# Young Women in Cities
# Urbanization and Gender-biased Migration


Yumi Koh
U of Seoul

Jing Li
Singapore Management U

Yifan Wu
Shanghai Trust

Junjian Yi
Peking U

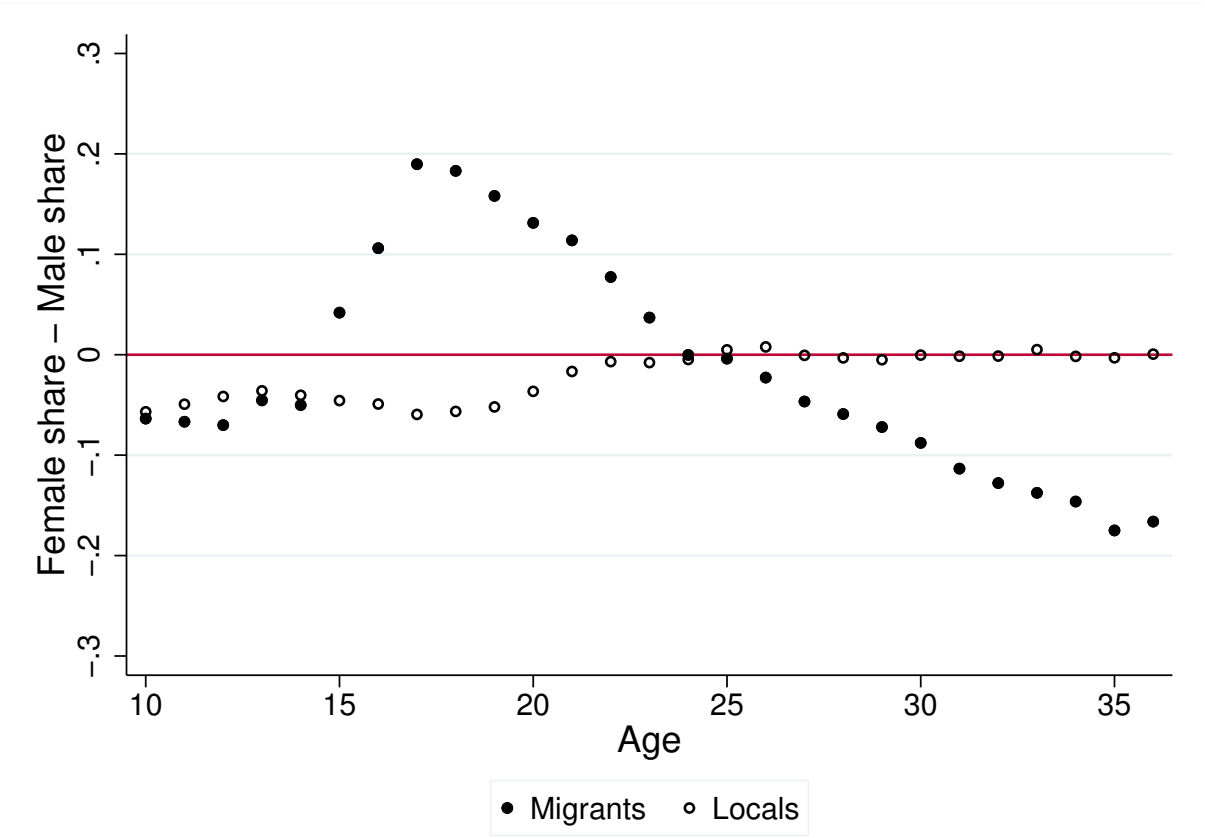
Hanzhe Zhang
Michigan State U



AEA
Saturday, January 4, 2025

# 1. Young women outnumber young men in Chinese cities

This is driven by migration (rather than birth): Young women in rural areas are more likely than their male counterparts to migrate to urban areas.




Data source: The 2000 China Population Census

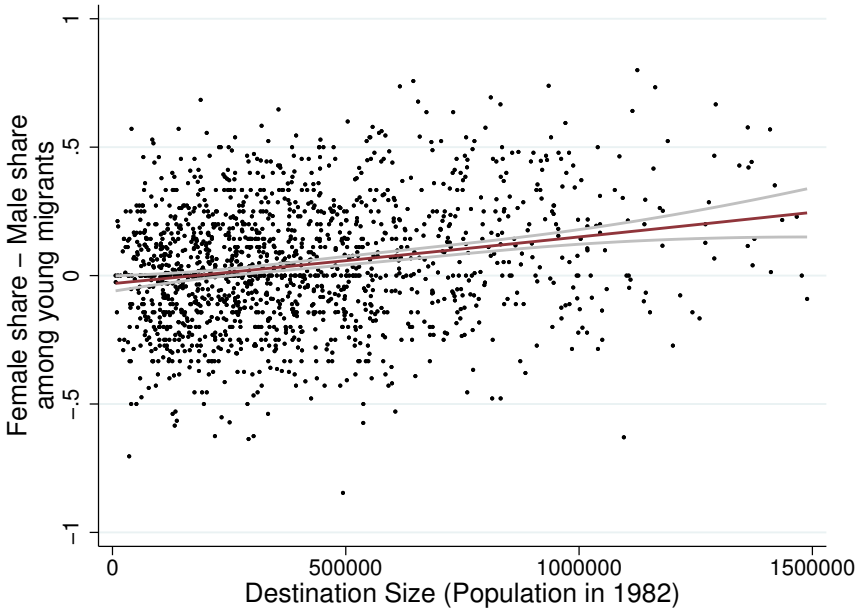
# 1. Young women outnumber young men in cities globally

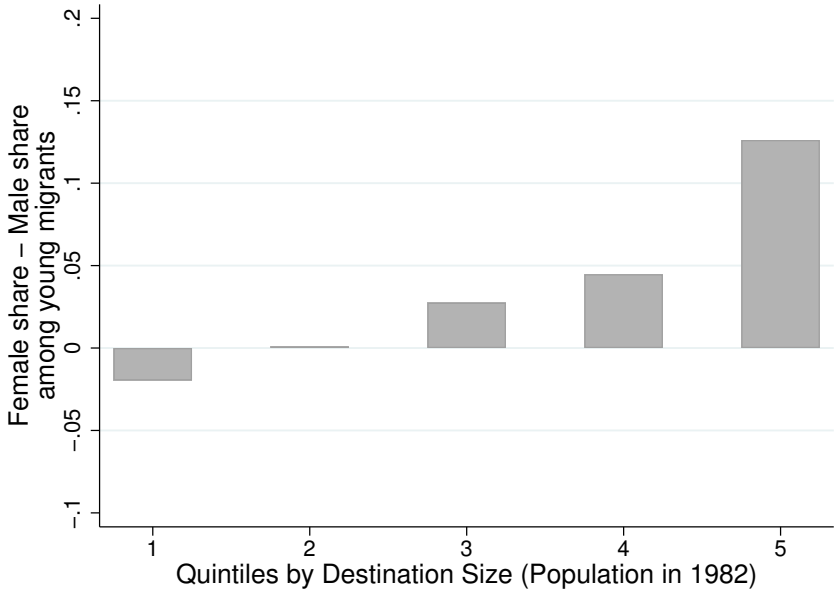
A similar pattern of female-dominant gender imbalance is observed in many other countries including but not limited to:

- all Central and South American countries (Tacoli 2012),
- most large cities in Germany and Russia (Wiest et al. 2013),
- Scandinavian countries (Pettay et al. 2021),
- India (PTI 2017), and
- Vietnam (Nguyen 2022).

# 2. Gender imbalance is more pronounced in larger cities

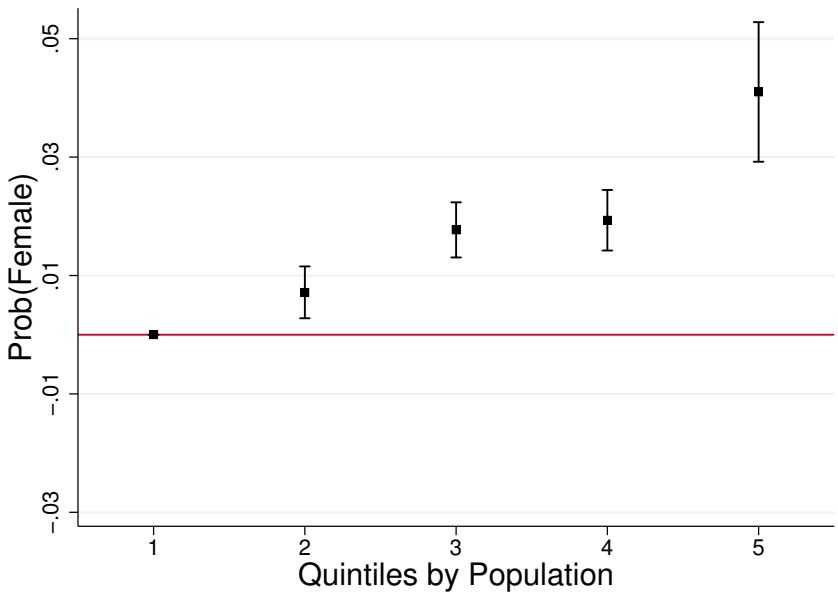
There are more females than males, especially in larger cities.

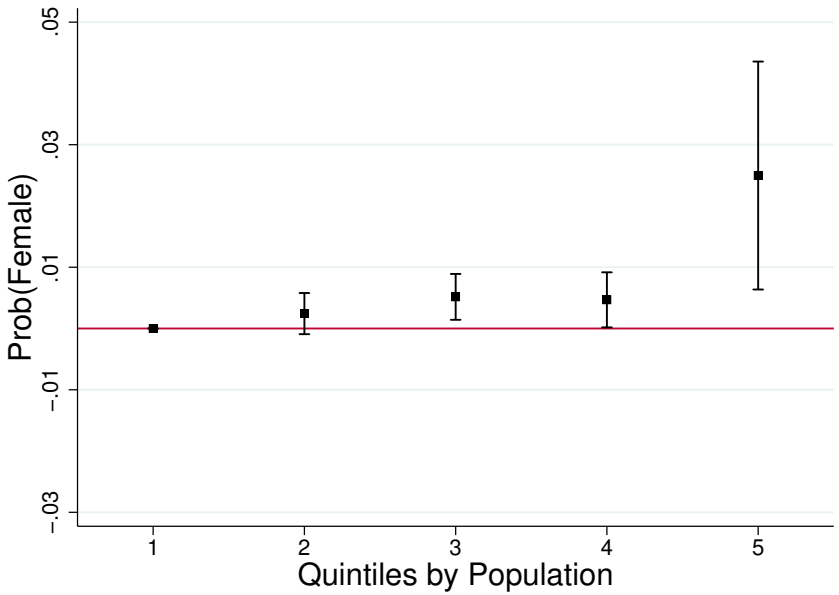


# 2. Gender imbalance is more pronounced in larger cities

The pattern remains consistent after accounting for variations in industrial composition through the inclusion of industry fixed effects.



# Main research goals

- We quantify gender disparity in rural-to-urban migration.
- We empirically investigate its potential causes.

# Identification

- We exploit the gradual roll-out of **special economic zones** (SEZs) across China in the early development stage as a quasi-experiment.
  - The SEZs create location-specific time variations in the extent of urbanization of the treated locations, which affect relative economic attractiveness across locations and individuals' migration incentives.
- We use the 2000 China Census data to estimate the impact of SEZ-induced economic shocks on population size and migration decisions between 1996 and 2000 (relaxed Hukou and pre-WTO).
- We quantify gender differences.

# Empirical Findings

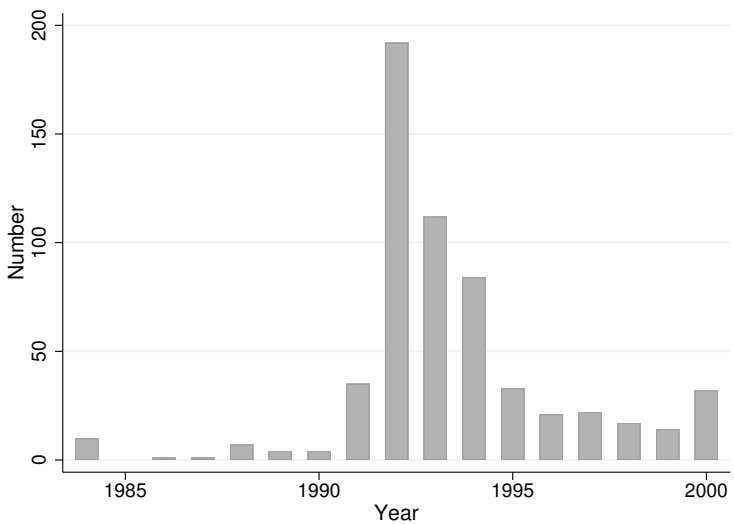
- A positive and statistically significant treatment effect of SEZ on the size of the population and inflow of young migrants.
  - The opening of an SEZ in a county increases the inflow of young female migrants by 43–48% and that of young male migrants by 35–38%.
- The gender difference in marital incentives plays a major role.
  - Young females' migration decisions are largely driven by higher (expected) returns from the marriage market in larger urban areas.
  - in addition to education, amenities, or industrial composition changes in labor market

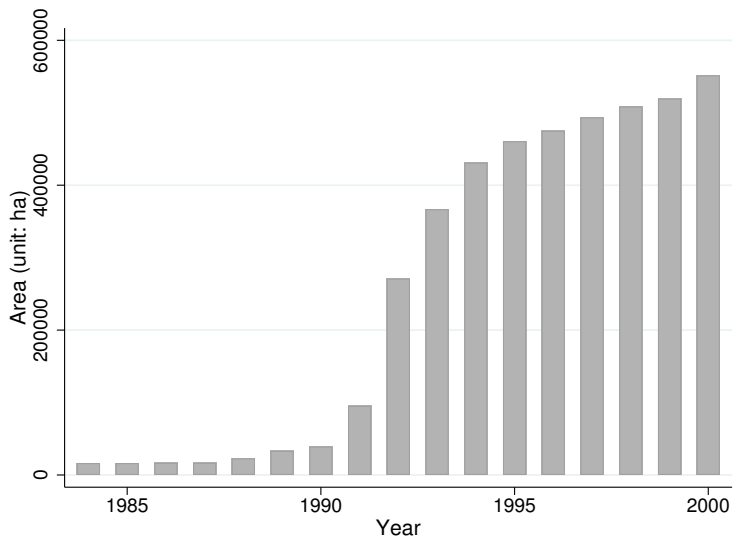
# Institutional Background – SEZs in China

- China's SEZs were established in the late 1970s as part of China's economic reform and opening up policy (Shirk et al. 1993).
- The first SEZ was established in Shenzhen in 1979, and was followed by three other SEZs in Zhuhai, Shantou, and Xiamen in 1980.
- These four cities were chosen because of their proximity to Hong Kong and Taiwan, and were intended to serve as pilot projects for China's economic reforms (Xu 2011).
- The SEZs were designed to attract foreign investment and promote exports, and were given special economic policies and incentives to facilitate economic growth.

# Establishment of SEZs Across Time

- The success of the initial four SEZs led to their proliferation in other cities and regions of China.
- In the 1990s, the central government embraced SEZ development as a national strategy, with the intention of achieving geographic diversity.



# Establishment of SEZs Across Space

# Institutional Background – The Hukou System and Migration in China

China witnessed unprecedented internal migration due to SEZs: The estimated migrant population reached 380 million.

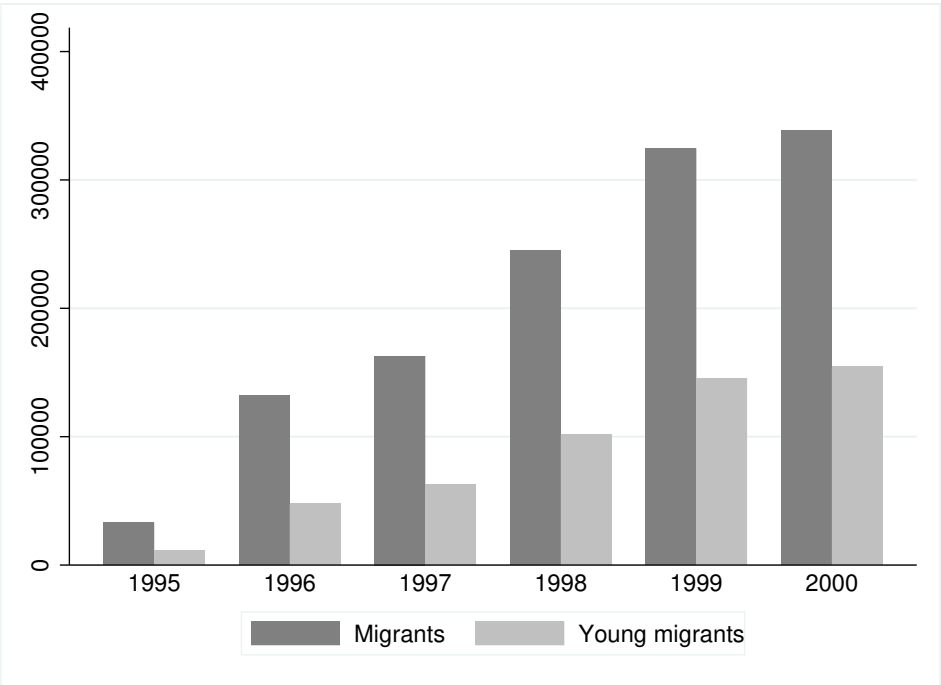



Figure: Total Number of Migrants (in 10,000s)

# Data Overview

1. National Population Census of China in 2000.
   - Information on sex, age, education level, marital status, migration status, county of residence, migration year, etc. of surveyed individuals.
2. Data on SEZs (especially 1996 to 2000 and 2001 to 2006)
   - List of SEZs published by the National Development and Reform Commission of China.
   - Information on name, code, approval date, approval authority, and major industries encouraged for each one of the SEZs.

# Data Construction

We construct our data at two different levels.

1. County-year level:
   - We track the changes in population size, migrants, and SEZ treatment status for each county across each year from 1996 to 2000.
2. Individual Level:
   - We form a panel dataset of young individuals aged 16 to 25 between 1996 and 2000 to track changes in their migration and marital status over time at the individual level.

We focus on cross-county migrants, because we track SEZ shocks at the county level.

# Empirical Research Designs – TWFE

- Exploiting the heterogeneous timing of SEZ establishments, we first analyze the aggregate-level data to see how population size changed in response to SEZ shocks using the TWFE model:

$$y_{c,t} = \alpha + \beta \cdot \mathbb{1}(\text{SEZ})_{c,t} + FE_c + FE_{c,t} + \epsilon_{c,t}$$

  - Subscripts $c$ and $t$ capture county and year, respectively.
  - $y_{c,t}$: the logarithm of population size for various groups of interest.
  - $\mathbb{1}(\text{SEZ})_{c,t}$: a binary indicator capturing whether county $c$ has ever been treated by a SEZ before.
  - $FE_c$: county fixed effects.
  - $FE_{c,t}$: prefecture-by-year fixed effects.
  - $\epsilon_{c,t}$: idiosyncratic error.

# Empirical Research Designs – TWFE

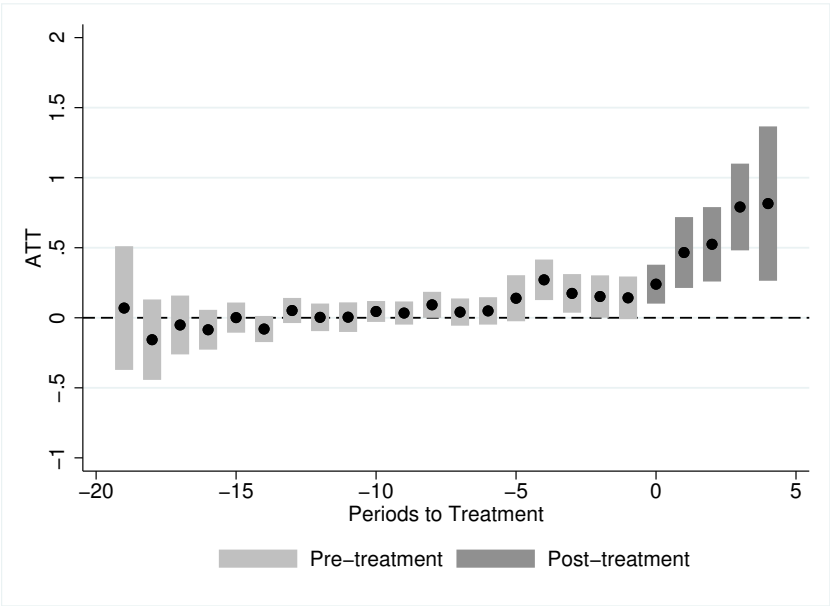
## Weighted by County Population

| | All Migrants | Young Migrants (Aged 16-25) | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dep. variable | Total | Females | Males |
| $\mathbb{1}(\text{SEZ})_{ct}$ | 0.4196*** | 0.4821*** | 0.3864*** |
| | (0.1198) | (0.0867) | (0.0820) |
| Observations | 6,876 | 6,876 | 6,876 |
| Control Group | Not-yet | Not-yet | Not-yet |

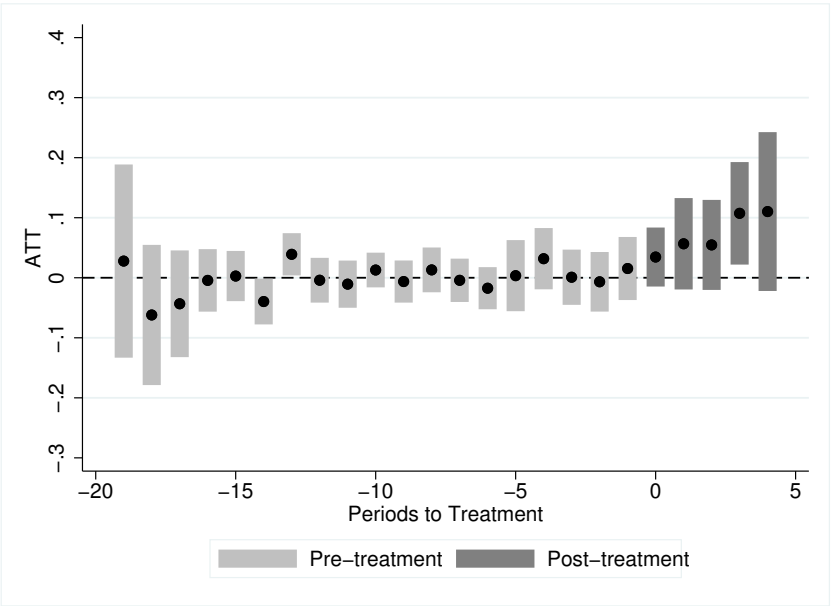
# Empirical Research Designs – Staggered DiD

## Weighted by County Population

To complement our TWFE model, we estimate a staggered DiD model (Callaway & Sant'Anna 2021; Sant'Anna & Zhao 2020). We compute the average treatment effect on the treated across different cohorts and years.




(a) Young Migrants




(b) Female Share in Young Migrants

# Empirical Research Designs – Staggered DiD

## Weighted by County Population

| | All Migrants | Young Migrants (Aged 16-25) | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dep. variable | Total | Total | Female Share |
| $\mathbb{1}(\text{SEZ})_{ct}$ | 0.4196*** | 0.4920*** | 0.0592** |
| | (0.1198) | (0.0887) | (0.0261) |
| Observations | 6,876 | 6,876 | 6,876 |
| Control Group | Not-yet | Not-yet | Not-yet |

# Empirical Research Designs – FD

$$\Delta(\textit{Migrate})_{i,c} = \beta\Delta(\text{Pull Factor})_c + \gamma\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{F})_i + \epsilon_{i,c}.$$

- Subscripts $i$ and $c$ denote individual and county, respectively.
- $\Delta(\textit{Migrate})_{i,c}$: a binary indicator that equals to 1 if individual $i$ migrated out of county $c$ to another county sometime between 1996 and 2000.
- $\Delta(\text{Pull Factor})_c$: Bartik-like changes in external urbanization intensity between 1996 and 2000, which attract residents of county c to migrate to elsewhere.
- $\Delta\epsilon_{i,c,p}$ is an idiosyncratic shock.

# External Urbanization Intensity: $\Delta$(Pull Factor)

- The incentives to migrate are affected by economic opportunities that exist outside the county of origin.
  - ex) When there is an economic boom due to SEZ outside the county of origin, people may migrate out in search of better lives.
- $\Delta$(Pull Factor) captures the degree of urbanization brought by SEZ shocks that occurred outside between 1996 and 2000

$$\Delta(\text{Pull Factor}) = \boldsymbol{M} \cdot \boldsymbol{V}$$

  - $\boldsymbol{M}$: the migration flow between provinces back in 1995 used as pre-determined weights;
  - $\boldsymbol{V}$: the count of counties in each province treated with SEZs for the first time between 1996 and 2000.
- Construct Expected $\Delta$(Pull Factor) using SEZs established 1996 to 2006 (Borusyak & Hull 2023).

# Empirical Research Designs – FD

| Dep. var: $\Delta$ Migrate$_{ic}$ | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\Delta$(Pull Factor)$_c$ | 0.0267*** | 0.0162*** | 0.0270*** | 0.0164*** |
| | (0.0006) | (0.0010) | (0.0006) | (0.0010) |
| $\Delta$(Pull Factor)$_c \times \mathbb{1}$(Female)$_i$ | 0.0036*** | 0.0016*** | 0.0036*** | 0.0017*** |
| | (0.0002) | (0.0004) | (0.0002) | (0.0004) |
| Expected $\Delta$(Pull Factor)$_c$ | | 0.0129*** | | 0.0130*** |
| | | (0.0011) | | (0.0011) |
| Expected $\Delta$(Pull Factor)$_c \times \mathbb{1}$(Female)$_i$ | | 0.0023*** | | 0.0023*** |
| | | (0.0004) | | (0.0004) |
| $\Delta$(SEZ)$_c$ | | | -0.0267*** | -0.0278*** |
| | | | (0.0090) | (0.0085) |
| $\Delta$(SEZ)$_c \times \mathbb{1}$(Female)$_i$ | | | -0.0047 | -0.0044 |
| | | | (0.0048) | (0.0046) |
| Observations | 2,562,835 | 2,562,835 | 2,562,835 | 2,562,835 |
| Root MSE | 0.302 | 0.301 | 0.302 | 0.301 |
| Mean(y) | 0.100 | 0.100 | 0.100 | 0.100 |

# Potential Explanations

- We now investigate the observed gender imbalance pattern among young migrants by exploring various possible interpretations.
- We examine each of the following potential explanations and conduct heterogeneity analyses on different subgroups to gain a deeper understanding of the phenomenon.
  - Education
  - Amenities
  - Industrial Composition Change
  - Marriage Market

# Potential Explanation: Education

- Females may place a higher value on education, and education quality may be better in larger cities, or the gender gap in returns to education is greater in larger cities, with women experiencing a higher increase in returns compared to men.
- To investigate such a possibility, re-estimate our FD model using the sample that excludes individuals who responded in the Census that they migrated for education-related purposes.

| Dep. var: $\Delta(\text{Migrate})_{i,c}$ | (1) Unadjusted OLS | (2) Controlled OLS | (3) Unadjusted OLS | (4) Controlled OLS |
|---|---|---|---|---|
| Panel (a) Education – Excluding Student Migrants | | | | |
| $\Delta(\text{Pull Factor})_c$ | 0.0022** | 0.0022** | 0.0024** | 0.0024** |
| | (0.0011) | (0.0011) | (0.0011) | (0.0011) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | 0.0035*** | 0.0018*** | 0.0035*** | 0.0019*** |
| | (0.0002) | (0.0004) | (0.0002) | (0.0004) |
| Expected terms | No | Yes | No | Yes |
| Own SEZ shocks | No | No | Yes | Yes |

# Potential Explanation: Amenities

- Another possible explanation is that larger cities may offer amenities that are more attractive to young females than to young males.
  - ex) Certain types of amenities that support family life may be valued more by young females and larger cities may offer more of such amenities.
- We estimate the FD model for the high-skilled and low-skilled samples separately and check whether the migration tendency is stronger among the high-skilled group, and particularly so for females.
  - High-skilled individuals are known to value amenities more than low-skilled individuals (Diamond 2016).
  - We proxy skill using education level.

# Potential Explanation: Amenities

| Dep. var: $\Delta(\text{Migrate})_{i,c}$ | (1) Unadjusted OLS | (2) Controlled OLS | (3) Unadjusted OLS | (4) Controlled OLS |
|---|---|---|---|---|
| Panel (b) Amenities – Sample of Less-educated Individuals | | | | |
| $\Delta(\text{Pull Factor})_c$ | -0.0003 | 0.0000 | -0.0001 | 0.0003 |
| | (0.0012) | (0.0012) | (0.0012) | (0.0012) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | 0.0051*** | 0.0029*** | 0.0052*** | 0.0029*** |
| | (0.0003) | (0.0005) | (0.0003) | (0.0005) |
| Panel (c) Amenities – Sample of Highly-educated Individuals | | | | |
| $\Delta(\text{Pull Factor})_c$ | 0.0068*** | 0.0069*** | 0.0070*** | 0.0071*** |
| | (0.0014) | (0.0015) | (0.0014) | (0.0015) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | -0.0016*** | -0.0015*** | -0.0016*** | -0.0015*** |
| | (0.0002) | (0.0004) | (0.0003) | (0.0004) |
| Expected terms | No | Yes | No | Yes |
| Own SEZ shocks | No | No | Yes | Yes |

# Potential Explanation: Industrial Composition Change

- There may be concerns that large urban areas offer a higher concentration of service sector jobs, which may rely more heavily on female labor.
  - ex) retail sales, education, personal care, health care.
- Gender imbalance in migration can result from changes in labor market conditions caused by SEZs, which may affect industries with different reliance on male and female labor.
- We construct measures $\Delta(\text{Labor Mkt})^F$ and $\Delta(\text{Labor Mkt})^M$ to control for possible changes in gender-specific labor market conditions driven by SEZ establishments.

# Potential Explanation: Industrial Composition Change

| Dep. var: $\Delta(\text{Migrate})_{i,c}$ | (1) Unadjusted OLS | (2) Controlled OLS | (3) Unadjusted OLS | (4) Controlled OLS |
|---|---|---|---|---|
| Panel (d) Industrial Composition Change – Including Labor Market Conditions | | | | |
| $\Delta(\text{Pull Factor})_c$ | 0.0026** | 0.0018 | 0.0029** | 0.0021* |
| | (0.0012) | (0.0012) | (0.0012) | (0.0012) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | 0.0016*** | 0.0012*** | 0.0016*** | 0.0012*** |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| Expected terms | No | Yes | No | Yes |
| Own SEZ shocks | No | No | Yes | Yes |

# Potential Explanation: Marriage Market

- When SEZs create jobs and stimulate labor market growth, local population can increase as a surge of workers are attracted to the area in search of employment opportunities.

- Such influx of people creates a larger pool of unmarried individuals, which can affect the dynamics of the local marriage market.
  - High-skilled males are drawn to large urban areas due to skill-biased technological progress.
  - This can have a ripple effect on young females' migration choices as prospects of marrying up increases in large urban areas.

- We thus check whether the gender imbalance pattern also exists for the subgroup of individuals who are likely to be already married.

# Potential Explanation: Marriage Market

| Dep. var: $\Delta(\text{Migrate})_{i,c}$ | (1) Unadjusted OLS | (2) Controlled OLS | (3) Unadjusted OLS | (4) Controlled OLS |
|---|---|---|---|---|
| Panel (e) Marriage Market – Sample of Older Individuals | | | | |
| $\Delta(\text{Pull Factor})_c$ | 0.0045*** | 0.0037*** | 0.0047*** | 0.0039*** |
| | (0.0006) | (0.0007) | (0.0006) | (0.0007) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | -0.0050*** | -0.0034*** | -0.0050*** | -0.0035*** |
| | (0.0001) | (0.0002) | (0.0001) | (0.0002) |
| Expected terms | No | Yes | No | Yes |
| Own SEZ shocks | No | No | Yes | Yes |

# Potential Explanation: Marriage Market

| | (1) Unadjusted OLS | (2) Controlled OLS | (3) Unadjusted OLS | (4) Controlled OLS |
|---|---|---|---|---|
| Panel (a) Dep. var: $\Delta(\text{Marry})_{i,c}$ | | | | |
| $\Delta(\text{Pull Factor})_c$ | -0.0076*** | -0.0069*** | -0.0076*** | -0.0068*** |
| | (0.0006) | (0.0007) | (0.0006) | (0.0006) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | 0.0132*** | 0.0104*** | 0.0133*** | 0.0104*** |
| | (0.0003) | (0.0005) | (0.0003) | (0.0005) |
| Panel (b) Dep. var: $\Delta(\text{Migrate \& Marry})_{i,c}$ | | | | |
| $\Delta(\text{Pull Factor})$ | -0.0000 | 0.0000 | 0.0001 | 0.0001 |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| $\Delta(\text{Pull Factor}) \times \mathbb{1}(\text{Female})_i$ | 0.0017*** | 0.0016*** | 0.0017*** | 0.0016*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Panel (c) Dep. var: $\Delta(\text{Marry up})_{i,c}$ | | | | |
| $\Delta(\text{Pull Factor})_c$ | -0.0049*** | -0.0026*** | -0.0049*** | -0.0026*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$ | 0.0105*** | 0.0053*** | 0.0106*** | 0.0054*** |
| | (0.0002) | (0.0004) | (0.0002) | (0.0004) |
| Expected Terms | No | Yes | No | Yes |
| Own SEZ shocks | No | No | Yes | Yes |

# An Equilibrium Marriage-Market Model

- Each person is endowed with one of two skill/education types, high and low: $H$ and $L$ in urban; $h$ and $\ell$ in rural
- Each rural person is endowed with a heterogeneous differential $y \in [-Y, Y]$ that indicates the (education + amenity + labor-market) gain from moving to the city.
- Total marriage surplus $s_{\theta_m \theta_w}$ is determined by husband's and wife's skill types $\theta_m$ and $\theta_w$ and is divided in equilibrium.
- **Claim**: Under mild assumptions, there is a unique equilibrium in which more skilled men and more unskilled women migrate.

# Conclusion

- This paper studies gender imbalance induced by urbanization.
- Our empirical study suggests that the interplay between labor and marriage markets plays an important role in migration.

# THANK YOU!

hanzhe@msu.edu