Factors associated with internal migration at the local level in the United States

Ernesto F. L. Amaral
Shih-Keng Yen
Jingqiu Ren

May 5, 2021

TEXAS A&M UNIVERSITY

www.ernestoamaral.com

Objective

- Several studies described associations of socioeconomic and demographic characteristics with internal migration rates in the United States
  - There is less focus on the profile and spatial distribution of internal migrants
- We investigate
  - Factors associated with internal migration in recent years
  - Local indicators of spatial association to understand clusters of internal migrants

Recent trends in migration

- The U.S. has been experiencing the lowest levels of internal migration since the late 1940s (Frey 2019)
  - 20% in 1950–1960
  - 9.8% in 2019

- Migration rates are higher for better educated, whites, African Americans, households without children, renters, unemployed (Molloy, Smith, Wozniak 2011; Moretti 2011)

Reasons for decline

- Robust economy in 1950–1960 (Frey 2019)
- In more recent decades (Frey 2019)
  - Older population
  - Labor market more homogeneous across country
  - Telecommuting, jobs from home
  - 2008 economic recession
- Neoclassical theory emphasizes that people move to places with more job opportunities
  - Fewer people are changing jobs, which seems to be related with the decline of internal migration (Molloy, Smith, Wozniak, 2017)

2008 economic recession

- Low-skilled Mexican immigrants were more responsive to the 2008 economic crisis than low-skilled U.S.-born workers (Cadena, Kovak 2016)
  - Reallocation of immigrants within the U.S. diminished spatial differences between local labor markets
  - Low-skilled U.S.-born workers in areas with many Mexican immigrants were shielded from the crisis
- Social networks (Motel, Patten 2012)
  - Communities with large proportions of Mexican immigrants are more likely to facilitate the flexibility of these groups in the labor market

Data and geographical areas

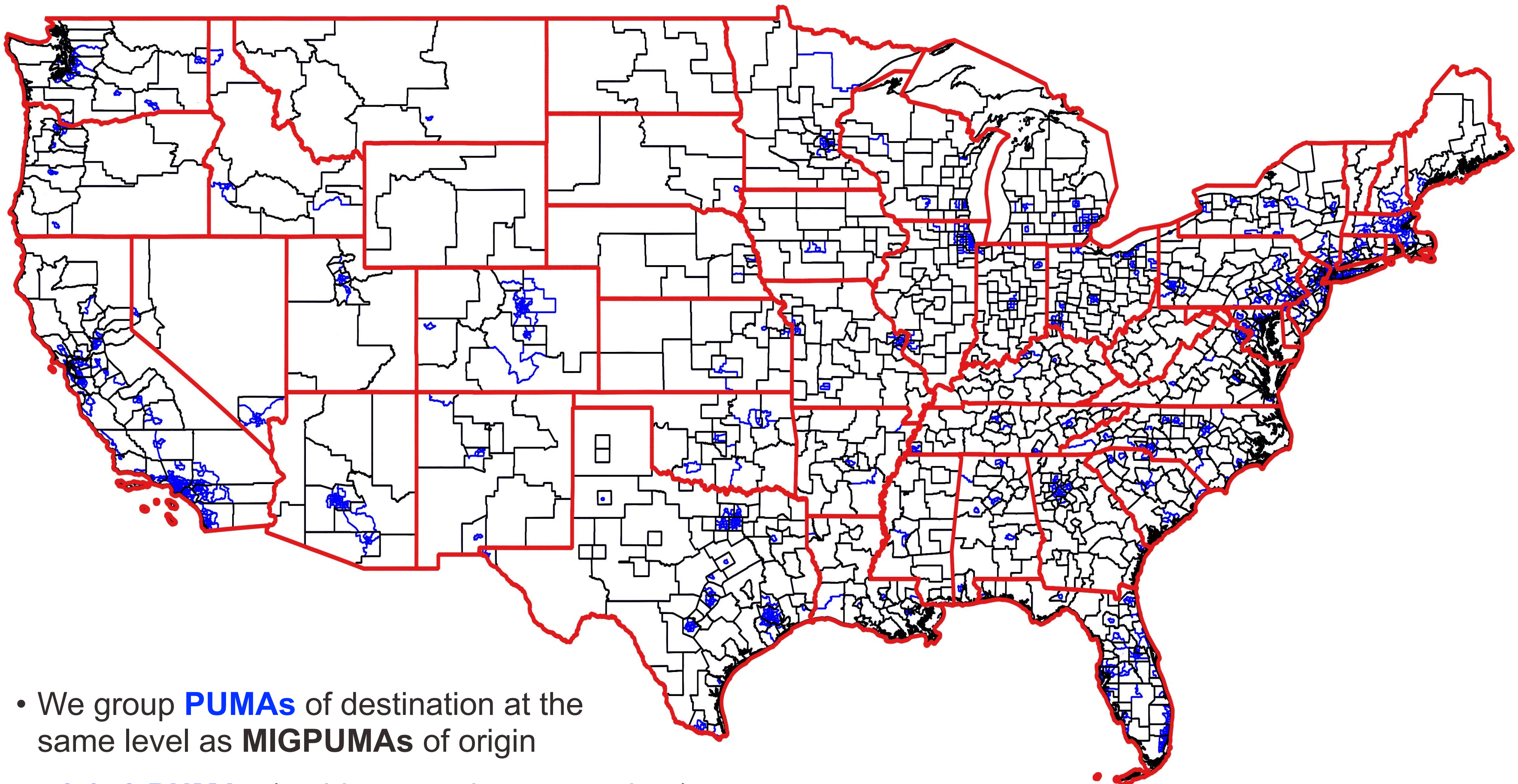
- We analyze spatial distributions of internal migrants with the 2005–2019 American Community Surveys
- Areas of destination (current residence)
  - Publicly available data has information on Public Use Microdata Areas (PUMAs) as the lowest level of geographic aggregation (100,000+ residents)
- Areas 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



- We group **PUMAs** of destination at the same level as **MIGPUMAs** of origin
  - **2,378 PUMAs** (residence at the survey date)
  - **1,005 MIGPUMAs** (residence one year before the survey)

Migration status

- Internal migrants
  - Those who resided in another 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)

Methods

- Estimate factors associated with internal migration flows
  - 2005–2019 American Community Surveys (ACS)
  - Logistics models
  - Dependent variable: internal migrants vs. non-migrants
  - Sample size: 36,039,390 (only people aged 18+)

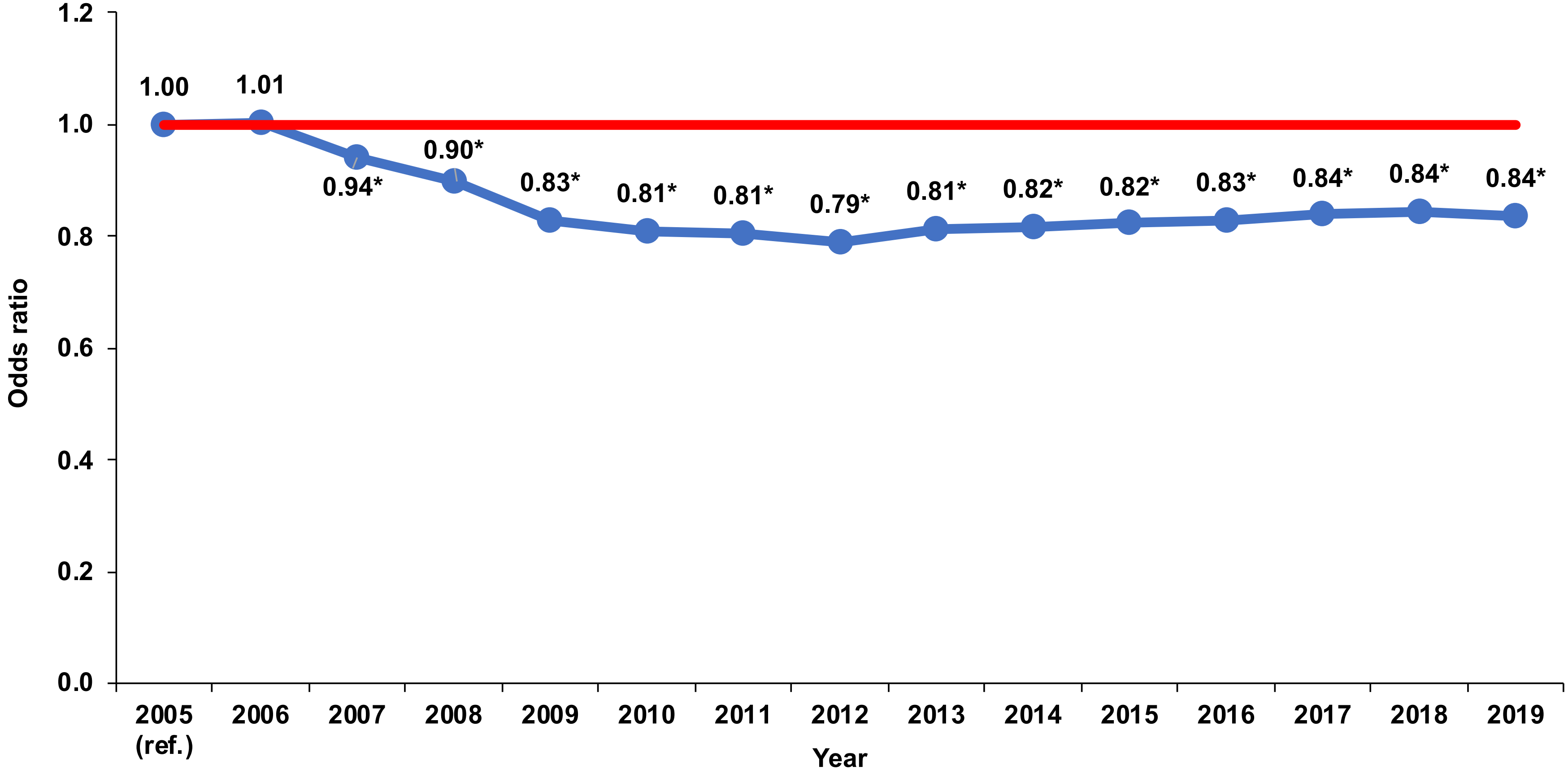
- Analysis of spatial distribution of proportion of internal migrants
  - 2019 ACS: focus on area of destination
  - Local indicators of spatial association (LISA)

Logistic regressions

- Independent variables
  - Year
  - Sex
  - Age group
  - Educational attainment
  - Marital status
  - Citizenship
  - Nativity (foreign born, U.S. born)
  - Race/ethnicity
  - At least one child in the household
  - Homeownership
  - Region of residence one year ago
- Interaction
  - Nativity * race/ethnicity
- For people 18+
  - In school
  - Speak English
  - Any disability
  - Occupation and employment status
  - Top 50% income

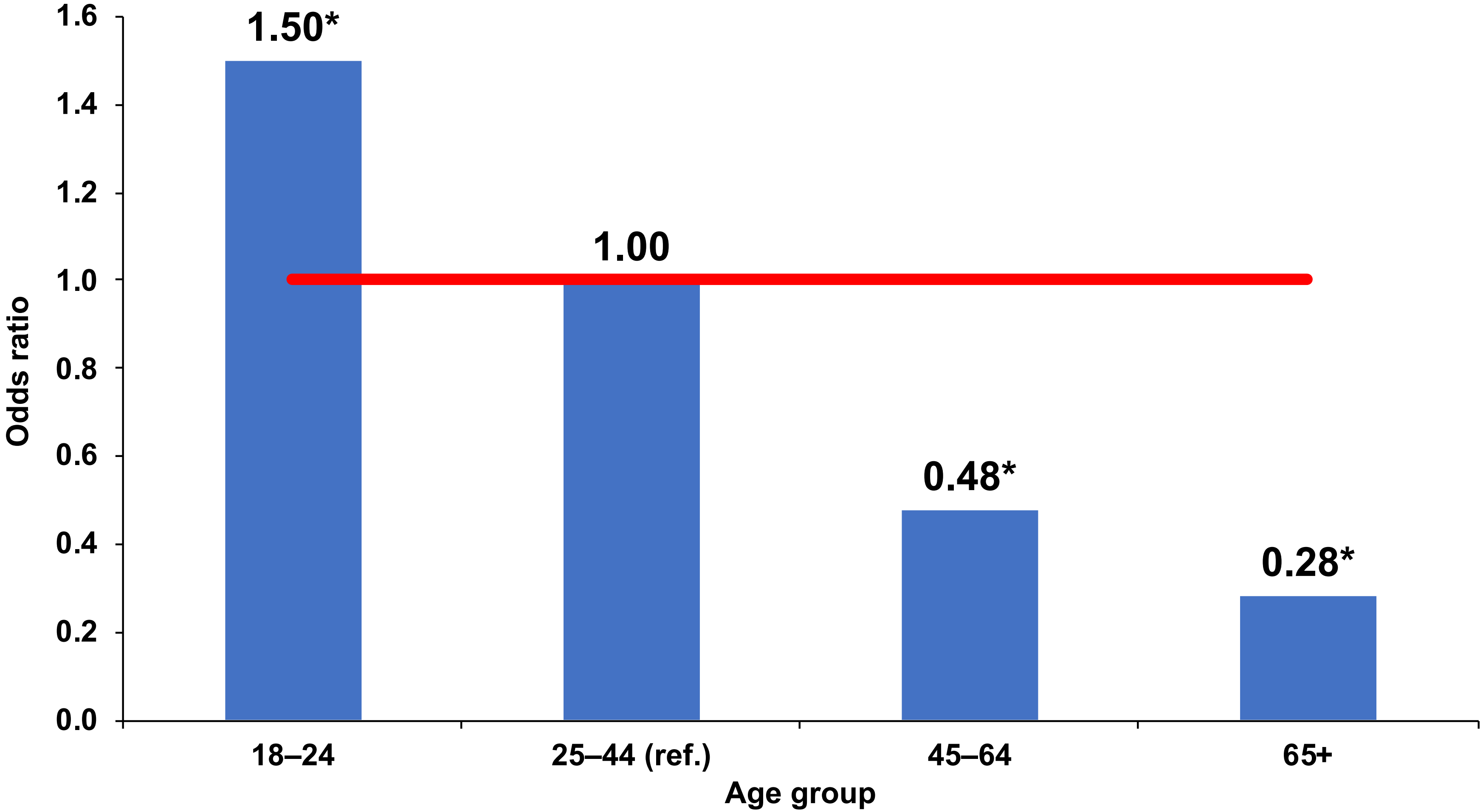
Note: Results for variables in red are presented in the following slides.
Variables selected based on Molloy, Smith, Wozniak (2011, 2017).

Odds ratios of being an internal migrant by year



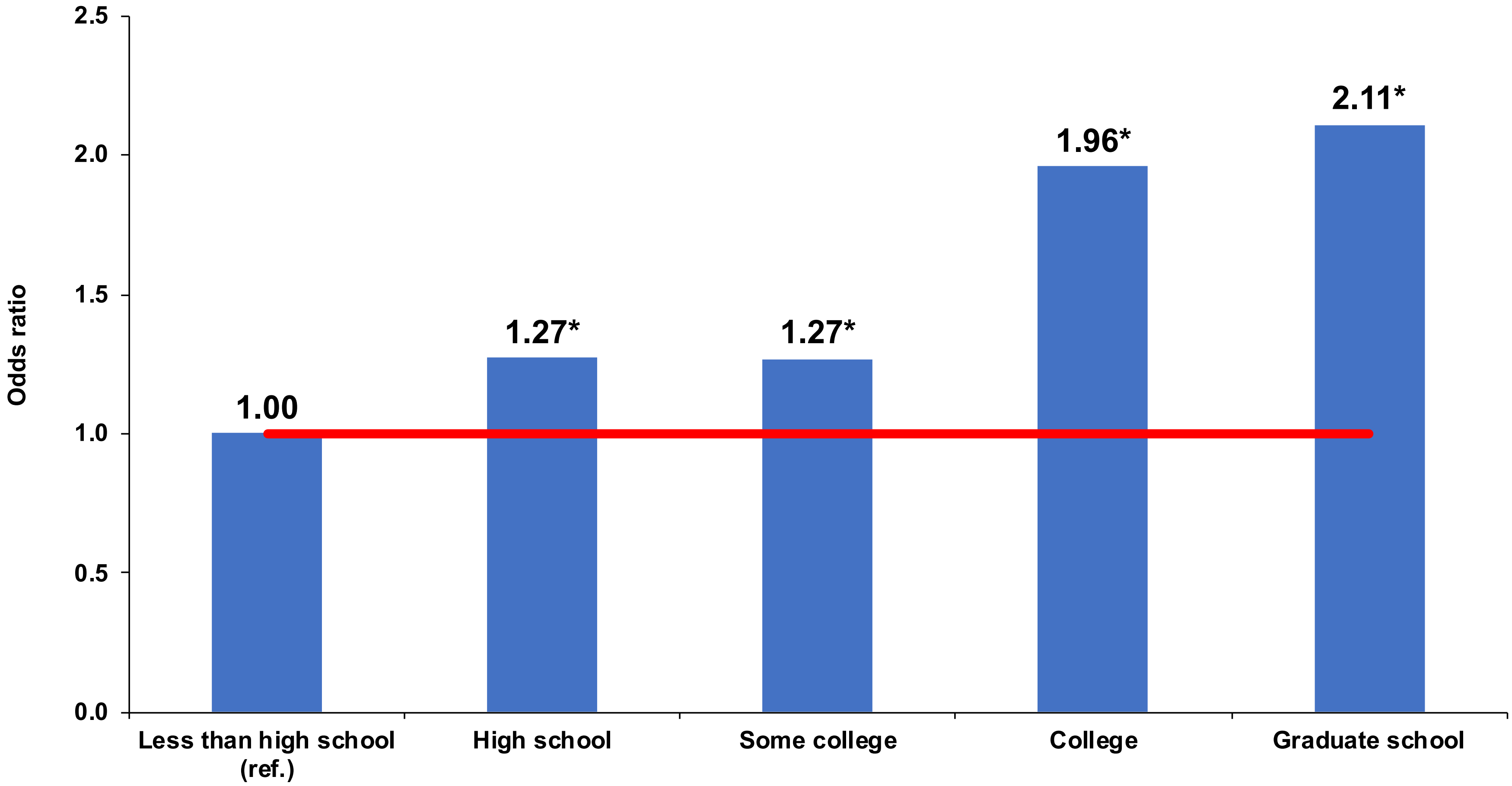
Note: Only for people aged 18+. * Significant at p<.01.
Source: 2005–2019 American Community Surveys.

Odds ratios of being an internal migrant by age group



Note: Only for people aged 18+. * Significant at p<.01.
Source: 2005–2019 American Community Surveys.

Odds ratios of being an internal migrant by educational attainment



Note: Only for people aged 18+. * Significant at p<.01.
Source: 2005–2019 American Community Surveys.

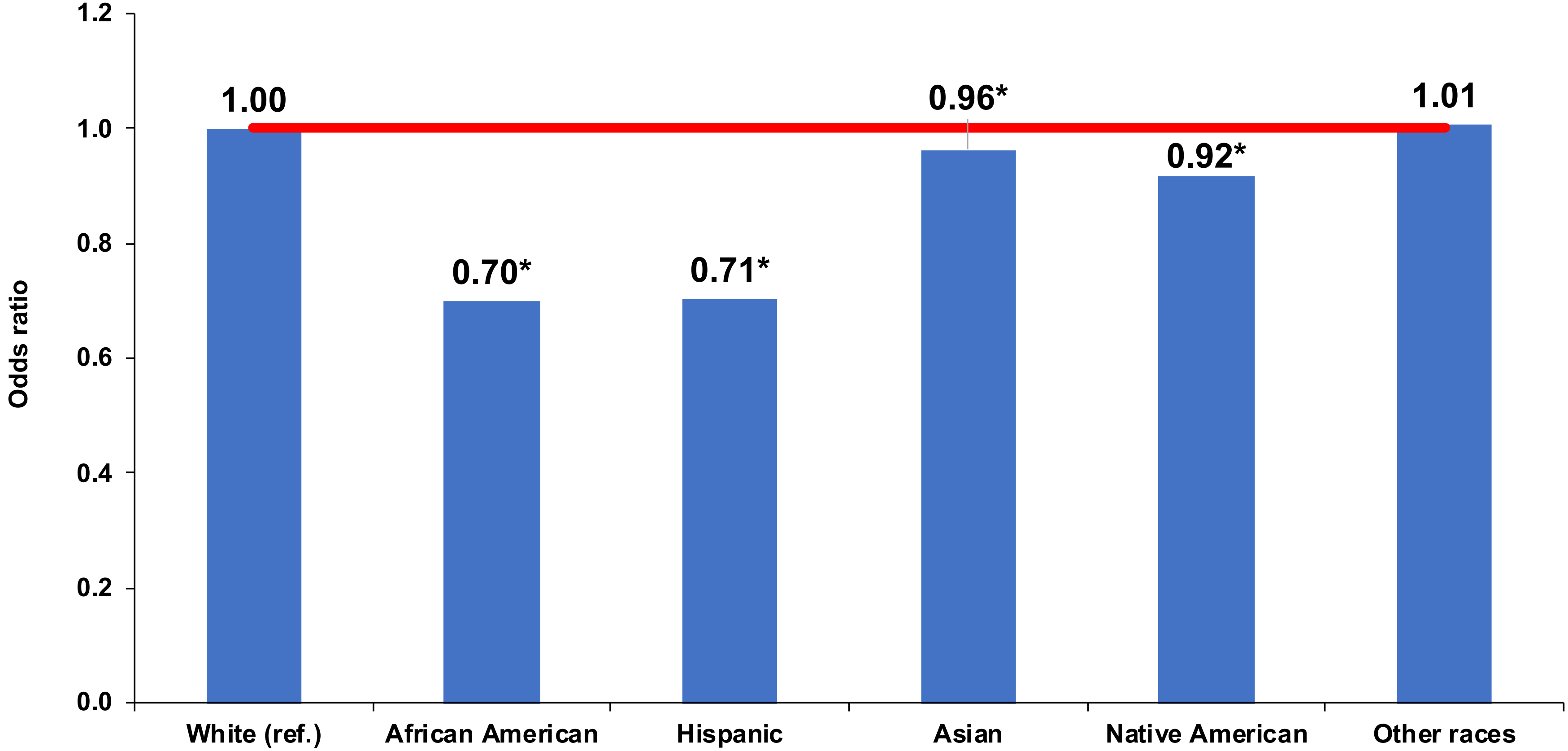
Odds ratios, selected variables

- Citizenship
  - Non-citizen (ref.): 1.00
  - Citizen: 1.07*

- Nativity
  - U.S. born (ref.): 1.00
  - Foreign born: 0.90*

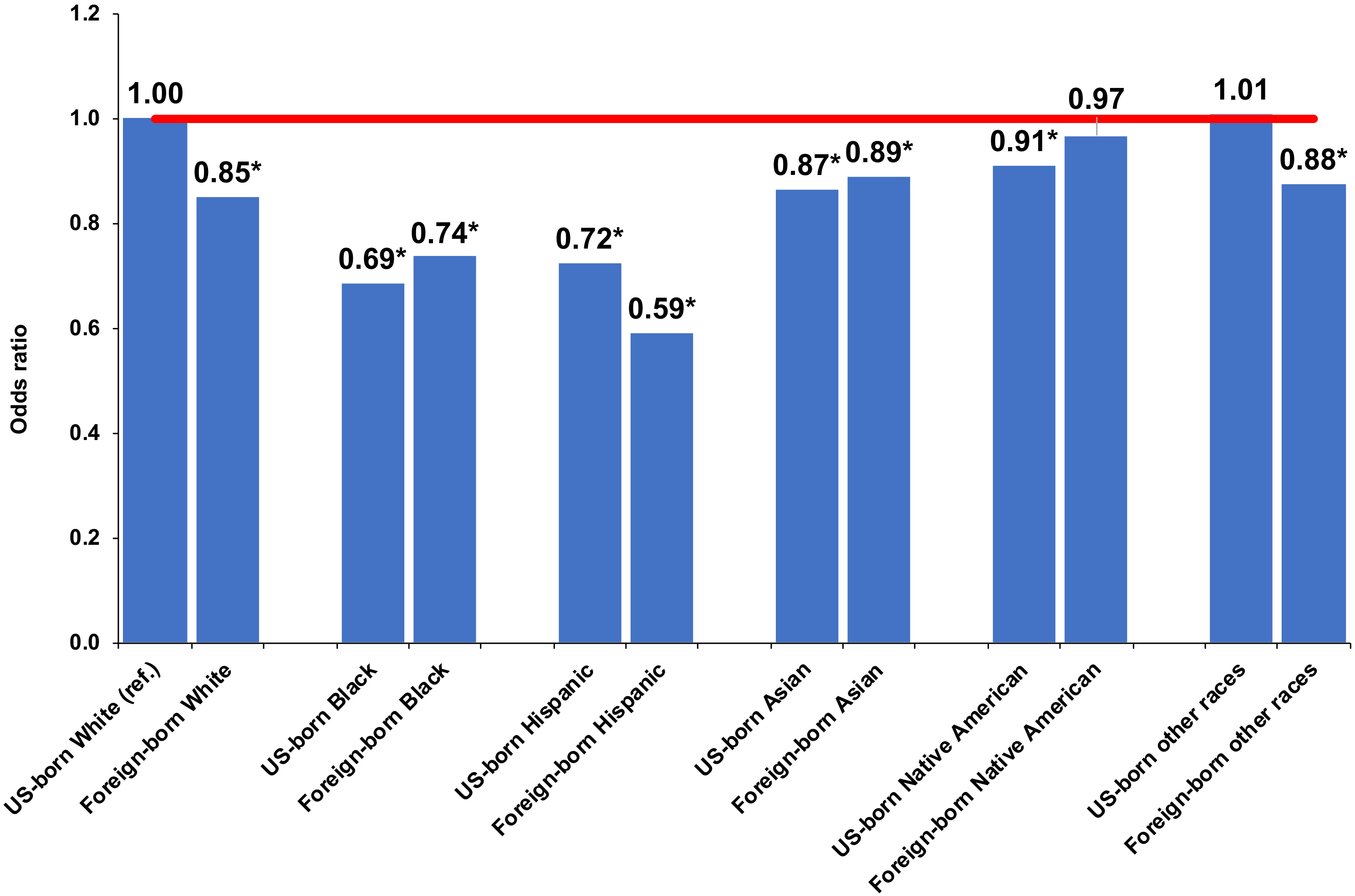
Note: Only for people aged 18+. * Significant at p<.01.
Source: 2005–2019 American Community Surveys.

Odds ratios of being an internal migrant by race/ethnicity



Note: Only for people aged 18+. * Significant at p<.01.
Source: 2005–2019 American Community Surveys.

Odds ratios of being an internal migrant by nativity and race/ethnicity



Note: Only for people aged 18+. * Significant at p<.01.
Source: 2005–2019 American Community Surveys.

Analysis of spatial association

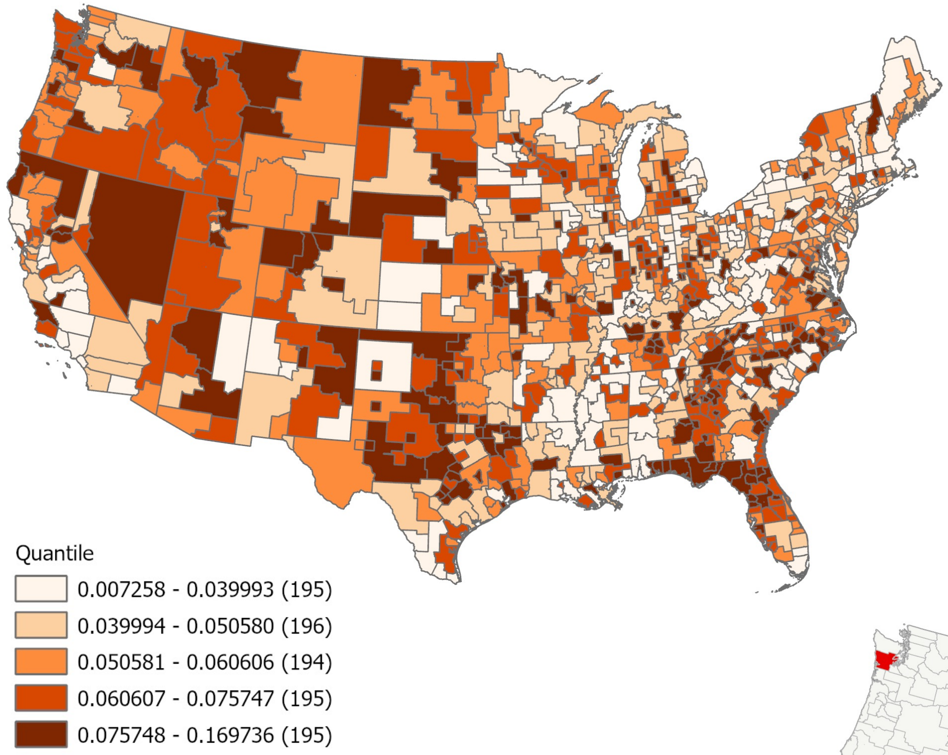
- In spatial association analysis, we recognize that people are not randomly distributed over space
- Local indicator of spatial association (LISA) identifies local clusters and spatial outliers
  - It estimates contributions of each area (Anselin 1995)
  - We considered neighbors as areas sharing a border (queen contiguity)
- We analyze concentration of internal migrants in areas of destination in the U.S.

Spatial clusters and outliers

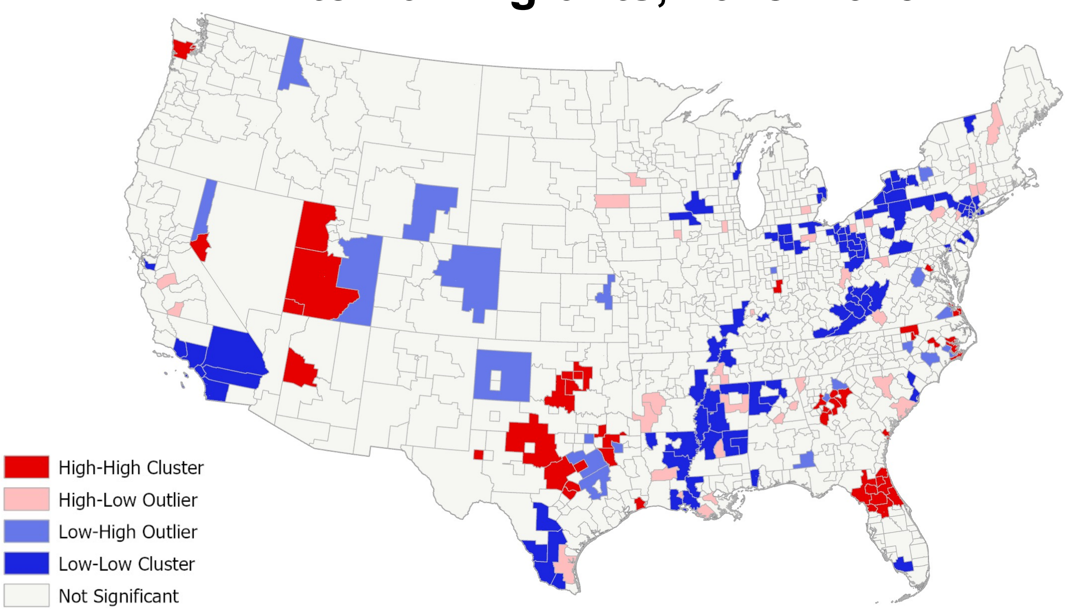
- **Spatial clusters**
  - **High-High**: areas with <u>high</u> levels of a specific indicator surrounded by areas with <u>high</u> levels for that indicator
  - **Low-Low**: areas with <u>low</u> levels of a specific indicator surrounded by areas with <u>low</u> levels for that indicator

- **Spatial outliers**
  - **High-Low**: areas with <u>high</u> levels of a specific indicator surrounded by areas with <u>low</u> levels for that indicator
  - **Low-High**: areas with <u>low</u> levels of a specific indicator surrounded by areas with <u>high</u> levels for that indicator

Proportion of internal migrants, 2018–2019



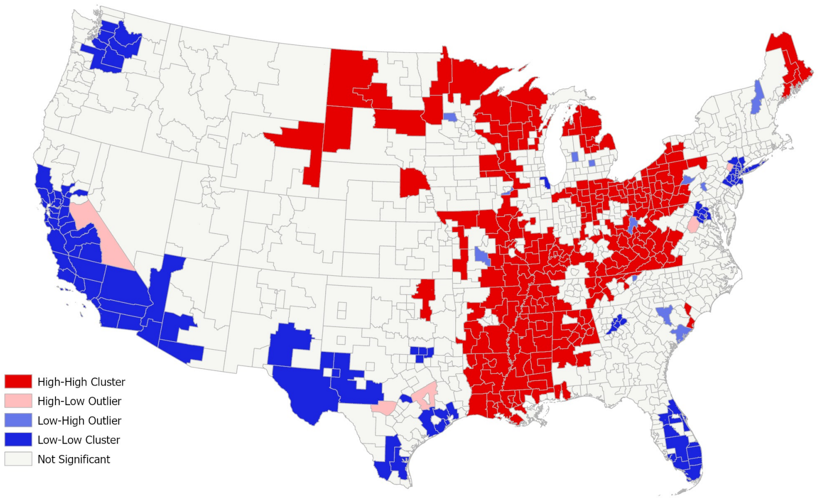
LISA of proportion of internal migrants, 2018–2019



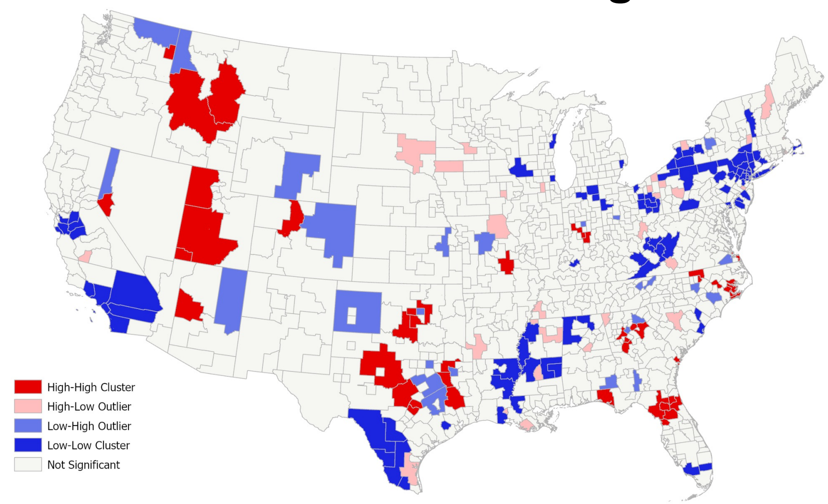
Source: 2019 American Community Survey.

Internal migrants are those who changed residence between 2018 and 2019

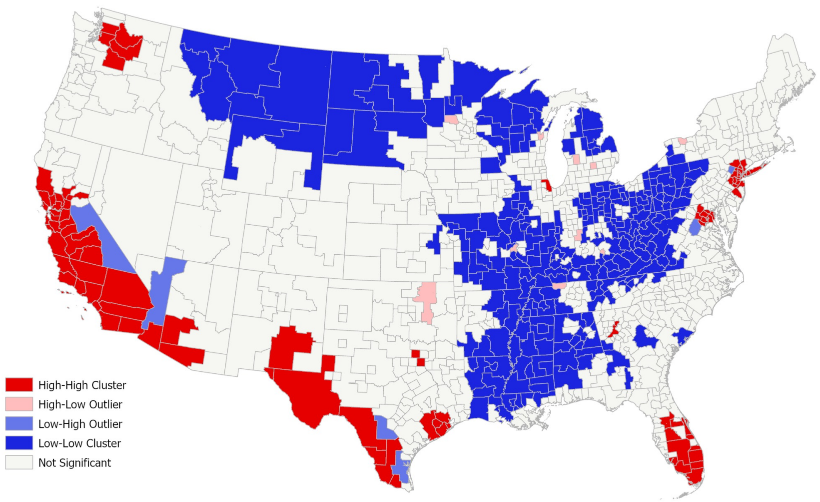
US-born non-migrants



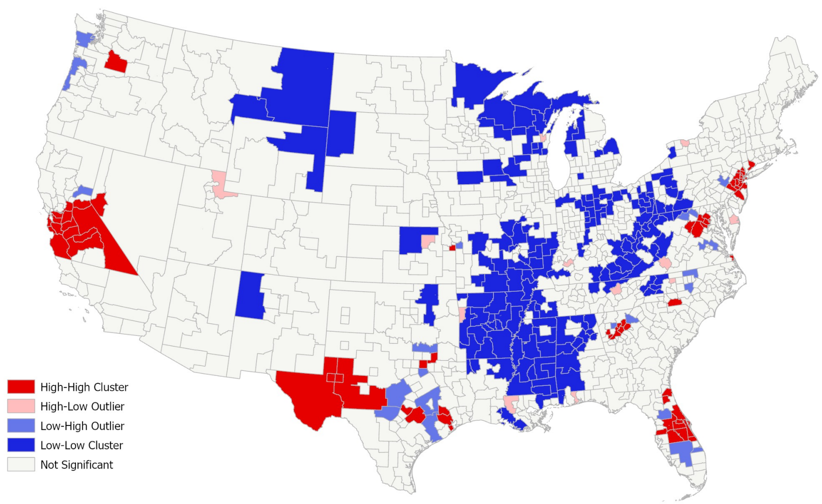
US-born internal migrants



Foreign-born non-migrants



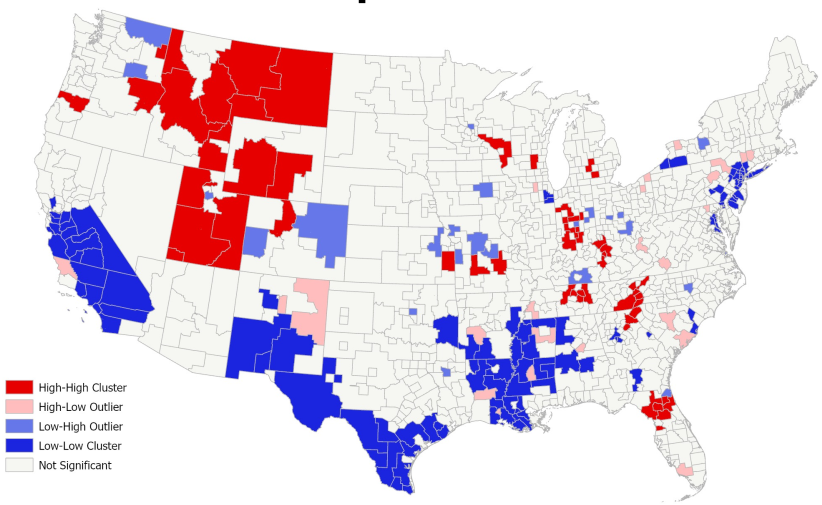
Foreign-born internal migrants



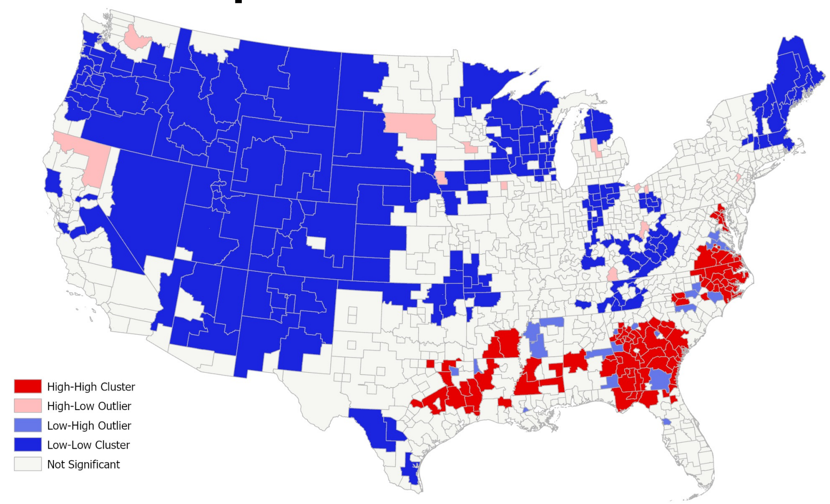
Source: 2019 American Community Survey.

All maps below are for internal migrants, 2018–2019

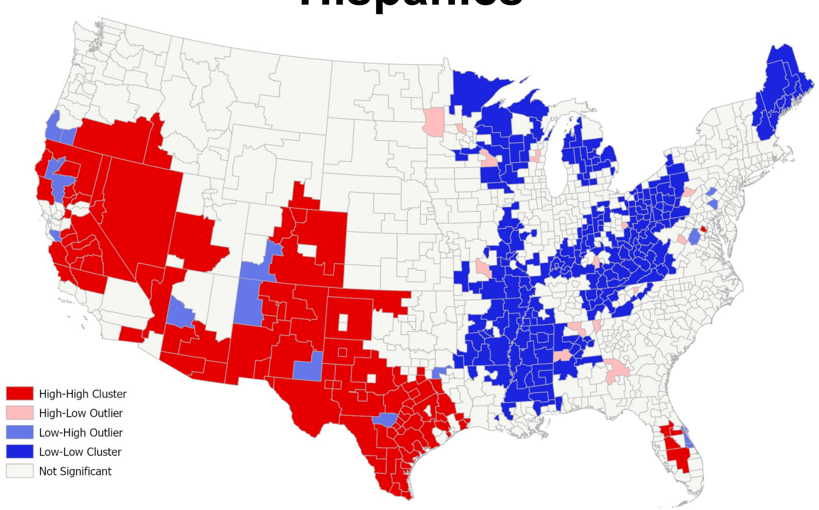
Non-Hispanic Whites



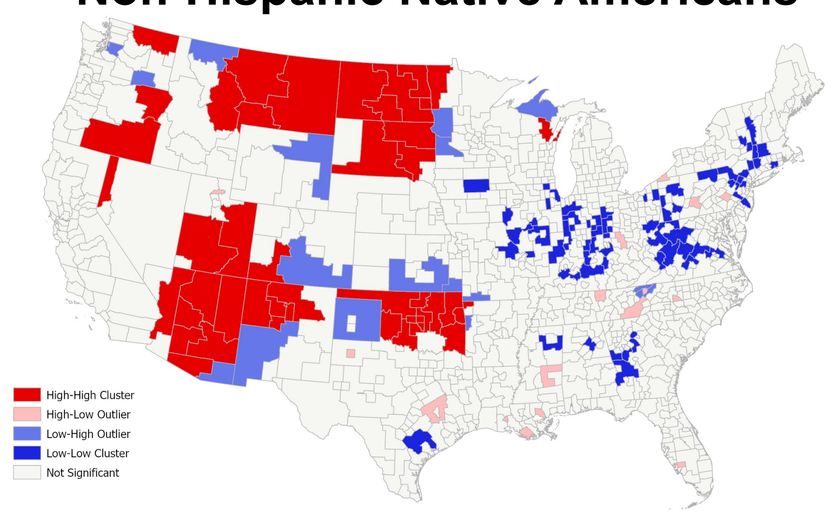
Non-Hispanic African Americans



Hispanics



Non-Hispanic Native Americans



Source: 2019 American Community Survey.

Final considerations

- Factors associated with migration rates similar to previous findings (Molloy, Smith, Wozniak 2011; Moretti 2011)
- Neoclassical theory (Molloy, Smith, Wozniak, 2017)
  - People move to areas with more jobs
  - Areas in Midwest with economic issues still have higher concentration of non-migrants
- Social networks (Motel, Patten 2012)
  - Spatial patterns of internal migration vary for different nativity and race/ethnicity groups
  - Areas with large proportions of specific race/ethnicity groups are attracting more of these groups

Next steps

- We will continue this analysis by incorporating 1950–2000 Censuses and 2005–2019 ACS
  - Analyze restricted data at the Texas Research Data Center (TXRDC) at Texas A&M University
- Models will estimate variations of area-level counts of migrants as the dependent variable
  - Integration of individual-level and area-level models
  - Distance and spatial terms will be introduced in the individual-level models as additional sets of predictors

Area-level models

- Gravity models
  - These models will have a set of independent variables, including distance between areas
- Autoregressive spatial models
  - Spatial dependence: influence of neighboring areas at origin and destination on the likelihood of migrating (Anselin, Rey 2014; LeSage, Pace 2008, 2009; Sardadvar, Vakulenko 2020)
  - Bayesian statistics approach: use prior knowledge based on other data sources and historical trends (LeSage, Fischer 2016; LeSage, Satici 2016)

Gravity models

- Poisson models will 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)

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

  - 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 in area of origin at the beginning of the period (P_i)
  - b_2: coefficient associated with the population in area of destination 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

Spatial models

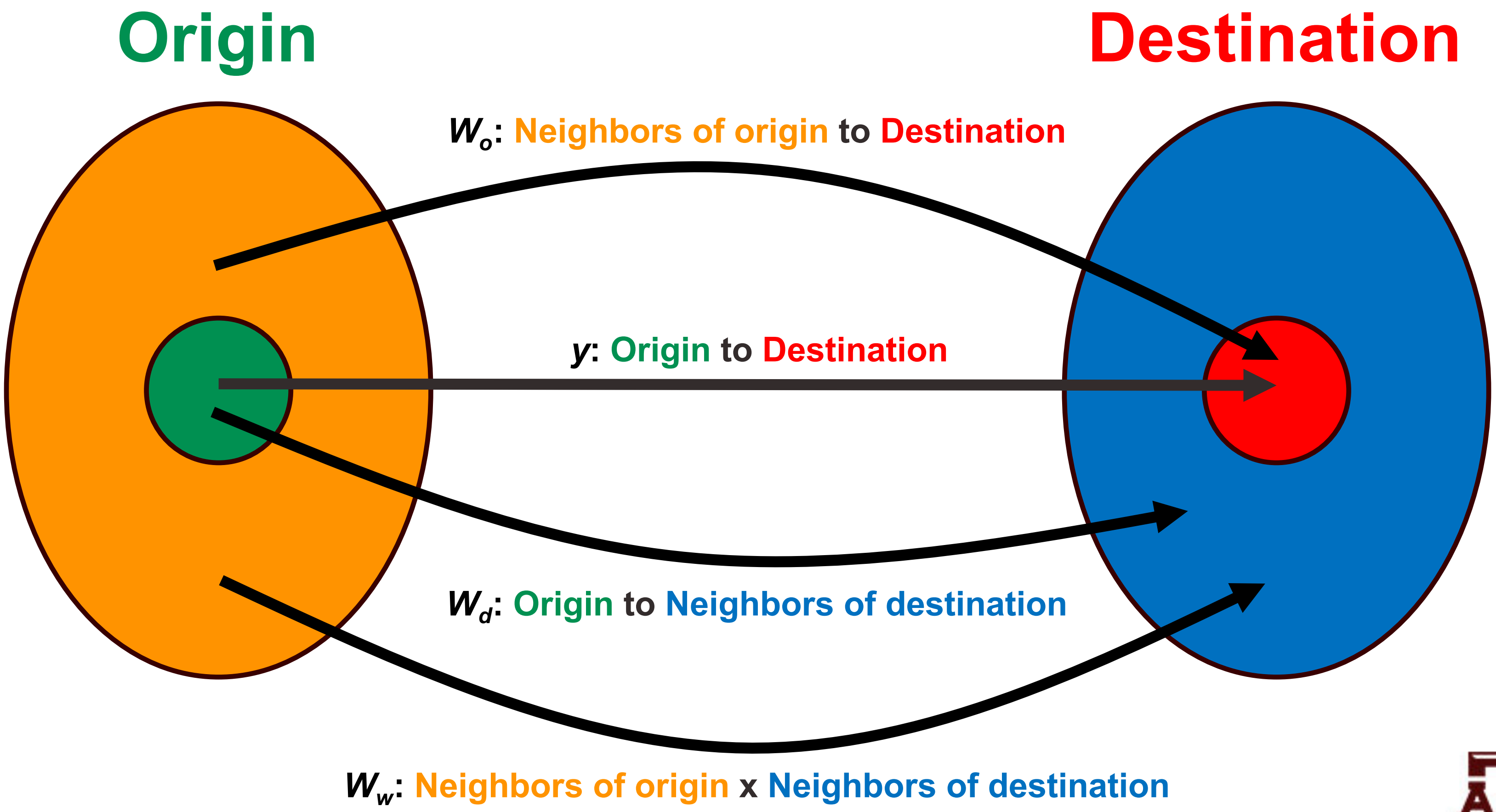
- The general spatial autoregressive model takes into account origin, destination, and origin-to-destination dependence (LeSage, Pace 2008, 2009)

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

  - W_o: spatial dependence at the origin
  - W_d: spatial dependence at the destination
  - W_w: interaction between origin and destination neighbors
  - X_o: characteristics for each of the regions of origin
  - X_d: characteristics for each of the regions of destination
  - Scalar γ: effect of distance g
  - α: constant term parameter on ι_N regions

Spatial dependence

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



Bayesian statistics approach

- Use IRS data to determine prior distributions
  - IRS sample size is much larger than ACS
- Then, we can estimate models with ACS
  - More detailed information about socioeconomic and demographic characteristics

Comparison between American Community Survey and IRS county-to-county migration data

Issue	ACS Migration Products	IRS Migration Data
Sample size	Approximately 2 million households per year	116 million+ households
Data universe	Sample is all US households	Universe is tax-filing households
Coverage period	2005–2016	1990–2016
Time period reported	Five-year average	Annual
Demographic characteristics	Each five-year product reports different sociodemographic characteristics (e.g., 2010–2014 contains relationship, household type, and tenure, 2011–2015 contains age/sex/race/Hispanic origin	No demographic characteristics

Source: Hauer, Byars 2019.

Research agenda

- Include a **longitudinal analysis** by linking individuals through time across censuses and surveys (Alexander et al. 2015; Leibbrand et al. 2019; Leibbrand et al. 2020; Logan, Stults, Xu 2016; Logan, Xu, Stults 2014; Wagner, Layne 2014)
- **Intergenerational mobility** among internal and international migrants (Leibbrand et al. 2019; Leibbrand et al. 2020)
- Estimate effects of our predicted migration flows on local **labor, health, and educational outcomes**
- Integrate **external data sources** to include other covariates
- Investigate **Mexico-U.S. migration** by merging other surveys
- Conduct **immigration policy simulations** to inform policymakers on the impacts of various policy options
- **Simulate future migration flows** under different hypothetical scenarios (Massey, Zenteno 1999; Klabunde, Willekens 2016)

Agent-based models

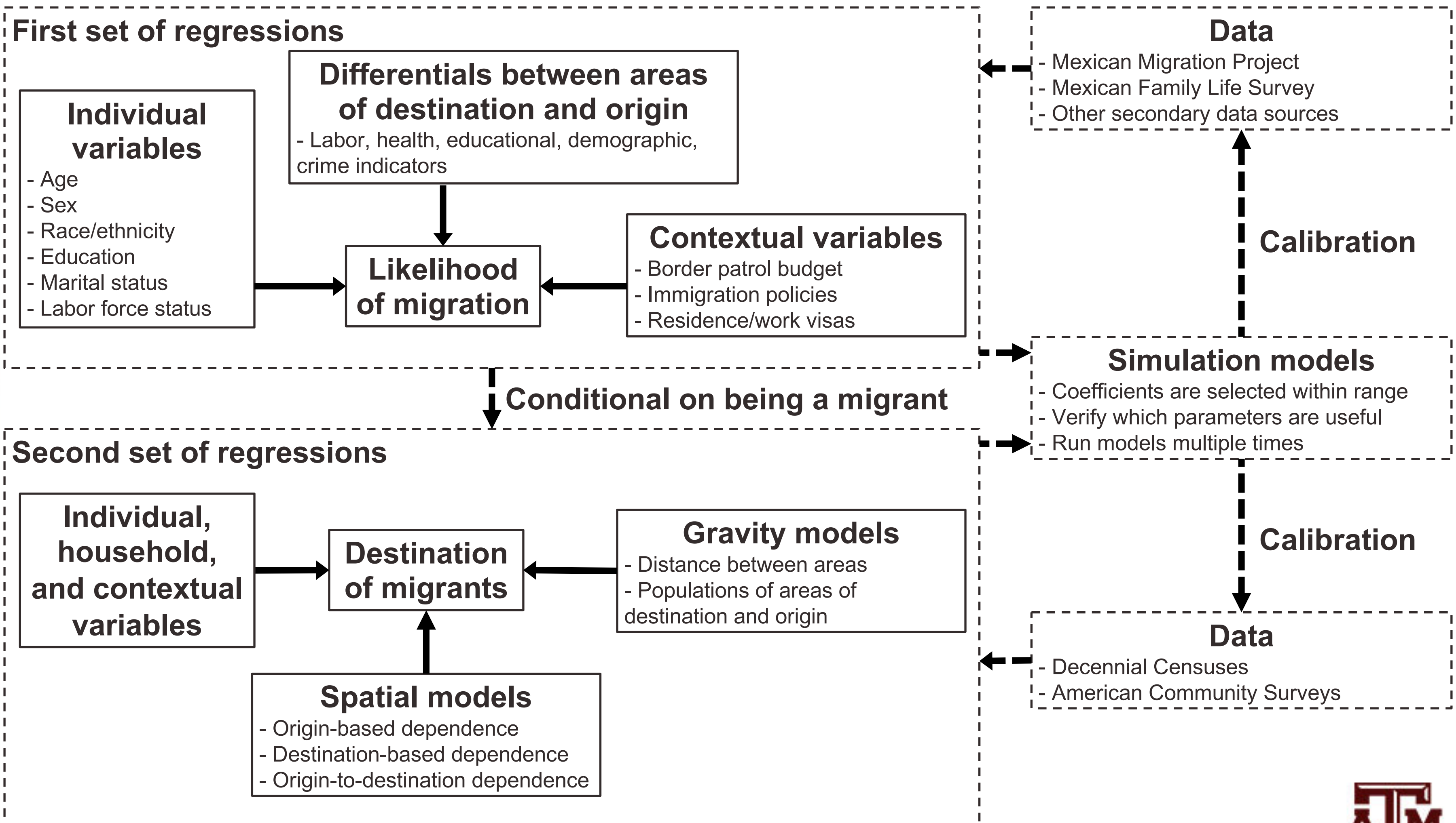
- Agent-based models can incorporate interactions between individual decisions, behavioral responses, and social networks related to migration outcomes (Massey, Zenteno 1999; Klabunde, Willekens 2016; Klabunde et al. 2017)
- These models can formalize interconnections and simulate potential feedback relationships between migration streams and several endogenous predictors
  - Education systems
  - Labor markets
  - Healthcare systems
  - Migration policies, border security
  - Social networks
- Agent-based models allow us to build different scenarios and simulate future population flows (Kabunde et al. 2017)

External data sources

- Combine textual data with demographic data with machine learning methods (Alburez-Gutierrez et al. 2019)

- Integrate data from other sources (e.g., social media, textual archives, private companies' data) to Census Bureau databases
  - Traditional datasets have the advantage of providing representative samples at the national, state, and local levels
  - Information from other sources tend to be more up to date (Alexander et al. 2019)

Mexico-U.S. migration



References

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.

Bijak J, Hilton J, Silverman E, Cao VD. 2013. “Reforging the Wedding Ring: Exploring a Semi-Artificial Model of Population for the United Kingdom with Gaussian process emulators.” Demographic Research, 29(27): 729–766.

Cadena BC, Kovak BK. 2016. “Immigrants equilibrate local labor markets: Evidence from the Great Recession.” American Economic Journal: Applied Economics 8(1): 257–290.

Cirillo P, Gallegati M. 2012. “The Empirical Validation of an Agent-based Model.” Eastern Economic Journal, 38: 525–547.

Frey WH. 2019. “For the first time on record, fewer than 10% of Americans moved in a year: Millennials are driving the trend.” The Brookings Institution (https://www.brookings.edu/blog/the-avenue/2019/11/22/for-the-first-time-on-record-fewer-than-10-of-americans-moved-in-a-year/).

Grazzini J, Richiardi M. 2015. “Estimation of ergodic agent-based models by simulated minimum distance.” Journal of Economic Dynamics & Control, 51: 148–165.

Hauer M, Byars J. 2019. “IRS county-to-county migration data, 1990–2010.” Demographic Research, 40(40): 1153–1166.

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.

Klabunde A, Zinn S, Willekens F, Leuchter M. 2017. “Multistate modelling extended by behavioural rules: An application to migration.” Population Studies, 71(S1): S51–S67.

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.

Massey DS, Zenteno RM. 1999. “The dynamics of mass migration.” Proceedings of the National Academy of Sciences (PNAS), 96: 5328–5335.

Molloy R, Smith CL, Wozniak A. 2011. “Internal Migration in the United States.” Journal of Economic Perspectives 25(3): 173–196.

Molloy R, Smith CL, Wozniak A. 2017. “Job changing and the decline in long-distance migration in the United States.” Demography 54: 631–653.

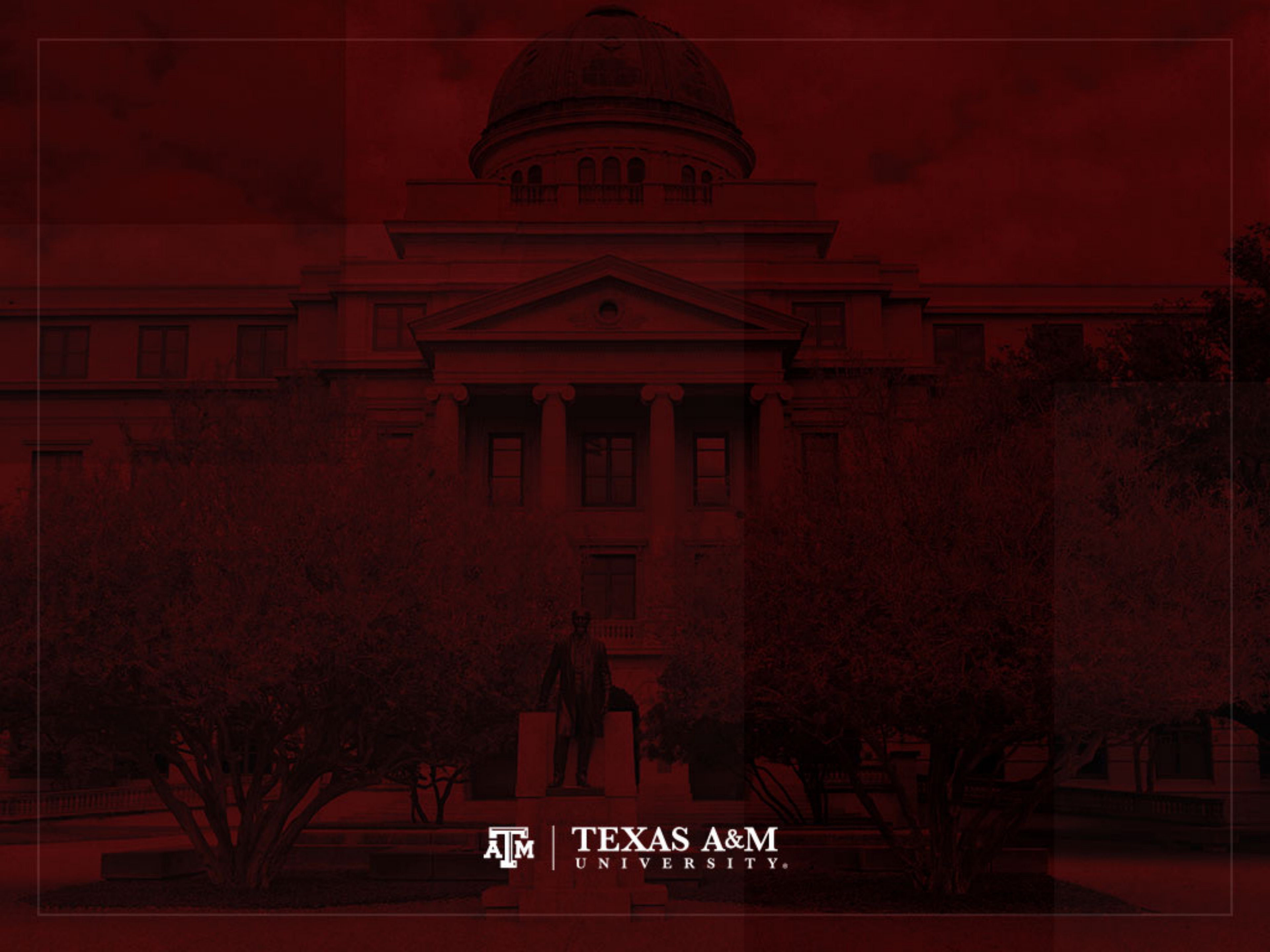
Moretti E. 2011. “Local labor markets.” In Handbook of Labor Economics, edited by O. Ashenfelter and D. Card, 1237-1313. Amsterdam: North-Holland.

Motel S, Patten E. 2012. “Characteristics of the 60 largest metropolitan areas by Hispanic population.” Pew Research Center, Hispanic Trends. http://www.pewhispanic.org/2012/09/19/characteristics-of-the-60-largest-metropolitan-areas-by-hispanic-population/.

Poole D, Raftery AE. 2000. “Inference for deterministic simulation models: The Bayesian melding approach.” Journal of the American Statistical Association, 95(452): 1244–1255.

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.


TEXAS A&M
UNIVERSITY