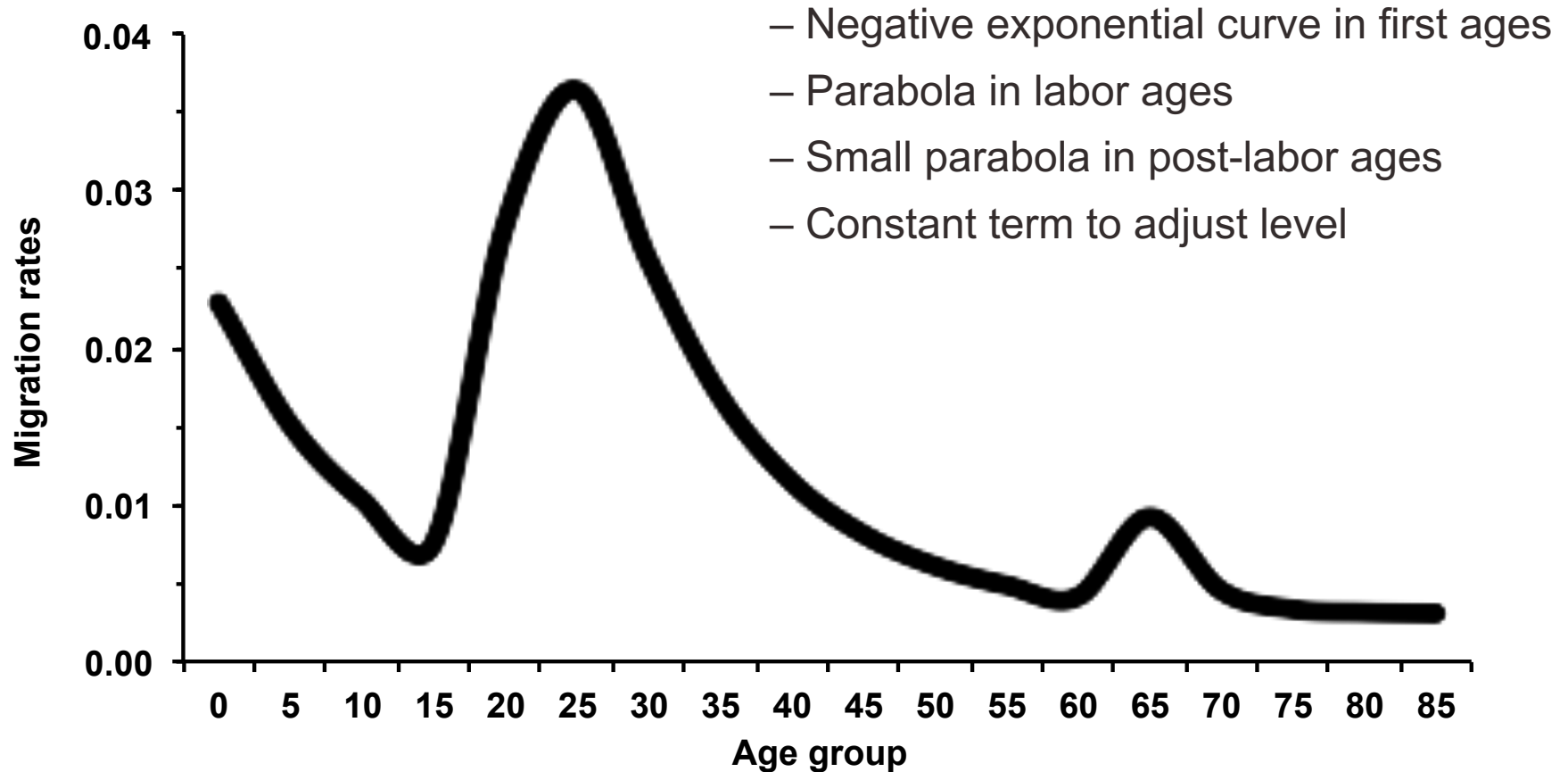
- It has a parabola in post-labor ages
- This equation has 11 parameters

$$\begin{aligned}M(x) = & a_1 * \exp(-\alpha_1 x) \\ & + a_2 * \exp\{-\alpha_2(x-\mu_2) - \exp[-\lambda_2(x-\mu_2)]\} \\ & + a_3 * \exp\{-\alpha_3(x-\mu_3) - \exp[-\lambda_3(x-\mu_3)]\} \\ & + c\end{aligned}$$

Source: Rogers, Castro 1981.



Basic migration model



Source: Rogers, Castro 1981.

Migration model schedule with an upward slope

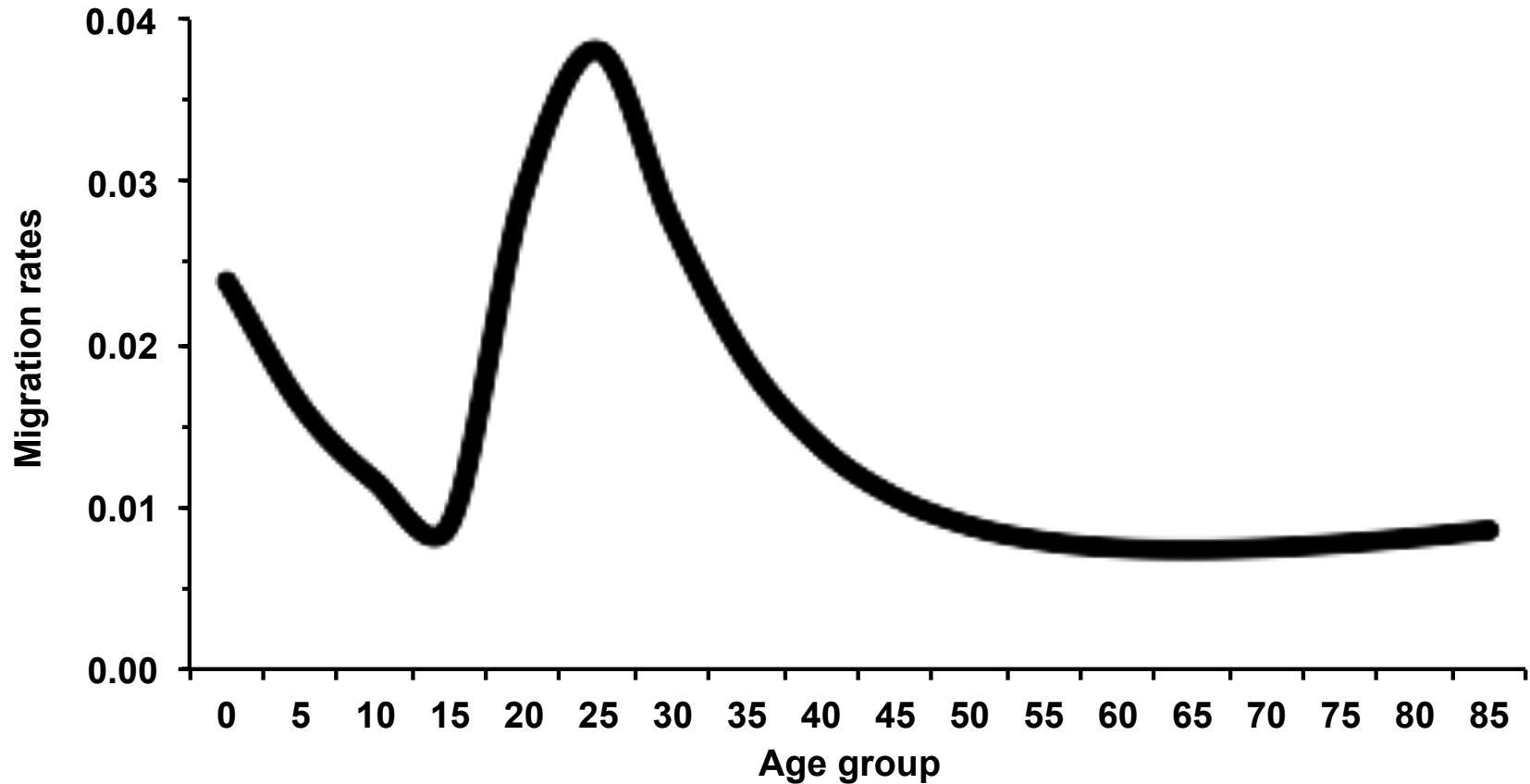
- It has a linear function in post-labor ages
- This equation has 9 parameters

$$\begin{aligned}M(x) = & a_1 * \exp(-\alpha_1 x) \\ & + a_2 * \exp\{-\alpha_2(x-\mu_2) - \exp[-\lambda_2(x-\mu_2)]\} \\ & + a_3 * \exp(\alpha_3 x) \\ & + c\end{aligned}$$

Source: Rogers, Castro 1981.



Migration model schedule with an upward slope



Source: Rogers, Castro 1981.



Reduced model

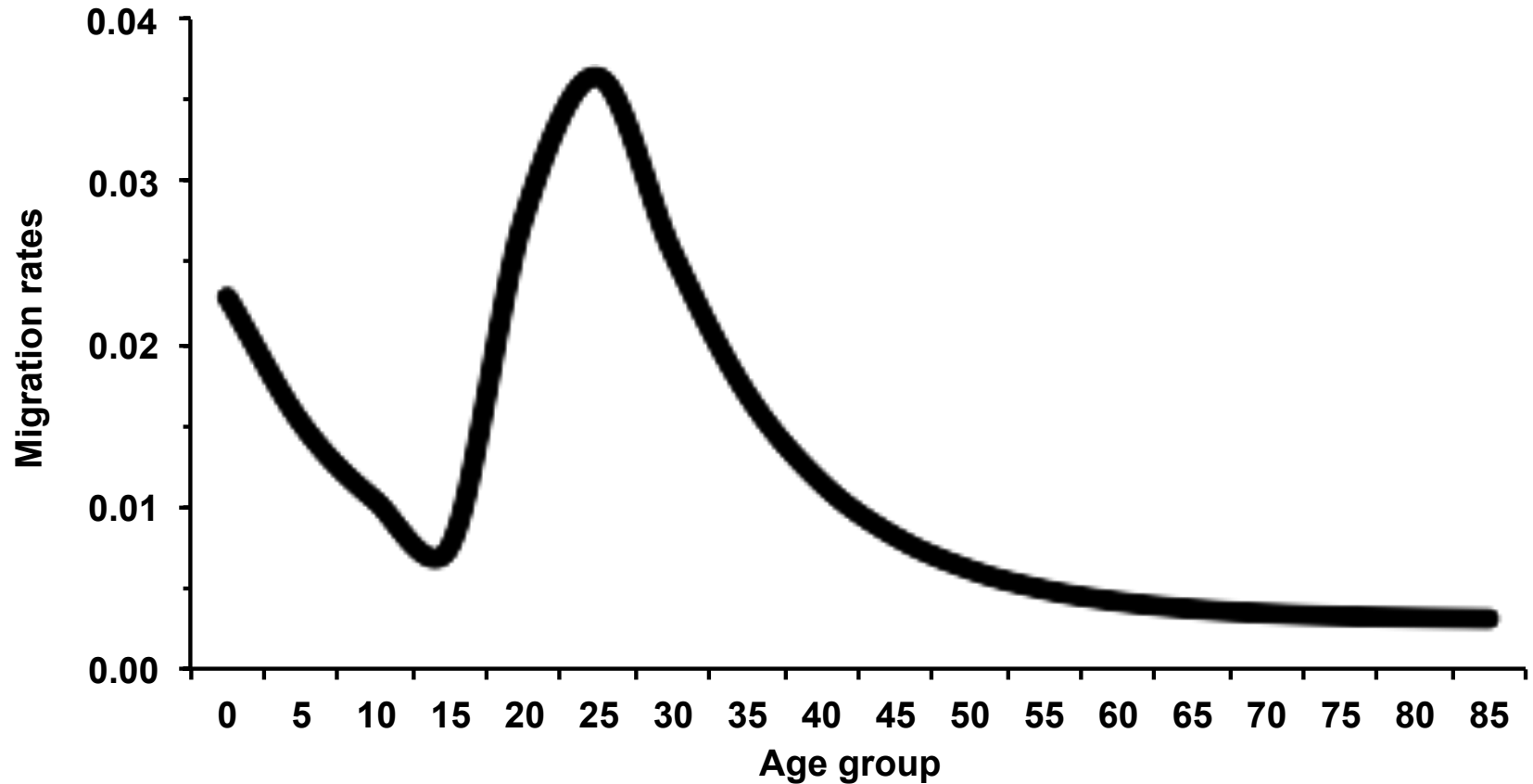
- It has a constant value in post-labor ages
- This equation has 7 parameters

$$M(x) = a_1 * \exp(-\alpha_1 x) \\ + a_2 * \exp\{-\alpha_2(x-\mu_2) - \exp[-\lambda_2(x-\mu_2)]\} \\ + c$$

Source: Rogers, Castro 1981.



Reduced model



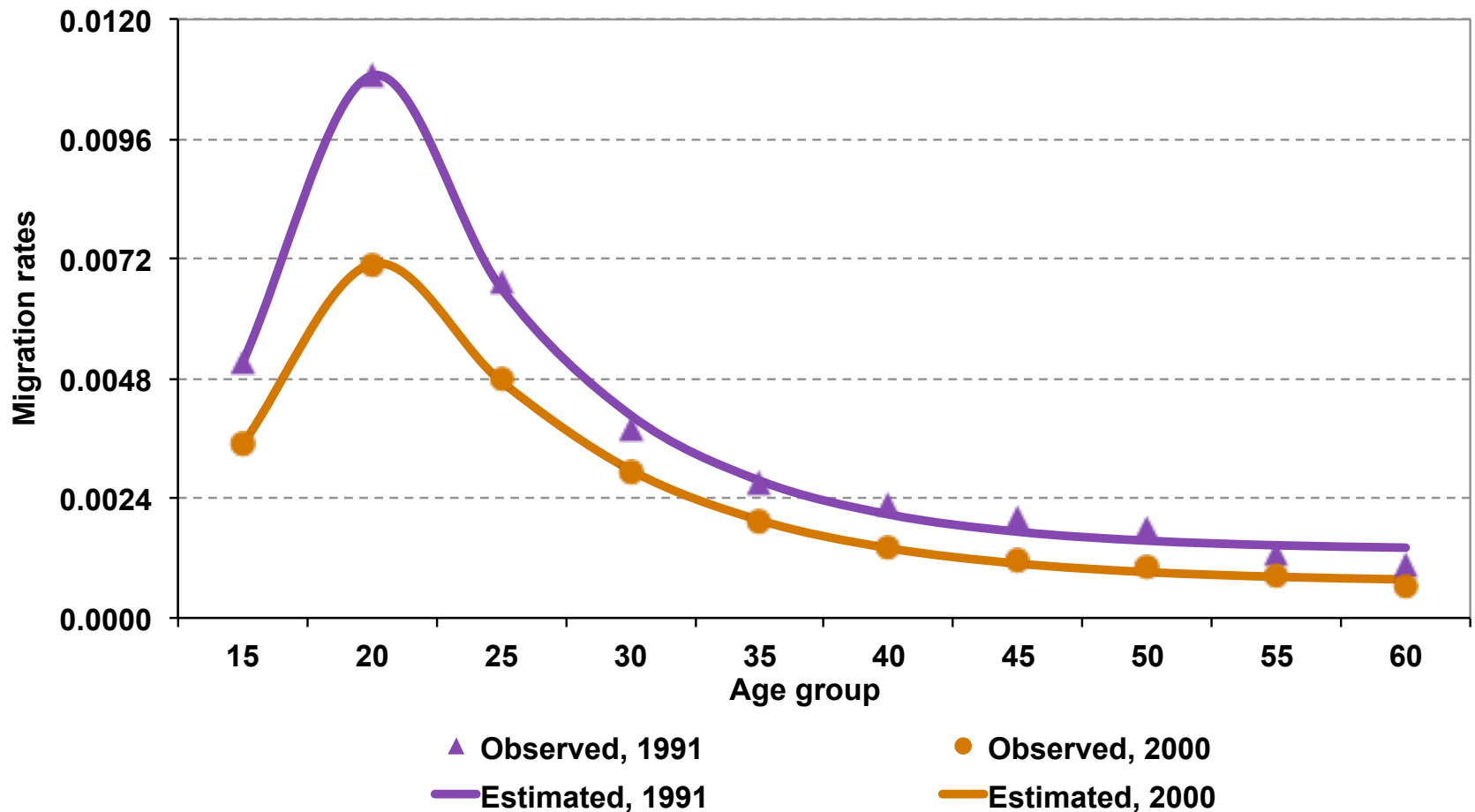
- Following example was done for men 15–64 years old...

Source: Rogers, Castro 1981.



Northeast to Southeast, Males, Brazil

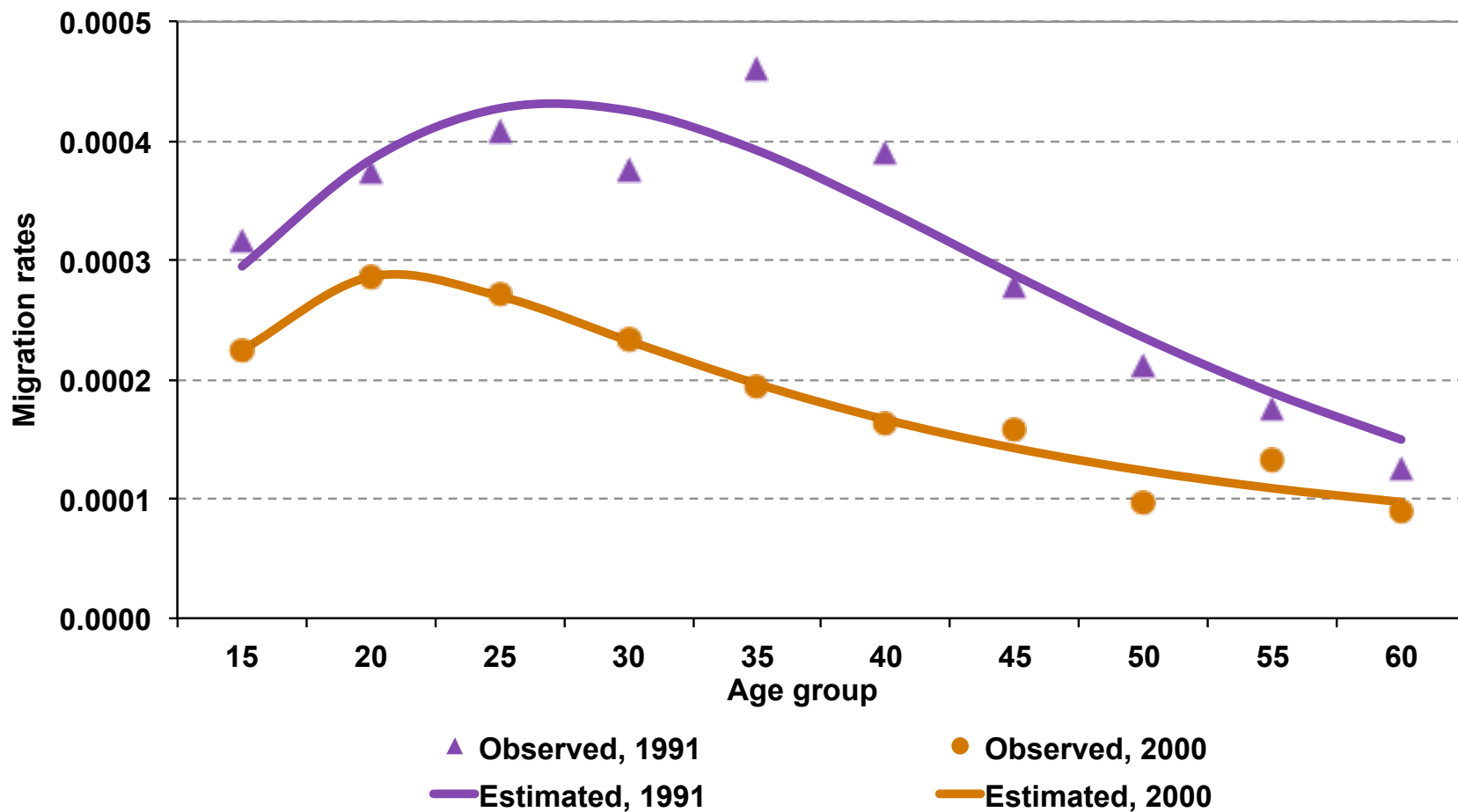
(place of residence 5 years before the census)



Source: Amaral et al. 2016.

North to Southeast, Males, Brazil

(place of residence 5 years before the census)



Source: Amaral et al. 2016.

Application of mathematical models

- SPSS
 - Non-linear regression. Menu: “Analyze”/“Regression”/“Non-linear”
 - Levenberg-Marquardt estimation method
- MATLAB
(<https://www.mathworks.com/products/matlab.html>)
- TableCurve 2D
(<https://systatsoftware.com/products/tablecurve-2d/>)
 - Graphical interface that helps the definition of initial values of parameters
 - We can use parameters from previous estimation, instead of maintaining initial values as SPSS
 - Test values for parameters before we estimate the model
- Application developed by Dr. Reinaldo dos Santos
(https://demometrics.shinyapps.io/RogersCastroModel_LCmetrics/)



Usefulness of migration schedules

- Model migration schedules by Rogers and Castro represent a powerful conceptual tool
- They provide a useful device for exploring the relationship between age and migration
- They have the singular strength that they capture the full variation in the propensity to migrate that occurs across the age profile
- They are applied widely in other areas, such as the development of inputs to population projections



Limitations of model schedule parameters

- Variability in the number and value of parameters
 - Since different numbers of parameters may be used to estimate model schedules, comparisons are compromised
- Sensitivity of parameter estimates to initial value selection
 - Change in initial parameters can result in varying final parameters
 - High level of correlation among parameters within a model
- Instability of parameter estimates
 - Parameters can be set to different values, but they can deliver similar degrees of goodness of fit (over-parameterization)
- Comparability of parameter estimates
 - Large set of parameters makes it difficult to compare countries
- Interpretability of parameter estimates is challenging
 - E.g., α_1 , α_2 , λ_2 do not best capture the slope of component curves. Rate of change (first derivative) provides more accurate measure



TEXAS A&M
UNIVERSITY.

Improved measures of age profiles of migration

Name of measure	Definition
Intensity at peak	The peak value of migration intensity.
Age at peak	The age at which migration intensity peaks.
Breadth of peak	The sum of the migration intensities for each of the five years of age before and after the age at peak and for the age at which migration peaks.
MURC	The maximum rate of change in the upward slope of the labour force curve.
MDRC	The maximum rate of change in the downward slope of the labour force curve.
Asymmetry	The ratio of the maximum upward to the maximum downward rate of change: MURC/MDRC



MURC & MDRC

- The maximum rate of change provides a more accurate measure of the migration slopes
 - Instead of exponential coefficients
- Maximum upwards rate of change (MURC)
 - Capture upward slope of labor force curve
- Maximum downwards rate of change (MDRC)
 - Capture downward slope of labor force curve



Estimating MURC & MDRC

- These measures ensure more consistent discrimination between countries whose migration profiles have different shapes
 - They can be calculated without estimating model schedules
 - Take difference in migration intensity between two consecutive ages
 - Repeat this sequentially across the relevant age range, and identify the age at which the maximum rate of change occurs



MURC & MDRC avoid problems

- Similar measures could also be computed by taking the first derivative of the model migration schedule
- However, the simplified approach ensures that results are not prejudiced by problems of parameter variability, sensitivity, or instability



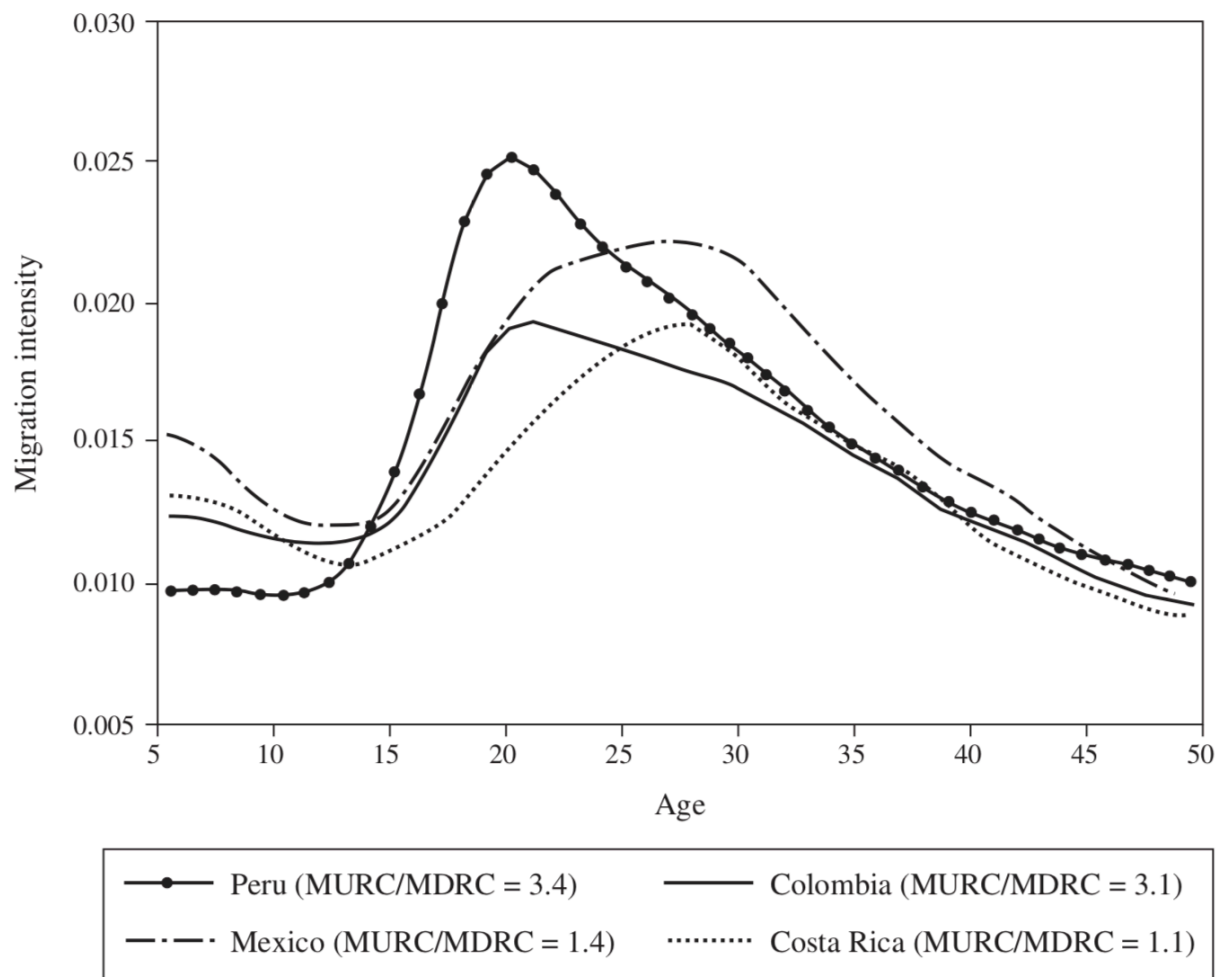


Figure 4 Age profiles of migration for selected countries showing different degrees of asymmetry around the age of peak migration

Note: Kernel-based smoothing. MURC/MDRC is the ratio of the maximum upward and downward rates of change; the higher the value the more asymmetrical the curve.

Source: IPUMS database: Peru (2002–07), Colombia (2000–05), Mexico (1990–95), and Costa Rica (1995–2000).

Principal characteristics of migration

- The complexity of the age profile of migration can be reduced to two principal characteristics
 - Associated with other features of the age profile
- **Age at which migration peaks (x-axis)**
 - It relates to the symmetry of the labor force curve
 - Symmetry increases as age at peak rises
- **Intensity of migration at the peak (y-axis)**
 - It shapes the slopes demarcating the labor force curve
 - As intensity increases, the upward and downward slopes progressively steepen
 - This relationship happens only when the slopes are expressed via the rate of change



Factor analysis

- Factor analysis shows that these two metrics (age and intensity at peak) account for 67% of inter-country variance across 25 countries

Table 4 Factor loadings of six metrics of migration age profile for 25 countries

Migration metric	Factor 1 Migration Concentration	Factor 2 Age Selectivity	Unique variance
Intensity at peak	0.99	0.10	0.01
MURC	0.94	0.30	0.03
MDRC	-0.96	0.12	0.07
Breadth of peak	0.94	0.03	0.12
Age at peak	-0.30	-0.78	0.30
Asymmetry	0.07	0.90	0.18
Proportion of total variance	0.63	0.26	

Note: Orthogonal rotation with Kaiser Normalization. Refer to [Table 2](#) for definitions of the measures. Values greater than 0.75 or lower than -0.75 are in bold.

Source: As for [Table 1](#).



Age at peak vs. Intensity at peak

- Plotting age at peak against intensity at peak provides evidence of regional variations in the age profile of internal migration among countries

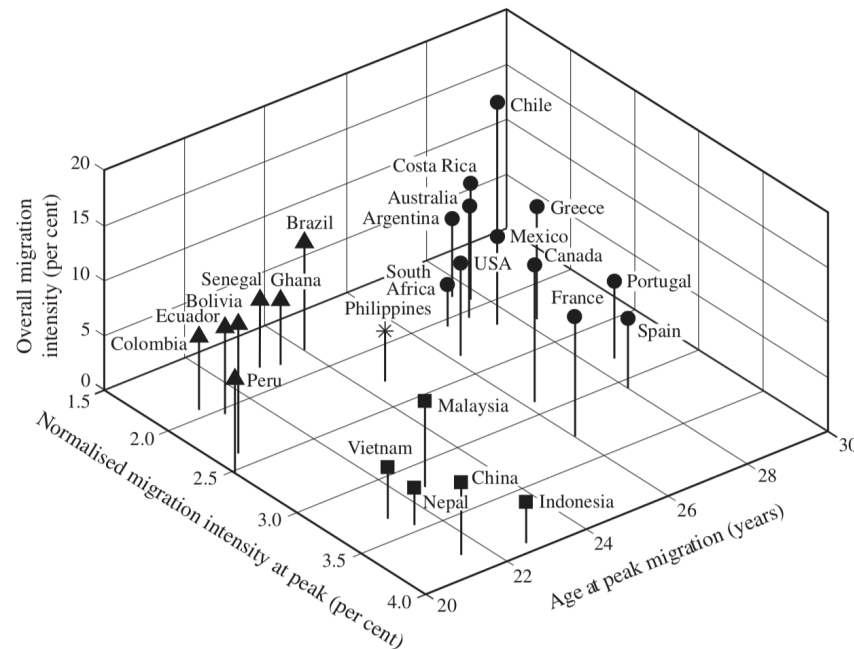
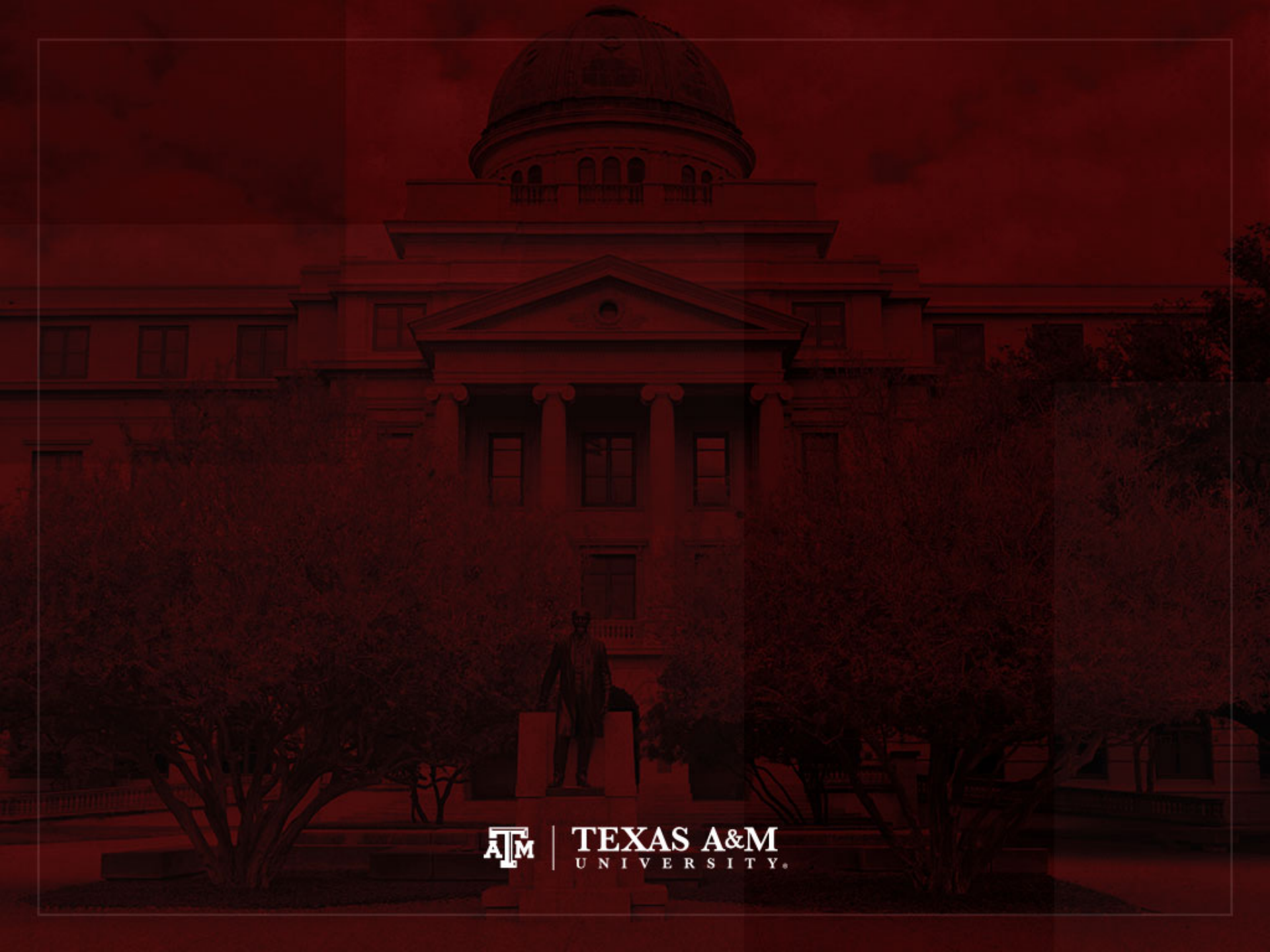


Figure 5 Age at peak, normalized intensity at peak, and overall migration intensity for 25 sample countries, indicating clusters of countries with similar age profiles of migration

Note: Overall migration intensity refers to the migration rates across the entire population. Clusters were defined according to a three-solution k -means cluster based on variables normalized to unit variance. The Philippines were excluded from the cluster analysis because their migration age patterns differ from those of the other sample countries in the same region, and also from patterns seen in the other major world regions.

Source: IPUMS database.





TEXAS A&M
UNIVERSITY.

Migration flows from stocks of infants

- When adequate data on migration are unavailable, we need to use indirect techniques
- Migration stocks of children (0–4 age group) is used to estimate rates for all other age groups
- We know that children who live in region j at census date and who were born in region i migrated during the previous five years
- **Assumption:** these children migrated only once during this period, because they were born on average two and a half years before the census



Model components

- $S_{ij}(+)$: fraction of persons of all ages who resided in region i at the start of time interval and in region j at the end of interval
- $S_{ij}(-5)$: fraction of all births born in region i during the past five years who survived to the census date to enter the 0–4 years age group resident in region j at that date
- ${}_iK_j(+)\%$: percentage of i -borns of all ages who are enumerated in region j at census time
- $S_{ij}(x)$: regression equation is used to estimate migration flows for other age groups x

$$S_{ij}(x) = a + b(x)S_{ij}(-5) + c(x){}_iK_j(+)\% + \text{error term}$$



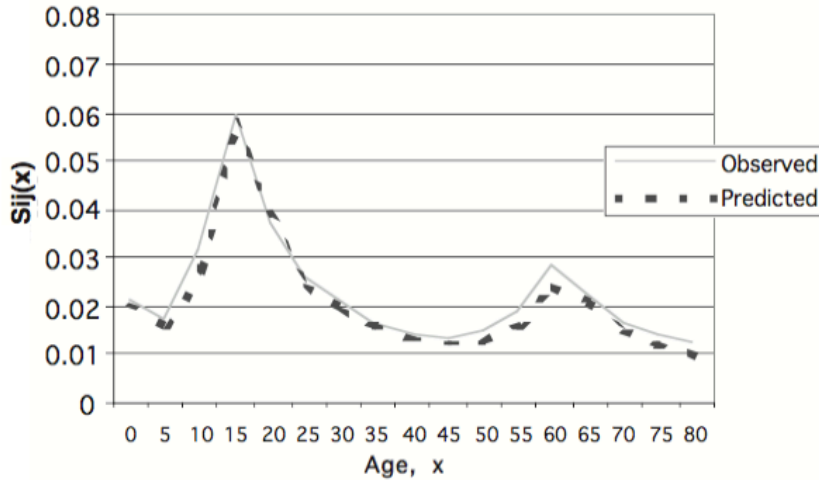
Logistic regression

- To ensure that the estimated conditional survivorship proportions are always non-negative, and range between zero and one unit, we estimate a logistic regression
- Thus, instead of predicting the survivorship proportions using a linear estimation approach, the logged odds of the survivorship are predicted, then converted back into probabilities

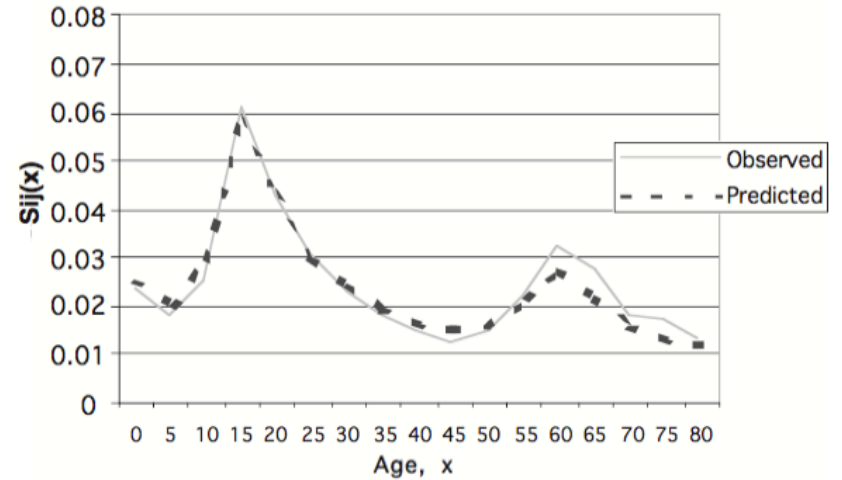
$$\ln\left(\frac{S_{ij}(x)}{1 - S_{ij}(x)}\right) = a(x) + b(x)S_{ij}(-5) + c(x)_i K_j(+) \% + error$$



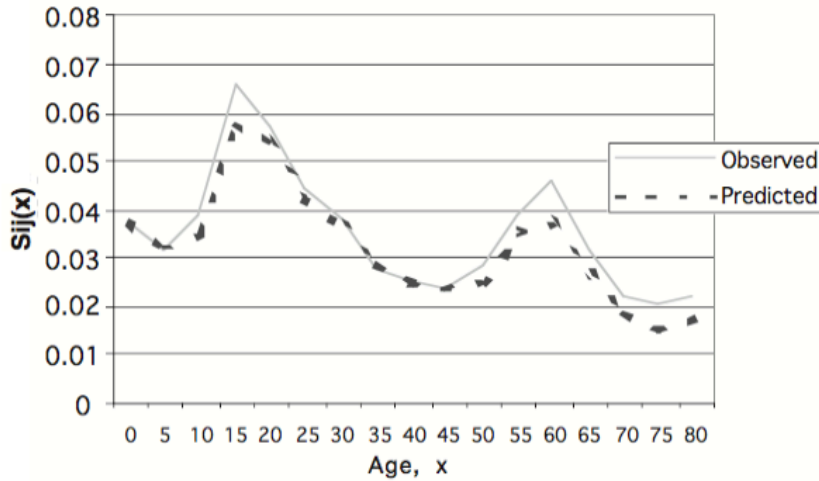
Predicted and Observed $S_{ij}(x)$ from NE to 1960



Predicted and Observed $S_{ij}(x)$ from NE to 1970



Predicted and Observed $S_{ij}(x)$ from NE to 1980



Predicted and Observed $S_{ij}(x)$ from NE to 1990

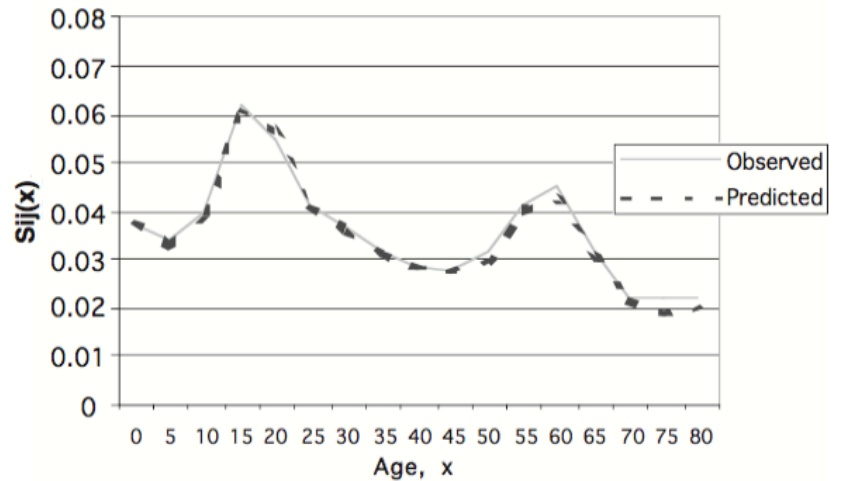
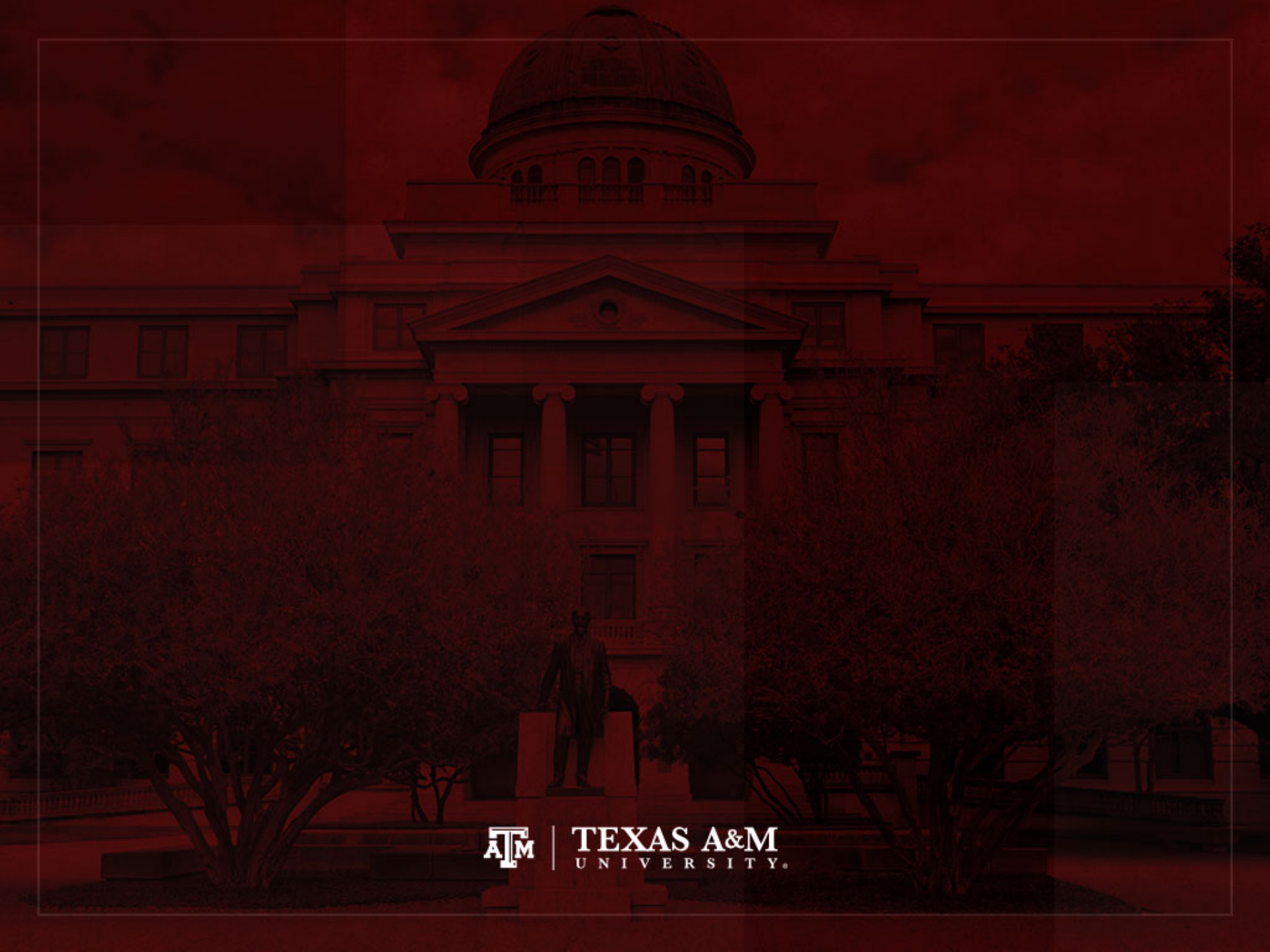


FIG. 4B. Predicted and Observed Migration Schedules from the Northeast to South

Quality of results

- Age patterns follow standard regularities observed in empirical schedules
 - But the flows tend to be underestimated by 9%
- 20–24 age group is the best predicted by the model
 - Mean average percent error (MAPE) of 11%
- 80–84 age group is the worst predicted by the model
 - MAPE of 34%
 - Infant migration better predicts the migration flow of parents instead of grandparents
- Indirect estimation of migration flows (numbers) and propensities (probabilities) is possible by the persistent regularities observed in demographic data





TEXAS A&M
UNIVERSITY.

Log-linear models for migration

- Raymer, Rogers (2007) propose the estimation of interregional migration flows using age and spatial structures
- Useful for countries with inadequate, inaccurate, or incomplete data-reporting systems
- Other countries might not have detailed information on migration in recent years
 - U.S. has migration information only on American Community Survey (not on Census since 2010)



Multiplicative component approach

- Interregional migration flows can be disaggregated into four separate components (without age)
- **Overall** component: level of migration
- **Origin** component: relative “pushes” from each region
- **Destination** component: relative “pulls” to each region
- **Two-way origin-destination** interaction component: impacts of physical or social distance between places, which are not explained by the overall and main effects



Multiplicative component model

$$n_{ijx} = (T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})(ODA_{ijx})$$

- n_{ijx} : observed flow of migration from region i to region j in age group x
- T : total number of migrants (n_{++})
- O_i : proportion of all migrants leaving from region i (n_{i+}/n_{++})
- D_j : proportion of all migrants moving to region j (n_{+j}/n_{++})
- OD_{ij} : ratio of observed migration to expected migration $n_{ij}/[(T)(O_i)(D_j)]$
- A_x : proportion of all migrants in age group x



Log-linear model

- The multiplicative component descriptive model...

$$n_{ijx} = (T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})(ODA_{ijx})$$

- ... can be expressed as a saturated log-linear statistical model

$$\ln(n_{ijx}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_x^A + \lambda_{ij}^{OD} + \lambda_{ix}^{OA} + \lambda_{jx}^{DA} + \lambda_{ijx}^{ODA}$$



Other considerations

- These models are hierarchical
 - For two-way interaction terms, the main effect parameters must be included
 - For three-way interaction terms, all the main effects and two-way interactions must be included
- Migration flow tables are complicated because they can mix migrants with non-migrants or intraregional migrants
 - Structural zeros can be inserted to remove non-migrants (quasi-independence model)



Example of migration among four regions of Mexico



Table 3. The Spatial Structure of Interregional Migration in Mexico, 1995–2000

Origin	Destination				
	Border	North Central	Central	South	Total
A. Observed Flows					
Border	0	122,915	69,709	20,883	213,507
North Central	308,712 n_{21}	0	134,961	28,589	472,262 n_{2+}
Central	278,185	219,251	0	199,803	697,239
South	89,973	89,041	201,156	0	380,170
Total	676,870 n_{+1}	431,207	405,826	249,275	1,763,178 n_{++}
B. Multiplicative Components					
Border	0.000	2.354	1.419	0.692	0.121
North Central	1.703 OD_{21}	0.000	1.242	0.428	0.268
Central	1.039	1.286	0.000	2.027	0.395
South	0.616	0.958	2.299	0.000	0.216
Total	0.384	0.245	0.230	0.141	1,763,178

Note: Numbers refer to Mexican-born persons.

$$n_{21} = (T)(O_2)(D_1)(OD_{21}) = n_{++} \left(\frac{n_{2+}}{n_{++}} \right) \left(\frac{n_{+1}}{n_{++}} \right) \left[\frac{n_{21}}{(n_{++}) \left(\frac{n_{2+}}{n_{++}} \right) \left(\frac{n_{+1}}{n_{++}} \right)} \right]$$

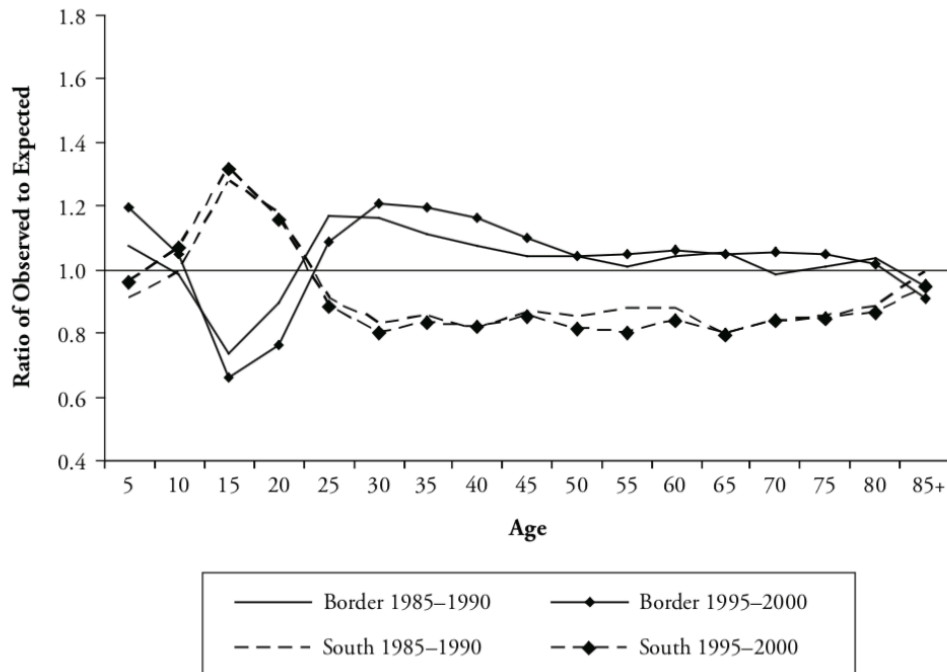
$$n_{21} = (1,763,178) \left(\frac{472,262}{1,763,178} \right) \left(\frac{676,870}{1,763,178} \right) \left(\frac{308,712}{181,452} \right)$$

$$n_{21} = (1,763,178)(0.268)(0.384)(1.703) = 308,712$$

OD_{21} (origin-destination association): ratio of observed migration to expected migration. There were 70% more migrants than expected.

Interregional Migration in Mexico, 1985–1990 and 1995–2000

A. Origin-Age Interaction Component

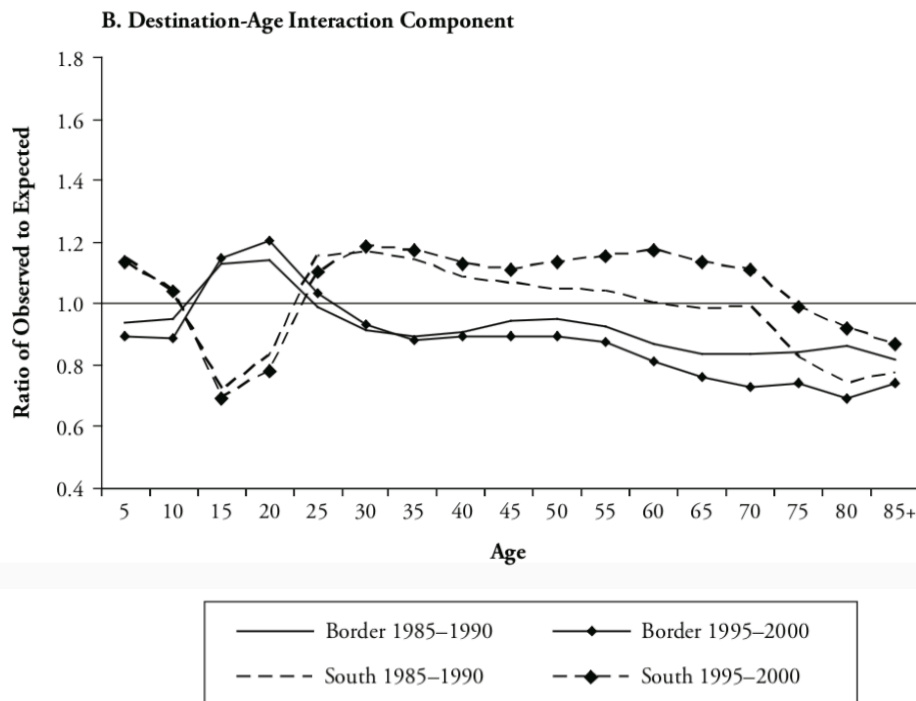


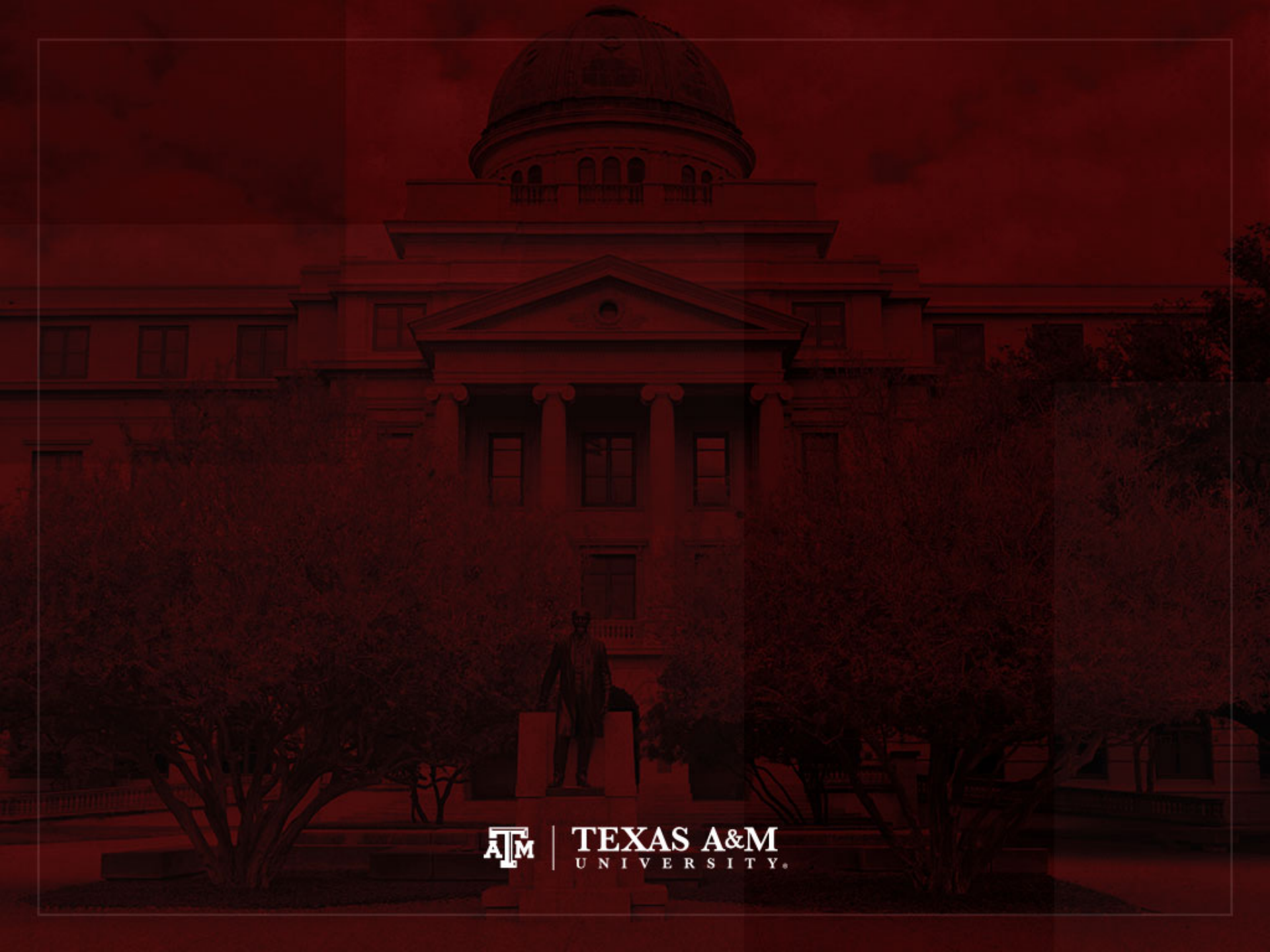
- Children (5–9) were more likely to migrate from Border region
- Young adults (15–24) were more likely to migrate from South region
- Young adults (15–24) were less likely to migrate from Border region
- Adults (25+) were more likely to migrate from Border region
- Adults (25+) were less likely to migrate from South region



Interregional Migration in Mexico, 1985–1990 and 1995–2000

- Children (5–9) were more likely to migrate to South region
- Young adults (15–24) were more likely to migrate to Border region
- Young adults (15–24) were less likely to migrate to South region
- Adults (25+) were more likely to migrate to South region
- Adults (30+) were less likely to migrate to Border region





TEXAS A&M
UNIVERSITY.

Gravity models

- Gravity models are usually implemented to predict the likelihood of migration, using distance as the main **exogenous factor** (Head 2000; Lowry 1966; Pöyhönen 1963; Stillwell 2005, 2009; Tinbergen 1962)
- Gravity models address the distance between areas, as well as the changing population in the areas over time
 - Distance is expected to play an intervening role on the levels of population streams
 - Distance is constant over time, but population growth affects out- and in-migration trends
- Based on the regional equilibrium framework, the idea behind these models is to use the distance between areas and population trends to estimate the level of migration between areas



Poisson regression

- Gravity models use population at the beginning of the period (P_i), population at the end of the period (P_j), and distance between areas (d_{ij}) to estimate migration flows (Head 2000; Lowry 1966; Pöyhönen 1963; Stillwell 2005, 2009; Tinbergen 1962)

$$M_{ij} = \exp(b_0 + b_1 \log P_i + b_2 \log P_j + b_3 \log d_{ij}) + \varepsilon_{ij}$$

- M_{ij} : counts of migrants at the end of the period between areas of origin (i) and destination (j)
- b_0 : constant
- b_1 : coefficient associated with the population at the beginning of the period (P_i)
- b_2 : coefficient associated with the population at the end of the period (P_j)
- b_3 : coefficient related to the distance between areas (d_{ij})
- ε_{ij} : random error term associated with all pairs of areas

Zero-inflated Poisson models

- Zero-inflated Poisson statistical regressions can generate gravity models for inter-regional migration flows (Stillwell 2005, 2009)
- Dependent variable is measured in discrete units (integer counts of migrants) and a discrete probability distribution
- These models are appropriate, because
 - They do not maintain error variances as constant for the different sizes of estimated flows, as is the case of “log-normal” models
 - The model is also recommended when there are a significant number of small flows among areas, no flows (zero migrants) among areas, and/or a small number of larger migration flows

Need to include covariates

- Anderson and Van Wincoop (2003) estimate gravity models to predict trade between countries
 - Distance is usually the main independent variable in gravity models
 - Authors make gravity models more appropriate for their subject by adding covariates that have influence on the dependent variable
 - They add national border barriers as a set of independent variables, which was not previously performed in their field
- We need to be aware of other factors that might influence population flows
 - E.g., year, age, sex, race/ethnicity, marital status, education, population in origin, population in destination



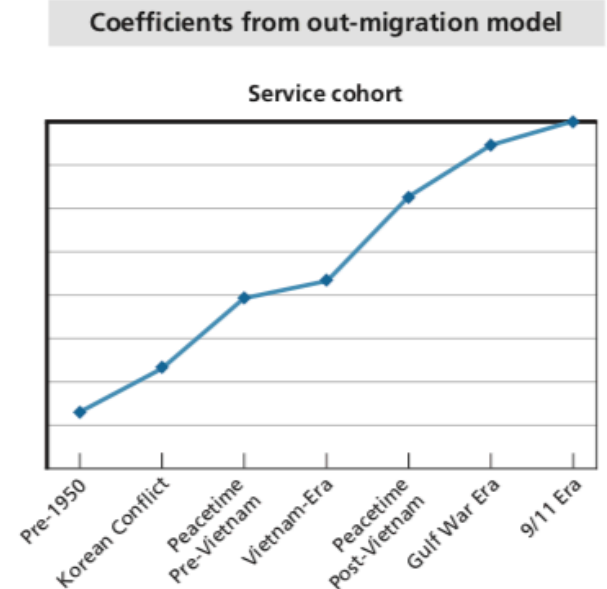
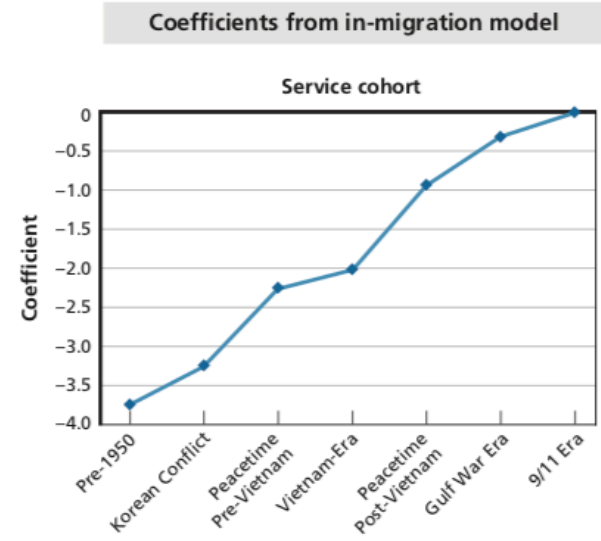
Example of projection of veteran population

- Project the geographic distribution of the veteran population from 2014 to 2024 by age, sex, race/ethnicity, and service era
- We considered migration among 2,351 Public Use Microdata Areas (PUMA)
 - Data: 2009–2013 American Community Survey (ACS)
- Gravity models estimated migration rates as a function of
 - Population in the area of origin (at the beginning of the period)
 - Population in the area of destination (at the end of the period)
 - Squared distance between areas
 - Age, sex, race/ethnicity, service era



Estimates from Zero-Inflated Poisson Regression Models for Number of Migrants (Dependent Variable)

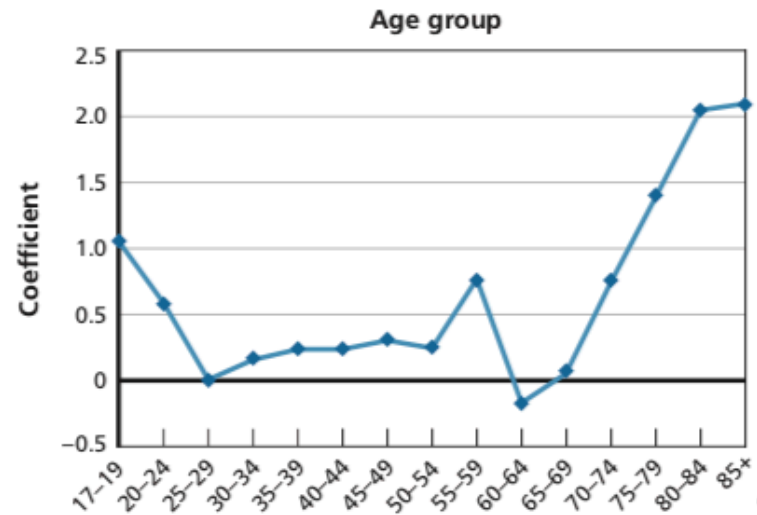
Independent Variables	In-Migration Model	Out-Migration Model
Constant	-2.254*** (0.0148)	-1.803*** (0.0138)
Female	ref.	ref.
Male	-1.168*** (0.0117)	-1.321*** (0.0108)
9/11 era	ref.	ref.
Gulf War era	-0.313*** (0.0154)	-0.264*** (0.0153)
Peacetime post-Vietnam	-0.930*** (0.0193)	-0.874*** (0.0190)
Vietnam era	-2.014*** (0.0304)	-1.836*** (0.0290)
Peacetime pre-Vietnam	-2.237*** (0.0419)	-2.031*** (0.0391)
Korean conflict	-3.241*** (0.0470)	-2.838*** (0.0446)
Pre-1950	-3.746*** (0.0594)	-3.356*** (0.0563)



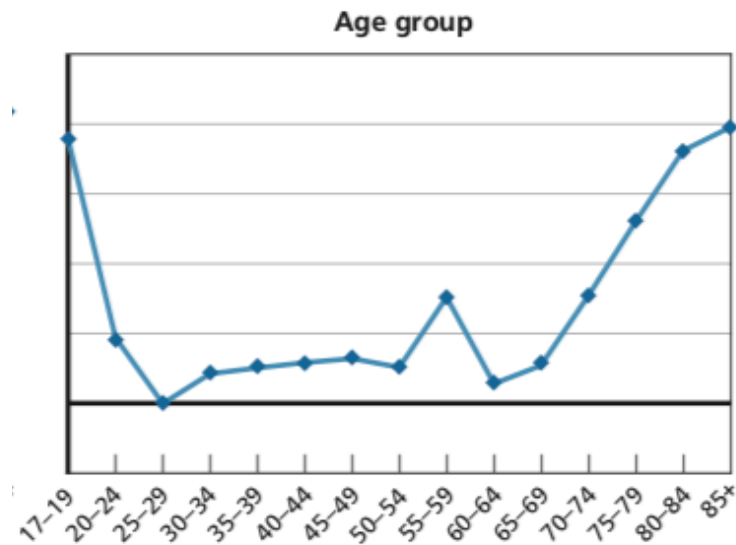
Source: Pollard et al. 2015, p.181-184.

Independent Variables	In-Migration Model	Out-Migration Model
17-19 years	1.059*** (0.0745)	1.891*** (0.0813)
20-24 years	0.574*** (0.0189)	0.464*** (0.0189)
25-29 years	ref.	ref.
30-34 years	0.166*** (0.0158)	0.222*** (0.0154)
35-39 years	0.240*** (0.0198)	0.257*** (0.0196)
40-44 years	0.240*** (0.0203)	0.295*** (0.0200)
45-49 years	0.317*** (0.0213)	0.327*** (0.0210)
50-54 years	0.254*** (0.0236)	0.265*** (0.0231)
55-59 years	0.759*** (0.0261)	0.762*** (0.0255)
60-64 years	-0.180*** (0.0352)	0.136*** (0.0333)
65-69 years	0.0651* (0.0349)	0.277*** (0.0337)
70-74 years	0.766*** (0.0415)	0.777*** (0.0392)
75-79 years	1.406*** (0.0449)	1.306*** (0.0423)
80-84 years	2.051*** (0.0492)	1.813*** (0.0470)
85+ years	2.091*** (0.0619)	1.982*** (0.0582)

Coefficients from in-migration model

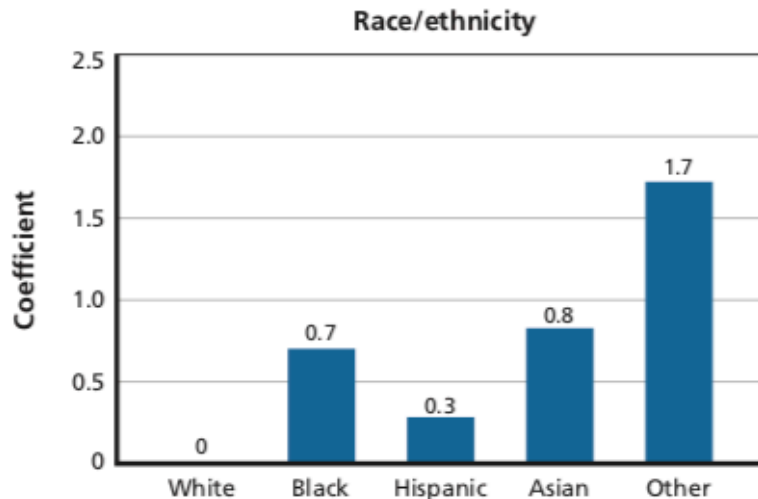


Coefficients from out-migration model

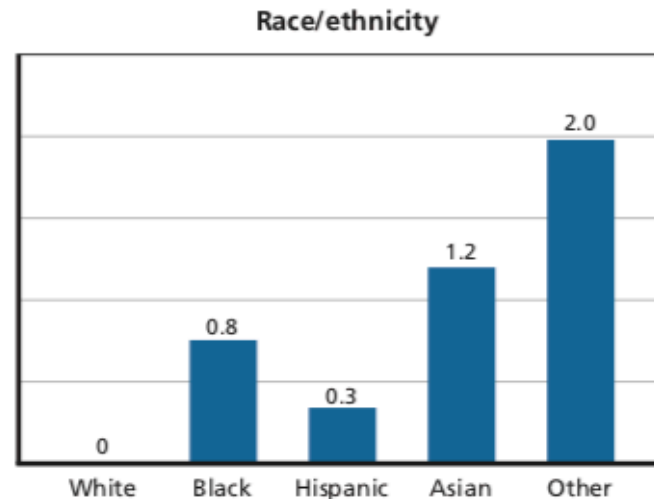


Independent Variables	In-Migration Model	Out-Migration Model
White	ref.	ref.
Black	0.688*** (0.01000)	0.752*** (0.00942)
Hispanic	0.285*** (0.0126)	0.339*** (0.0120)
Asian	0.815*** (0.0250)	1.187*** (0.0218)
Other	1.718*** (0.0218)	1.968*** (0.0206)

Coefficients from in-migration model



Coefficients from out-migration model



Independent Variables	In-Migration Model	Out-Migration Model
Squared distance	-0.0000000222*** (0.00000000233)	-0.000000012*** (0.00000000191)
Population in origin at the beginning of period	0.000963*** (0.0000236)	
Population in destination at the end of period		-0.0000036 (0.000022)
Exposure variable	Pop. in destination at the end of period	Pop. in origin at the beginning of period
Inflate model		
Constant	-31.38*** (0.00114)	-33.88*** (0.00114)
Indicator of cells without migrants	62.55*** (0.00305)	67.48*** (0.00270)
Non-zero observations	776,082	776,082
Zero observations	2,133,534	1,132,194
Total observations	2,909,616	1,908,276

SOURCE: 2009–2013 ACS five-year estimates.

NOTES: Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.



Example to deal with reverse causality (details on next lecture)

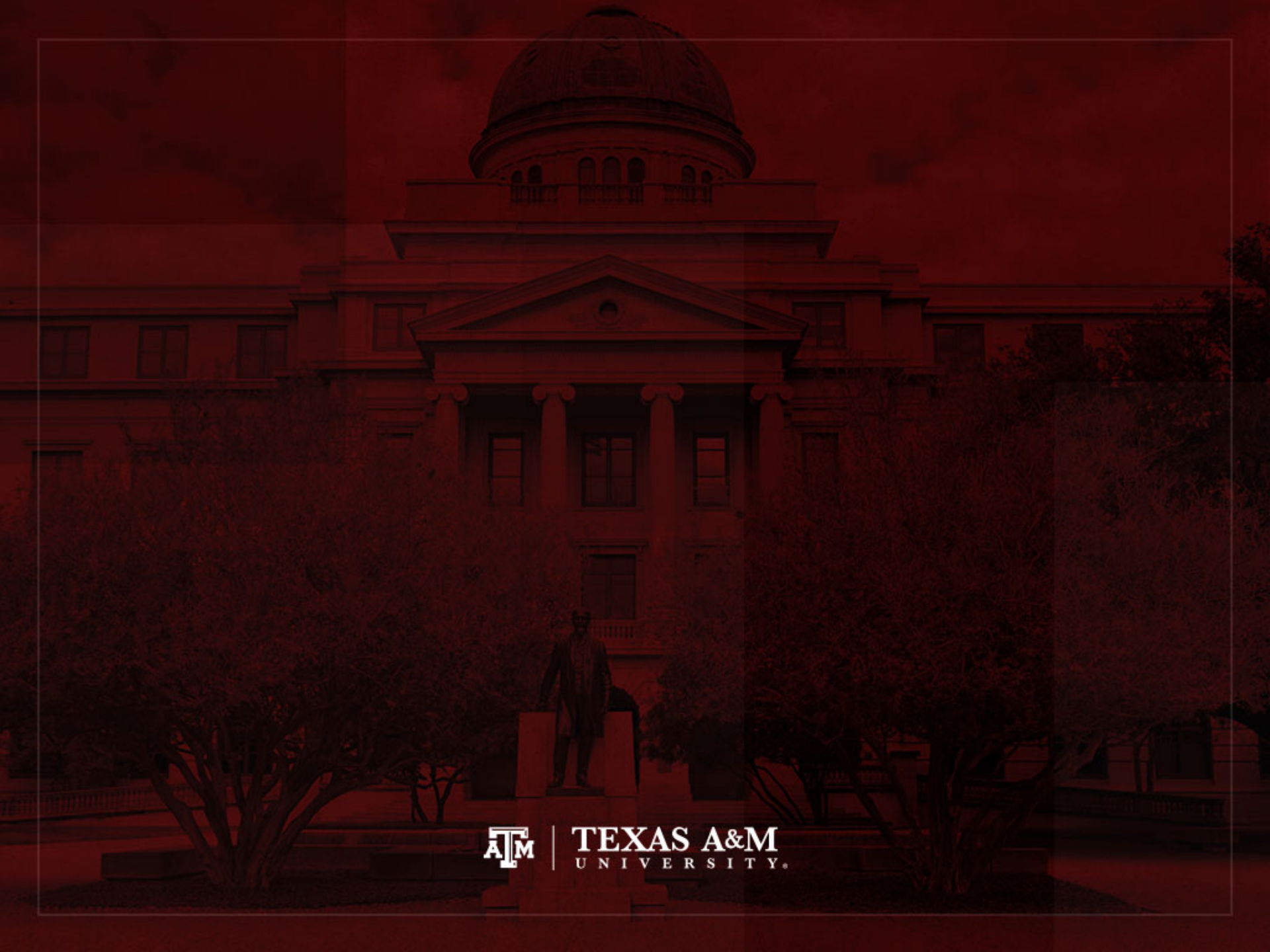
- Gravity models can be used to estimate exogenous measures of migration
 - Example: reverse causality between migration and earnings

Migration \longleftrightarrow **Earnings**

- Immigration increases competition and affects earnings
 - Availability of jobs and income levels influence migration
- Distances among areas
 - Used as an instrumental variable for predicting migration
 - Related to migration levels, but not to earnings

Distance \rightarrow **Migration** \rightarrow **Earnings**





TEXAS A&M
UNIVERSITY.

Analysis of spatial association

- In spatial association analysis, we recognize that people are not randomly distributed over space
 - Neighboring areas tend to be more similar to each other than areas situated at greater distances
- We can utilize the local indicator of spatial association (LISA) to identify spatial clusters and spatial outliers
 - LISA allows for the decomposition of global indicators (e.g., Moran's I) into the contribution of each individual area (Anselin 1995)
 - GeoDa: An introduction to spatial data analysis
 - <https://spatial.uchicago.edu/geoda>



Local spatial autocorrelation

- LISA classifies areas into spatial clusters or outliers, considering information of indicators of surrounding areas
- Spatial clusters
 - Areas of high levels of a specific indicator surrounded by areas with high levels for that indicator (**high-high**)
 - Areas of low levels of a specific indicator surrounded by areas with low levels for that indicator (**low-low**)
- Spatial outliers
 - Areas of high levels of a specific indicator surrounded by areas with low levels for that indicator (**high-low**)
 - Areas of low levels of a specific indicator surrounded by areas with high levels for that indicator (**low-high**)



LISA example

- We analyze spatial distributions of internal migrants with the 2017 American Community Survey
- Areas of destination (current residence)
 - Publicly available data has information on Public Use Microdata Areas (PUMAs) as the lowest level of geographic aggregation (100,000+ residents)
- Areas of origin (previous residence)
 - Data relates to PUMAs or, for confidentiality issues, groups of PUMAs (also known as MIGPUMAs)

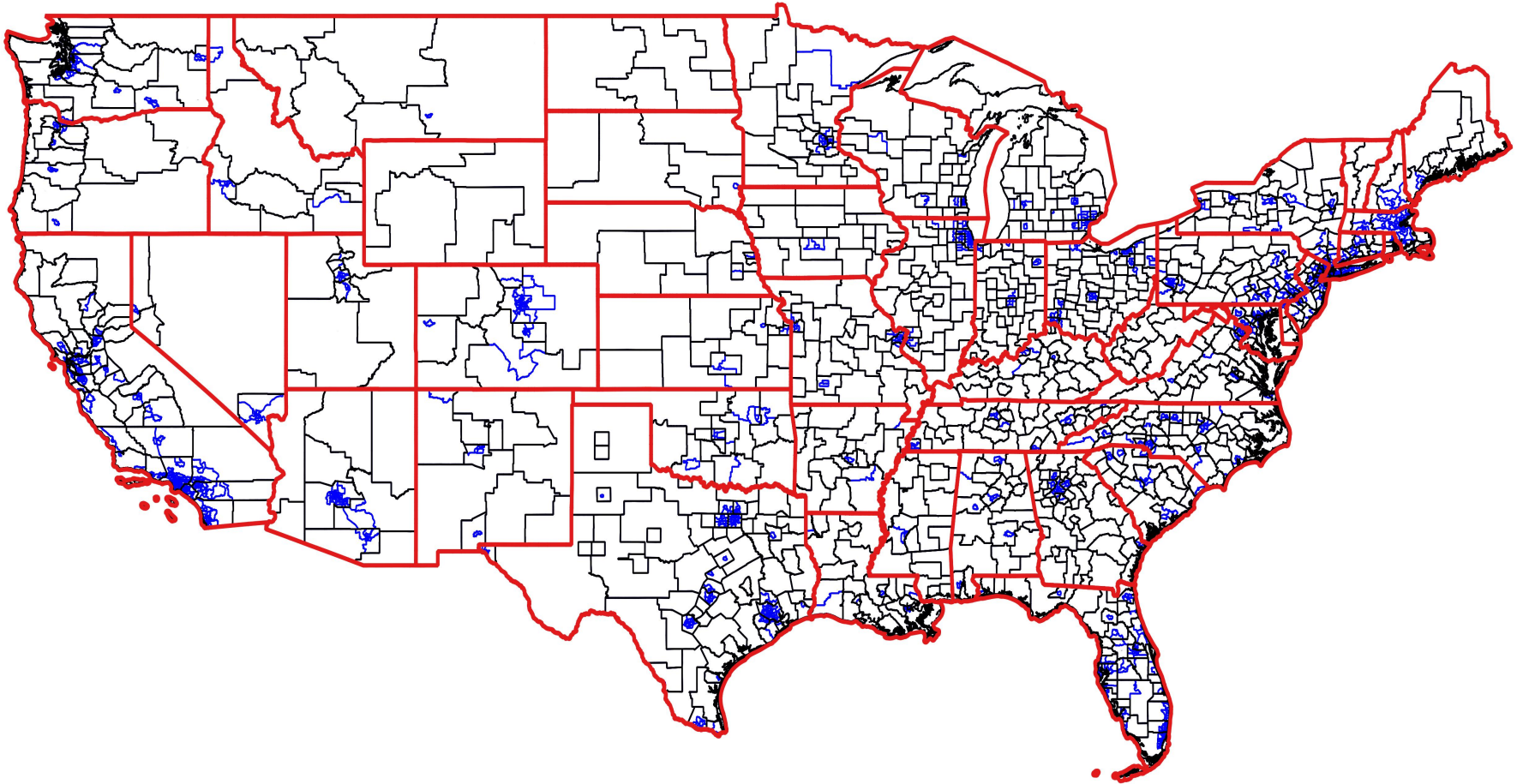


Homogenize areas

- We group PUMAs of destination at the same geographic level as MIGPUMAs of origin
 - 2,378 PUMAs (current residence)
 - 1,005 MIGPUMAs (previous residence)
- This is a strategy to homogenize areas of previous and current residence



State, MIGPUMA, PUMA

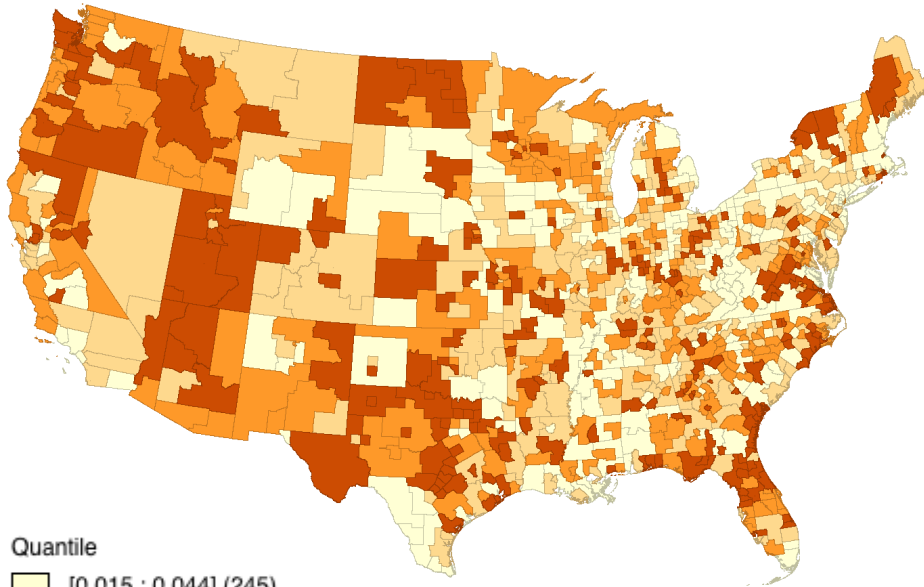


Migration status

- Internal migrants
 - Those who resided in another PUMA (or MIGPUMA) one year before the survey
- Non-migrants
 - Those who resided in the same area in the previous year
- International migrants
 - Those who resided in another country one year before the survey (not included in our analysis)



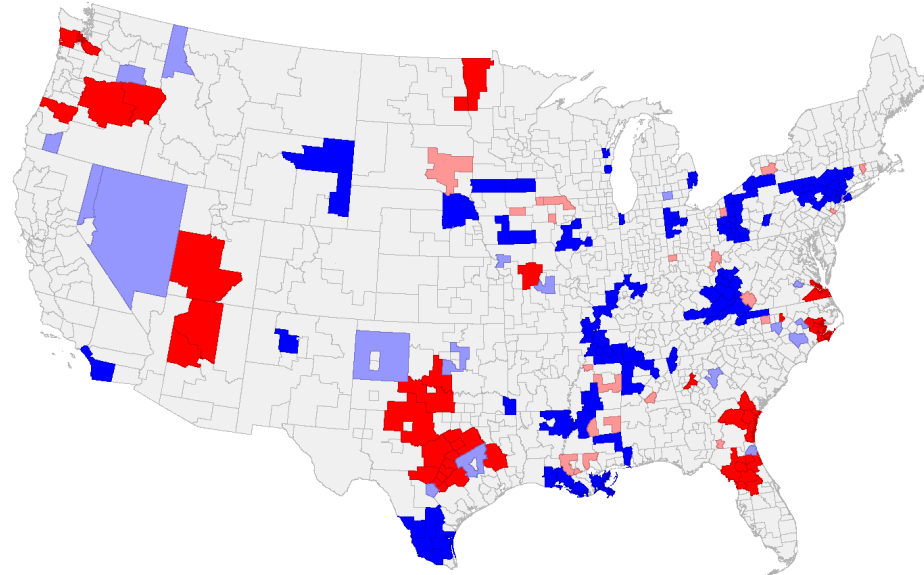
Proportion of internal migrants, 2016–2017



Quantile

- [0.015 : 0.044] (245)
- [0.044 : 0.055] (246)
- [0.055 : 0.071] (246)
- [0.071 : 0.184] (245)
- undefined (23)

LISA of proportion of internal migrants, 2016–2017



LISA Cluster Map

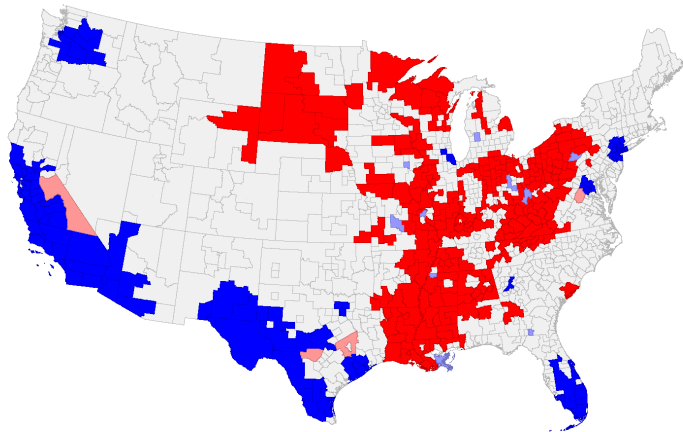
- Not Significant (797)
- High-High (50)
- Low-Low (91)
- Low-High (18)
- High-Low (23)
- Neighborless (3)
- Undefined (23)



Internal migrants are those who changed residence between 2016 and 2017

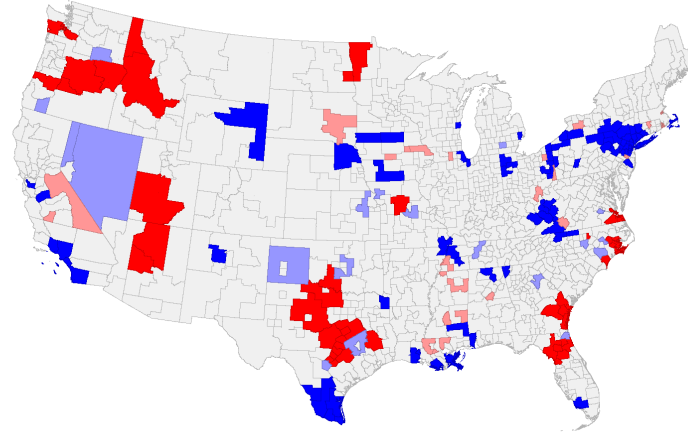
US-born non-migrants

USA Cluster Map
Not Significant (830)
High-High (222)
Low-Low (115)
Low-High (15)
High-Low (4)
Neighborhoods (2)
Undefined (2)



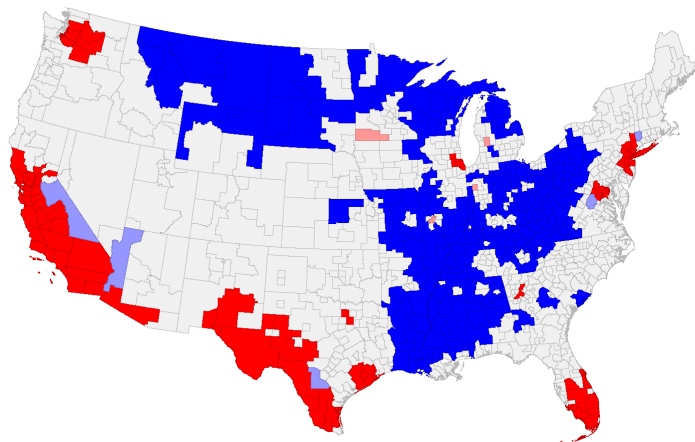
US-born internal migrants

USA Cluster Map
Not Significant (297)
High-High (51)
Low-Low (87)
Low-High (13)
High-Low (25)
Neighborhoods (2)
Undefined (2)



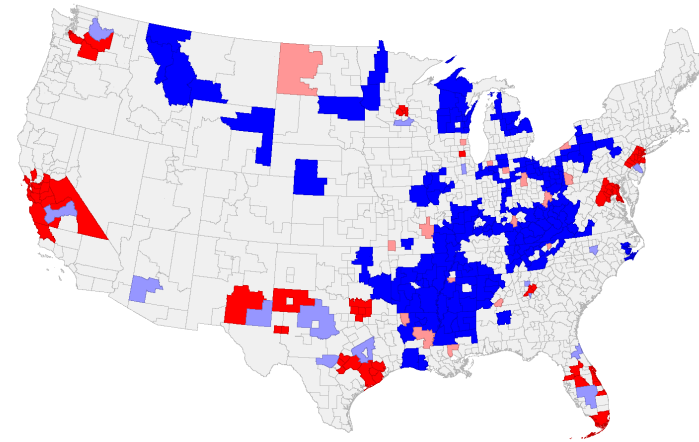
Foreign-born non-migrants

USA Cluster Map
Not Significant (818)
High-High (116)
Low-Low (265)
Low-High (2)
High-Low (4)
Neighborhoods (3)
Undefined (2)



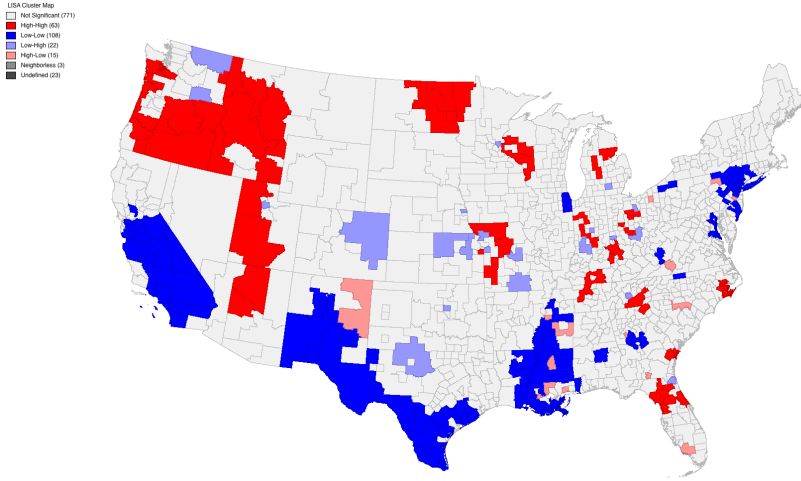
Foreign-born internal migrants

USA Cluster Map
Not Significant (209)
High-High (62)
Low-Low (160)
Low-High (15)
High-Low (17)
Neighborhoods (2)
Undefined (2)

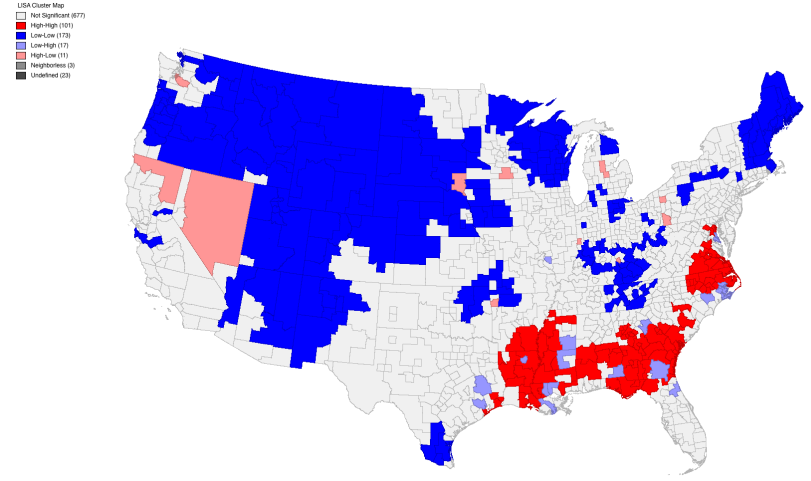


All maps below are for internal migrants, 2016–2017

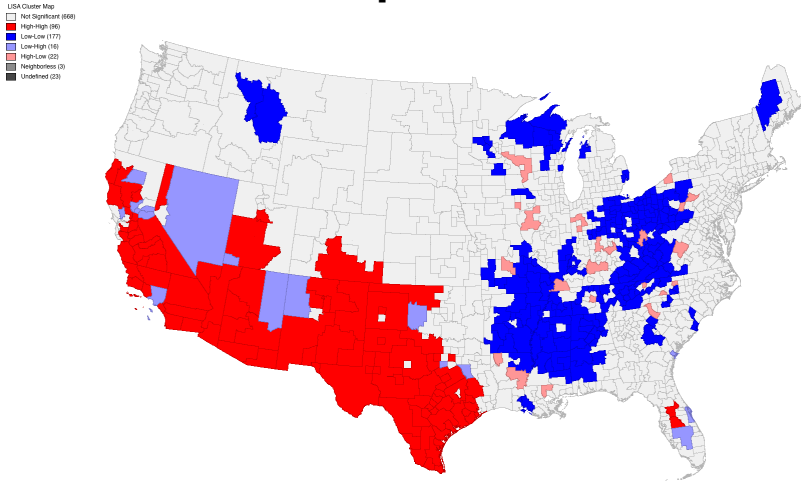
Non-Hispanic Whites



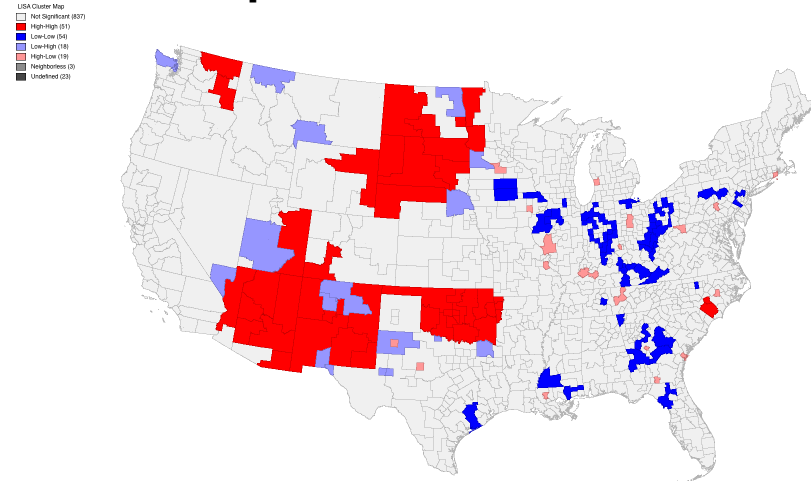
Non-Hispanic African Americans

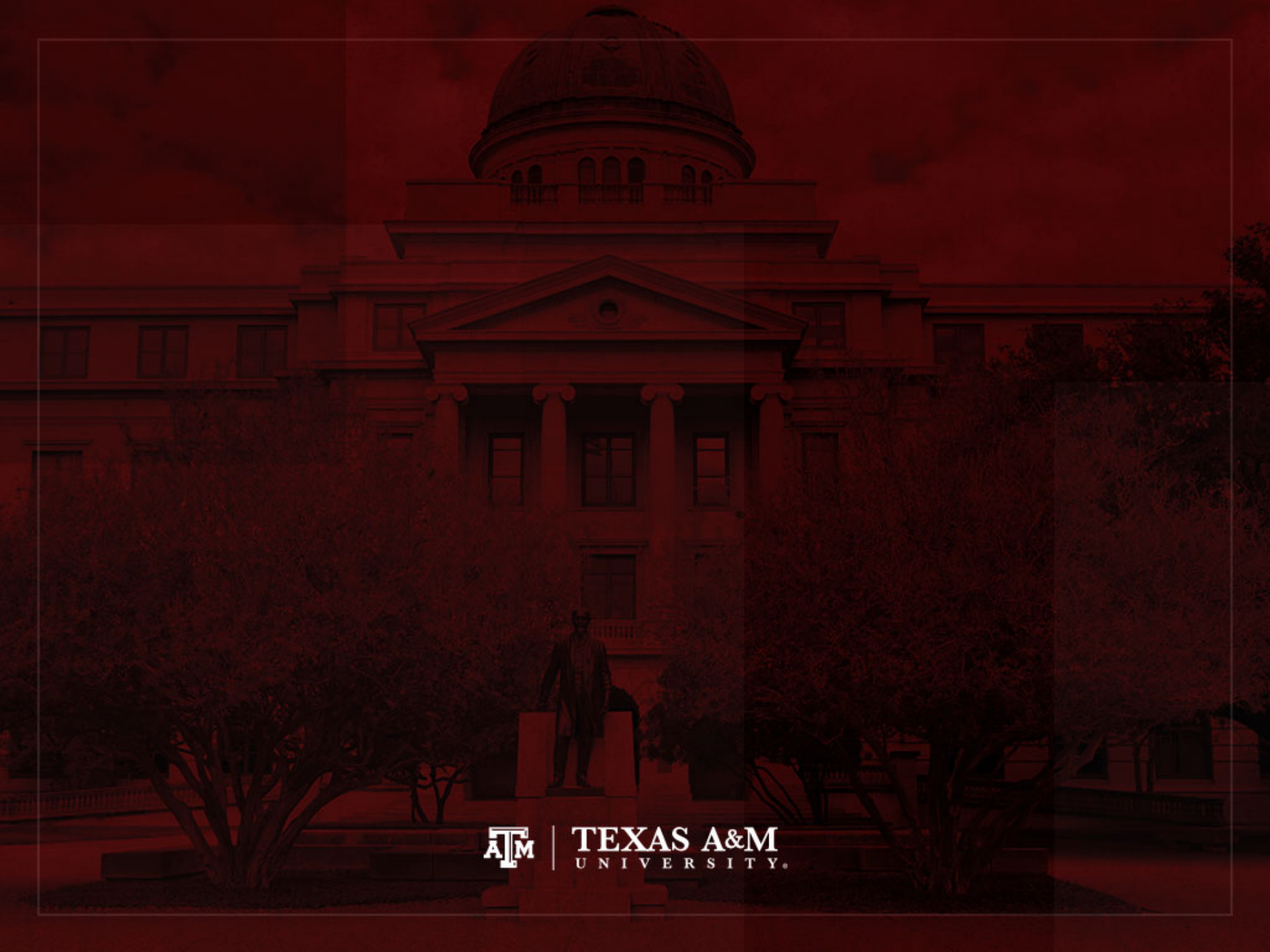


Hispanics



Non-Hispanic Native Americans





TEXAS A&M
UNIVERSITY.

Spatial analysis

- Spatial models can estimate multivariate models to verify the association of several independent variables (explanatory variables) with a specific dependent variable (explained variable)
- These models deal with spatial dependence by measuring the influence of neighboring areas at origin and destination for several variables at the same time (Anselin, Rey 2014, LeSage, Pace 2009)
- These multivariate models that consider effects of neighboring areas are also known as spatial autoregressive models



Components of spatial analysis

- **Origin-based dependence:** individuals moving from a locality are influenced by the levels of migration of neighboring areas of origin
- **Destination-based dependence:** individuals moving to a locality are influenced by the levels of migration of neighboring areas of destination
- **Origin-to-destination dependence:** individuals are influenced by both neighboring localities of origin and destination



Spatial model

- The general spatial autoregressive model that takes into account origin, destination, and origin-to-destination dependence is expressed by the following equation (LeSage, Pace 2008, 2009)

$$y = \rho_d W_d y + \rho_o W_o y + \rho_w W_w y + \alpha \iota_N + X_d \beta_d + X_o \beta_o + \gamma g + \varepsilon$$

- W_d : spatial dependence at the destination
- W_o : spatial dependence at the origin
- W_w : dependence based on interaction between origin and destination neighbors

...



Spatial model (cont.)

- This model contains a conventional gravity model least-squares regression approach to explaining variation in the vector of origin-destination (OD)
 - Flows relies on an n by k matrix of explanatory variables labeled X with k characteristics for each of the n regions

$$y = \rho_d W_d y + \rho_o W_o y + \rho_w W_w y + \alpha t_N + X_d \beta_d + X_o \beta_o + \gamma g + \varepsilon$$

- X_d : contains k characteristics for each of the n regions of destination and the associated k by 1 parameter vector is β_d
- X_o : contains k characteristics for each of the n regions of origin and the associated k by 1 parameter vector is β_o
- Scalar γ : effect of distance g
- α : constant term parameter on t_N regions



Issues with spatial models

- Some issues arise when we estimate spatial models for population flows at the local level
- One problem of dealing with small areas is the presence of **zero flow** magnitudes between origin-destination areas
 - We can treat the zero flows using a threshold Tobit model that contains spatial lags of the dependent variable (Ranjan, Tobias 2007, LeSage, Pace 2009)
- Another problem is the presence of **large intraregional migration** (diagonal of the origin-destination matrix) relative to interregional migration
 - An approach is to add a separate intercept and explanatory variables for intraregional flows, which have non-zero observations for the intraregional observations and zero elsewhere (LeSage, Pace 2008, 2009)



Application to state-level migration

- State-to-state flows from 1995 to 2000 for the population 5 years and over
- 1990 Census characteristics of the states
- Flows were all nonzero
- Log transformation produced a dependent variable that was nearly normal
- Two explanatory variables
 - State population in 1995
 - State unemployment rate in 1995



Variables used in the model

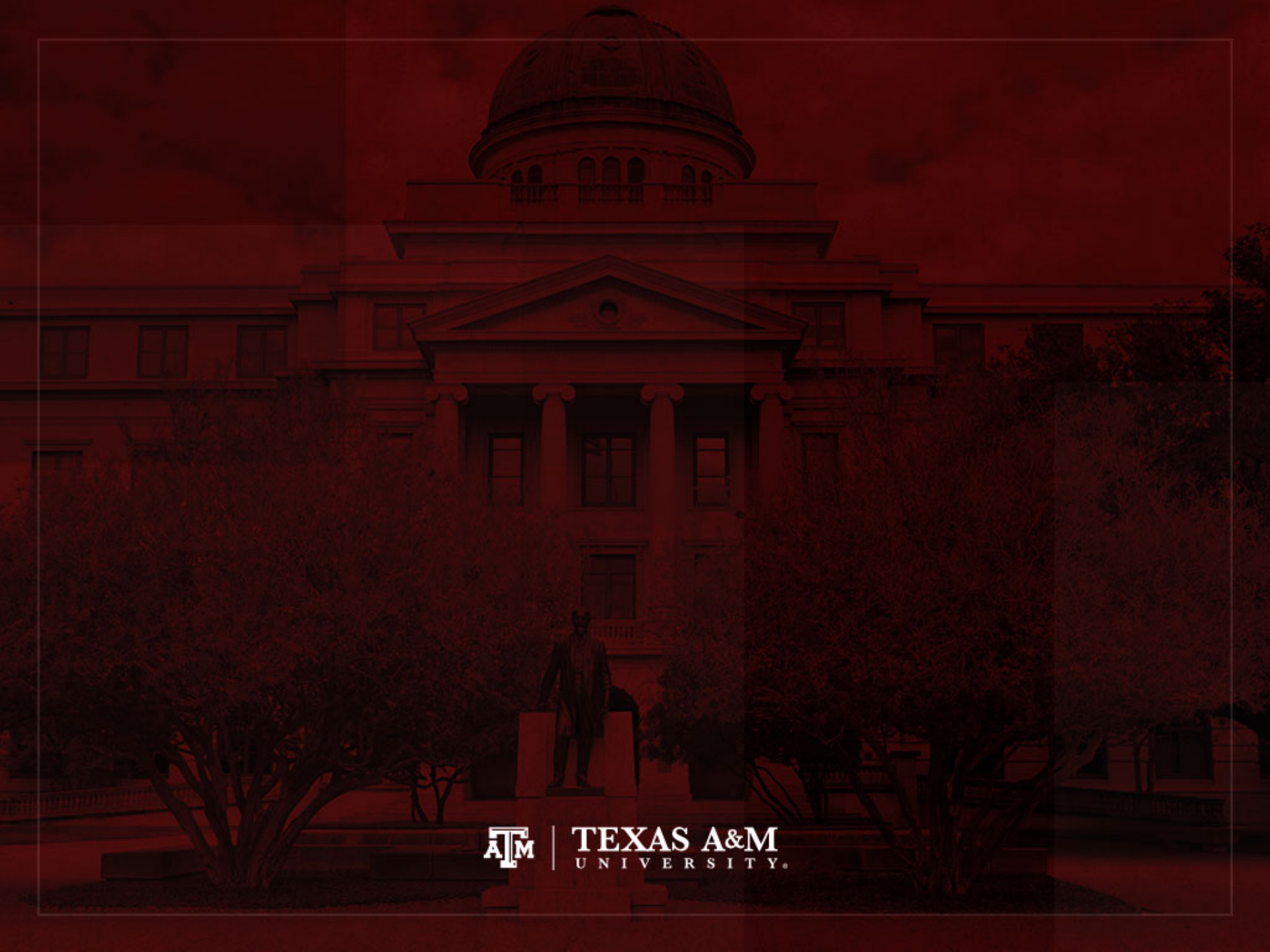
Variable name	Description
y	Log (interstate and intrastate migration flows 1995–2000)
Population 1995	log (state population in 1995)
Area	Log (state area)
College	Proportion of population > age 25, college degree as highest
Born in state	Proportion of population > age 5 born in the state in 1990
Unemployment rate 1995	Unemployment rate in 1995
Mortgage	Log (median mortgage costs in 1990)
Executive	Proportion of employment in executive and managerial occupations
Sales	Proportion of employment in sales occupations
Local Govt.	Proportion of employment in local government
State Govt.	Proportion of employment in state government
Federal Govt.	Proportion of employment in federal government
Farming	Proportion of employment in farming
Distance	Log (distance between origin and destination state centroids)



Estimates from Least-squares and spatial model

Variable	Least-Squares		Spatial Model	
	Coefficient	t-statistic(plevel)	Coefficient	t-statistic(plevel)
Constant	-12.5281	-8.21 (0.0000)	-5.8195	-15.27 (0.0000)
D_pop95	0.9281	38.85 (0.0000)	0.4296	22.88 (0.0000)
D_area	0.2273	12.66 (0.0000)	0.0868	7.14 (0.0000)
D_college	8.5537	7.89 (0.0000)	4.5691	6.20 (0.0000)
D_borninstate	-2.9131	-22.10 (0.0000)	-1.0983	-11.65 (0.0000)
D_urate95	-0.5264	-10.55 (0.0000)	-0.1883	-5.67 (0.0000)
D_mortgage	-0.5958	-4.66 (0.0000)	-0.4149	-5.12 (0.0000)
D_exec	-0.4993	-1.76 (0.0781)	-0.0978	-0.53 (0.5941)
D_sales	0.4464	1.86 (0.0617)	0.3939	2.59 (0.0096)
D_local_govt	-2.5698	-3.83 (0.0001)	-0.5446	-1.23 (0.2156)
D_state_govt	-1.3228	-2.07 (0.0385)	-1.2894	-2.97 (0.0030)
D_fed_govt	0.8228	1.63 (0.1026)	0.6782	2.04 (0.0410)
O_pop95	0.8108	33.94 (0.0000)	0.4185	22.71 (0.0000)
O_area	0.2680	14.92 (0.0000)	0.1172	9.52 (0.0000)
O_college	6.0058	5.54 (0.0000)	4.5046	6.07 (0.0000)
O_borninstate	-1.5466	-11.73 (0.0000)	-0.4946	-5.50 (0.0000)
O_urate95	-0.4208	-8.43 (0.0000)	-0.1367	-4.19 (0.0000)
O_mortgage	0.0628	0.49 (0.6225)	-0.0616	-0.88 (0.3759)
O_exec	0.3886	1.37 (0.1702)	0.1198	0.68 (0.4945)
O_sales	-0.0039	-0.01 (0.9868)	0.2819	1.83 (0.0660)
O_local_govt	1.7072	2.54 (0.0108)	1.5082	3.47 (0.0005)
O_state_govt	-0.3572	-0.55 (0.5760)	-1.0193	-2.37 (0.0176)
O_fed_govt	-0.6902	-1.36 (0.1709)	0.4101	1.23 (0.2175)
OD_pop95	1.0150	8.68 (0.0000)	0.4468	5.79 (0.0000)
OD_area	0.2260	2.29 (0.0218)	0.1159	1.84 (0.0653)
OD_urate95	0.6557	2.61 (0.0089)	0.5319	3.54 (0.0004)
OD_farming	-0.6185	-3.34 (0.0008)	-0.3431	-3.22 (0.0013)
Log(distance)	-0.5759	-75.22 (0.0000)	-0.1671	-16.96 (0.0000)
ρ_d			0.4581	33.36 (0.0000)
ρ_o			0.5175	39.28 (0.0000)
ρ_w			-0.3863	-23.56 (0.0000)
σ^2	0.2438		0.1042	





TEXAS A&M
UNIVERSITY.

Agent-based models

- Agent-based models can estimate the relationship between migration and several individual and contextual variables (Massey, Zenteno 1999; Klabunde, Willekens 2016)
- They include modules of endogenous predictors
- They allow researchers to incorporate interactions between individual decisions, behavioral responses, and social networks related to migration outcomes
- These models are useful for projects that do not simply require descriptive current or project information about migration flows
- Researchers can seek to incorporate consequences of immigration into a given research question

Simulations

- Agent-based model is able to formalize interconnections and simulate potential feedback relationships between migration streams and several endogenous predictors
 - Education systems
 - Labor markets
 - Healthcare systems
 - Migration policies
 - Border security
 - Social networks
- These models allow researchers to build different scenarios and simulate future population flows



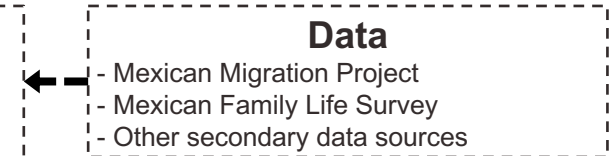
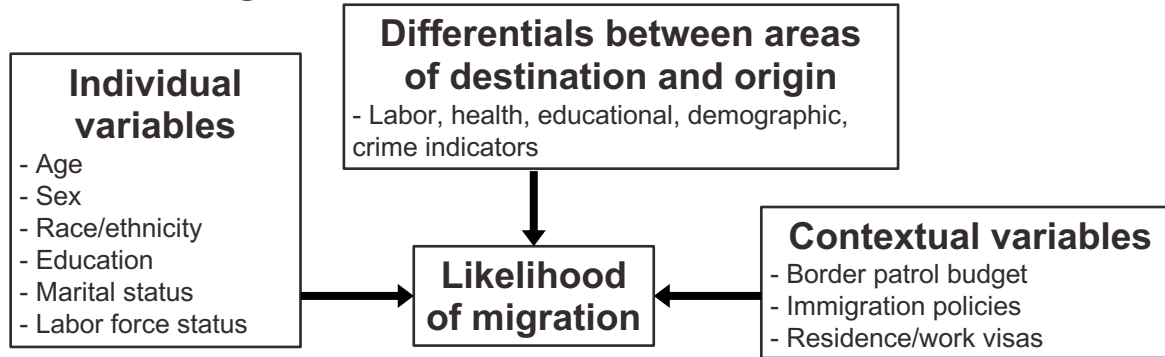
Ideas for proposals

- **Modeling immigration of Central American children**
 - Department of Homeland Security (DHS): interviews of immigrant children conducted after apprehension
 - Machine-learning (big data): social media conversations about immigration
- **Developing immigration policy scenarios/simulations**
 - Inform policymakers on the various immigration policy options
- **Job polarization, migration, and earnings in the U.S.**
 - Increase in low-skill and high-skill jobs, which affects inequality
- **Micro-simulation models of international migration...**



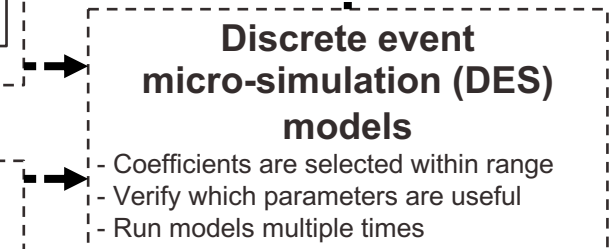
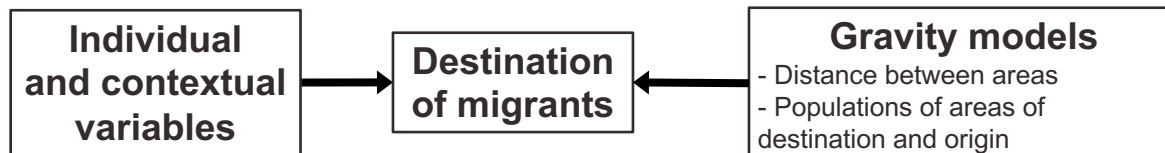
Model international migration to the U.S.

First set of regressions

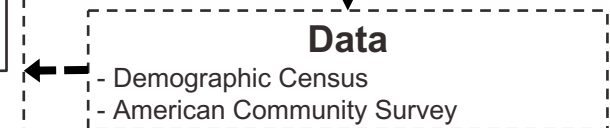


Calibration

Second set of regressions



Calibration

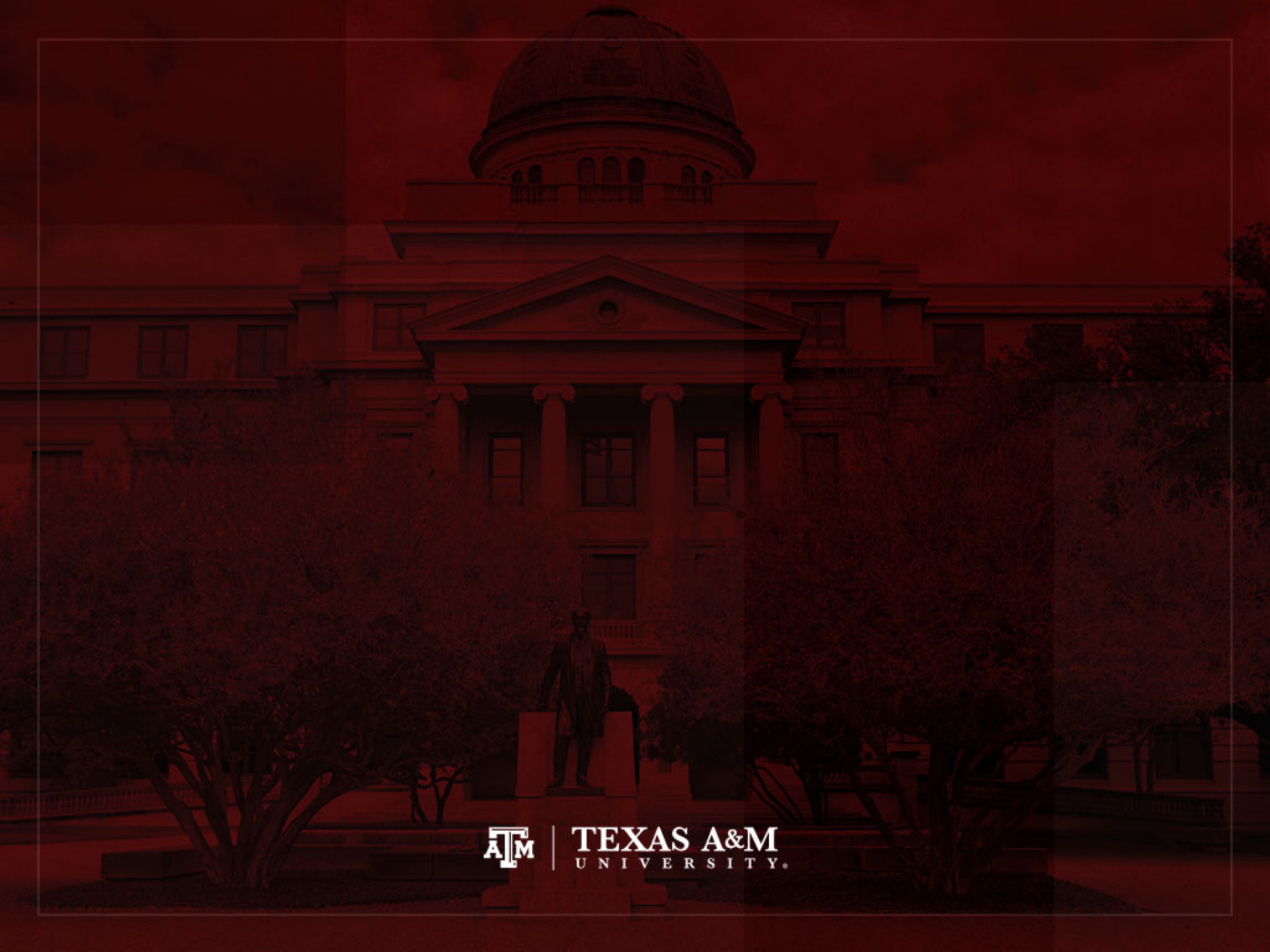


Conditional on being a migrant



References

- Amaral EFL, Rios-Neto ELG, Potter JE. 2016. "The influence of internal migration on male earnings in Brazil, 1970–2000." *Migration and Development* 5(1): 55–78.
- Amaral EFL, Yen SK, Harris C, Cisneros Franco CL. 2020. "Factors associated with internal migration at the local level in the United States." In: 115th American Sociological Association (ASA) Annual Meeting. Virtual Engagement Event: ASA. August 8–11.
- Anderson JE, Van Wincoop E. 2003. "Gravity with gravitas: A solution to the border puzzle." *The American Economic Review*, 93(1): 170–192.
- Anselin L. 1995. "Local indicators of spatial association – LISA." *Geographical Analysis*, 27(2): 93–115.
- Anselin L, Rey SJ. 2014. *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*. Chicago: GeoDa Press LLC.
- Bernard A, Bell M, Charles-Edwards E. 2014. "Improved measures for the cross-national comparison of age profiles of internal migration." *Population Studies*, 68(2): 179–195.
- Head K. 2000. "Gravity for beginners." *Rethinking the Line: The Canada-U.S. Border Conference*, Vancouver, British Columbia.
- Klabunde A, Willekens F. 2016. "Decision-making in agent-based models of migration: State of the art and challenges." *European Journal of Population*, 32(1): 73–97.
- LeSage JP, Pace RK. 2008. "Spatial econometric modeling of origin-destination flows." *Journal of Regional Science*, 48(5): 941–967.
- LeSage JP, Pace RK. 2009. *Introduction to Spatial Econometrics*. Boca Raton: CRC Press, Taylor & Francis Group.
- Lowry IS. 1966. *Migration and Metropolitan Growth: Two Analytical Models*. San Francisco: Chandler Publishing.
- Massey DS, Zenteno RM. 1999. "The dynamics of mass migration." *Proceedings of the National Academy of Sciences (PNAS)*, 96: 5328–5335.
- Pollard M, Amaral EFL, Mendelsohn J, Cefalu M, Kress A, Ross R. 2015. "Current and future demographics of the Veteran population." In: *Current and Projected Characteristics and Unique Health Care Needs of the Patient Population Served by the Department of Veterans Affairs*. (RAND Health, ed.). Santa Monica: RAND Corporation, RR-1165/1-VA, 31–56.
- Pöyhönen P. 1963. "A tentative model for the volume of trade between countries." *Weltwirtschaftliches Archiv*, 90: 93–100.
- Ranjan R, Tobias JL. 2007. "Bayesian inference for the gravity model." *Journal of Applied Econometrics*, 22(4): 817–838.
- Raymer J, Rogers A. 2007. "Using age and spatial flow structures in the indirect estimation of migration streams." *Demography*, 44(2): 199–223.
- Rogers A, Castro LJ. 1981. *Model Migration Schedules*. Laxenburg: International Institute for Applied Systems Analysis.
- Rogers A, Jordan L. 2004. "Estimating migration flows from birthplace-specific population stocks of infants." *Geographical Analysis*, 36(1): 38–53.
- Stillwell, J. 2005. "Inter-regional migration modelling: A review and assessment." 45th Congress of the European Regional Science Association, Vrije Universiteit Amsterdam, The Netherlands, August, 23–27.
- Stillwell, J. 2009. "Inter-regional migration modelling: A review." In *Migration and Human Capital*, edited by Jacques Poot, Brigitte Waldorf and Leo van Wissen, 29–48. Cheltenham: Edward Elgar Publishing.
- Tinbergen, J. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.



TEXAS A&M
UNIVERSITY.