Young Women in Cities: Urbanization and Gender-biased Migration[*]

Yumi Koh[†] Jing Li[‡] Yifan Wu[§] Junjian Yi[¶] Hanzhe Zhang[‖]

January 4, 2024

Abstract

Young women outnumber young men in cities in many countries during periods of economic growth and urbanization. This gender imbalance among young urbanites is more pronounced in larger cities. We use the gradual rollout of Special Economic Zones across China as a quasi-experiment to establish the causal impact of urbanization on gender-differential incentives to migrate. We highlight the role of the marriage market in increasing rural women's chance of marrying and marrying up in urban areas during rapid urbanization.

Keywords: urbanization, migration, gender imbalance, marriage market

JEL classifications: O15, J12

[*]We thank Kristian Behrens, Janet Currie, Scott Hegerty, Ben Zou, and the audience at various seminars and conferences. Zhang acknowledges the National Science Foundation.

[†]School of Economics, University of Seoul

[‡]School of Economics, Singapore Management University

[§]Strategic Development Department, Shanghai Trust

[¶]China Center of Economic Research, National School of Development, Peking University

[‖]Department of Economics, Michigan State University

1 Introduction

Women outnumber men in urban areas in most countries (World Bank 2023).[1] This gender imbalance is largely influenced by migration, rather than by factors present at birth. Figure 1 depicts the net female share (female minus male share) by age for internal migrants and locals, based on Chinese Population Census 2000. The net female share is close to zero across all ages for locals, indicating a balanced gender distribution. In contrast, the net female share is positive for young migrants, suggesting an excess of young female migrants.

Moreover, the gender imbalance among young urbanites is more pronounced in larger cities. Figure 2 shows that young female migrants are predominantly moving to larger cities. Figures 2a and 2b show that the surplus of young females among migrants is more prominent in destinations with a larger size, measured by population size as reported in Chinese Population Census 1982. Consequently, the overall gender imbalance among young individuals tends to increase with city size (Figure 2c). This trend persists even when accounting for heterogeneities in industrial composition across cities by controlling for industry-fixed effects (Figure 2d).

In this paper, we aim to (i) identify gender-differential migration patterns among young cohorts within a developing country during periods of rapid urbanization and (ii) investigate various migration incentives that may contribute to gender-specific migration patterns. Our empirical analysis centers on internal migration in China between 1995 and 2000. This provides an ideal setting because it was a time marked by rapid urbanization, active internal migration, and a growing gender imbalance across regions. Although our analysis is grounded in the specific context of China, the implications discussed later are broadly applicable to other developing countries that experience similar rapid urbanization.

To causally identify the impact of urbanization on gender-varying migration incentives, we exploit the gradual rollout of Special Economic Zones (hereinafter SEZs) across China. The SEZs create location-time-specific variations in the extent of urbanization, influencing economic

[1] Urban areas in many countries exhibit a notable surplus of young women, such as all Central and South American countries (Tacoli 2012); Germany and Russia (Wiest et al. 2013); India (PTI 2017); Scandinavian countries (Pettay et al. 2021); Vietnam (Nguyen 2022); and China, as we will document in this paper.

attractiveness across different locations and consequently impacting individuals' migration decisions. Our identification leverages the numerous SEZs established across China in a sporadic manner during our sample period and the quasi-random establishment timing driven by the bureaucratic process. These features are particularly important because identifying the causal link between urbanization and migration is subject to endogeneity concerns: The birth and growth of a city depend on a comprehensive set of location-specific factors such as culture, geography, transportation, natural resources, climate, and history (Mumford 1961), which are often difficult to observe and measure. To the extent that these factors correlate with gender-differential preferences for living in urban areas, it could give rise to an endogeneity problem. Our empirical strategy that exploits the rollout of SEZs helps address address such endogeneity concerns.

We use two sets of data: (i) Chinese Population Census 2000 and (ii) the list of SEZs published by National Development and Reform Commission of China. By aligning the migration timing of individuals in the census with the establishment timing of SEZs, we estimate the impact of SEZ openings on gender-specific volumes of migration for each year between 1995 and 2000. Our estimates from the staggered difference-in-differences (hereinafter DiD) model reveal that the opening of an SEZ increases the inflow of young female migrants by 47% and that of young male migrants by 38% at the county level.

To further account for individual heterogeneity, we estimate a first-difference (hereinafter FD) model at the individual level. We construct a Bartik-like composite explanatory variable that captures changes in the external attraction force in other locations between 1996 and 2000. This variable acts as a "pull force" that motivates people to relocate when other locations treated with SEZ openings become more economically attractive. Its impact is weighted by the pre-existing migration network in 1995. We find that a one standard deviation increase in the constructed pull factor leads to a 0.26 percentage-point increase in the probability of emigrating to other counties for males and a 0.46 percentage-point increase for females. These estimates not only exhibit statistical differences between males and females but also demonstrate a substantial magnitude of difference, particularly notable considering that approximately 10%

of the population migrated during our sample period. These results remain robust even after addressing potential biases that arise from the endogeneity of the pre-existing migration network in constructing the pull factor, using the method proposed by Borusyak and Hull (2023).

To further investigate potential mechanisms driving gender-differential migration among young individuals, we develop a theoretical framework that articulates marital and nonmarital incentives to migrate. Each person is endowed with a type and resides in either an urban or rural area. A person's payoff comprises an exogenous nonmarital component and an endogenously determined marital component. Individuals in rural areas decide whether to migrate to urban areas, and the urban and rural marriage markets clear based on these migration decisions. The division of total surplus from marriage is determined in a general equilibrium. While the model accommodates various migration patterns, we focus on an equilibrium consistent with observed trends: increased migration of rural women to cities. The model highlights the role of both marital and nonmarital factors. On the marital front, given the patrilocality and hypergamy context in China, more rural women may be attracted to migrate in search of more competent and wealthier husbands. On the nonmarital aspect, better labor market opportunities, educational opportunities, and increased benefits from amenities may also attract women.

We empirically analyze the marital and nonmarital incentives, respectively. Our estimates from the FD model, based on various subsamples and additional marital market outcomes, align with explanations centered on marital incentives. We find that a stronger pull force not only increases the likelihood of marriage but also the likelihood of marrying up for young females, in comparison to their young male counterparts. In contrast, in the sample of older individuals who are likely to be already married, females show a statistically lower magnitude of emigration in response to a larger pull factor. As for nonmarital incentives, we explore alternative potential mechanisms related to the industrial shifts in the labor market, educational opportunities, and amenities. While these nonmarital factors may partially account for our observations, they do not seem to provide a complete explanation for our findings.

Our study carries weight for both scholars and policymakers. The disparities that accompany rapid urbanization are notable between urban and rural areas and between males and females. From the perspective of the marriage market, the presence of relatively more young females in urban areas benefits males and disadvantages females in cities; the reverse is true in rural areas. One implication of our study is that such widening spatial inequality and gender divide may have far-reaching social implications on marriage and birth outcomes. Spatial mismatch in gender may further cause a decline in overall social and family stability. Another implication of our study is that the gender imbalance in migration driven by marital market incentives can further have a reciprocal effect on the labor market. Specifically, the presence of more women in urban areas can potentially affect the types of jobs available, competition for job search, and the gender wage gap in cities.

This study contributes to two strands of literature. First, it adds to the literature on gender differences in urbanization. Starting from Marshall (1890), researchers have documented cities' advantages in higher productivity, higher wages, and better amenities, which provide incentives to migrate and settle (Rosenthal and Strange 2004; Combes and Gobillon 2015; Diamond 2016; Couture and Handbury 2017; Fan and Zou 2021). A few studies highlight gender differences in response to such urban advantages, leading to variations between urban and rural areas in terms of gender gaps in labor participation, wages, and entrepreneurship (Phimister 2005; Rosenthal and Strange 2012; Bacolod 2017). In this paper, we conclude that the gender difference in migration incentives of rural youths is also likely driven by gender-differential returns from the marriage market in cities—an important perspective that has not been fully studied in the literature.

Second, this paper contributes to the literature on gender differences in premarital investment and consequent marriage and labor market outcomes. Previous studies mainly focus on premarital investment in the form of wealth, education, or their interactions (Zhang 1994; Peters and Siow 2002; Chiappori et al. 2009; Zhang 2021; Bhaskar et al. 2023; Zhang and Zou 2023). Similar to Dupuy (2021) and Ahn et al. (2023), our paper considers migration as premarital investment. In contrast to studies that mostly suggest the theoretical importance of premarital investments, we employ carefully designed empirical methods to causally identify their importance. Our research thus contributes to the classic question of who marries whom, as explored by Choo and Siow (2006) and Choo (2015), with a specific focus on the context of economic development, urbanization, and migration.

The rest of the paper is organized as follows. Section 2 details the institutional background. Section 3 describes the data and variables. Section 4 lays out the empirical design. Section 5 reports empirical results. Section 6 discusses potential mechanisms, and Section 7 concludes. The remaining proofs and results can be found in the appendices.

2 Institutional Background

2.1 SEZs in China

China's SEZs were first 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 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). SEZs were designed to attract foreign investments and promote exports. These zones were equipped with special economic policies and incentives that aimed to facilitate economic growth (Alder et al. 2016).[2] The economic performance of the four initial SEZs was remarkable. For example, between 1980 and 1990, Shenzhen's gross domestic product grew at an average rate of approximately 28% per year (National Bureau of Statistics of China 2021).

The success of the four initial SEZs led to their proliferation in other regions. In the 1990s,

[2]Such incentives included preferential tax policies that lower or exempt corporate taxes; simplified customs procedures to facilitate trade; reduced bureaucratic procedures that ease business operations; preferential land use, such as lower land-use fees and priority access to land; access to credit and other financial resources; and openness to foreign investment by allowing foreign investors to own their entire enterprises without the need for a Chinese partner.

the central government embraced SEZ development as a national strategy, with the intention of achieving geographic diversity. Figure 3a depicts the number of SEZs established per year from 1984 to 2000 and Figure 3b shows the cumulative area of SEZs established across years in the bottom panel. The number and area of SEZs increased significantly in the 1990s. Figure 4 illustrates the temporal geographic expansion of SEZs with four panels representing different years (1990, 1995, 2000, and 2005). Between 1995 and 2005, SEZs were established across China in a dispersed manner and played a significant role in promoting urbanization in China.

Our identification strategy relies on the quasi-random variation in the timing of SEZ establishment between 1996 and 2005, given their selection, as outlined in detail in Section 4. In the 1980s, SEZs were primarily located in China's eastern coastal regions at the initial stages of economic reform. The establishment and success of these SEZs have contributed to the prosperity of the coastal regions but also exacerbated economic disparities between different regions. Consequently, efforts were made in the 1990s to establish SEZs in a more balanced manner across the country to reduce regional disparities. Especially since 1995, the SEZs were scattered across the country almost evenly (Swerts et al. 2021). For example, SEZs established between 1996 and 2005 share very similar longitudes. During this stage, the time of their establishment can be considered quasi-random as it varied primarily due to the bureaucratic processes involved in approving the SEZs (Crane et al. 2018).

2.2 The Hukou System and Internal Migration in China

Before the economic reform in 1979, migration within China was rare under the *hukou* system, which has been in place since the 1950s (Young 2013). The system is based on household registration, which assigns every Chinese citizen a place of origin that is recorded on their hukou or household registration document. The hukou system divides the population into two categories: rural and urban. A rural (resp., an urban) hukou is assigned to individuals who were born and raised in the countryside or smaller towns (resp., in cities). The hukou system serves various purposes, including the determination of access to public services such as education and healthcare, and tracking population movement. The system restricts access to public services and job opportunities based on an individual's place of origin, which makes it difficult for people to move from rural to urban areas and receive the same level of services as they would in their place of origin.

China's economic reform in 1979 brought significant changes to the hukou system. With the establishment of SEZs and the transition to a market-oriented economy, there was growing demand for labor in urban areas and many rural residents began to migrate to cities in search of work. However, the hukou system remained a significant barrier to migration and mobility, since rural residents were not allowed to obtain an urban hukou; this restricted their access to social welfare benefits and public services. To address this issue, the government began to implement reforms to the hukou system in the 1980s. One of the key changes was the introduction of a temporary residence system, which allowed rural residents to obtain temporary urban residence permits. These permits granted them access to certain social welfare benefits and public services in urban areas, although they were not equivalent to an urban hukou.

In the 1990s, the government introduced further reforms to the hukou system to promote greater social inclusion and mobility. One of the significant changes was the introduction of the "floating population" concept, which recognized the existence of migrant workers and granted them certain legal rights and protections. The government has been extending social welfare benefits and public services to nonlocal migrants, while fostering increased labor mobility.

With economic growth and the relaxation of the hukou system, China witnessed an unprecedented wave of internal migration in the 1990s and 2000s. The urban population share increased from 19.39% in 1980 to 26.22% in 2000 (National Bureau of Statistics of China 2021). In particular, based on our calculations using Chinese Population Census 2000, the total number of cross-county migrants aged between 16 and 25 grew about 13-fold from 1995 to 2000 (Figure 5). According to the population census conducted by the National Bureau of Statistics in 2020, the number of migrants was estimated to be 376 million.

2.3 Hypergamy and Patrilocal Practices in China

Two enduring marital traditions—status hypergamy and patrilocal practices—continue to prevail in China. As we will elaborate in our subsequent analysis, understanding these practices is crucial, as they play a pivotal role in generating asymmetries between males and females in migration and marital decisions.

Status hypergamy describes the tendency of a woman to form a relationship with a man of higher social, economic, or educational status and is common across various regions and cultures. Although hypergamy has endured as a long-standing tradition in China, it has been further fueled by the increasing economic pressures resulting from reforms.[3] Table 1 presents evidence of hypergamy based on age gaps, as older men are more likely to be associated with greater economic resources compared to their younger counterparts. Analyzing all married couples with an age below 55 from the Chinese Population Census 2000, we find that majority of women are matched to older husbands, implying age hypergamy. Moreover, a higher prevalence of married couples with an age gap of 4 years or more is observed among migrant wives, as compared to their non-migrant counterparts. Education level serves as another proxy for men's higher status, and we examine marriage matching patterns in Table 2, where education levels are categorized into five groups for both husbands and wives. Level 1 (level 5) represents the lowest (highest) level of education. White cells indicate marriages with wives having higher education, the light-grey diagonal cells denote equal education levels, and dark-grey cells indicate marriages where husbands have higher education. Ninety-one percent of couples involve husbands who are equally or more educated than their wives.

Another important marriage tradition in China is the patrilocal practice, which refers to a residence pattern in which a married couple lives with or near the husband's family or the husband's relatives.[4] Its practice is still resilient. Our analysis using the 2010 China Family

[3]More specifically, "intensified labor market pressure, rising consumerism, and skyrocketing costs of living acted to promote marriages of older men to younger women on the basis of a need or preference for status hypergamy" (Mu and Xie 2014, p.151).

[4]Patrilocality is "a core aspect of the traditional Chinese kinship system and is deeply rooted in Confucianism" (Gruijters and Ermisch 2019, p.562).

Panel Studies shows that, for couples with partners born in different cities, 78.1% live in the husband's city after marriage, and 20.5% live in the wife's city. The share is more unbalanced if one of the spouses was born in a rural area: among those couples, 92.8% live in the husband's birth city, and merely 5.6% live in the wife's birth city.

3 Data and Variables

3.1 Data Overview

Our empirical analysis primarily relies on two datasets. The first is a 1% sample of Chinese Population Census 2000. From the sample, we use information on various demographic and migration-related variables from each surveyed individual: gender, year of birth, education level, marital status, marriage year, migration status, migration year, county of residence, as well as the county of origin and destination for those who migrated.

We focus on the 2000 census sample for three reasons. First, it contains one of the largest samples—1% of the population—compared with census samples available for analysis in later years. Second, it provides information on both the origin and the destination of a migrant, which is an advantage over the census data collected in earlier years. Third, it allows us to accurately identify the migration year up to 1995.[5] For this reason, our focus is on the changes that took place between 1995 and 2000 or between 1996 and 2000. The specific time frame depends on whether we need to exclude the initial year, 1995, to use predetermined conditions. Fourth, during the sample period, hukou restrictions on migration were significantly relaxed compared with earlier years, which allows us to observe a large sample of migrants. Last, the country has not yet been affected by its WTO accession in 2001.

The second dataset is the list of SEZs published by the National Development and Reform Commission of China.[6] The list provides information on the SEZ ID, name, approval date,

[5]The census asked respondents whether they had always lived in their birth town. If not, they were asked to provide the destination county and the year of their move. Due to the questionnaire design, all years before 1995 were grouped as "moved before 1995," so we can only determine the exact moving year up to 1995.

[6]The list was first published in 2006 (NDRC 2006) and later updated in 2018 (NDRC 2018). To ensure that

approval authority, and targeted industries for each SEZ. To identify counties that were treated with SEZs, we track the geographic boundaries of each SEZ established during our sample period and match them with the corresponding counties.

We construct our data at two levels. First, at the county-year level, we track the changes in migration size and SEZ treatment status for each county across each year from 1995 to 2000. Throughout this paper, we define individuals as migrants if they move across counties. We focus on cross-county migrants because we track SEZ shocks at the county level. In panel (a) of Table 3, we present the summary statistics on demographics in 2000 at the county level. All population and migrant count-related variables are reported after undergoing an inverse hyperbolic sine transformation. While there were more male migrants overall, there were more female migrants in the age group of 16 to 25. Specifically, we show that on average, 54% of young migrants in a county were females. Among all counties in our sample, 16% have at least one SEZ as of 2000.

Second, we form a panel dataset of individuals who were between ages 16 and 25 at any point from 1996 to 2000 to track changes in their migration status and marriage outcome over time at the individual level. In panel (b) of Table 3, we present individual-level summary statistics. On average, 49% of the sample were females. The average age is 22.73, 72% had an education level of middle school or below, 10% migrated across counties between 1996 and 2000, and 25% got married during the period. $\Delta(\text{Marry up})_{i,c}$ is set to one if someone married to a partner who has a strictly higher level of education during the years 1996 to 2000 and zero otherwise. Among the married, 16% married someone of strictly higher education level.

3.2 Urbanization-Induced Attractiveness: Δ(Pull Factor)

When other counties are designated as SEZs, it signifies that individuals are exposed to increased urbanization and enhanced economic prospects beyond their current county of residence. Consequently, the opportunities that stem from urban development in other areas can entice

we use the full information, we combine both lists and track SEZ establishments during our sample period.

individuals to consider relocating from their current residence. Such urbanization-induced attractiveness in other locations can act as a pull factor. The term "pull factor" is commonly used in the literature, and our usage aligns with established conventions.[7]

To quantify this pull factor, we construct the following composite variable:

$$\Delta(\text{Pull Factor}) = \boldsymbol{M} \cdot \boldsymbol{V} \tag{1}$$

The square matrix $\boldsymbol{M}$ represents predetermined weights derived from observed migration flows during the initial year of our dataset, 1995. Each row i of the matrix $\boldsymbol{M}$ corresponds to a specific province. The matrix's elements capture various facets of migration patterns. Entries $\boldsymbol{m}_{ij}$'s, where $i \neq j$, represent the count of individuals who migrated from province i to province j, signifying inter-provincial migrations. The diagonal elements $\boldsymbol{m}_{ii}$'s account for migrations within the same province i, but across different counties. To ensure meaningful comparisons, each element in row i is normalized by considering the total count of people who originally resided in province i but subsequently moved across counties in 1995. Additionally, vector $\boldsymbol{V}$ summarizes the SEZ shocks that occurred between 1996 and 2000. Its element i corresponds to the count of counties in province i designated as SEZs for the first time between 1996 and 2000.

The Δ(Pull Factor) combines $\boldsymbol{M}$ and $\boldsymbol{V}$ through multiplication. Its element i captures the level of urbanization influenced by SEZ shocks, encompassing migrations within province i but spanning different counties, as well as accounting for migrations originating from outside province i between 1996 and 2000. These influences are weighted by the predetermined migration network. A higher value implies a stronger urbanization-induced attraction, acting as a pull force that encourages people to relocate.

The Δ(Pull Factor) values constructed using the method outlined above apply only to individuals residing in counties that did not experience SEZ shocks for the first time from 1996 to 2000. For those living in counties that underwent their own internal SEZ shocks between

[7]For instance, in a review of the Great Migration literature, Aaronson et al. (2021) identify increased labor demand and better economic opportunities in the industrial North as pull factors that attract people from southern states to migrate to the northern and western states in the United States.

1996 and 2000, a slight calculation adjustment is required to net out the influence of these internal shocks on vector $\boldsymbol{V}$. To be specific, we subtract 1 from the count of SEZ-treated counties within the corresponding province element of vector $\boldsymbol{V}$ to net out the count of internal shock while keeping all other elements unchanged. This is because we want Δ(Pull Factor) to capture the external attraction pull force in a consistent manner. Consequently, Δ(Pull Factor) values are county-specific, even within the same province, depending on whether internal SEZ shocks occurred or not.[8]

4 Empirical Design

4.1 County-Level Analysis

Using the varying timing of SEZ establishments, we first analyze county-level data to observe how the size of the migrant population changed in response to SEZ shocks. Given the staggered treatment timing, we estimate a staggered DiD model, following the approach of Sant'Anna and Zhao (2020) and Callaway and Sant'Anna (2021). de Chaisemartin and D'Haultfoeuille (2020) and Goodman-Bacon (2021) find that a two-way fixed effects model does not yield an interpretable causal parameter when there are cross-sectional variations in treatment timing and heterogeneous treatment effects.

Among the individuals surveyed in Chinese Population Census 2000, we track the number of people who migrated to a particular county from other counties each year between 1995 and 2000. Thus, the outcome variable is the count of individuals migrating to a specific county from other counties in a given year. As there are cases with zero migrants in certain years for some

[8] An alternative method for constructing Δ(Pull Factor) involves defining $\boldsymbol{M}$ and $\boldsymbol{V}$ at the county level. In this scenario, $\boldsymbol{M}$ represents the normalized cross-county migration flows in 1995, while $\boldsymbol{V}$ is a vector consisting of ones and zeros, indicating whether each county received SEZ shocks during the 1996-2000 period. However, in our empirical context with 5,298 counties, 65.35% of the elements in matrix $\boldsymbol{M}$ are zeros due to the absence of migration flows in 1995 for 3,462 county pairs. To address this sparsity issue and to capture broader migration patterns and trends, we opt not to pursue this approach. In our dataset, 53.88% of migrations were between provinces, while the remaining involved cross-county and intra-province migrations. Because our construction of Δ(Pull Factor) encompasses both intra-province and cross-county migrations within a province, it consistently captures the attractiveness induced by urbanization.

counties, we employ an inverse hyperbolic sine transformation to handle these instances. This concave log-like transformation allows us to retain zero-valued observations, and the coefficient estimate yields a similar interpretation to that of a standard logarithmic specification (Bellemare and Wichman 2020).

To estimate a staggered DiD model, we categorize counties into cohorts based on the initial year an SEZ was established within a county. For example, one cohort includes counties where SEZs were first established in 1996, and another comprises those with SEZ establishments in 1997. In our estimation, we use a pool of counties that have not yet experienced SEZ establishment shocks by the end of our sample period (i.e. year 2000) as a comparison group. Hereinafter, we denote this control group as the "not-yet-treated" group. Considering the nationwide initiation of SEZ establishments in the late 1990s and the bureaucratic procedures affecting the timing of SEZ shocks, we find the not-yet-treated group to be an appropriate comparison. The identification of causal effects hinges on the assumption of parallel trends, justified by the quasi-random timing of SEZ establishments during the sample period.

4.2 Individual-Level Analysis

Whereas the staggered DiD model enables us to measure the impact of SEZs on migration flows, it is limited to capturing aggregate volumes of migration at the county level. To delve deeper into migration patterns using individual-level information, we employ an FD model described as follows:

$$\Delta(Migrate)_{i,c} = \beta\Delta(\text{Pull Factor})_c + \gamma\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i + \delta_i + \epsilon_{i,c}. \tag{2}$$

For individual i living in county c in 1996, the binary indicator $\Delta(Migrate)_{i,c}$ equals one if they migrated from county c to another county between 1996 and 2000. $\Delta(\text{Pull Factor})_c$ captures the changes in urbanization-induced attractiveness elsewhere between 1996 and 2000, which motivates residents of county c to emigrate. The coefficient of interest is γ, which captures whether females respond differently to the pull factor than males. We also include age dummies, denoted as δ_i, to account for the variation in age among individuals used in the estimation for running the FD model. Lastly, $\epsilon_{i,c}$ is an idiosyncratic shock. Standard errors are corrected for heteroskedasticity and clustered at county level.

The FD model estimator helps to analyze how SEZ-induced changes in attractiveness affect the migration decisions of individuals while accounting for cross-individual age heterogeneity. In an alternative specification, we also control for $\Delta(SEZ)_c$, which is a binary indicator that equals one if county c received its own internal SEZ shock for the first time between 1996 and 2000 and zero otherwise. The additional control variable reflects the internal urbanization shock generated by the county's own SEZ establishment, which acts as a "staying" force.

There may be bias concerns related to $\Delta(\text{Pull Factor})_c$, which combines the changes in SEZ shocks and exposure to such shocks using predetermined migration flows. Even if the SEZ shocks may be as good as random, omitted variable bias may still arise if *exposure* to exogenous shocks is not random (Adao et al. 2019; Goldsmith-Pinkham et al. 2020; Borusyak et al. 2022). To address this concern, we follow the estimation strategy proposed by Borusyak and Hull (2023). The idea is to randomly draw shocks that may plausibly have occurred, recompute the composite index, and repeat this across many simulations to compute their averaged value, denoted as the "expected $\Delta(\text{Pull Factor})_c$." Then we can avoid omitted variable bias by including the expected $\Delta(\text{Pull Factor})_c$ as a control variable in the regression. To compile a set of SEZ shocks that could have plausibly occurred, we incorporate 518 additional SEZ shocks that happened from 2001 to 2006, later years beyond our sample period. From this expanded set, we draw counterfactual SEZ shocks in our simulations. We elaborate on the simulation algorithm in Appendix A.[9] Therefore, identification of the FD model relies on the assumption that out of all SEZs established between 1996 and 2006, whether they are established before or after 2000 is as good as random. Moreover, conditional on the expected pull factor, the composite Bartik-like pull factor based on realized SEZs between 1996 and 2000 is not correlated with the remaining unobserved residual.

[9]In Figure C.1, we depict the actual values of $\Delta(\text{Pull Factor})c$ in panel (a) and the expected $\Delta(\text{Pull Factor})c$ values calculated from simulations in panel (b) for each county using a map.

5 Gender Imbalance in Migration

We begin by analyzing the impact of SEZ shocks on the size of the migrant population and whether the extent of migration varied between young males and young females. We utilize county-level data from 1995 to 2000 to estimate a staggered DiD model. Table 4 reports the average treatment effect on the treated (hereinafter ATT) for all cohorts across all years. The dependent variable in each column is in its logarithmic-like inverse hyperbolic sine transformation. In column (1), we find that the establishment of an SEZ has a positive and statistically significant impact on the influx of the entire migrant population size at the county level. In columns (2) and (3), we separately estimate the effects for young female and young male migrants, respectively. The estimated coefficient is 0.4673 and statistically significant at the 1% level for young female migrants and is 0.3829 and statistically significant at the 1% level for young male migrants. Thus, young females appear to exhibit a stronger response to the SEZ shock than their male counterparts. In Table D.1 in the Appendix, we also present the estimated effects for each year during our sample period, across all cohorts. Although the coefficients do not exhibit statistically significant differences, we observe that, except for the year 1996, the estimates for young female migrants consistently surpass those for young male migrants. In 1996, the estimate for young male migrants is slightly larger, albeit only by 0.0069.

In Figure 6, we present the dynamic ATTs. Specifically, we follow the specifications reported in columns (2) and (3) in Table 4 to estimate the ATTs for each year relative to the first year of treatment, across all cohorts. Panel (a) displays the results for young female migrants, and panel (b) displays the results for young male migrants. While a few coefficients show a slight positive trend a few periods before the treatment, almost all coefficients during the pre-treatment period are statistically insignificant in both panels. The estimated coefficients increase sharply with the establishment of an SEZ, and the post-treatment coefficients appear to be larger for young female migrants than for young male migrants.

To gain deeper insights using the individual-level data, we further examine the FD model estimates. In Table 5, we present the baseline estimates from the FD model using the sample of all individuals who were between the ages of 16 and 25 at any point between 1996 and 2000. Hereinafter, we refer to this sample as the baseline sample. Column (1) shows that when $\Delta(\text{Pull Factor})_c$ becomes larger, indicating a heightened urbanization-induced pull in other areas, young individuals are more likely to migrate to other counties. Moreover, such a tendency is more pronounced among young females. The statistical significance of the interaction term, denoted as $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$, at the 1% level indicates a substantial gender difference in response to the pull factor. We find that females exhibit a significantly stronger inclination to migrate in response to a heightened pull factor compared to males. Moreover, this additional effect for females capturing the gender difference is consistently statistically significant at the 1% level across all specifications in Table 5.

In column (2), we address endogeneity concerns by including the expected terms obtained via simulations, following the method of Borusyak and Hull (2023). Even after addressing such endogeneity concerns, we find that the main effects are still consistent and statistically significant, although the magnitude difference has slightly diminished. Specifically, a one standard deviation increase in the pull factor leads to a $(0.0022 \times 1.17 \times 100 \approx)$ 0.26 percentage point increase in the propensity for young males to emigrate from county c. The effect is $(0.0017 \times 1.17 \times 100 \approx)$ 0.20 percentage points larger for young females.

In columns (3) and (4), we extend our analysis to include a control variable, $\Delta(\text{SEZ})_c$, which captures the internal SEZ shock that serves as a staying force. We find that when the county of residence establishes an SEZ, this reduces the likelihood of migrating out. Even with this additional control variable, we document robust evidence that young females are more responsive to $\Delta(\text{Pull Factor})_c$ to emigrate out of their original place of residence than young males: The gender difference in the effects is still 0.20 percentage points in column (4). In Table D.2 in the Appendix, we replicate our analysis, narrowing down our sample to encompass only small-sized counties that had a population of less than 700,000 in Chinese Population Census 1982. We find that the main estimates' magnitude and their statistical significance remain consistent and largely similar.

6 Potential Mechanisms

In this section, we investigate potential mechanisms that drive the observed gender imbalance among young migrants. We begin with a theoretical framework that articulates rural individuals' marital and nonmarital incentives to migrate.

6.1 A Theoretic Framework

Our theoretical framework is as follows. A person derives utility from two components: a nonmarital payoff and a marital payoff. The nonmarital payoff encompasses moving costs and benefits from nonmarital aspects such as the labor market, educational opportunities, and amenities. While the nonmarital payoff may vary by gender due to differences in preferences and opportunities, our model simplifies the nonmarital payoff as exogenous and mainly serves to highlight how gender disparities in marital payoffs may differ due to equilibrium effects. The marital payoff is determined in a general equilibrium marriage market, with gender differences arising endogenously (Chiappori et al. 2009). Intuitively, when there are more desirable men in the cities, the urban marriage market becomes more attractive for women in rural areas. Then women have higher incentives to migrate to the cities to capture the extra marital benefits.

Each person is endowed with one of two types, high and low, which may encapsulate a bundle of attributes such as education level, physical attractiveness, fecundity, and noncognitive skills (Dupuy and Galichon 2014). Denote by h and ℓ the types in the rural marriage market and by H and L the types in the urban marriage market. Each rural person is endowed with a heterogeneous nonmarital net gain of moving to the city, such as gains from educational opportunities, amenities, and/or labor market and costs of moving. Note that the net gain may be negative. Distributions of nonmarital gains can also vary by gender and type. Suppose there are initially equal masses of men and women in both rural and urban areas. The total surplus $s_{\theta_m \theta_w}$ is determined by husband's and wife's types θ_m and θ_w, but the division of this surplus between the couple is determined in equilibrium. Assume strict surplus supermodularity to guarantee positive assortative matching: $s_{HH} + s_{LL} > s_{HL} + s_{LH}$ and $s_{hh} + s_{\ell\ell} > s_{h\ell} + s_{\ell h}$. In the Chinese context, this is consistent with hypergamy and patrilocality documented in Section 2.3.

The marital component of a person's payoff is endogenously determined in a matching market à la Becker (1973). A *stable outcome* of the marriage market (G_m, G_w), where G_m and G_w denote the mass of males and females, respectively. The stable outcome consists of *stable matching* and *stable marriage payoffs*. Stable matching G satisfies *feasibility*: $\sum_{\theta_w \in \Theta_w} G_{\theta_m \theta_w} \leq G_{m\theta_m}$ for any $\theta_m \in \Theta_m$ and $\sum_{\theta_m \in \Theta_m} G_{\theta_m \theta_w} \leq G_{w\theta_w}$ for any $\theta_w \in \Theta_w$. Stable marriage payoffs u_θ and v_θ, where $\theta \in \Theta = \Theta_m = \Theta_w = \{H, L, h, \ell\}$, satisfy (i) *individual rationality*: $u_{\theta_m} \geq 0$ for any $\theta_m \in \Theta_m$ and $v_{\theta_w} \geq 0$ for any $\theta_w \in \Theta_w$ (every person receives at least as much as they would have if they had remained single); (ii) *pairwise efficiency*: $u_{\theta_m} + v_{\theta_w} = s_{\theta_m \theta_w}$ (every married couple divides the entire marriage surplus); and (iii) *Pareto efficiency*: $u_{\theta_m} + v_{\theta_w} \geq s_{\theta_m \theta_w}$ for all $\theta_m \in \Theta_m$ and $\theta_w \in \Theta_w$. That is, no man-woman pair not married to each other can simultaneously improve their marriage payoffs by marrying each other.

Let $\phi_{g\theta}$, where $g \in \{m, w\}$ and $\theta \in \{H, L, h, \ell\}$, denote the original population mass in the urban and rural areas. Assume balanced gender ratios to start with. In addition, although we could allow any migration patterns, we focus on analyzing the cases that are the closest to the observed pattern in China. We explicate the exact assumptions in Appendix E so that in equilibrium on net more women end up in urban areas and more men end up in rural areas, and that there are more high-type men than high-type women in rural areas. Note that there is indeterminacy and flexibility regarding whether more high-type men or more high-type women are in cities.

All rural individuals decide whether to migrate to cities, assuming without loss of generality that urban individuals do not consider moving to rural areas. Urban and rural marriage markets clear based on the migration decisions. An *equilibrium* is one in which (i) each individual maximizes their payoff based on rationally expected marital payoffs and (ii) marital payoffs are stable with respect to the marriage markets formed after migration. More specifically, the equilibrium is characterized by marriage payoffs u^* and v^* and nonmarital net gain cutoffs

y^* such that (i) $y^*_{mh} + u^*_H = u^*_h$, $y^*_{m\ell} + u^*_L = u^*_\ell$, $y^*_{wh} + v^*_H = v^*_h$, and $y^*_{w\ell} + v^*_L = v^*_\ell$, and (ii) $(u^*_H, u^*_L, v^*_H, v^*_L)$ and $(u^*_h, u^*_\ell, v^*_h, v^*_\ell)$ are stable in the urban and rural marriage markets, respectively, after rural men and women with sufficiently high nonmarital gains move to cities.

Both marital and nonmarital factors may incentivize more rural women migrating to the cities. When there are more men in the cities who become more competent and wealthier, the marriage market mechanism would be in effect and attract more rural women to migrate to cities. On the nonmarital aspect, increasing economic and educational opportunities and increasing incomes, especially for women, such as due to more service or more teaching jobs, would encourage more women to move to cities. In addition, increased benefits from amenities and other non-labor-market gains may also attract women.

Subsequently, we separately examine plausible channels: (i) increasing gains for women in the marriage-market, partly resulting from the growing economic outcomes of men in the cities; and (ii) increasing labor and educational opportunities and amenities for women.

6.2 Evidence on Marital Incentives

To investigate empirical evidence supporting the marital incentives, we initiate by re-estimating our FD model using various subsamples. If individuals are motivated by marital incentives to migrate to more urban areas, the gender imbalance in the migration tendency, as documented in our baseline findings, should not be applicable to individuals likely to be already married. Based on our calculations using Chinese Population Census 2000, the average age for females who got married in the 1990s was 22.70, while for males, it was 24.47. Therefore, we re-estimate our FD model using individuals who were aged 26-35 in 1996, the majority of whom are already likely to be married. The estimates in panel (a) of Table 6 reveal that for individuals within this age group, females exhibit a weaker migration tendency compared to their male counterparts. This result starkly contrasts with our baseline findings for young individuals.

In panels (b) and (c) of Table 6, we conduct further examinations by focusing on different subgroups conditional on their marital status. In panel (b), we analyze a subset of individuals in the baseline sample who remained single throughout our sample period from 1996 to 2000.[10] In panel (c), we shift our attention to the sample of older individuals who were aged 26-35 in 1996 and were already married by that year. While we acknowledge that the samples used in panels (b) and (c) are subject to selection bias, we analyze whether there are any meaningful gender-differential migration patterns within the respective unmarried and married groups. Panel (b) shows that the young singles' response to the pull factor is indeed stronger among the singles cohort, compared to the baseline estimated coefficients for $\Delta(\text{Pull Factor})_c$ in Table 5. Moreover, single females show a stronger migration tendency compared to their single male counterparts. However, panel (c) shows contrasting results. Married females display a significantly lower likelihood of migrating in response to the pull factor when compared to married males, and such effect is statistically significant at 1% across all specifications.

In panel (d) of Table 6, we use the subset of individuals in the baseline sample who migrated between 1996 and 2000. The dependent variable, denoted as $\Delta(\text{Migrate for marital reasons})_{i,c}$, equals 1 for those who migrated out of county c and explicitly stated in the census that they migrated for marital reasons, while it equals 0 for all other migrants. In response to the pull factor, we find that females are much more likely to state that they migrated for marital reasons compared to their male counterparts.

In Table 7, we examine outcomes related to marriage. In panel (a), we use the baseline sample to estimate the likelihood of getting married during our sample period. The dependent variable, denoted as $\Delta(\text{Marry})_{i,c}$, equals 1 for all individuals who resided in county c in 1996 and married between 1996 and 2000, and 0 otherwise. We observe a decrease in the probability of marriage for men but an increase for women in response to the pull factor. In panel (b), we use the baseline sample and the dependent variable is set to 1 for individuals who migrated out of their origin county c and got married between 1996 and 2000, and 0 otherwise. The probability of both migrating and getting married in response to the pull factor is higher for women. Lastly in panel (c), we present the estimated impact of the pull factor on the probability

[10] Among the individuals in the baseline sample, 50.64% of young females and 64.67% of young males remained single during our sample period.

of marrying up. We utilize a subset of individuals from the baseline sample who fall into one of two categories: (i) those who remained continuously single between 1996 and 2000, or (ii) those who married during this period, with their spouse's education level identified in the census data. To investigate hypergamy tendencies, we compare the education levels of couples and define $\Delta(\text{Marry up})_{i,c}$ as one for a person residing in county c in 1996 if they married someone with a higher education level than their own. We observe that the pull factor decreases the probability of males engaging in hypergamous marriages but increases the likelihood for females. Considering that the age group used in the baseline sample may not encompass individuals with higher education degrees, we broaden our sample to include individuals aged between 20-35 at any point from 1996 to 2000. The results in Table D.3 in the Appendix confirm our consistent findings.

In summary, our empirical findings support the notion that young females are significantly influenced by marital incentives to migrate to large urban areas in pursuit of improved marriage prospects. When urbanization brought about by SEZs results in skill-biased technological progress and a male-dominant skill pool in urban areas, it seems to create a ripple effect on young females' migration choices. Essentially, the prospect of marrying and marrying up increases in large urban areas. This interplay enhances the expected returns from living in urban areas and strengthens the incentives for young females to migrate to large urban areas. While marital incentives may also exist for young males, hypergamy and patrilocal patterns in China's context make young females' marital incentives particularly more pertinent.

6.3 Evidence on Nonmarital Incentives

6.3.1 Labor Market

SEZs often involve the implementation of policies and incentives that are designed to attract investment and promote economic development in specific industries. Therefore, SEZ establishment can potentially change the industrial composition of the affected area and result in non-gender-neutral growth in labor demand. That is, gender imbalance among migrants may arise due to increased demand for young female workers, given the SEZ-induced industrial composition changes.[11]

We address possible changes in gender-specific labor market conditions driven by SEZ establishments in our FD model by constructing two measures, $\Delta(\text{Labor Mkt})^F$ and $\Delta(\text{Labor Mkt})^M$. In essence, these two measures capture the changes in labor market conditions that attract female and male workers to migrate, respectively. We detail the variable construction in Appendix B. In panel (a) of Table 8, we present the FD model estimates that further include $\Delta(\text{Labor Mkt})^F$ and $\Delta(\text{Labor Mkt})^M$. Compared to the baseline estimates, we observe a slight reduction in the estimated impact of the pull factor on emigration incentives for females across all specifications. This suggests that the increased migration of women to urban areas can be partially attributed to shifts in labor market conditions influenced by evolving industrial composition in urban areas brought by SEZs, making these areas more attractive to female workers. At the same time, the evidence suggests that even when labor market conditions are considered, young females are still more responsive to the pull factor than their male counterparts.

Table 9 reports the distribution of targeted industries within the SEZs during our sample period.[12] For each industry, we calculate proportion of female workers using data from Chinese Population Census 1990. The industries heavily targeted do not appear to be those that were more reliant on female labor in particular. Moreover, the weighted average of female labor share across the targeted industries equals to 42.33%, suggesting that the targeted industries were not more reliant on female workers.

6.3.2 Educational Opportunities

One potential explanation is that females may place a higher value on educational opportunities, and education quality may be better in larger urban areas. Another possibility is that the gender

[11] For instance, "the feminization of SEZ production is attributed to three broad factors in the literature: women's relative "cheapness" owning to the gender wage gap, rising international competition, and gendered norms and stereotypes that segment work by sex and assign women to low-skill and low-paying work" (Farole and Akinci 2011, p.251).

[12] The National Development and Reform Commission of China specifies targeted industries for each SEZ.

gap in returns to education is greater in larger urban areas, with women experiencing a higher return than men. To investigate this possibility, we re-estimate our FD model using the baseline sample that excludes individuals who responded that they migrated for education-related purposes. If the gender disparity in migration is primarily driven by educational opportunities, it is likely that our main findings would no longer hold or would be substantially weakened if we exclude respondents who migrated for educational purposes. In our sample of young migrants, 7.29% responded that they migrated for educational reasons.[13] Moreover, 6.59% of young female migrants and 8.07% of young male migrants responded that they migrated for educational reasons. While there is not a substantial difference between males and females, males rather demonstrated a slightly higher tendency to migrate for educational purposes. Panel (b) of Table 8 shows that the estimates based on the sample that excludes individuals who reported migrating for education-related purposes. We find that the estimates remain mostly unchanged compared with the baseline estimates in Table 5. Thus, while educational opportunities may have driven migration for some individuals, it does not appear to be the primary cause of gender imbalance in migration in our context. This is not surprising, given that around 95% of our rural sample have an education level of middle school or below.

6.3.3 Amenities

Another possible explanation is that larger urban areas may offer amenities that are more attractive to young females than to young males. For example, certain types of amenities that support family life, such as access to vibrant community centers or green parks, may be valued more by young females and larger urban areas may offer more of such amenities.[14] To analyze

[13]In Chinese Population Census 2000, reasons for relocation were reported at the individual level, so responses could vary within the same household. The primary reason for moving among young migrants was for work, accounting for 74.74%. The second most common reason was for study. Other responses, in descending order of frequency, included relocating for a joint marriage location, moving in with family members, and various other reasons.

[14]Based on U.S. data, Reynolds and Weinstein (2021) find that males and females largely share preferences for natural amenities (e.g. coastal location, climate), but there are important gender differences among young singles regarding their importance relative to nonnatural amenities (e.g. public transportation, safety, progressive gender-role attitudes).

the role of 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. This is because high-skilled individuals are known to value amenities more than low-skilled individuals (Diamond 2016). We use education level to proxy for skills. Individuals with an education level of middle school or below are classified as having low skills, and those with an education level above middle school are classified as having high skills. In panels (c) and (d) of Table 8, we re-estimate the values for the sample of 1,845,444 less educated individuals and the sample of 717,391 educated individuals, respectively.[15] We find that less-educated females exhibit a much stronger tendency than their male counterparts to emigrate in response to the pull factor. However, in the highly-educated individuals sample, females demonstrate less responsiveness to the pull factor compared to their male counterparts and such difference is statistically significant at 1% across all specifications. Therefore, our evidence indicate that gender imbalance in migration is not more pronounced among the high-skilled group, a demographic known to place greater value on amenities (Diamond 2016).[16]

7 Conclusion

In this paper, we analyze the extent of gender imbalance in migration among young individuals and examine explanations that can rationalize this phenomenon. Using the gradual rollout of SEZs across China between 1996 and 2000, we find that more young women than young men migrated from rural areas to urban areas. Such gender-differential migration patterns are shown

[15] The education level is determined by the highest level of educational attainment as of the year 2000. Among individuals with a lower level of education, 50.33% were females. For those with a higher level of education, 46.98% were females.

[16] Note that education level is part of the "type" defined in the theoretical framework, which is a combination of various attributes valued in the marriage market, such as education level, physical attractiveness, fecundity, and noncognitive skills (Dupuy and Galichon 2014). This distinction is particularly pertinent in the context of China during our sample period, in which a vast majority of women were under-educated, with only 6.79% of young females during this period having an education level of junior college or higher. The results in panels (c) and (d) of Table 8 further suggest that education level does not appear to be the primary distinguishing factor of migration incentives. Our theory suggests that more high-type women tend to move to urban areas. The theoretical and empirical results together suggest that factors besides education are main components of marriage type and play significant roles in migration incentives.

to be largely driven by marital incentives. Our study suggests that such a gender imbalance in migration can further trigger growing disparities in the marriage market between urban and rural areas and between males and females. Moreover, widening inequality and gender divide in the marriage market may have far-reaching implications for social and family stability.

References

Aaronson, D., Mazumder, B., Sanders, S. G., and Taylor, E. J. (2021). Estimating the Effect of School Quality on Mortality in the Presence of Migration: Evidence from the Jim Crow South. *Journal of Labor Economics*, 39(2):527–558.

Adao, R., Kolesár, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Ahn, S. Y. et al. (2023). Matching Across Markets: An Economic Analysis of Cross-Border Marriage. *Journal of Labor Economics*, forthcoming.

Alder, S., Shao, L., and Zilibotti, F. (2016). Economic Reforms and Industrial Policy in a Panel of Chinese Cities. *Journal of Economic Growth*, 21(4):305–349.

Bacolod, M. (2017). Skills, the Gender Wage Gap, and Cities. *Journal of Regional Science*, 57(2):290–318.

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, 81(4):813–846.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61.

Bhaskar, V., Li, W., and Yi, J. (2023). Multidimensional Pre-Marital Investments with Imperfect Commitment. *Journal of Political Economy*, 131(10):2893–2919.

Borusyak, K. and Hull, P. (2023). Nonrandom Exposure to Exogenous Shocks. *Econometrica*, 91(6):2155–2185.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies*, 89(1):181–213.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2):200–230.

Chiappori, P.-A., Iyigun, M., and Weiss, Y. (2009). Investment in Schooling and the Marriage Market. *American Economic Review*, 99(5):1689–1713.

Choo, E. (2015). Dynamic Marriage Matching: An Empirical Framework. *Econometrica*, 83(4):1373–1423.

Choo, E. and Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, 114(1):175–201.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. *Handbook of Regional and Urban Economics*, pages 247–348.

Couture, V. and Handbury, J. (2017). Urban Revival in America, 2000 to 2010. Technical report, National Bureau of Economic Research.

Crane, B., Albrecht, C., Duffin, K. M., and Albrecht, C. (2018). China's Special Economic Zones: An Analysis of Policy to Reduce Regional Disparities. *Regional Studies, Regional Science*, 5(1):98–107.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–96.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*, 106(3):479–524.

Dupuy, A. (2021). Migration in China: To Work or To Wed? *Journal of Applied Econometrics*, 36(4):393–415.

Dupuy, A. and Galichon, A. (2014). Personality Traits and the Marriage Market. *Journal of Political Economy*, 122(6):1271–1319.

Fan, J. and Zou, B. (2021). The Dual Local Markets: Family, Jobs, and the Spatial Distribution of Skills. Mimeo.

Farole, T. and Akinci, G. (2011). *Special Economic Zones: Progress, Emerging Challenges, and Future Directions*. World Bank Group, Washington DC.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2):254–277.

Gruijters, R. J. and Ermisch, J. (2019). Patrilocal, Matrilocal, or Neolocal? Intergenerational Proximity of Married Couples in China. *Journal of Marriage and Family*, 81(3):549–566.

Marshall, A. (1890). *Principles of Economics*. Macmillan, London.

Mu, Z. and Xie, Y. (2014). Marital Age Homogamy in China: A Reversal of Trend in the Reform Era? *Social Science Research*, 44:141–157.

Mumford, L. (1961). *The City in History: Its Origins, Its Transformations, and Its Prospects*, volume 67. Houghton Mifflin Harcourt.

National Bureau of Statistics of China (2021). *China Statistical Yearbook 2021*. China Statistics Press.

NDRC (2006). Catalog of Announcement Reviews for Chinese Development Zones (2006).

NDRC (2018). Catalog of Announcement Reviews for Chinese Development Zones (2018).

Nguyen, M.-N. (2022). Sex Ratio Among Urban and Rural Population in Vietnam 2010-2021. Accessed: 2023-05-23.

Peters, M. and Siow, A. (2002). Competing Premarital Investments. *Journal of Political Economy*, 110(3):592–608.

Pettay, J. E., Lummaa, V., Lynch, R., and Loehr, J. (2021). Female-Biased Sex Ratios in Urban Centers Create a "Fertility Trap" in Post-War Finland. *Behavioral Ecology*, 32(4):590–598.

Phimister, E. (2005). Urban Effects on Participation and Wages: Are There Gender Differences? *Journal of Urban Economics*, 58(3):513–536.

PTI (2017). Sex-Ratio in Urban Areas Worse than Rural Areas: Survey. Accessed: 2023-05-23.

Reynolds, C. L. and Weinstein, A. L. (2021). Gender Differences in Quality of Life and Preferences for Location-Specific Amenities Across Cities. *Journal of Regional Science*, 61(5):916–943.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, pages 2119–2177. Elsevier.

Rosenthal, S. S. and Strange, W. C. (2012). Female Entrepreneurship, Agglomeration, and A New Spatial Mismatch. *Review of Economics and Statistics*, 94(3):764–788.

Sant'Anna, P. H. and Zhao, J. (2020). Doubly Robust Difference-in-Differences Estimators. *Journal of Econometrics*, 219(1):101–122.

Shirk, S. L. et al. (1993). *The Political Logic of Economic Reform in China*, volume 24. Univ. of California Press.

Swerts, E., Pumain, D., and Xie, L. (2021). 经济开发区与中国城市增长. *Cybergeo: European Journal of Geography*.

Tacoli, C. (2012). *Urbanization, Gender and Urban Poverty: Paid Work and Unpaid Carework in the City*. Human Settlements Group, International Institute for Environment and Development.

Wiest, K., Leibert, T., Johansson, M., Rauhut, D., Ponnikas, J., Timar, J., Velkey, G., and Győrffy, I. (2013). Selective Migration and Unbalanced Sex Ratio in Rural Regions. *SEMIGRA. ESPON & Leibniz Institute for Regional Geography, Leibniz.*

World Bank (2023). World Bank Gender Statistics DataBank. Online. Accessed on: May 23, 2023.

Xu, C. (2011). The Fundamental Institutions of China's Reforms and Development. *Journal of Economic Literature*, 49(4):1076–1151.

Young, J. (2013). *China's Hukou System: Markets, Migrants and Institutional Change*. Palgrave Macmillan London.

Zhang, H. (2021). An Investment-and-Marriage Model with Differential Fecundity: On the College Gender Gap. *Journal of Political Economy*, 129(5):1464–1486.

Zhang, H. and Zou, B. (2023). A Marriage-Market Perspective on Risk-Taking and Career Choices. *European Economic Review*, 152:104379.

Zhang, J. (1994). Bequest as a Public Good within Marriage: A Note. *Journal of Political Economy*, 102(1):187–193.

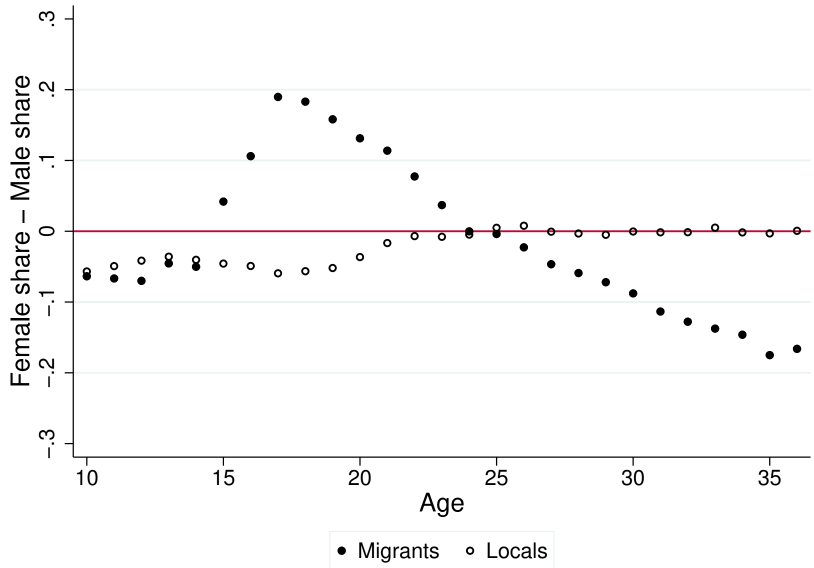
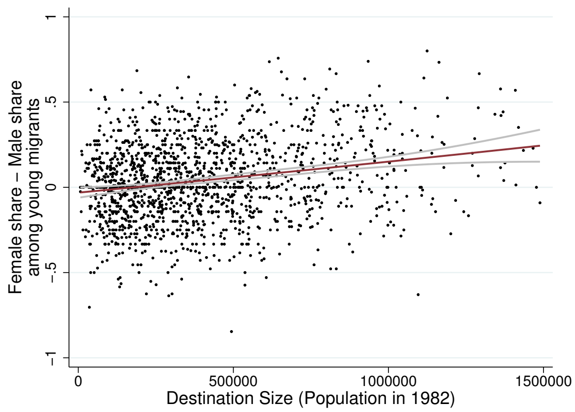
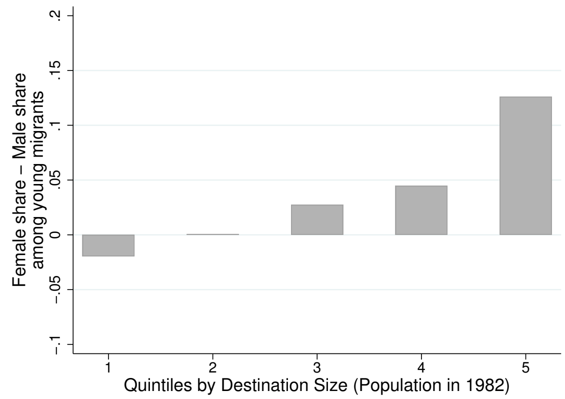


Figure 1: Gender Disparity by Age for Migrants and Locals in China in 2000

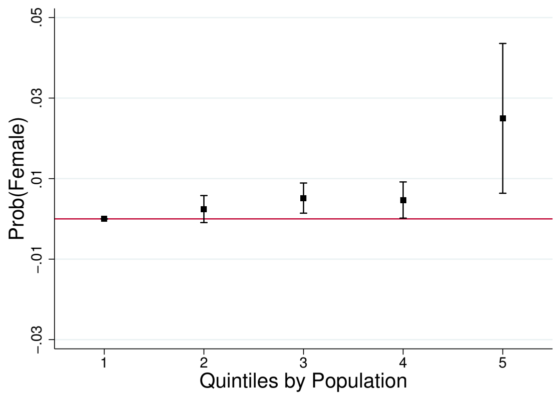
Data source: Chinese Population Census 2000.
Notes: A person is defined as a migrant if they moved across counties and is defined as a local otherwise. We calculate the difference between the female share and the male share for a given age. This value is zero when the gender distribution is perfectly balanced, positive when there is an excess of females, and negative when there is an excess of males.



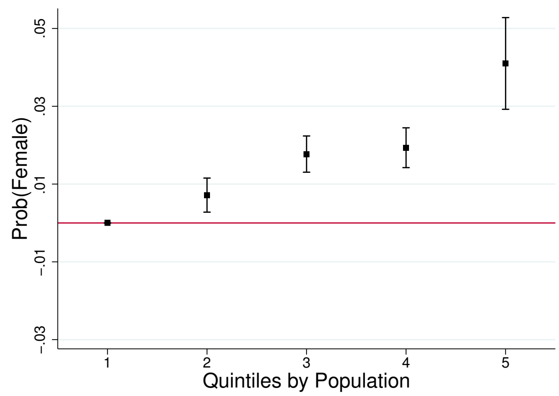
(a) By Destination Size



(b) By Quintile



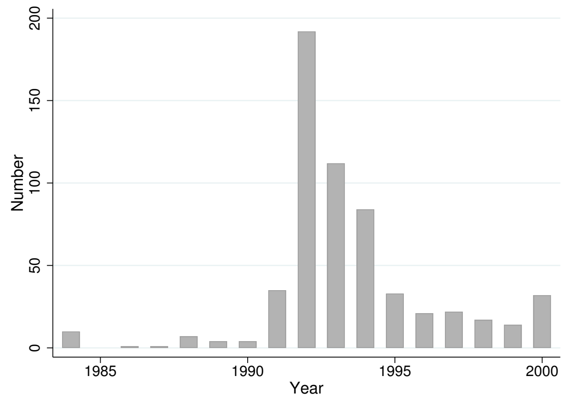
(c) Coefficient Plot



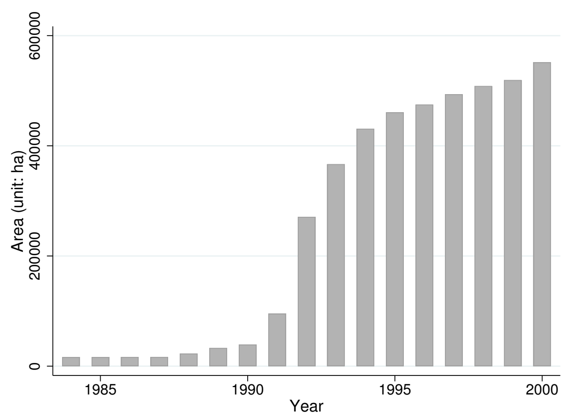
(d) Coefficient Plot, with Industry Fixed Effects

Figure 2: Gender Imbalance among Young Individuals in Larger Urban Areas

Data source: Chinese Population Census 2000.
Notes: In panels (a) and (b), we use a sample of young migrants aged 16 to 25 in the year 2000. In panel (a), the scatter plot displays counties represented by markers, along with a quadratic fitted line and corresponding confidence intervals. In panel (b), each bar represents the average net female share among young migrants across counties within a specific quintile. Quintile 1 (quintile 5) consists of the smallest (largest) counties in terms of population in 1982. In panel (c), we use a sample of all individuals aged 16 to 25 in the year 2000 and estimate the following: $\mathbb{1}(\text{Female})_i = \alpha_1 + \alpha_2 \mathbb{1}(Q2) + \alpha_3 \mathbb{1}(Q3) + \alpha_4 \mathbb{1}(Q4) + \alpha_5 \mathbb{1}(Q5) + \epsilon_i$. Quintiles are created using the county population in 2000, where quintile 1 (quintile 5) consists of the smallest (largest) counties. Standard errors are clustered at county level. Coefficients and vertical confidence intervals are displayed for each quintile, with quintile 1 serving as the reference category. In panel (d) we conduct a similar analysis, but further include fixed effects for individuals' labor-market industry.



(a) Number of Counties with SEZs



(b) Total Area of SEZs

Figure 3: Establishment of SEZs Across Years

Data source: The National Development and Reform Commission of China (NDRC 2006).
Notes: Panel (a) displays the total number of SEZs that were newly established for the first time in a county during the specified year. Panel (b) shows the cumulative area of SEZs in a given year (unit: ha).

(a) Year 1990

(b) Year 1995

(c) Year 2000

(d) Year 2005

Figure 4: Geographic Spread of SEZs over Time in China

Data source: The National Development and Reform Commission of China.
Notes: The map displays China and its county borders. For each specified year (1990, 1995, 2000, and 2005), we denote SEZ-designated counties during that year using a darker shade of gray.

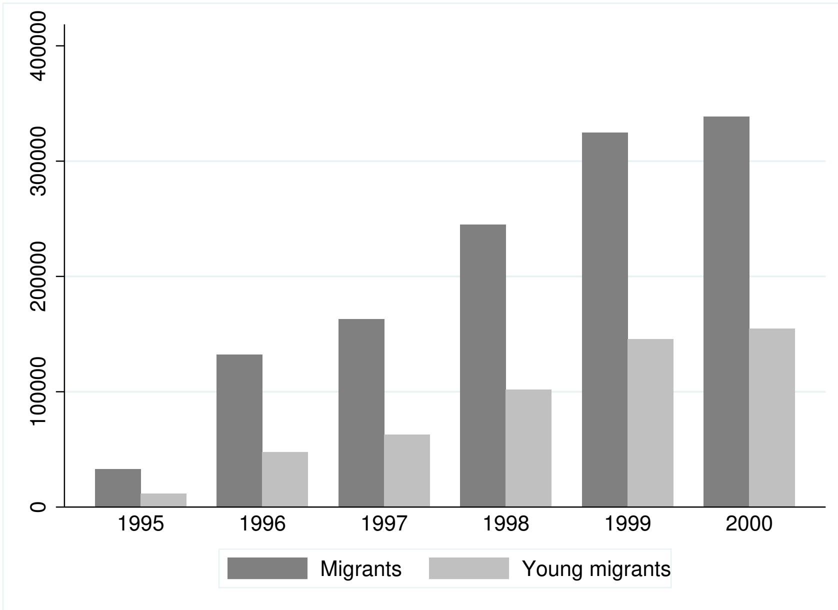
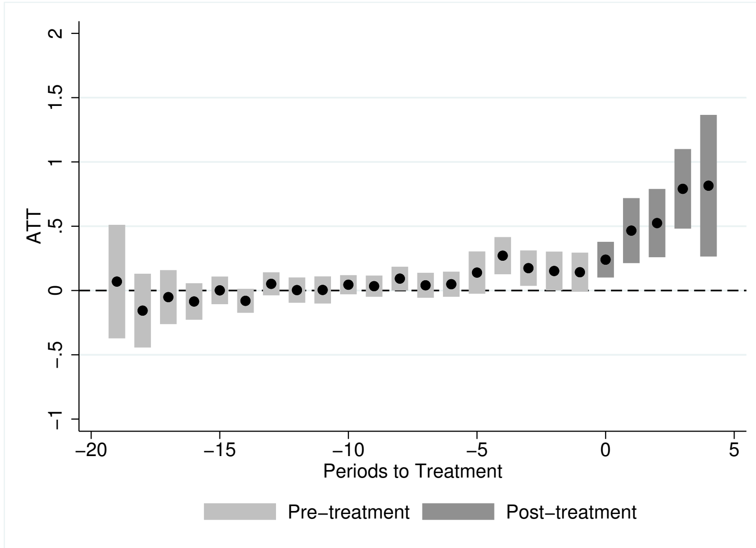
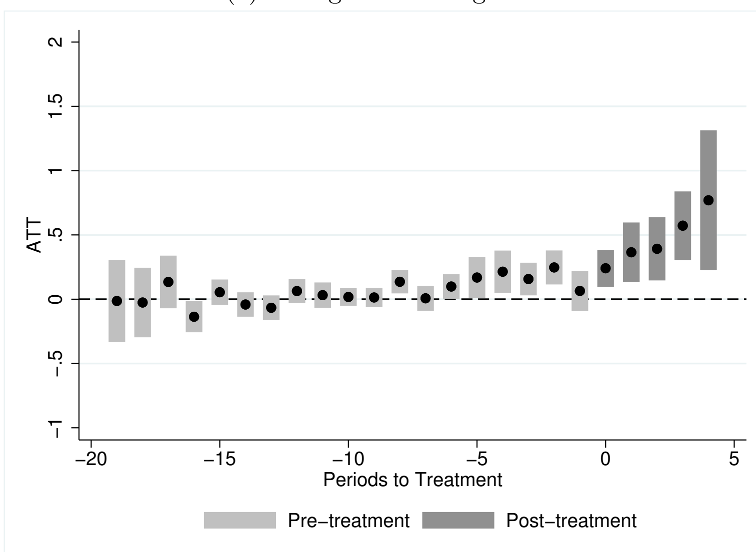


Figure 5: Total Number of Migrants

Data source: Chinese Population Census 2000.
Notes: Darker gray bars indicate the count of all migrants who moved across counties in the specified year. The lighter gray bars indicate the count of young migrants aged between 16 and 25 for the specified year.



(a) Young Female Migrants



(b) Young Male Migrants

Figure 6: Time-varying Effects of SEZ Establishment on Young Migrants

Data source: Chinese Population Census 2000.
Notes: In panel (a), we present dynamic ATTs (i.e., average treatment on the treated) obtained from estimating the specification in column (2) in Table 4. The dependent variable is young female migrants in its inverse hyperbolic sine transformation and we follow Callaway and Sant'Anna (2021) in our estimation, using the not-yet-treated as the control group. In panel (b), we similarly present dynamic ATTs when the dependent variable is young male migrants in its inverse hyperbolic sine transformation. The specification corresponds to column (3) in Table 4.

Table 1: Spousal Age Difference

	Husband's age - Wife's age			
	-1 or below	0 to 1 years	2 to 3 years	4+ years
Migrant wife	15.00%	31.93%	26.35%	26.72%
Non-migrant wife	17.70%	34.93%	26.12%	21.25%

Notes: We focus on 2,076,139 married couples in which both spouses are identified and are aged under 55 years. In each row, the percentage shares collectively add up to 100%. A wife is classified as a migrant if she ever moved across counties, and a non-migrant otherwise.

Table 2: Marriage Matching Patterns by Education Level (Unit: %)

Wife \ Husband	1	2	3	4	5
1	25.44	19.07	2.96	0.67	0.03
2	4.15	25.11	5.42	1.88	0.16
3	0.46	2.79	3.49	1.63	0.27
4	0.06	0.69	0.94	3.04	0.87
5	0.00	0.02	0.04	0.25	0.57

Notes: We focus on 2,076,139 married couples in which both spouses are identified and are aged under 55 years. In the table, each cell represents the percentage of couples falling into a specific category, which is determined by the educational levels of the wife and husband. 1 stands for primary school and below. 2 stands for junior high school. 3 stands for senior high school. 4 stands for technical secondary school and junior college. 5 stands for bachelor's degree and above. Cells where the husband has a lower education level are denoted in white, while those where the husband has a higher education level than the wife are highlighted in dark grey. The diagonal cells represent instances where the wife and husband share the same education level and are presented in light grey.

Table 3: Summary Statistics

Panel (a) County-level	Mean	SD	Obs.
Population	8.70	0.89	2,870
All migrants	4.95	1.48	2,870
Female migrants	4.18	1.49	2,870
Male migrants	4.30	1.46	2,870
Female migrants aged 16-25	2.81	1.56	2,870
Male migrants aged 16-25	2.65	1.61	2,870
Female share	0.49	0.01	2870
Female share among migrants	0.47	0.11	2,846
Female share among migrants aged 16-25	0.54	0.20	2,765
$\mathbb{1}$(Received SEZ shock by 2000)	0.16	0.37	2,870

Panel (b) Individual-level	Mean	SD	Obs.
$\mathbb{1}$(Female)	0.49	0.50	2,562,835
Age	22.73	4.18	2,562,835
$\mathbb{1}$(Middle school or below)	0.72	0.45	2,562,835
Δ(Migrate)	0.10	0.30	2,562,835
Δ(Marry)	0.25	0.44	2,562,835
Δ(Marry up)	0.04	0.20	2,141,339

Data source: Chinese Population Census 2000.
Notes: In panel (a), we report the county-level values in year 2000. All population and migrant count-related variables are reported after undergoing an inverse hyperbolic sine transformation. In panel (b), we use the same sample as that for the regression analysis in Table 3: observations of young individuals who were between the ages of 16 and 25 at any point during the period 1996 to 2000.

Table 4: Effects of SEZ on Migrants – A County-level Analysis

	All Migrants	Young Migrants (Aged 16-25)	
	(1)	(2)	(3)
Dep. variable	Total	Females	Males
$\mathbb{1}(\text{SEZ})_{c,t}$	0.4153***	0.4673***	0.3829***
	(0.1184)	(0.0868)	(0.0801)
Observations	6,975	6,975	6,975
Control Group	Not-yet	Not-yet	Not-yet

Notes: We utilize the yearly sample of counties in China from 1996 to 2000. Across all specifications, each dependent variable is subjected to its inverse hyperbolic sine transformation. We estimate a staggered DiD model following Callaway and Sant'Anna (2021) and report the estimates of the average treatment effect on the treated for all groups across all periods. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote $p<0.01$, $p<0.05$, $p<0.1$, respectively.

Table 5: Migration Outcomes – An Individual-level Analysis

$\Delta(\text{Migrate})_{i,c}$	(1) Unadjusted OLS	(2) Controlled OLS	(3) Unadjusted OLS	(4) Controlled OLS
$\Delta(\text{Pull Factor})_c$	0.0020*	0.0022**	0.0022**	0.0024**
	(0.0011)	(0.0011)	(0.0011)	(0.0011)
$\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$	0.0034***	0.0017***	0.0035***	0.0017***
	(0.0002)	(0.0004)	(0.0002)	(0.0004)
Expected $\Delta(\text{Pull Factor})_c$		0.0037***		0.0038***
		(0.0013)		(0.0013)
Expected $\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$		0.0021***		0.0021***
		(0.0004)		(0.0004)
$\Delta(\text{SEZ})_c$			-0.0239***	-0.0245***
			(0.0081)	(0.0080)
$\Delta(\text{SEZ})_c \times \mathbb{1}(\text{Female})_i$			-0.0037	-0.0038
			(0.0047)	(0.0046)
Observations	2,562,835	2,562,835	2,562,835	2,562,835
Root MSE	0.299	0.299	0.299	0.299
Mean(y)	0.100	0.100	0.100	0.100

Notes: The dependent variable $\Delta(\text{Migrate})_{i,c}$ equals one for individual i if they emigrated from county c to another county between 1996 and 2000. The sample consists of all individuals who were between the ages of 16 and 25 at any point during the period 1996 to 2000. Across all specifications, we include fixed effects for age in year 2000. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

Table 6: Evidence on Marital Incentives: Migration Outcomes

	(1) Unadjusted OLS	(2) Controlled OLS	(3) Unadjusted OLS	(4) Controlled OLS
Panel (a) Dep. var: $\Delta(\text{Migrate})_{i,c}$ Sample of individuals aged 26-35 in 1996.				
$\Delta(\text{Pull Factor})_c$	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)
Panel (b) Dep. var: $\Delta(\text{Migrate})_{i,c}$ Subset of individuals within the baseline sample who remained single 1996-2000.				
$\Delta(\text{Pull Factor})_c$	0.0028**	0.0023*	0.0032**	0.0027**
	(0.0013)	(0.0013)	(0.0013)	(0.0013)
$\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$	0.0038***	0.0041***	0.0039***	0.0042***
	(0.0003)	(0.0005)	(0.0003)	(0.0005)
Panel (c) Dep. var: $\Delta(\text{Migrate})_{i,c}$ Sample of individuals aged 26-35 in 1996 who were already married.				
$\Delta(\text{Pull Factor})_c$	0.0047***	0.0039***	0.0048***	0.0040***
	(0.0006)	(0.0007)	(0.0006)	(0.0007)
$\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$	-0.0050***	-0.0035***	-0.0050***	-0.0035***
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Panel (d) Dep. var: $\Delta(\text{Migrate for marital reasons})_{i,c}$ Subset of individuals within the baseline sample who migrated between 1996-2000.				
$\Delta(\text{Pull Factor})_c$	-0.0138***	-0.0108***	-0.0139***	-0.0109***
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
$\Delta(\text{Pull Factor})_c \times \mathbb{1}(\text{Female})_i$	0.0222***	0.0191***	0.0223***	0.0192***
	(0.0007)	(0.0014)	(0.0007)	(0.0014)
Expected terms	No	Yes	No	Yes
Own SEZ shocks	No	No	Yes	Yes

Notes: In panels (a) to (c), the dependent variable $\Delta(\text{Migrate})_{i,c}$ equals one for individual i if they emigrated from county c to another county between 1996 and 2000. In panel (d), the dependent variable Δ(Migrate for marital reasons) equals to 1 if an individual migrated for marital reasons, and 0 for all other reasons. There are 2,186,997; 1,258,041; 2,119,399; and 257,484 observations in panels (a), (b), (c), and (d) respectively. Across all specifications in every panel, we include fixed effects for age in year 2000. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote $p<0.01$, $p<0.05$, $p<0.1$, respectively.

Table 7: Evidence on Marital Incentives: Marital Outcomes

	(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: $\Delta(\text{Marry up})_{i,c}$				
$\Delta(\text{Push Factor})_c$	-0.0049***	-0.0026***	-0.0049***	-0.0026***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\Delta(\text{Push Factor})_c \times \mathbb{1}(\text{Female})_i$	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

Notes: In panel (a), the dependent variable $\Delta(\text{Marry})_{i,c}$ equals one for individual i who was living in county c in 1996 and married sometime between 1996 and 2000. The sample is the same as the baseline sample. In panel (b), the dependent variable Δ(Migrate & Marry) equals to one for individual i who migrated and got married between 1996-2000, and 0 otherwise. The sample is the same as the baseline sample. In panel (c), the dependent variable $\Delta(\text{Marry up})_{i,c}$ equals one for individual i who was living in county c if they married sometime between 1996 and 2000 and married a partner who has a higher education level. The sample is a subset of individuals within the baseline sample: (i) continuously single 1996-2000; or (ii) married during this period, with identified spouse's education. There are 2,562,835; 2,562,835; and 2,141,339 observations in panels (a), (b), and (c) respectively. Across all specifications in both panels, we include fixed effects for age in the year 2000. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***, **, * denote $p < 0.01$, $p < 0.05$, $p < 0.1$.

Table 8: Evidence on Nonmarital Incentives

Dep. var: $\Delta(\text{Migrate})_{i,c}$	(1) Unadjusted OLS	(2) Controlled OLS	(3) Unadjusted OLS	(4) Controlled OLS
Panel (a) Labor Market				
The baseline sample with labor market conditions added.				
$\Delta(\text{Pull Factor})_c$	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)
Panel (b) Educational Opportunities				
Subset of individuals within the baseline sample, 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)
Panel (c) Amenities				
Subset of individuals within the baseline sample who are less-educated.				
$\Delta(\text{Pull Factor})_c$	-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 (d) Amenities				
Subset of individuals within the baseline sample who are highly-educated.				
$\Delta(\text{Pull Factor})_c$	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

Notes: The dependent variable $\Delta(\text{Migrate})_{i,c}$ equals one for individual i if they emigrated from county c to another county between 1996 and 2000. There are 2,562,835; 2,429,582; 1,845,444; and 717,391 observations in panels (a), (b), (c), and (d), respectively. Across specifications, we include fixed effects for age in year 2000. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote $p <0.01$, $p <0.05$, $p <0.1$.

Table 9: SEZ Targeted Industries and Female Labor Shares

Industry	Targeted Share (Unit: %)	Female Labor Share (Unit: %)
Machinery Manufacturing	16.03	35.90
Food, Beverage & Tobacco	15.61	40.73
Biotechnology & Pharmaceuticals	13.92	50.04
Chemicals	13.50	35.68
Electrical Equipment	8.02	56.78
Textiles & Apparel	8.02	65.95
Metals & Mining	6.75	31.53
Construction Materials	6.33	32.54
Automobiles	5.49	31.32
Paper & Forest Products	2.53	57.19
Household Durables	1.69	16.41
Containers & Packaging	0.84	58.42
Leisure Products	0.84	42.59
Communications Equipment	0.42	26.37
Weighted Average		42.33

Notes: We focus on the SEZs that were established for the first time between 1996 and 2000 in our data. For targeted shares in the second column, we calculate the individual shares of each industry. For calculating the female labor share within each industry, we utilize Chinese Population Census 1990 and compute the proportion of females among all workers in each respective industry.

Appendices

A Simulation Algorithm for Expected Δ(Pull Factor)

A total of 87 counties received SEZ shocks between 1996 and 2000 and another 518 counties received SEZ shocks between 2001 and 2006. To compute the expected $\Delta(\text{Pull Factor})_c$, we run simulations following Borusyak and Hull (2023). For each simulation s, we go through the following steps:

1. Given a set of 605 counties in which SEZs were established for the first time between 1996 and 2006, we randomly draw a set of 87 counties X^s.

2. Compute corresponding $\boldsymbol{V}^s$ according to the set of shocks X^s drawn.

3. Compute $\Delta(\text{Pull Factor})^s = \boldsymbol{M}\boldsymbol{V}^s$. For simulation s, we separately compute the values for counties that received or did not receive their own internal SEZ shock and reflect this in the value of $\Delta(\text{Pull Factor})^s$.

4. Repeat steps 1 to 3 5,000 times.

5. Compute the expected instruments:

$$\text{Expected } \Delta(\text{Pull Factor}) = \frac{1}{N^s} \sum_{s=1}^{N^s} \Delta(\text{Pull Factor})^s$$

We contrast the realized Δ(Pull Factor) and the expected Δ(Pull Factor) in Figure C.1. As expected, individuals from Shandong, Hubei, Jiangxi, and Fujian are more affected by the pull of urbanization taking place in nearby fast-growing provinces or metropolitan cities, such as Beijing, Tianjin, Shanghai, Zhejiang, and Guangdong. The realized Δ(Pull Factor) is positively correlated with the expected Δ(Pull Factor), but also presents obvious discrepancies; e.g., Yunnan province.

B Gender-Specific Labor Market Opportunities

We construct $\Delta(\text{Labor Mkt})^F$ and $\Delta(\text{Labor Mkt})^M$ to represent the changes in labor market conditions specific to females and males, respectively. We explain the construction of $\Delta(\text{Labor Mkt})^F$ below, and $\Delta(\text{Labor Mkt})^M$ follows analogously.

$$\Delta(\text{Labor Mkt})^F = \boldsymbol{M}\boldsymbol{B}^F, \tag{B.1}$$

where $\boldsymbol{M}$ is the normalized matrix of migration flows at the province level, as defined in Section 3.2. $\boldsymbol{B}^F$ is a vector whose size is equal to the number of provinces and is constructed in a manner that is similar to a Bartik instrument. Specifically, element $b_p^F \in \boldsymbol{B}^F$ for province p is computed as $b_p^F = \sum_{k=1}^{K} b_{pk}^F$, where k indexes an industry, K is the total number of industries, and

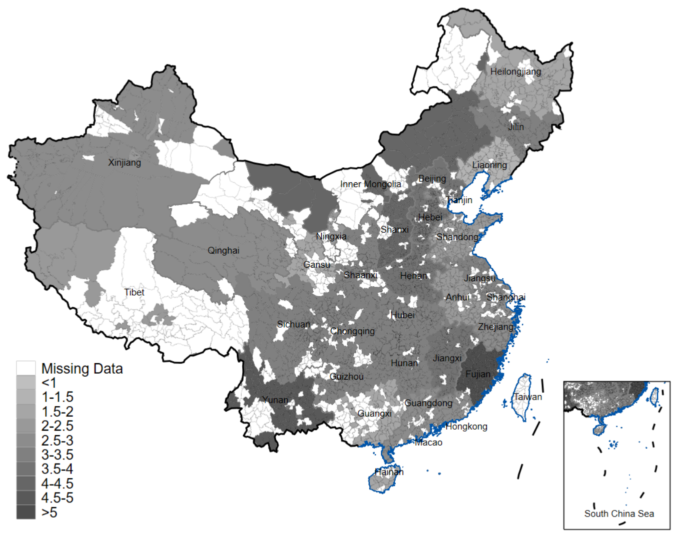
$$b_{pk}^F = z_{pk} \cdot \{g_k^{treat} w_p^{treat} + g_k^{not} w_p^{not}\} \cdot f_k. \tag{B.2}$$

In equation (B.2), z_{pk} is industry k's share in province p in 1990, measured as the number of workers in industry k in province p divided by the number of all workers in province p in 1990. The term in braces shows the employment growth between 1990 and 2000 in industry k as a weighted average between treated and not-yet-treated groups. Specifically,

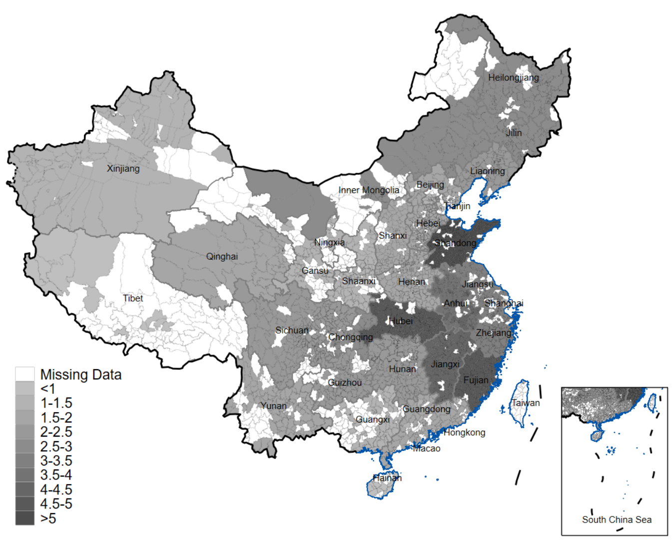
$$g_k^{treat} = \log(\text{employment in } 2000)_k^{treat} - \log(\text{employment in } 1990)_k^{treat} \tag{B.3}$$

is the employment growth in industry k for the treated at national level and g_k^{not} is defined analogously for the not-yet-treated. Weight w_p^{treat} is the share of the population in province p living in counties that were treated with SEZs as of year 1990, and $w_p^{not} = 1 - w_p^{treat}$. Lastly, the f_k term is the percentage of female workers in industry k in 1990, which is shown in Figure C.2. As $\Delta(\text{Labor Mkt})^F$ is constructed at the province level, we map its values to individuals within specific counties in our FD model, depending on the province in which a county is located.

C Additional Figures



(a) Realized Δ(Pull Factor)$_c$



(b) Expected Δ(Pull Factor)$_c$

Figure C.1: Δ(Pull Factor) – Realized and Expected

Notes: Panel (a) displays the values of Δ(Pull Factor)$_c$ calculated from the method outlined in Section 3.2 using actual data. Counties in darker gray exhibit higher values of Δ(Pull Factor)$_c$, implying a stronger pull force that encourages residents to emigrate to other locations. Panel (b) shows the expected values of Δ(Pull Factor)$_c$ derived from implementing the simulation algorithm outlined in Appendix Section A.

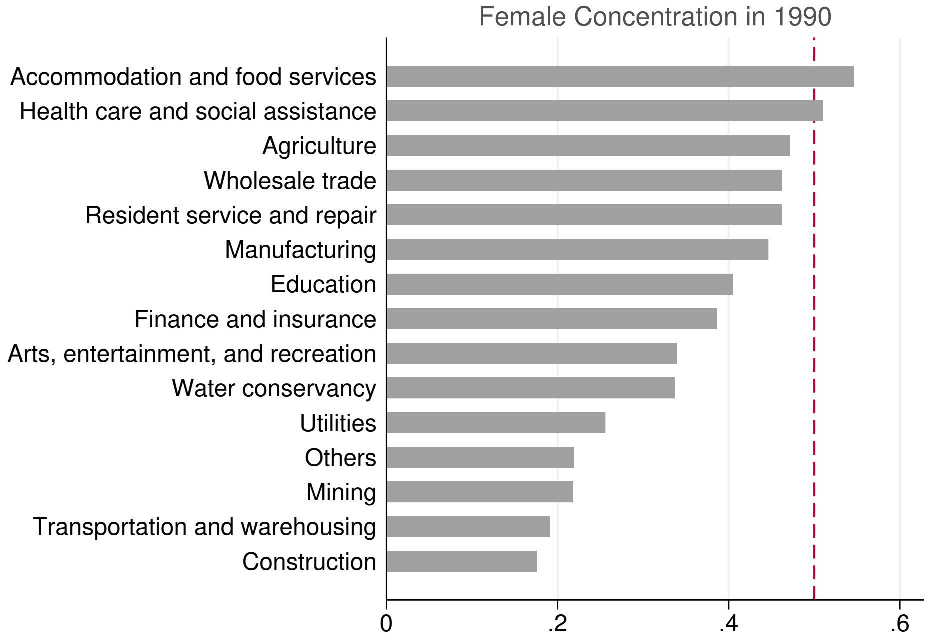


Figure C.2: Industry-specific Female Share in 1990

Data source: Chinese Population Census 1990.
Notes: For each industry, we compute the share of female workers using Chinese Population Census 1990 data. The vertical dashed line indicates the value 0.5, which suggests a perfectly balanced ratio of male to female workers.

D Additional Tables

Table D.1: ATT Estimates for Each Year, Across All Groups

	(1)	(2)
Dep. variable	Young female migrants	Young male migrants
Calendar Average	0.4374***	0.3597***
	(0.0793)	(0.0771)
Year 1996	0.3630***	0.3699***
	(0.1393)	(0.1430)
Year 1997	0.2656*	0.1881
	(0.1379)	(0.1216)
Year 1998	0.4987 ***	0.3348***
	(0.1307)	(0.1221)
Year 1999	0.6158***	0.4867***
	(0.1331)	(0.1183)
Year 2000	0.4440***	0.4189***
	(0.0984)	(0.0974)
Observations	6,975	6,975
Control Group	Not-yet	Not-yet

Notes: We utilize the yearly sample of counties in China from 1995 to 2000. Across all specifications, each dependent variable is subjected to its inverse hyperbolic sine transformation. The term 'young' refers to individuals between the ages of 16 and 25. We estimate a staggered DiD model following Callaway and Sant'Anna (2021) and report the estimates of the average treatment effect on the treated for each calendar year across all groups. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

Table D.2: Migration Outcomes – A Sample of Individuals Living in Small-sized Counties

$\Delta(\text{Migrate})_{i,c}$	(1) Unadjusted OLS	(2) Controlled OLS	(3) Unadjusted OLS	(4) Controlled OLS
$\Delta(\text{Push Factor})_c$	0.0027**	0.0018*	0.0029***	0.0021*
	(0.0011)	(0.0011)	(0.0011)	(0.0011)
$\Delta(\text{Push Factor})_c \times \mathbb{1}(\text{Female})_i$	0.0035***	0.0016***	0.0036***	0.0016***
	(0.0002)	(0.0003)	(0.0002)	(0.0004)
Expected $\Delta(\text{Push Factor})_c$		0.0091***		0.0092***
		(0.0015)		(0.0015)
Expected $\Delta(\text{Push Factor})_c \times \mathbb{1}(\text{Female})_i$		0.0025***		0.0025***
		(0.0004)		(0.0004)
$\Delta(\text{SEZ})_c$			-0.0235***	-0.0255***
			(0.0074)	(0.0077)
$\Delta(\text{SEZ})_c \times \mathbb{1}(\text{Female})_i$			-0.0051*	-0.0055**
			(0.0027)	(0.0026)
Observations	1,571,397	1,571,397	1,571,397	1,571,397
Root MSE	0.283	0.283	0.283	0.283

Notes: The dependent variable $\Delta(\text{Migrate})_{i,c}$ equals one for individual i if they emigrated from county c to another county between 1996 and 2000. The sample consists of all individuals who were between the ages of 16 and 25 at any point during the period 1996 to 2000 and living in counties that had a population of less than 700,000 as of 1982. Across all specifications, we include fixed effects for age in year 2000. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote $p<0.01$, $p<0.05$, $p<0.1$, respectively.

Table D.3: Tendency to Marry Up – Individuals Aged 20-35 Years

Δ(Marry up)$_{i,c}$	(1) Unadjusted OLS	(2) Controlled OLS	(3) Unadjusted OLS	(4) Controlled OLS
Δ(Push Factor)$_c$	-0.0033***	-0.0017***	-0.0033***	-0.0017***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Δ(Push Factor)$_c$ $\times$ $\mathbb{1}$(Female)$_i$	0.0063***	0.0028***	0.0063***	0.0028***
	(0.0001)	(0.0003)	(0.0001)	(0.0003)
Observations	3,327,375	3,327,375	3,327,375	3,327,375
Root MSE	0.177	0.177	0.177	0.177
Mean(y)	0.034	0.034	0.034	0.034

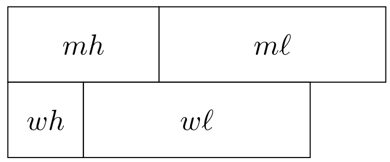
Notes: The dependent variable Δ(Marry up)$_{i,c}$ equals one for individual i if they married sometime between 1996 and 2000 and married a partner who has a higher education level. The sample comprises 3,327,375 individuals who meet the following criteria: (1) they were between the ages of 20 and 35 at any time from 1996 to 2000 and (2) they were either consistently single throughout the period from 1996 to 2000 or they married between 1996 and 2000, with their spouse's education type identified. Across all specifications, we include fixed effects for age in year 2000. Standard errors in parentheses are corrected for heteroskedasticity and clustered at county level. Asterisks ***, **, * denote p <0.01, p <0.05, p <0.1

E Omitted Proofs

Theorem 1. *There exists a unique equilibrium.*

Proof of Theorem 1. For expositional convenience, we define net cost c as the negative of net gain y, and the distributions of net costs as $F_{g\theta}$, where $g \in \{m, w\}$ and $\theta \in \{h, \ell\}$.

The distribution of types in the rural marriage market by assumption takes the following form, that is, $G_{mh} > G_{wh}$ and $G_{mh} + G_{m\ell} > G_{wh} + G_{w\ell}$.



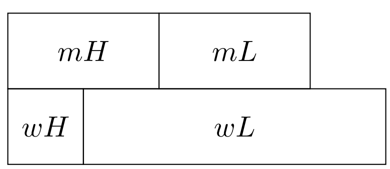
Hence, the stable payoff conditions are $u_\ell = 0, u_\ell + v_\ell = s_{\ell\ell}, u_h + v_\ell = s_{h\ell}, u_h + v_h = s_{hh}$, which imply payoffs

$$u_\ell = 0, v_\ell = s_{\ell\ell}, u_h = s_{h\ell} - s_{\ell\ell}, v_h = s_{hh} - (s_{h\ell} - s_{\ell\ell}).$$

There are three cases with respect to the urban marriage market. They differ in whether the mass of high-type men is (i) strictly more than, (ii) strictly less than, or (iii) equal to that of high-type women.

Case (i). The mass of high-type men is strictly more than that of high-type women.

The distribution of types in the marriage market takes the following form, that is, $G_{mH} > G_{wH}$ and $G_{mH} + G_{wH} < G_{wH} + G_{wL}$.

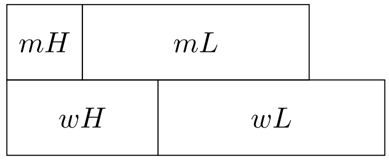


The stable payoff conditions are $u_H + v_H = s_{HH}, u_H + v_L = s_{HL}, u_L + v_L = s_{LL}, v_L = 0$, which imply payoffs

$$v_L = 0, u_L = s_{LL}, u_H = s_{HL}, v_H = s_{HH} - s_{HL}.$$

Case (ii). The mass of high-type men is strictly less than that of high-type women.

The distribution of types in the marriage market satisfies $G_{mH} < G_{wH}$ and $G_{mH} + G_{wH} < G_{wH} + G_{wL}$:



The stable payoff conditions are

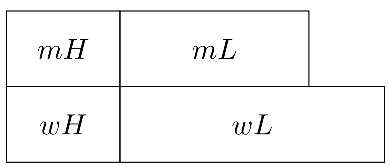
$$u_H + v_H = s_{HH}, u_L + v_H = s_{LH}, u_L + v_L = s_{LL}, v_L = 0,$$

which imply payoffs

$$v_L = 0, u_L = s_{LL}, v_H = s_{LH} - s_{LL}, u_H = s_{HH} - (s_{LH} - s_{LL}).$$

Case (iii). The mass of high-type men is equal to that of high-type women.

The distribution of types in the marriage market satisfies $G_{mH} = G_{wH}$ and $G_{mH} + G_{wH} < G_{wH} + G_{wL}$:



Stability conditions are

$$u_H + v_H = s_{HH}, u_L + v_L = s_{LL}, v_L = 0, u_H + v_L \geq s_{HL}, u_L + v_H \geq s_{LH}.$$

Together, they imply

$$v_L = 0, u_L = s_{LL}, v_H = \lambda(s_{HH} - s_{HL}) + (1-\lambda)(s_{LH} - s_{LL}), u_H = \lambda s_{HL} + (1-\lambda)(s_{HH} - (s_{LH} - s_{LL}),$$

where $\lambda \in [0, 1]$.

In fact, in summary, in all three cases,

$$v_L = 0, u_L = s_{LL}, v_H = \lambda(s_{HH} - s_{HL}) + (1-\lambda)(s_{LH} - s_{LL}), u_H = \lambda s_{HL} + (1-\lambda)(s_{HH} - (s_{LH} - s_{LL}),$$

where $\lambda \in [0,1]$. In case (i), $\lambda = 1$, and in case (ii), $\lambda = 0$.

The marital gains from migration are

$$\begin{aligned}
u_L - u_\ell &= s_{LL}, \\
u_H - u_h &= (u_H - u_h)_\lambda := \lambda s_{HL} + (1-\lambda)(s_{HH} - (s_{LH} - s_{LL})) - (s_{h\ell} - s_{\ell\ell}), \\
v_L - v_\ell &= -s_{\ell\ell}, \text{ and} \\
v_H - v_h &= (v_H - v_h)_\lambda := \lambda(s_{HH} - s_{HL}) + (1-\lambda)(s_{LH} - s_{LL}) - (s_{hh} - (s_{h\ell} - s_{\ell\ell})).
\end{aligned}$$

Note that $(u_H - u_h)_\lambda$ is strictly decreasing in λ and $(v_H - v_h)_\lambda$ is strictly increasing in λ.

Case (i) holds if $\phi_{mH} + F_{mh}((u_H - u_h)_1) > \phi_{wH} + F_{wh}((v_H - v_h)_1)$, namely,

$$\phi_{mH} + F_{mh}(s_{HL} - (s_{h\ell} - s_{\ell\ell})) > \phi_{wH} + F_{wh}(s_{HH} - s_{HL} - (s_{hh} - (s_{h\ell} - s_{\ell\ell}))).$$

Case (ii) holds if $\phi_{mH} + F_{mh}((u_H - u_h)_0) < \phi_{wH} + F_{wh}((v_H - v_h)_0)$, namely,

$$\phi_{mH} + F_{mh}(s_{HH} - (s_{LH} - s_{LL}) - (s_{h\ell} - s_{\ell\ell})) < \phi_{wH} + F_{wh}(s_{LH} - s_{LL} - (s_{hh} - (s_{h\ell} - s_{\ell\ell}))).$$

Case (iii) holds if $\phi_{mH} + F_{mh}((u_H - u_h)_1) \leq \phi_{wH} + F_{wh}((v_H - v_h)_1)$, namely,

$$\phi_{mH} + F_{mh}(s_{HL} - (s_{h\ell} - s_{\ell\ell})) \leq \phi_{wH} + F_{wh}(s_{HH} - s_{HL} - (s_{hh} - (s_{h\ell} - s_{\ell\ell})))$$

and $\phi_{mH} + F_{mh}((u_H - u_h)_0) \geq \phi_{wH} + F_{wh}((v_H - v_h)_0)$, namely,

$$\phi_{mH} + F_{mh}(s_{HH} - (s_{LH} - s_{LL}) - (s_{h\ell} - s_{\ell\ell})) \geq \phi_{wH} + F_{wh}(s_{LH} - s_{LL} - (s_{hh} - (s_{h\ell} - s_{\ell\ell}))).$$

To prove equilibrium existence and uniqueness, define function

$$\psi(\lambda) := [\phi_{mH} + F_{mh}((u_H - u_h)_\lambda)] - [\phi_{wH} + F_{wh}((v_H - v_h)_\lambda)],$$

which is the LHS minus the RHS in the inequalities of the conditions for the three cases. Note that $\psi(\lambda)$ is strictly decreasing in λ and crosses 0. Define function

$$\Gamma(\lambda) = \begin{cases} 1 & \text{if } \psi(1) > 0 \\ 0 & \text{if } \psi(0) < 0 \\ \lambda^* \text{ s.t. } \psi(\lambda^*) = 0 & \text{otherwise} \end{cases}$$

The three conditions in the function correspond to those in the three cases. Note that a λ uniquely identifies payoff differences $u_L - u_\ell$, etc. Hence, a λ characterizes an equilibrium payoff. Finding solutions to $\Gamma(\lambda) = \lambda$ is equivalent to finding all equilibria. By Kakutani's fixed-point theorem, there exists a solution, hence proving equilibrium existence. In addition, given the monotonicity of ψ and that $\psi(\lambda) = 0$ has a unique solution when it exists, $\Gamma(\lambda) - \lambda = 0$ has a unique solution.

In addition, to sustain the fixed marriage market structure in the rural area, we have the following restrictions: (i) $G_{mh} > G_{wh}$ even with the most departures of high-type men and fewest departures of high-type women, i.e.,

$$\text{(A1)} \qquad \phi_{mh} - F_{mh}(s_{HH} - (s_{LH} - s_{LL}) > \phi_{wh} - F_{wh}(s_{LH} - s_{LL})$$

and (ii) $G_{mh} + G_{m\ell} > G_{wh} + G_{w\ell}$ under the most departures of men and fewest of women, i.e.,

$$\text{(A2)} \qquad F_{mh}(s_{HL}) + F_{ml}(s_{LL}) < F_{wh}(s_{HH} - s_{HL}) + F_{w\ell}(-s_{\ell\ell}).$$

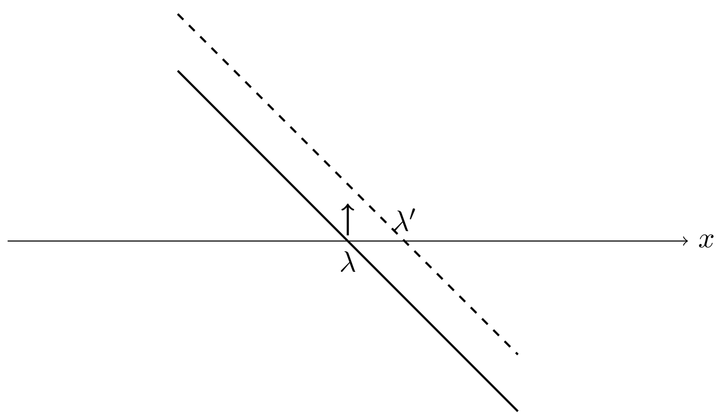
$\square$

Proposition 1. *More women than men move to cities if (i) the mass of high-type urban men increases, and/or (ii) rural women's distributions of nonmarital gains shift up first-order stochastically.*

Proof of Proposition 1. To prove claim (i), consider

$$\psi(\lambda) := [\phi_{mH} + F_{mh}((u_H - u_h)_\lambda)] - [\phi_{wH} + F_{wh}((v_H - v_h)_\lambda)],$$

which is defined in the proof of Theorem 1. It is a strictly decreasing function of λ. It crosses x-axis at a higher point now, hence leading to a higher λ^* solution and consequently more high-type women and more women overall moving. The figure below illustrates the shift of $\psi(\lambda)$ when ϕ_{mH} increases.



To prove claim (ii), note that the mass of female migrants is $F_{wh}(c^*_{wh}) + F_{w\ell}(-s_{\ell\ell})$, which is determined by, as shown in the proof of Theorem 1,

$$c^*_{wh} = v^*_H - v^*_h = \lambda^*(s_{HH} - s_{HL}) + (1 - \lambda^*)(s_{LH} - s_{LL}) - s_{hh} + (s_{h\ell} - s_{\ell\ell}),$$

where $\lambda^* \in [0, 1]$. The mass of female migrants increases F_{wh} shifts first-order stochastically. $\square$