Out of Male-land: Imbalanced Sex Ratios and Migration *

Zhang, Chongjiu †

Beijing Normal University

January 1, 2024

Abstract

Male-skewed sex ratio imbalances have a wide range of negative repercussions and may serve as a driving force behind migration. I investigate the extent to which imbalanced sex ratios contribute to local exodus in the Chinese context. On the aggregate level, I find that regions with higher sex ratios in 2010 experienced higher rates of out-migration flows in the following decade. On the individual level, I find that individuals from regions with higher sex ratios are more inclined to migrate, with this effect being particularly pronounced for males, especially those from lower socioeconomic backgrounds. I further note that these implications extend to families, as migrant parents from areas with higher sex ratios are more inclined to bring their children along rather than leaving them behind.

*I would like to express my heartfelt thanks to ...

†Address: 19 Xinjiekouwai St, Haidian District, Beijing, 100875, China. Email: emisrogan@outlook.com. Homepage: https://zhangchongjiu.github.io/

1 Introduction

Male-skewed sex ratio imbalance stands as one of the cruelest manifestations of gender inequality in our world. It occurs when there is an excessive number of males and a significant deficit of females within a population. While a balanced sex ratio should ideally be around 100 (males per 100 females), imbalances prevail, resulting in millions of "missing women" (Qian, 2008; Sen, 1992). As of 2020, the global population consists of 50.30% males and 49.70% females, with a surplus of 46.27 million males. The imbalance is most pronounced in many Asian populations, such as China, India, South Korea, or even countries like Singapore that are often considered to have achieved relatively small gender gaps. This persistent phenomenon is often attributed to cultural preferences for sons, along with various forms of pre- or postnatal sex-selective practices (Chen, Li and Meng, 2013; Hesketh and Xing, 2006; Li, Yi and Zhang, 2011).

The consequences of imbalanced sex ratios [1] extend far beyond gender disparities. On one hand, they give rise to intense competition for marriage among males, resulting in various physical and psychological challenges for those experiencing such pressures (Jiang and Sánchez-Barricarte, 2012; Jin et al., 2013). On the other hand, they perpetuate patriarchal norms and limit opportunities for females (Edlund, 1999). The scarcity of women further contributes to an increase in human trafficking, forced marriages, and violence against women (D'Alessio and Stolzenberg, 2010; Xiong, 2022*a*). These consequences extend to numerous social and economic domains. Both men and women exposed to such imbalances face heightened health risks and increased mortality rates (Chang, Kan and Zhang, 2021). The aggregation of single men poses a significant threat to public safety, as higher sex ratios can directly contribute to higher levels of crime and violence (Cameron, Meng and Zhang, 2019; Edlund et al., 2013). In the long term, these imbalances can undermine social welfare and impede economic growth (Bhaskar, 2011; Eggleston et al., 2013).

In response to these circumstances, individuals employ various strategies to navigate the disadvantages caused by sex ratio imbalances. Numerous studies have documented how people adapt to high sex ratios. For instance, men, who face direct pressure to find a partner in the marriage market, strive to increase their desirability for marriage. They may increase investments in education before marriage, intensify their work efforts, take on more dangerous or undesirable jobs, or even engage in criminal activities (Cameron, Meng and Zhang, 2019; Lafortune, 2013; Wei and Zhang, 2011*b*). Furthermore, families with sons

[1] It bears mentioning that sex ratios are not always male-skewed. There are also cases for female-skewed sex ratios, or male scarcities. See, for example, Abramitzky, Delavande and Vasconcelos (2011) and Brainerd (2017) for discussions of the effects of female-skewed sex ratio imbalances, both of which caused by war. It is, however, a relatively less common demographic phenomenon. In the context of this paper, sex ratio imbalance specifically refers to a male-skewed imbalance.

may also attempt to enhance their sons' competitiveness by saving more money or investing in risky assets (Li et al., 2022; Wei and Zhang, 2011*a*).

However, individuals facing adverse circumstances have another option besides adapting: leaving. This idea is exemplified in the study of "push factors" of migration, where people may decide to migrate due to multiple unfavorable circumstances in their environment (Lee, 1966). Migration decisions are complex and multifaceted, often involving a myriad of factors that may not be immediately apparent. While much of the literature on this topic focuses on direct factors that may induce migration, such as war and violence (Buggle et al., 2023; Padilla-Romo and Peluffo, 2023), natural disasters and extreme weather events (Deschenes and Moretti, 2009), and pollution (Chen, Oliva and Zhang, 2022), there is also recognition that broader contextual factors can play a crucial role in pushing migration (Van Hear, Bakewell and Long, 2018). For instance, Guo, Zhang and Zhou (2023) documents the significant impact of demographic factors, particularly fertility and cohort sizes, on internal migration flows across different regions of China.

I investigate the extent to which sex ratios (hereafter SR) drive local exodus in the Chinese context. Over the past decades, China has witnessed huge migration flows and pronounced male-skewed SR, making it an ideal setting for this study. The empirical analysis proceeds in two lines. On the aggregate level, I document a pattern on the relationship between a region's SR and its subsequent out-migration in the following decade. I employ the residual approach to calculate the net out-migration ratio for each region (prefecture or county), which is the proportion of the population leaving the region, accounting for deaths and new arrivals (Chen, Oliva and Zhang, 2022; Feng, Krueger and Oppenheimer, 2010). On the individual level, I examine the SR of each individual's place of origin and their likelihood of migrating. I leverage on China's household registration (hukou) system to trace each migrant back to their place of origin, and calculate SRs adjusting for previous migration flows. The findings demonstrate a clear role of SR imbalance as a push factor of internal migration, with this effect being particularly pronounced for males. I document that imbalanced SRs create a hostile environment that, once becomes detectable, pushes people to leave.

2 Contextual setting: two demographic trends in post-reform China

China experienced a remarkable surge in economic growth and urbanization following its reform in 1978. It coincides with two significant demographic changes: extensive internal migration and rapidly increasing SRs. This section provides a brief overview of the contextual setting in China.

2.1 Internal migration

Unlike countries with more relaxed migration policies, China had historically implemented strict planned economic policies along with a rigid household registration (*hukou*) system, which documents and registers individuals to their region of origin from birth until death. Between the 1950s and 1980s, the number of internal migrants in China was relatively small. However, starting from the mid-1980s, as the economic system underwent reforms and the government gradually eased restrictions on population movements, the scale of migration started to grow rapidly. As of 2020, the number of migrants in China has soared to 385.82 million, accounting for 26.04% of the national population. This phenomenon stands as the largest migration in human history (Liang, 2016).

China's great migration has garnered much attention in recent decades, leading to numerous researches exploring the driving factors and implications behind it. One of the key perspectives attributes the rise of internal migration to the coaction of the hukou system relaxing, market transition, and rapid economic development and urbanization processes (Lin, 1988; Liu and Xu, 2017; Zhao, 1999). Changes in the economic landscape and labor market have led many rural residents migrate to urban regions in search of better living conditions and job opportunities. Other studies bring attention to less apparent factors, such as air pollution in the migrant's place of origin (Chen, Oliva and Zhang, 2022; Lai et al., 2021), as well as the impact of marriage prospects (Xiong, 2023). These factors highlight the complex interplay between economic, demographic, and social factors in shaping migration patterns.

2.2 Imbalanced SRs

Another characteristic of Chinese demographic is an extremely male-skewed SR. In 2020, China's SR at birth is 111.3, though lower than that in 2010, is still much higher than a normal value of 103-106. Age cohorts with the most severe SR imbalance are 10-14 year olds (119.1) and 15-19 year olds (118.4), indicating that the imbalance will likely endure as this generation matures. While many associate the primary cause of the rising SR with sex-selective abortion (Chen, Li and Meng, 2013), it is widely recognized that the underlying root cause can be traced back to entrenched sexism, a culture of favoring males, and the systemic disadvantages that women face in terms of opportunities for development (Qian, 2008; Xue, 2016).

This imbalance poses seriously issues for males as they encounter intense competition and find themselves squeezed in the marriage market. According to Guilmoto (2012), even if China's SR at birth returns to normal levels by 2020 (which obviously has not), the unmarried rate of all men under 50 in China will reach 15% by 2055. In China where the practice of marriage is almost universal and being unmarried is stigmatized (Ji, 2015), males may resort to various means to enhance their competitiveness and desirability (Cameron, Meng and Zhang, 2019). They may also experience psychological anomie and mental disorders, particularly those in rural or underprivileged areas (Jiang and Sánchez-Barricarte, 2012; Jin et al., 2013).

The effects of high SRs on women are more multifaceted. While some argue that female scarcity may contribute to the enhancement of women's relative status and bargaining power, more have elaborated on how this imbalance can lead to the reinforcement of disadvantages for women (see, e.g., Edlund, 1999, for an excel discussion). In both the short and long run, SR imbalance will reduce female labor force participation and lead to fewer women in high-reward or high-status jobs (Angrist, 2002; Grosjean and Khattar, 2019). The shortage of women has led to an increase in human trafficking and forced marriage (Xiong, 2022*a*), and there is also a clear link between SRs and violence, both in general (Dreze and Khera, 2000; Messner and Sampson, 1991) and against women (D'Alessio and Stolzenberg, 2010). While the Chinese government has taken steps to address the issue, the long-term effects of SR imbalance remain to be seen. It is clear that this imbalance has the potential to cause significant social, economic, and demographic upheaval in China and other countries where the problem persists.

3 Data

The primary source of data is China's 6^{th} and 7^{th} Population Census, conducted by the State Council in 2010 and 2020 respectively. The 2010 Census includes two datasets: a micro dataset at the individual level (referred to as the *Micro Census*) that includes demographic characteristics such as gender, age, educational level, locations of both current residence and hukou registration for over 4 million Chinese citizens; and the second one is the *Tabulation on China Population Census by County* (known as the *Tabulation by County*). It contains the population structure and a slew of other indicators for county-level units across the nation in 2010. For the 2020 Census, only the *Tabulation by County* is available. Both censuses neglect to include information on Macau, Hong Kong, and Taiwan, thus limiting the analysis within mainland China.

The empirical analysis is conducted at both the aggregate level and individual level. At the aggregate level, the primary data source is the *Tabulation by County* from two censuses. The focus is on examining the relationship between a region's SR in 2010 and its out-migration patterns in the following decade. At the individual-level, the chief sample is the 2010 *Micro Census*.

I further rely on multiple Statistical Yearbooks to take into account a series of indicators of social and economic development. For cities and prefectures, I consult the *China City Statistical Yearbook* (CCiSY) 2010 compiled by the Department of Urban Surveys of the National Bureau of Statistics of China. The CCiSY provides comprehensive information on the urbanization and socio-economic development for Chinese cities and prefectures. Other sources of data, such as 2015 China 1% Population Sample Survey, 2017 China Migrants Dynamic Survey, and multiple other yearbooks are also employed for robustness checks and further discussion.

4 Aggregate analysis: SRs and out-migration

I begin the empirical investigation by examining the pattern between SRs and population outflows on an aggregate level. The unit of analysis is regions, including prefectures or counties. Prefectures form the second level division within China's administrative system, falling under provinces. Counties function as the administrative division beneath prefectures. In mainland China, as of 2010, there are a total of 31 provinces, 341 prefectures, and 2,893 counties [2].

4.1 Method and identification

4.1.1 Dependent variable: Net out-migration ratio

I employ the residual approach to calculate the net out-migration ratio for each region (Chen, Oliva and Zhang, 2022; Feng, Krueger and Oppenheimer, 2010). This approach assesses the population change between 2010 and 2020 by calculating the proportion of the population leaving a region, accounting for deaths and new arrivals. It represents the balance between the number of individuals ever entering a region and the number of individuals ever leaving it (Siegel and Hamilton, 1952). Mathematically, the ratio can be

[2] The administrative regions in China, as outlined by the Constitution of the People's Republic of China, consist of three de jure levels: province, county, and town. However, given the substantial number of counties nationwide (approximately 2,900), and the inherent challenges associated with direct jurisdiction by provincial governments, the establishment of prefecture-level administrative divisions has occurred to oversee these counties. I also include the four province-level municipalities (Beijing, Tianjin, Shanghai, and Chongqing) for analyses at the prefecture level as they function akin to large cities, with their respective counties being directly administered by them.

represented as:

$$Out\_Migr_{r,2010-2020} = \frac{POP_{[15,50),r,2010} - POP_{[25,60),r,2020} - Death_{[15,50),r,2010-2020}}{POP_{[15,50),r,2010}} \tag{1}$$

$Out\_Migr_{r,2010-2020}$ denotes the net out-migration ratio of region r between 2010 and 2020. $POP_{[15,50),r,2010}$ represents the baseline population consisting of individuals aged 15-50 within the region in 2010. $POP_{[25,60),r,2020}$ denotes the total population aged 25-60 in the same region ten years later, and $Death_{[15,50),r,2010-2020}$ refers to the total number of deaths among individuals of this age group during the decade. A larger net out-migration ratio indicates a higher level of out-migration experienced by a region. I also calculate the net out-migration ratio separately for the male and female populations to explore the gender-specific patterns between regional SR and population outflows.

The calculation requires age-specific deaths or mortality rates for each region. However, this data is not publicly available for most regions in non-census years. I rely on the *China Population Census Yearbook* to estimate regional mortality rates. This involves calculating mortality rates for each province, adjusting these rates using region-specific weights, and estimating the number of deaths by age group in each region. Appendix B explains this calculation in detail.

4.1.2 Explanatory variable: SR of local residents

The explanatory variable is the SR of the current residents in each region in 2010, sourced from the 2010 *Tabulation by County* and calculated as follows:

$$Sex\_Ratio_r = \frac{Male\_POP_r}{Female\_POP_r} \times 100 \tag{2}$$

where $Male\_POP_r$ denotes the number of male population and $Female\_POP_r$ denotes the number of female population. Descriptive statistics for SRs at the prefecture level based on current residence are presented in Panel A of Table 1. For the main analyses, I employ SR of individuals aged 15-50 as the key explanatory variable. However, further investigations reveal that the results are not sensitive to the chosen age groups.

4.1.3 Analytic strategy

I run the following regression:

$$Out\_Migr_{r,\ 2010-2020} = \alpha_0 + \alpha_1 \times Sex\_Ratio_{r,\ 2010} + X \times \beta + \epsilon_r \tag{3}$$

The dependent variable is the net out-migration ratio from 2010-2020 for individuals aged between 15 and 50 residing in prefecture or county r as of 2010. X includes a series of regional characteristics. Descriptive statistics of these variables at the prefecture level can be found in Appendix A. I incorporate regional fixed effects to account for trends at a level higher than the unit of analysis. Specifically, analyses on prefectures control for province-level fixed effects, and analyses on counties control for prefecture-level fixed effects.

4.2 Results: SRs and out-migration

Table 2 presents the regression results. Columns (1) and (2) of the table display the results for the total population, while columns (3)-(4) and (5)-(6) show the results for males and females, respectively.

Panel A of the table focuses on prefectures as the unit of analysis. For the total population, for every one unit increase in the SR (1 man per 100 women), the population out-migration will increase by 0.005, which represents 15.20% of the sample mean. A one standard deviation increase in the SR (5.906) is associated with an 18.57% standard deviation increase in the net out-migration ratio. For the male population, the effect of SRs on out-migration is even more pronounced. Each standard deviation increase in the SR corresponds to a 27.15% standard deviation increase in the net out-migration ratio. On the other hand, the impact of SRs on out-migration for the female population remains positive but becomes statistically insignificant when controlling for province fixed effects.

Panel B additionally examines counties as the unit of analysis. On average, the net out-migration ratios of counties tend to be higher than that of prefectures due to the large number of migrants moving across counties but remaining within the same prefecture. SRs at the county level still significantly predict population outflow, with a coefficient size similar to that observed in the prefecture-level regressions. For the entire population, each one unit increase in the SR in 2010 is associated with a 0.0042 increase in county out-migration, equivalent to 4.46% of the sample mean, and a one standard deviation increase in the SR (9.698) leads to a 20.78% standard deviation increase in the net out-migration ratio of the region. Similar to results from prefectures, the impact of SRs on population outflow is more significant for the male population than the female population.

These results show that local SRs are strong predictors of population outflows. I conduct additional analyses using the net out-migration ratio for individuals aged 15-60 as the dependent variable, as well as exploring the effects of SRs for other age groups (including 15-60 years old, 20-50 years old, and SR of the total population, among others) as explanatory variables, and they yield consistent results.

5 Individual analysis: SRs and migration decisions

While the aggregate analyses offer a simple and straightforward perspective regarding SRs and population outflows, the do not provide a nuanced view of whether every individual in an environment with SR imbalance will choose to leave. In this section, I shift to an individual-level analysis and investigate how SRs influence the likelihood of each individual becoming a migrant.

5.1 Method and identification

5.1.1 Dependent variable: Migrant (Yes or No)

I define migrants as individuals who *move across prefectures*, following NBS (2023)'s definition of *floating populations*. Specifically, these are individuals whose hukou is registered in a different prefecture from their place of residence in the 2010 *Micro Census*. I focus mainly on cross-prefecture migration for two main reasons. Firstly, provinces, the administrative division above prefectures, are limited in numbers and generally have an extensive geographical and demographic scale, with populations of numerous provinces in China exceeding 50 or even 100 million, contributing to a great deal of internal heterogeneity. On the other hand, counties, the division under prefectures, are too small and ill-suited for studying migration and calculating regional SR. This limitation arises from the fact that people's daily activities normally extend beyond the boundaries of a single county. Analyses at the county level may not capture the comprehensive shared environment and connectivity that occur among populations within a prefecture.

It is important to acknowledge that the definition of migrants used here gives rise to certain concerns. One fundamental concern is that individuals who undergo a change in their hukou registration are excluded from the definition. This means that if a person moves to a different location and registers their hukou there, they cannot be identified as migrants. However, this situation should be relatively rare, as China imposes stringent and intricate restrictions on changing one's hukou registration. Meanwhile, the census only includes information on Chinese residents, making it impossible to identify international immigrants who have departed from China.

5.1.2 Explanatory variable: SR for the place of origin

The explanatory variable is the SR for each prefecture. While in the aggregate analyses, SRs are calculated based on current residence in 2010, one must bear in mind that these values represent the *aftermath* of migration that had occurred before 2010, rather than the *cause* of it. In the individual analyses, the focus is on whether a person was a migrant as of 2010, and SR based on current residence no longer function as an appropriate measure. To truly understand how SRs affect migration, it's necessary to calculate the SR of the original population in a given region.

China's hukou system allows for tracing each migrant back to their place of origin, and I calculate SRs based on the prefecture where one's hukou is registered. These values represent the local SRs assuming that everyone returns to the prefecture of their hukou registration - that is, if no migration has ever happened [3] . Descriptive statistics are presented in Panel B of Table 1.

However, SR based on hukou registration is not the *de facto* SR one face when making migration decisions due to the varying timing of migration across different locations. For instance, individuals born in prefecture h who leave that place in year t should not be included in the calculation of SR for h after t. Conversely, individuals born outside of h who arrive in t will affect the SR of h from that point onwards. Therefore, when considering any specific prefecture in any given year, it is crucial to take into account the population inflow and outflow that occurred before that year, while excluding migration that happened after that year.

I therefore propose a third calculation of local SRs which considers the timing of migration. In the *Micro Census*, migrants were asked to specify the duration of their migration, and the responses were categorized into various timeframes ranging from less than 6 months to more than 6 years. Individuals who had migrated for more than 6 years were not further classified. Using the information gathered, I am able to obtain the population for each prefecture for each year between 2004 to 2010 and calculate SRs based on the timing of migration. Consider an example: in 2005, the population of prefecture h includes the following groups: (1) individuals whose hukou is registered in h and have always resided there; (2) individuals whose hukou is registered in h, but currently reside outside of h (as of 2010 during the census), and have left h for less than 5 years; (3) individuals whose hukou is not registered in h, but currently live in h and have resided there for at least 5 years. Consequently, the SR of h in 2005 is determined by the number of males compared to the number of females in the population mentioned above.

Admittedly, this calculation has its limitations. Firstly, since the highest response option for the duration of migration is "six years and above", it is unable to differentiate long-term migrants who have migrated for six years or more. Consequently, it cannot accurately calculate the SR for various locations before the year 2004. Additionally, this calculation inherently assumes that each individual migrates only once, while it is possible for a migrant to have made multiple stops along the way. For instance, a migrant may have originated from

[3] This, of course, cannot take into account those who have changed their hukou registration.

region h, currently reside in region r, but may have briefly stayed in other places during the migration process. Unfortunately, the Population Census does not include questions regarding migration history, so we must assume a direct migration from region h to r [4]. Lastly, this calculation does not account for return migration. A person who resides in r in both 2005 and 2010 may have been absent for a portion of those five years, while this calculation assumes that the person has been present throughout.

Table 1 summarizes the three calculations of SRs. The first is based on *current residence* for each region in 2010. The second is based on *hukou registration*, which assumes that each migrant returns to their place of origin. The third is based on *the timing of migration*, which considers the duration of migration and calculates SR for each prefecture in each year. The second and third sets of SRs come from the 2010 *Micro Census* and are used for individual analyses.

5.1.3 Analytic strategy

The baseline regression equation is as follows:

$$Migrant_{i,h} = \alpha_0 + \alpha_1 \times Sex\_Ratio_h + X \times \beta + \epsilon_{i,h} \tag{4}$$

The regression considers $Migrant_{i,h}$, a binary variable equaling 1 if individual i, who originates from prefecture h, qualifies as a migrant. This occurs when h differs from where they currently reside. The vector X contains personal characteristics, including age, gender, ethnic group, hukou status, and educational level. To account for regional variations, the model further controls for a set of regional characteristics specific to the home prefecture h. Standard errors allow for clustering within home prefectures for all regressions.

In order to address any potential influence of unobservable regional characteristics or variations in population migration patterns across different years, I incorporate additional controls including province fixed effects, cohort fixed effects, or an interaction between the two. This aims to capture and account for any unique attributes or dynamics that may exist within specific provinces or over different time periods. The above equation is therefore estimated by a Linear Probability Model (LPM), as it is particularly advantageous when dealing with high-dimensional fixed effects.

[4] This issue may not be overly serious. While it may be tempting to assume that many migrants are highly mobile, studies have actually revealed that most migrants in China tend to remain relatively stable. For instance, Zhai, Duan and Bi (2007) finds that the average number of cities a migrant have lived in is only 1.56. Duan, Lv and Zou (2013) shows that 70.3% of migrants have only migrated to one single city, and 18.32% of migrants have changed their destination only once. Those who have moved between three or more cities account for only 3.77% of all migrants. These findings indicate that the majority of migrants have clear and stable destinations, debunking the notion that they are restless individuals continuously wandering between regions.

The explanatory variable is the SR of each individual's place of origin, calculated using the similar way as equation 2. I utilize two sets of SRs: one based on hukou registration and the other based on the timing of migration, to exclude the impact of changes in regional SRs caused by migration. Similar to the aggregate analysis, SRs of the population aged 15-50 serves as the primary explanatory variable.

5.2 Results: SRs and migration decisions

5.2.1 Main results

Figure 1 illustrates a scatter plot showcasing the changes in regional SRs for migrants before and after migration. The calculation of SRs in this case is based on the timing of migration. The X-axis indicates the age for each migrant, while the Y-axis represents the variation in regional SRs, specifically, the deviation between the SRs in the place of origin and the destination. A significant majority of migrants (67.48%) relocated from a prefecture with a higher SR to one with a lower SR.

Table 3 presents the regression results. In all cases, the coefficients of SRs on migration exhibit positive and statistically significant values, indicating a strong relationship. Those in column (4) are considered the preferred specifications since they incorporate the full set of controls and fixed effects.

Panel A of the table presents regressions using SRs based on hukou registration as the explanatory variable. Panel B and C focus on SRs based on the timing of migration. In Panel C, long-term migrants who have been outside their prefecture of origin for six years or more (which takes about 32.28% of all migrants in 2010) are excluded from the analysis due to the challenge of accurately determining the SR at the time of their migration. Despite using different measures of SR, the results remain consistent and robust. While there are slight variations in the estimated coefficients across these three panels, the overall findings align. The stability and similarity of these coefficients across different measures suggest that any potential bias in the estimation is likely to be minimal.

To interpret the coefficients and carry out further analyses, my preferred approach is SRs based on hukou registration, as presented in Panel A of Table 3. This is because, firstly, compared to those based on the timing of migration, SRs based on hukou registration exhibit fewer extreme values and are closer to the actual SRs of local residents. It also allows for retaining the sample size without excluding long-term migrants, and the coefficient value falls within the medium range relative to those in Panel B and C. The results indicate that an increase of one unit in the SR of origin leads to a 0.42% increase in the likelihood of someone becoming a migrant, which corresponds to 4.36% of the sample mean. A one standard deviation increase in the SR (8.889) translates into a 3.73% increase in the likelihood of someone becoming a migrant.

5.2.2 Heterogeneity

While the main analyses capture the average effect of SRs on migration, one may wonder how it may differ. Because, though the repercussions of SR imbalance can be detrimental for all local residents, the nature and extent of these effects can vary. As a result, different subgroups within the same prefecture may respond differently. One crucial aspect to explore is the gender differences in these effects. While the cause of SR imbalance is often - and rightfully - attributed to deeply ingrained sexism and male preference, empirical evidence have suggested that it is men, particular those of lower socioeconomic status, that bear the brunt of the negative effects of SR imbalance (Jiang and Sánchez-Barricarte, 2012; Hesketh and Xing, 2006). This is especially evident in the issue of marriage squeeze, where Chinese men, particularly those with a rural background or lower education levels, face increasing difficulties in finding marriage partners. In a society where marriage holds high value and is almost universally pursued, the absence of a spouse is regarded as a significant personal and social issue (Ji, 2015), therefore exposing unmarried men to serious psychological and physical risks.

Table 4 presents regression results examining the differential effect of SRs based on gender and socioeconomic status. The overall effect of SRs on migration is greater for males than females. As shown in column (1) of the table, a one unit increase in the local SR leads to a 0.44% increase in the likelihood of males becoming migrants, while for females, a one unit increase in the local SR leads to a 0.40% increase in the likelihood of becoming migrants. In addition, the results in columns (2)-(3) focus on individuals with lower socioeconomic status, specifically those with either a rural hukou or an educational level of high school or below. Among this group, men experience a more pronounced negative impact from SR imbalance, leading to an even stronger driving force for them to leave. However, among individuals with higher socioeconomic status, as indicated in columns (4)-(5), it is females who display a higher intolerance towards high SR environments and are more likely to leave.

Furthermore, I investigate the variations in the impact of SRs on migration across different regions. The findings reveal that these effects are particularly pronounced in less developed regions characterized by limited development and economic opportunities, such as those that exhibit lower levels of urbanization, lower average wages, lower per capita GDP, and lower average years of education. These results are presented in Appendix A.

5.3 Robustness checks

I hereby perform a series of tests for robustness that encompass: (1) the use of exogenous variations in SRs as instruments, (2) the substitution of 2015 1% Population Sample Survey as an alternate sample, (3) the introduction of other SRs as explanatory variables, (4) different definition of migrants as individuals who have migrated for over six months or a year, (5) the inclusion of a broader range of covariates and fixed effects, and (6) the use of counties as the basis of analyses.

An instrumental variable approach I leverage on regional variations in the effect of China's One-Child Policy as instruments for regional SRs, following Cameron, Meng and Zhang (2019). Appendix C illustrates the details of this approach. These analyses yield results highly consistent with the main regressions.

An alternative sample I use the micro data set from China's 1% Population Sample Survey in 2015 as a supplementary sample [5] and conduct two sets of checks. In the first set, I follow the same methods as the main analysis, using the 2015 sample to calculate SR based on hukou registration and examining its impact on each individual's migrant status as of 2015. For the second set, I use the SR of local residents in 2010 as an explanatory variable, exclude individuals who had migrated for 5 years or more from the 2015 sample, and examine the impact of the SR in 2010 on whether an individual migrated between 2010 and 2015. The results remain robust.

Additional SRs I consider SRs of other age-groups as explanatory variables, including those of the total population, 15-60 years old, 20-50 years old, and SRs of those who have never been married. The results are highly consistent with the main findings. Furthermore, I calculate SRs of local residents for prefectures in 2000, gathered from the 5^{th} national Population Census in that year. Using the 2010 *Micro Census* as the main sample, I use the SR of local residents in 2000 to predict migration between 2004 and 2010 . All results remain robust.

Other definitions of migrants The main analyses define migrants based on a discrepancy between their current place of residence and place of hukou registration. However, some studies only consider people as migrants if they have been away from their place of origin for a certain duration. In these checks, I adopt two stricter definitions of migrants, excluding short-term migrants who have migrated for less than 6 months or less than a year. The impact of SR on migration remains consistent and strong.

[5] The 1% Population Sample Survey is often referred to as a Mini-Census, where 1% of the population across the country was selected for investigation. From this selected population, a micro data set was created by randomly choosing 10% of the samples. This micro data set include basic demographic information of over 1.2 million individuals in 2015 and has a data structure almost identical to the *Micro Census*.

More covariates and fixed effects I include more covariates and fixed effects that could potentially confound the findings. I also consider SRs at birth as a control for male preference across different regions. They do not change the results. Moreover, while this study examines the factors that potentially push individuals away from their origin, it may be important to consider the role the destination plays in attracting people to migrate. To address this, I include additional controls related to the characteristics of the current place of residence. Whether controlling for a range of characteristic of the current residence or directly including prefecture-of-residence fixed effects, the results remain highly robust.

Counties as the basis of analyses Lastly, I shift the analytical focus from prefectures to counties and carry out tests following the same empirical framework. In the main regressions, migrants are refined to inter-prefecture ones, and SRs are calculated at the prefecture level. I further narrow the analysis to counties, treating individuals who cross county boundaries as migrants and using SR of the county of origin as the explanatory variable. The results remain robust [6].

6 Implications for family migration and children left-behind

Previous analyses focus on whether each *individual* is more likely to leave under high SRs, while it's important to note that the decision to move is often a *family* decision rather than an individual one (Mincer, 1978; Dustmann et al., 2023). One phenomenon accompanied by China's great migration is the huge amount of children left-behind. When men and women - mostly from rural areas - swarm into big cities seeking better lives and opportunities, many of their families are left behind. A considerable number of these families are children who must navigate their upbringing without the presence or support of their migrating parents. According to UNICEF [7] , in 2020, China has a staggering 66.93 million left-behind children, comprising 22.5% of its total child population. The absence of parents in these children's lives gives rise to a range of challenges concerning their personal safety, education, and overall physical and psychological development (Li, Liu and Zang, 2015; Lu, 2012). Gaining insights into the factors that influence parents' decisions to migrate with their children is crucial for understanding the complexities surrounding the movement of entire families and its implications for the well-being of left-behind children (Gao et al., 2023; Huang, Jiang and Sun, 2024).

This section delves deeper into the effects of SRs on family migration decisions. In particular, I investigate whether parents migrating from regions characterized by high SRs are more likely to relocate with their children instead of leaving them behind.

[6] More information regarding the data used in these analyses can be found in Appendix D.

[7] See https://www.unicef.cn/media/24496/file/.

6.1 Data and method

I use data from the 2017 *China Migrant Dynamics Survey* (CMDS) as the primary sample. The CMDS is a comprehensive nationwide survey on migrants and their family members conducted by the National Health Commission of China. I focus on children within migrant families, examining whether they relocate with their migrant parents or remain in their original area as left-behind children. The dependent variable, whether the child moves along, is coded as 1 if the children currently reside with their parents and 0 if they do not.

The explanatory variable is the SR of each child's prefecture of origin in 2010, obtained from the 2010 *Tabulation by County*. I use the hukou registration of the head of the household to determine the prefecture of origin for each child and match it with the census data. As in the main analyses, I focus specifically on migration across prefectures, intra-prefecture migrants are therefore excluded from the sample. Additionally, children who migrated before 2010 are excluded to study the impact of the 2010 SRs on the migration of migrant children between 2010 and 2017. The sample is further limited to children under 18 years old. The basic regression model aligns with Equation 4, where the dependent variable is replaced by ”whether the child moves along.” The regression controls for various child, parental, family, regional characteristics and fixed effects.

6.2 Results: SR and children left-behind

The results are presented in Table 5. On average, 57.9% of children migrate with their parents, with slightly higher migration rates observed for boys compared to girls. The findings are consistent with previous findings, where a one unit increase in the local SR leads to a 0.31% increase in the likelihood of children moving along, indicating that SR imbalance not only push individuals to leave but also affect migrant parents' decisions on family relocation. Furthermore, the heterogeneity observed in this effect aligns with the heterogeneity found in the individual analyses, where SRs have a more pronounced impact on pushing the migration of boys. For girls, the estimated coefficient of SR remain positive but no longer statistically significant.

6.3 Robustness checks

I conduct a series of robustness checks that follow a similar approach to the individual analyses, and all of them yield robust and consistent results. The checks include:

Additional SRs I use different SR for different age groups, SR of those who have never married, and SR based on hukou registration as the explanatory variable. The results remain robust.

Expanded covariates, fixed effects, and other factors I consider additional covariates and fixed effects that account for public services, educational resources, and other relevant factors in both the prefecture of origin and destination. These adjustments are made to address the parents' motivation to bring their children along in pursuit of improved education opportunities. Furthermore, I take into account any hardships or difficulties faced by the migrant family in their place of origin, such as economic challenges or the absence of suitable childcare options if the child was left behind. The results remain robust.

Counties as the basis of analyses I also shift the analytical focus from prefectures to counties, thereby including cross-county migrant families in the analyses. SRs of the county of origin serve as the explanatory variable. The results remain robust.

Families as the unit of analyses In addition to treating each child as a unit of analysis, I perform an additional set of checks where each migrant family is treated as a unit. The dependent variable is adjusted to assess whether the family brought at least one child when moving or left all children behind. The results remain robust.

7 Possible mechanism

In this section, I explore the potential mechanism by which imbalanced SRs can lead to migration. I propose that SR should be considered as a *predisposing* driver of migration (Van Hear, Bakewell and Long, 2018) as they create a structural context that increases the likelihood of migration. More specifically, I posit that high SRs create a hostile environment that pushes people to leave.

This perspective builds upon a crucial assumption that individuals are consciously aware of their surroundings. If individuals lack awareness of the SRs in their environment, they are less likely to consider leaving as a means to escape this unfavorable situation. To address this concern, I analyze how the ability to perceive the surrounding environment affects the relationship between SRs and migration decisions.

In China, there is considerable variation in population density across different areas. Urban agglomerations, for example, can have population densities exceeding 2,000 people per square kilometer, while some regions remain extremely sparsely populated, with only a few individuals residing within square kilometers and vast stretches of uninhabited land for miles. If SR imbalance truly motivates local residents to leave by creating a hostile environment, it may not necessarily have an impact on residents in these sparsely populated areas, regardless of how severe the imbalance may be, as individuals in such areas might not be able to detect the demographic structure surrounding them.

I divide all prefectures into three groups based on their population density: low-density, medium-density, and high-density groups, and perform the analyses separately for each group. The results are presented in Table 6. Panel A of the table represents the aggregate analysis, with prefectures as the unit of analysis and the net out-migration ratio of each prefecture as the dependent variable. Panel B represents the individual analysis, with the migration status of each person as the dependent variable. Columns (1)-(3) of the table present the results for the total population. In regions with low population density, SRs have no discernible impact on migration. However, in regions with medium and high population density - where local residents are more likely to directly experience and feel the effects of high SRs, the impact of SRs on population outflow becomes significantly positive. Columns (4)-(6) of the table present the effects for males, while columns (7)-(9) focus on females. The results are consistent with previous findings: the impact of SR is larger for males than females, and is not observed in areas with low population density.

However, living in a densely populated area does not necessarily guarantee a comprehensive awareness of the overall population structure within that environment. After all, individuals living in the same area may perceive local SRs differently due to variations in their daily activities and social interactions. Even in places with severe SR imbalances, it is possible to identify specific locations, industries that are predominantly occupied by women for instance, where females are more concentrated. While high population density allows for people to perceive their surroundings, it is crucial to go beyond that and assess whether there is a high SR and the severity of such imbalance within that environment.

Therefore, in addition to considering population density, I examine the severity of SR imbalance that individuals may detect, particularly from the perspective of the workplace environment they are engaged in. The 2010 Census collected information on the industry of employment for its participants, categorizing them into 19 groups according to the *Industrial Classification for National Economic Activities* (*GB/T 4754*). I calculate the SR for each industry, group the sample based on industry and population density, running separate regressions for each group. The rationale behind this approach is that population density is directly linked to an individual's likelihood of being able to perceive their surroundings, and workplace dynamics play a role in shaping individuals' perceptions and what they actually become aware of. For instance, if someone works in a predominantly male industry, they may be more aware of the hostile environment created by high SRs and experience a greater impact from it.

I depict the coefficients estimated by this into a scatter plot shown by Figure 2. In the plot, each scatter point represents an α_1 coefficient from equation 4, which signifies the impact of SRs on individuals leaving. The X axis represents the industry-specific SR in which the individuals work. The size of each point corresponds to the number of individuals in that particular group. Industries with the highest SRs are building and construction, transportation, and mining. Industries such as health and social security, and education tend to have lower SRs.

The findings are twofold. Firstly, the impact of SR is more pronounced in areas with higher population density regardless of industries, while generally no discernible effect of SR is observed in low-density regions. More importantly, in densely populated areas, a positive relationship is observed between the size of the coefficient and the SR of each industry. Specifically, industries with higher SRs also exhibit larger coefficients, indicating a stronger impact of SR imbalance on migration. These results support the hypothesis that the influence of SR imbalance on individuals leaving becomes apparent only when they are aware of the pressures in their immediate environment - and the more severe the imbalance, the more evident this effect manifests itself.

8 Conclusion

In this paper I investigate how regional sex ratio imbalances drive internal migration. China serves as a valuable case study to explore the broader implications of imbalanced sex ratios and the underlying dynamics of migration decisions. The findings are clear: high sex ratios push people away. This effect is particularly prominent among males from lower socioeconomic backgrounds and females from higher socioeconomic backgrounds. Moreover, sex ratios not only shape individual migration decisions but also influence the choices made by migrant parents regarding whether or not to bring their children along. I demonstrate that imbalanced sex ratios create an unfavorable environment for all locals. The more severe the imbalance, the stronger it pushes people away.

These findings offer insights into the complex nature of human migration. They allow for a more comprehensive understanding of the broader context that influence migration patterns. Much of the literature examining the regional characteristics in migration focus on more direct factors that may induce or prevent migration, such as wages, employment, and housing market (Dustmann, 2003; Zabel, 2012), war and violence (Buggle et al., 2023; Padilla-Romo and Peluffo, 2023), natural disasters and extreme weather events (Deschenes and Moretti, 2009), migration policies (Gao et al., 2023), among others. This paper adds to the body of knowledge and shows that alongside these direct triggers, broader contextual factors such as population sex ratios play a crucial role in driving migration. I argue that sex ratio imbalances create a hostile environment that, once becomes detectable, increases the likelihood of people leaving.

The findings also suggest that imbalanced sex ratios have the potential to self-correct through the course of internal migration. Regions with an excess of males experience a higher rate of male out-migration, which may eventually contribute to a gradual balance of regional sex ratios over time. This pattern is supported by Xiong (2022*b*), who suggests that regions boasting a remarkably high sex ratio at birth encounter a more substantial decline in the sex ratio of marriageable population 20 years later. This paper leans heavily on Xiong (2022*b*)'s findings and identifies the complementary mechanism that underlies this self-balancing dynamic. These findings are consistent with the homeostatic principle in demography, which asserts that the population system has the ability to adjust and stabilize itself to maintain long-term equilibrium (Hansson and Stuart, 1990; Lee, 1987, 2002).

This paper aligns closely with existing literature that highlights the extensive negative consequences associated with male-skewed sex ratio imbalances. They indicate that, while traditional notions of male dominance and preference may seem advantageous to males initially, the long-term effects of male-preference can create an environment in which males bear significant burdens. High sex ratios create intense competition for partners, increased social unrest, and higher rates of undesirable behaviors like crime and violence (Cameron, Meng and Zhang, 2019; Edlund et al., 2013; Jin et al., 2013). These imbalances do not position females favorably either. The scarcity of females can lead to increased marital and reproductive pressures, greater vulnerability to gender-based violence, and limited access to resources and opportunities (Edlund, 1999; D'Alessio and Stolzenberg, 2010). These findings present a crucial implication that, over time, a society that perpetuates male-preference and sexism fails to benefit anyone, regardless of their gender.

References

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos. 2011. "Marrying up: the role of sex ratio in assortative matching." *American Economic Journal: Applied Economics*, 3(3): 124–57.

Angrist, Josh. 2002. "How do sex ratios affect marriage and labor markets? Evidence from America's second generation." *The Quarterly Journal of Economics*, 117(3): 997–1038.

Bhaskar, Venkataraman. 2011. "Sex selection and gender balance." *American Economic Journal: Microeconomics*, 3(1): 214–44.

Brainerd, Elizabeth. 2017. "The lasting effect of sex ratio imbalance on marriage and family: Evidence from World War II in Russia." *Review of Economics and Statistics*, 99(2): 229–242.

Buggle, Johannes, Thierry Mayer, Seyhun Orcan Sakalli, and Mathias Thoenig. 2023. "The Refugee's Dilemma: Evidence from Jewish Migration out of Nazi Germany." *The Quarterly Journal of Economics*, 138(2): 1273–1345.

Cameron, Lisa, Xin Meng, and Dandan Zhang. 2019. "China's sex ratio and crime: Behavioural change or financial necessity?" *The Economic Journal*, 129(618): 790–820.

Chang, Simon, Kamhon Kan, and Xiaobo Zhang. 2021. "Too many men, too-short lives: The effect of the male-biased sex ratio on mortality." *Journal of Human Resources*.

Chen, Shuai, Paulina Oliva, and Peng Zhang. 2022. "The effect of air pollution on migration: evidence from China." *Journal of Development Economics*, 156: 102833.

Chen, Yuyu, Hongbin Li, and Lingsheng Meng. 2013. "Prenatal sex selection and missing girls in China: Evidence from the diffusion of diagnostic ultrasound." *Journal of Human Resources*, 48(1): 36–70.

Chu, Junhong. 2001. "Prenatal sex determination and sex-selective abortion in rural central China." *Population and Development Review*, 27(2): 259–281.

D'Alessio, Stewart J, and Lisa Stolzenberg. 2010. "The sex ratio and male-on-female intimate partner violence." *Journal of Criminal Justice*, 38(4): 555–561.

Deschenes, Olivier, and Enrico Moretti. 2009. "Extreme weather events, mortality, and migration." *The Review of Economics and Statistics*, 91(4): 659–681.

Dreze, Jean, and Reetika Khera. 2000. "Crime, gender, and society in India: Insights from homicide data." *Population and Development Review*, 26(2): 335–352.

Duan, Chengrong, Lidan Lv, and Xiangjiang Zou. 2013. "Major challenges for China's floating population and policy suggestions: An analysis of the 2010 population census data [in Chinese]." *Population Research [Renkou yanjiu]*, 37(2): 17–24.

Dustmann, Christian. 2003. "Return migration, wage differentials, and the optimal migration duration." *European Economic Review*, 47(2): 353–369.

Dustmann, Christian, Francesco Fasani, Xin Meng, and Luigi Minale. 2023. "Risk attitudes and household migration decisions." *Journal of Human Resources*, 58(1): 112–145.

Edlund, Lena. 1999. "Son preference, sex ratios, and marriage patterns." *Journal of Political Economy*, 107(6): 1275–1304.

Edlund, Lena, Hongbin Li, Junjian Yi, and Junsen Zhang. 2013. "Sex ratios and crime: Evidence from China." *Review of Economics and Statistics*, 95(5): 1520–1534.

Eggleston, Karen, Jean C Oi, Scott Rozelle, Ang Sun, Andrew Walder, and Xueguang Zhou. 2013. "Will demographic change slow China's rise?" *Journal of Asian Studies*, 72(3): 505–518.

Feng, Shuaizhang, Alan B Krueger, and Michael Oppenheimer. 2010. "Linkages among climate change, crop yields and Mexico–US cross-border migration." *Proceedings of the National Academy of Sciences*, 107(32): 14257–14262.

Gao, Xuwen, Wenquan Liang, Ahmed Mushfiq Mobarak, and Ran Song. 2023. "Migration Restrictions Can Create Gender Inequality: The Story of China's Left-Behind Children." National Bureau of Economic Research.

Grosjean, Pauline, and Rose Khattar. 2019. "It's raining men! Hallelujah? The long-run consequences of male-biased sex ratios." *The Review of Economic Studies*, 86(2): 723–754.

Gu, Baochang, Wang Feng, Guo Zhigang, and Zhang Erli. 2007. "China's local and national fertility policies at the end of the twentieth century." *Population and Development Review*, 129–147.

Guilmoto, Christophe Z. 2012. "Skewed sex ratios at birth and future marriage squeeze in China and India, 2005–2100." *Demography*, 49(1): 77–100.

Guo, Rufei, Junsen Zhang, and Minghai Zhou. 2023. "The demography of the great migration in China." *Journal of Development Economics*, 103235.

Hansson, Ingemar, and Charles Stuart. 1990. "Malthusian selection of preferences." *The American Economic Review*, 529–544.

Hesketh, Therese, and Zhu Wei Xing. 2006. "Abnormal sex ratios in human populations: causes and consequences." *Proceedings of the National Academy of Sciences*, 103(36): 13271–13275.

Huang, Zibin, Xu Jiang, and Ang Sun. 2024. "Fertility and delayed migration: How son preference protects young girls against mother–child separation." *Journal of Development Economics*, 166: 103191.

Jiang, Quanbao, and Jesús J Sánchez-Barricarte. 2012. "Bride price in China: The obstacle to 'Bare Branches' seeking marriage." *The History of the Family*, 17(1): 2–15.

Jin, Xiaoyi, Lige Liu, Yan Li, Marcus W Feldman, and Shuzhuo Li. 2013. """ Bare branches" and the marriage market in rural China: Preliminary evidence from a village-level survey." *Chinese Sociological Review*, 46(1): 83–104.

Ji, Yingchun. 2015. "Between tradition and modernity:"Leftover" women in Shanghai." *Journal of Marriage and Family*, 77(5): 1057–1073.

Lafortune, Jeanne. 2013. "Making yourself attractive: Pre-marital investments and the returns to education in the marriage market." *American Economic Journal: Applied Economics*, 5(2): 151–178.

Lai, Wangyang, Hong Song, Chang Wang, and Huanhuan Wang. 2021. "Air pollution and brain drain: Evidence from college graduates in China." *China Economic Review*, 68: 101624.

Lee, Everett S. 1966. "A theory of migration." *Demography*, 3(1): 47–57.

Lee, Ronald. 1987. "Population dynamics of humans and other animals." *Demography*, 443–465.

Lee, Ronald. 2002. "The demographic transition: three centuries of fundamental change." *Journal of Economic Perspectives*, 17(4): 167–190.

Liang, Zai. 2016. "China's great migration and the prospects of a more integrated society." *Annual Review of Sociology*, 42: 451–471.

Li, Hongbin, Junjian Yi, and Junsen Zhang. 2011. "Estimating the effect of the one-child policy on the sex ratio imbalance in China: Identification based on the difference-in-differences." *Demography*, 48(4): 1535–1557.

Lin, Justin Yifu. 1988. "The household responsibility system in China's agricultural reform: a theoretical and empirical study." *Economic Development and Cultural Change*, 36(S3): S199–S224.

Li, Qiang, Gordon Liu, and Wenbin Zang. 2015. "The health of left-behind children in rural China." *China Economic Review*, 36: 367–376.

Liu, Ye, and Wei Xu. 2017. "Destination choices of permanent and temporary migrants in China, 1985–2005." *Population, Space and Place*, 23(1): e1963.

Li, Wenchao, Changcheng Song, Shu Xu, and Junjian Yi. 2022. "High sex ratios and household portfolio choice in China." *Journal of Human Resources*, 57(2): 465–490.

Lu, Yao. 2012. "Education of children left behind in rural China." *Journal of Marriage and Family*, 74(2): 328–341.

McElroy, Marjorie, and Dennis Tao Yang. 2000. "Carrots and sticks: fertility effects of China's population policies." *American Economic Review*, 90(2): 389–392.

Messner, Steven F, and Robert J Sampson. 1991. "The sex ratio, family disruption, and rates of violent crime: The paradox of demographic structure." *Social Forces*, 69(3): 693–713.

Mincer, Jacob. 1978. "Family migration decisions." *Journal of Political Economy*, 86(5): 749–773.

NBS. 2023. "How to differentiate between permanent population and floating population?" http://www.stats.gov.cn/zs/tjws/tjbz/202301/t20230101_1903796.html, Last accessed 29/10/2023.

Padilla-Romo, María, and Cecilia Peluffo. 2023. "Violence-induced migration and peer effects in academic performance." *Journal of Public Economics*, 217: 104778.

Qian, Nancy. 2008. "Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance." *The Quarterly Journal of Economics*, 123(3): 1251–1285.

Sen, Amartya. 1992. "Missing women." *BMJ: British Medical Journal*, 304(6827): 587.

Siegel, Jacob S, and C Horace Hamilton. 1952. "Some considerations in the use of the residual method of estimating net migration." *Journal of the American Statistical Association*, 47(259): 475–500.

Van Hear, Nicholas, Oliver Bakewell, and Katy Long. 2018. "Push-pull plus: reconsidering the drivers of migration." *Journal of Ethnic and Migration Studies*, 44(6): 927–944.

Wei, Shang-Jin, and Xiaobo Zhang. 2011*a*. "The competitive saving motive: Evidence from rising sex ratios and savings rates in China." *Journal of Political Economy*, 119(3): 511–564.

Wei, Shang-Jin, and Xiaobo Zhang. 2011*b*. "Sex ratios, entrepreneurship, and economic growth in the People's Republic of China." National Bureau of Economic Research.

Xiong, Wanru. 2022*a*. "Does the shortage of marriageable women induce the trafficking of women for forced marriage? Evidence from China." *Violence Against Women*, 28(6-7): 1441–1463.

Xiong, Wanru. 2022*b*. "Dynamics between Regional Sex Ratios at Birth and Sex Ratios at Prime Marriageable Ages in China." *Population and Development Review*, 48(2): 545–578.

Xiong, Wanru. 2023. "Love is Elsewhere: Internal Migration and Marriage Prospects in China." *European Journal of Population*, 39(1): 6.

Xue, Melanie Meng. 2016. "High-value work and the rise of women: The cotton revolution and gender equality in China." *Available at SSRN 2389218*.

Yang, Dennis Tao, and Dandan Chen. 2004. "Transformations in China's population policies and demographic structure." *Pacific Economic Review*, 9(3): 269–290.

Zabel, Jeffrey E. 2012. "Migration, housing market, and labor market responses to employment shocks." *Journal of Urban Economics*, 72(2-3): 267–284.

Zhai, Zhenwu, Chengrong Duan, and Qiuling Bi. 2007. "The floating population in Beijing: an update [in Chinese]." *Population Research* [*Renkou yanjiu*], 31(2): 30.

Zhao, Yaohui. 1999. "Leaving the countryside: rural-to-urban migration decisions in China." *American Economic Review*, 89(2): 281–286.

Table 1: Summary of prefecture-level SRs

Variable	Obs.	Mean	Std. dev.	Min	Max	Source of data	What they represent	Where they are used
Panel A: SRs based on current residence								
Total sex ratio	275	105.145	4.12	89.651	132.578	2010	SR of local residents	Aggregate
Sex ratio of [15,60)	275	103.732	5.906	85.916	140.335	*Tabulation*	for regions in 2010	analyses
Sex ratio of [15,50)	275	103.853	5.365	86.486	139.048	*by County*		
Panel B: SRs based on hukou registration								
Total sex ratio	275	105.022	8.451	68.182	150.980	2010	SR if all migrants return	Individual
Sex ratio of [15,60)	275	103.970	8.659	57.895	156.250	*Micro Census*	to their place of origin	analyses
Sex ratio of [15,50)	275	103.895	8.889	62.50	154.167			
Panel C: SRs based on the timing of migration								
Total sex ratio	1,960	105.018	10.177	20	200	2010	SR that each migrant	Individual
Sex ratio of [15,60)	1,960	103.915	10.719	20	200	*Micro Census*	faced when they left	analyses
Sex ratio of [15,50)	1,960	103.944	10.764	25	191.667			

Table 2: Aggregate evidence: SRs and out-migration

Dep. Var.	Net Out-migration Ratio 2010-2020					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Unit of analyses: Prefectures					
Sample:	*Total population*		*Male population*		*Female population*	
Sex ratio	0.0064***	0.0050**	0.0083***	0.0074***	0.0044***	0.0027
	(5.032)	(2.400)	(5.618)	(3.205)	(3.688)	(1.365)
Mean D.V.	0.0308	0.0329	0.0110	0.0119	0.0313	0.0342
Std. dev. D.V.	0.155	0.159	0.158	0.161	0.152	0.156
Obs.	275	271	275	271	275	271
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects						
Prov.	No	Yes	No	Yes	No	Yes
# Clusters	25	25	25	25	25	25
	Panel B: Unit of analyses: Counties					
Sample:	*Total population*		*Male population*		*Female population*	
Sex ratio	0.0042***	0.0042**	0.0065***	0.0075***	0.0023*	0.0015
	(3.657)	(2.487)	(7.140)	(5.363)	(1.924)	(0.856)
Mean D.V.	0.0945	0.0942	0.0666	0.0662	0.0983	0.0981
Std. dev. D.V.	0.195	0.196	0.201	0.202	0.194	0.194
Obs.	1,862	1,849	1,862	1,849	1,862	1,849
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects						
Prefec.	No	Yes	No	Yes	No	Yes
# Clusters	299	286	299	286	299	286

Notes: All regressions estimated using OLS. Robust t-statistics in parentheses. For regressions in Panel A, standard errors allow for clustering within provinces; for regressions in Panel B, prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

Table 3: Individual evidence: SRs and migration decisions

Dep. Var.	Migrant: Yes or No			
	(1)	(2)	(3)	(4)
	Panel A: SRs based on hukou registr.			
Sex ratio	0.0072***	0.0082***	0.0042***	0.0042***
	(4.101)	(3.303)	(3.135)	(3.280)
Mean D.V.	0.0964	0.0964	0.0964	0.0964
Std. dev. D.V.	0.295	0.295	0.295	0.295
Obs.	3,655,894	3,655,894	3,655,894	3,655,894
	Panel B: SRs based on the timing of migr.			
Sex ratio	0.0070***	0.0082***	0.0056***	0.0046***
	(3.897)	(3.401)	(4.848)	(4.042)
Mean D.V.	0.0964	0.0964	0.0964	0.0964
Std. dev. D.V.	0.295	0.295	0.295	0.295
Obs.	3,655,894	3,655,894	3,655,894	3,655,894
	Panel C: SRs based on the timing of migr., w/o long-term migrants			
Sex ratio	0.0056***	0.0060***	0.0040***	0.0038***
	(3.698)	(2.896)	(3.488)	(3.339)
Mean D.V.	0.0770	0.0770	0.0770	0.0770
Std. dev. D.V.	0.267	0.267	0.267	0.267
Obs.	3,542,609	3,542,609	3,542,609	3,542,609
Controls	Yes	Yes	Yes	Yes
Fixed effects				
Prov.	No	Yes	Yes	absorbed
Cohort	No	No	Yes	absorbed
Prov. × Cohort	No	No	No	Yes
# Clusters	275	275	275	275

Notes: Robust t-statistics in parentheses. Standard errors allow for clustering within home prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

Table 4: Heterogeneity: by gender and SES

Dep. Var.	Migrant: Yes or No				
	(1)	(2)	(3)	(4)	(5)
		Low SES		*High SES*	
Sample:	*ALL*	*Rural hukou*	*High school & below*	*Urban hukou*	*College & above*
Sex ratio × Female	-0.0004***	-0.0003***	-0.0005***	-0.0001	0.0009***
	(-3.863)	(-2.911)	(-4.055)	(-0.410)	(2.720)
Sex ratio	0.0044***	0.0048***	0.0043***	0.0027***	0.0052***
	(3.356)	(3.339)	(3.393)	(2.809)	(3.220)
Mean D.V.	0.0964	0.105	0.0866	0.0732	0.179
Std. dev. D.V.	0.295	0.307	0.281	0.261	0.384
Obs.	3,655,894	2,764,417	3,460,096	1,110,358	414,678
Controls	Yes	Yes	Yes	Yes	Yes
Fixed effects					
Prov. × Cohort	Yes	Yes	Yes	Yes	Yes
# Clusters	277	277	277	277	277

Notes: Regression in Panel A are estimated using OLS. Regressions in Panel B are estimated using instrumental variables. Robust t-statistics in parentheses. Standard errors allow for clustering within home prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

Table 5: Implications for family relocation

Dep. Var.	Whether the child moves along: Yes or No					
	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	*All*		*Boys*		*Girls*	
Sex ratio	0.0030**	0.0031**	0.0042***	0.0044***	0.0016	0.0016
	(2.206)	(2.551)	(1.214)	(1.846)	(2.240)	(0.769)
Mean D.V.	0.579	0.579	0.581	0.581	0.578	0.578
Std. dev. D.V.	0.494	0.494	0.493	0.493	0.494	0.494
Obs.	83,647	83,629	45,662	45,635	37,985	37,964
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects						
Prov.	Yes	absorbed	Yes	absorbed	Yes	absorbed
Cohort	Yes	absorbed	Yes	absorbed	Yes	absorbed
Prov. $\times$ Cohort	No	Yes	No	Yes	No	Yes

Notes: Regressions in columns (1)-(3) are estimated using OLS. Regressions in columns (4)-(6) are estimated using instrumental variables. Robust t-statistics in parentheses. Standard errors allow for clustering within home prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

Table 6: By POP density & gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Aggregate level - Prefectures								
Dep. Var.	Net out-migration ratio 2010-2020								
Sample:	*Total population*			*Male population*			*Female population*		
POP density	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
Sex ratio	0.0013	0.0051	0.0077**	0.0030	0.0064*	0.0108**	-0.0008	0.0042	0.0049
	(0.666)	(1.467)	(2.236)	(1.487)	(1.845)	(2.885)	(-0.421)	(1.173)	(1.480)
Mean D.V.	0.103	0.0227	0.0214	0.0853	-0.00232	0.000569	0.0984	0.0269	0.0236
Std. dev. D.V.	0.109	0.135	0.172	0.108	0.140	0.174	0.111	0.128	0.170
Obs.	63	62	132	63	62	132	63	62	132
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects									
Prov.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Clusters	14	15	16	14	15	16	14	15	16
	Panel B: Individual level								
Dep. Var.	Migrant: Yes or No								
Sample:	*All*			*Males*			*Females*		
POP density	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
Sex ratio	0.0013	0.0057**	0.0057***	0.0011	0.0059**	0.0059***	0.0014	0.0054**	0.0055***
	(1.468)	(2.487)	(2.739)	(1.287)	(2.432)	(2.807)	(1.627)	(2.541)	(2.658)
Mean D.V.	0.0966	0.0988	0.0956	0.0989	0.105	0.103	0.0942	0.0929	0.0881
Std. dev. D.V.	0.295	0.298	0.294	0.299	0.306	0.304	0.292	0.290	0.283
Obs.	538,824	762,311	2,352,898	273,710	387,006	1,193,701	265,114	375,304	1,159,197
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects									
Prov. $\times$ Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Clusters	69	69	138	69	69	138	69	69	138

Notes: All regressions estimated using OLS. Robust t-statistics in parentheses. For regressions in Panel A, standard errors allow for clustering within provinces; for regressions in Panel B, prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

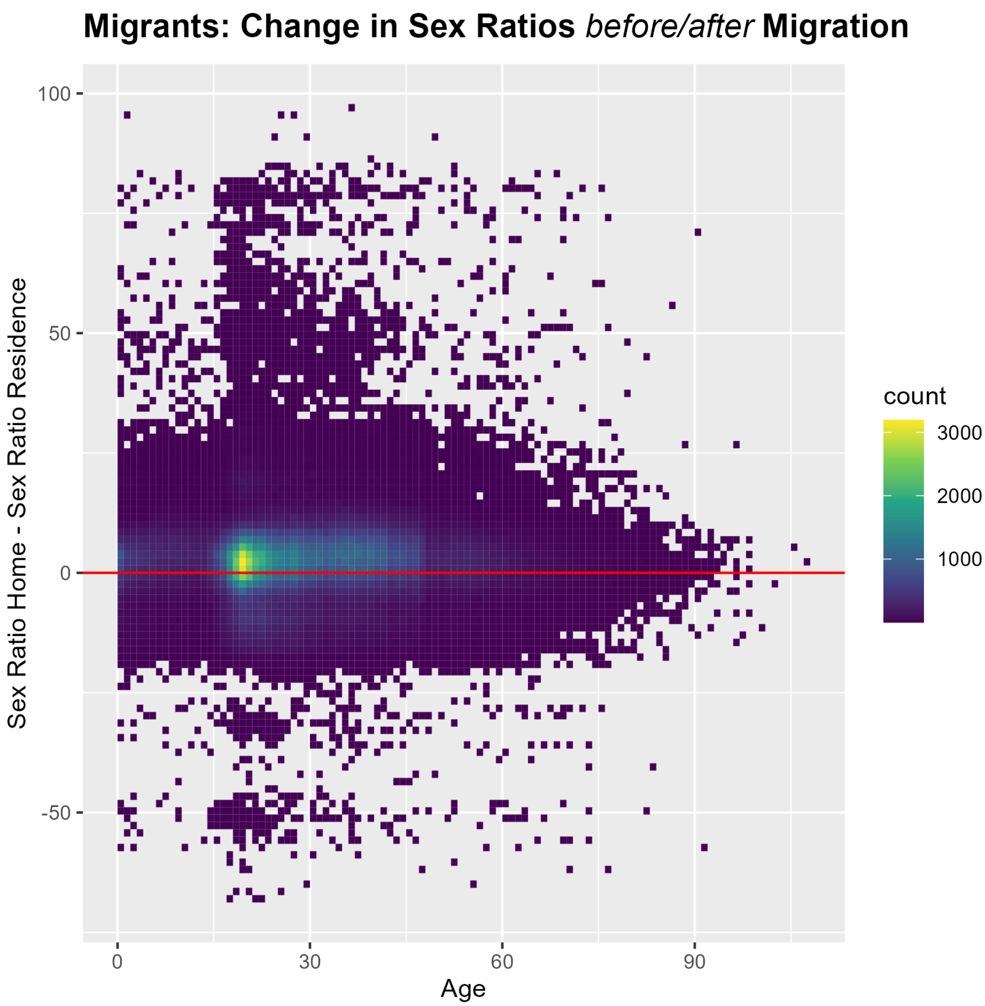

Figure 1: Changes in SRs for migrants before/after migration

Notes: This figure depicts a scatter plot of the variations in regional SRs for migrants. Each scatter point represents a migrant - defined as those who crossed prefectures. On the Y-axis is the difference between the SR of the place of origin (from which the migrant left) and the SR of the destination (to which the migrant traveled). The data used in this plot is sourced from the 2010 *Micro Census*. The different colors or heat displayed in the plot indicate the density of scatter points. More than two-thirds of cross-prefecture migrants migrated from areas with a higher SR to areas with a lower SR.

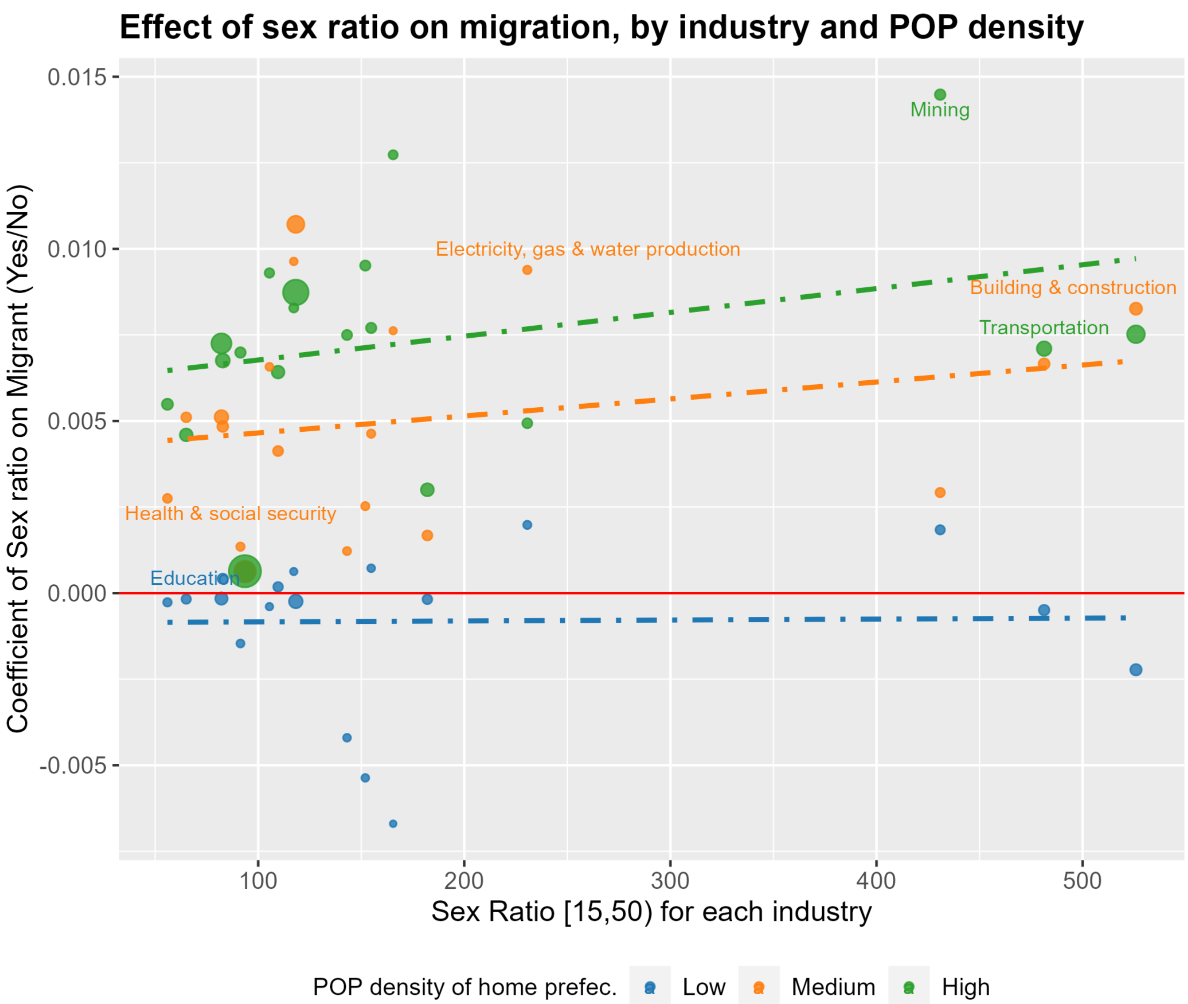

Figure 2: Regression coefficients of SRs on migration, by industry and population density

Notes: This figure depicts a scatter plot of the regression coefficients of SRs on migration. Each scatter point represents an α_1 coefficient from equation 4, which signifies the impact of SRs on individuals leaving, categorized by different industries and levels of population density. The X-axis represents the SR $[15,50)$ for each of 19 industries, classified according to the *Industrial Classification for National Economic Activities* $(GB/T\ 4754)$.

The industries are further differentiated by different colors, representing the population density of each prefecture. The data utilized in this plot is sourced from the 2010 *Micro Census*. In areas with low population density, the SR has minimal influence on migration, regardless of the SR within the respective industry. However, in regions with higher population density, the effect of SR on migration becomes apparent. Furthermore, this impact increases with higher SR within an industry.

A Additional tables and figures

Table A1: Summary of prefecture characteristics

Variable	Obs.	Mean	Std. dev.	Min	Max
	Panel A: Net out-migration ratio of [15,50)				
Total out-migration	275	0.041	0.153	-0.657	0.910
Male out-migration	275	0.021	0.154	-0.624	0.894
Feamle out-migration	275	0.040	0.152	-0.713	0.907
	Panel B: Other characteristics				
Total population	275	4,256,129	2,885,189	231,853	28,846,170
Population density	275	419.32	320.16	4.97	2,474.47
% migrants	275	0.18	0.13	0.04	0.82
% urban hukou	275	29.82	14.87	8.48	80.30
Birth rate	275	10.34	3.03	4.89	20.23
Natural growth rate	275	5.97	3.90	-3.31	21.66
GDP (10m, log)	275	6.66	0.89	4.48	9.62
GDP per capita (log)	275	10.02	0.71	4.60	11.81
Average wage (log)	275	10.20	0.23	9.45	11.06

Table A2: Summary of personal characteristics

Variable	Obs.	Mean	Std. dev.	Min	Max
	Panel A: Dependent variable				
Migrant across counties	3,655,894	0.129	0.335	0	1
Migrant across prefectures	3,655,894	0.096	0.295	0	1
Migrant across provinces	3,655,894	0.062	0.241	0	1
	Panel B: Non-migrants				
Age	3,303,576	38.366	18.862	6	110
Female	3,303,576	0.496	0.500	0	1
Ethnic minority	3,303,576	0.051	0.219	0	1
Rural hukou	3,303,576	0.712	0.453	0	1
Educational level	3,303,576	3.010	1.114	1	7
	Panel C: Migrants				
Age	352,318	30.713	12.900	6	106
Female	352,318	0.460	0.498	0	1
Ethnic minority	352,318	0.036	0.187	0	1
Rural hukou	352,318	0.786	0.410	0	1
Educational level	352,318	3.483	1.203	1	7

Table A3: Heterogeneity: by regional characteristics

Dep. Var.	Migrant: Yes (1) or No (0)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Panel A: OLS											
Sample:	*All*				*Males*				*Females*			
Var. interacted	*% urban hukou*	*Ave. wage (log)*	*GDP PC (log)*	*Ave. year of education*	*% urban hukou*	*Ave. wage (log)*	*GDP PC (log)*	*Ave. year of education*	*% urban hukou*	*Ave. wage (log)*	*GDP PC (log)*	*Ave. year of education*
Sex ratio × VAR	-0.0002***	-0.0077*	-0.0021**	-0.0017**	-0.0002***	-0.0080*	-0.0021**	-0.0017**	-0.0002***	-0.0075*	-0.0021**	-0.0016**
	(-3.096)	(-1.816)	(-2.513)	(-2.326)	(-3.072)	(-1.831)	(-2.435)	(-2.267)	(-3.104)	(-1.801)	(-2.582)	(-2.377)
Sex ratio	0.0086***	0.0827*	0.0253***	0.0184***	0.0088***	0.0853*	0.0252***	0.0188**	0.0084***	0.0798*	0.0253***	0.0179***
	(3.493)	(1.882)	(2.715)	(2.603)	(3.498)	(1.898)	(2.652)	(2.558)	(3.473)	(1.865)	(2.766)	(2.636)
	Panel B: Two-stage											
Sample:	*All*				*Males*				*Females*			
Var. interacted	*% urban hukou*	*Ave. wage (log)*	*GDP PC (log)*	*Ave. year of education*	*% urban hukou*	*Ave. wage (log)*	*GDP PC (log)*	*Ave. year of education*	*% urban hukou*	*Ave. wage (log)*	*GDP PC (log)*	*Ave. year of education*
Sex ratio × VAR	-0.0002**	-0.0080*	-0.0002	-0.0017**	-0.0002**	-0.0079*	-0.0001	-0.0018**	-0.0001**	-0.0081**	-0.0004	-0.0016**
	(-2.467)	(-1.891)	(-0.265)	(-2.370)	(-2.433)	(-1.787)	(-0.096)	(-2.326)	(-2.471)	(-1.991)	(-0.434)	(-2.369)
Sex ratio	0.0129***	0.0917**	0.0124	0.0168**	0.0126***	0.0902*	0.0107	0.0171**	0.0132***	0.0929**	0.0140*	0.0163**
	(3.572)	(2.056)	(1.417)	(2.370)	(3.384)	(1.939)	(1.166)	(2.267)	(3.733)	(2.168)	(1.668)	(2.439)
Mean D.V.	0.0964	0.0964	0.0964	0.0964	0.103	0.103	0.103	0.103	0.0900	0.0900	0.0900	0.0900
Std. dev. D.V.	0.295	0.295	0.295	0.295	0.303	0.303	0.303	0.303	0.286	0.286	0.286	0.286
Obs.	3,655,894	3,655,894	3,655,894	3,655,894	1,854,418	1,854,418	1,854,418	1,854,418	1,799,615	1,799,615	1,799,615	1,799,615
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects												
Prov. × Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Clusters	277	277	277	277	277	277	277	277	277	277	277	277

Notes: Regressions in Panel A are estimated using OLS. Regressions in Panel B are estimated using instrumental variables. Robust t-statistics in parentheses. Standard errors allow for clustering within prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level. VAR denotes the regional characteristic interacted with SR to examine how the main effects of SR on migration differ. This variable is presented at the head of the column in each Panel.

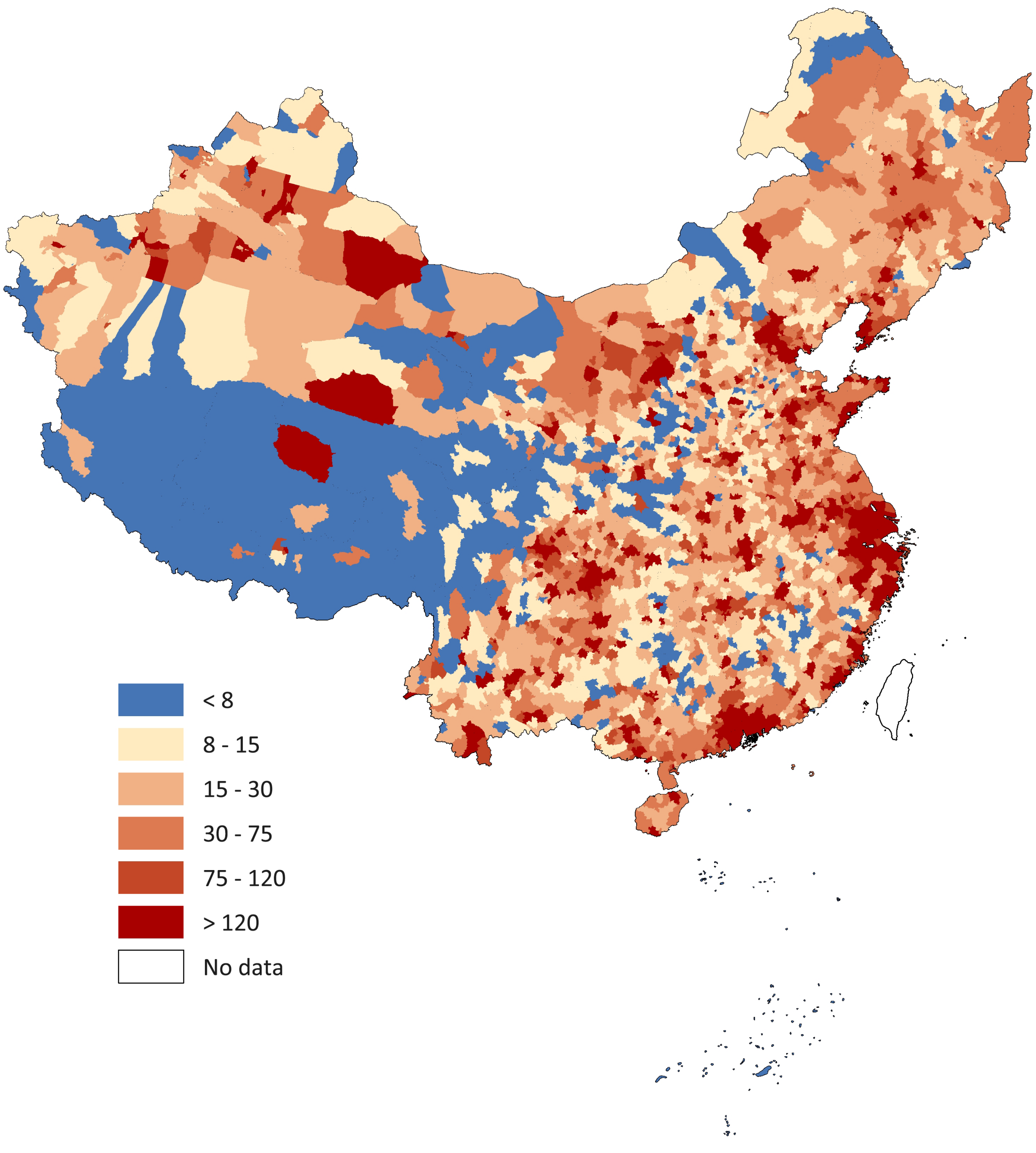

Figure A1: Number of migrants (Thousands) by county, 2020

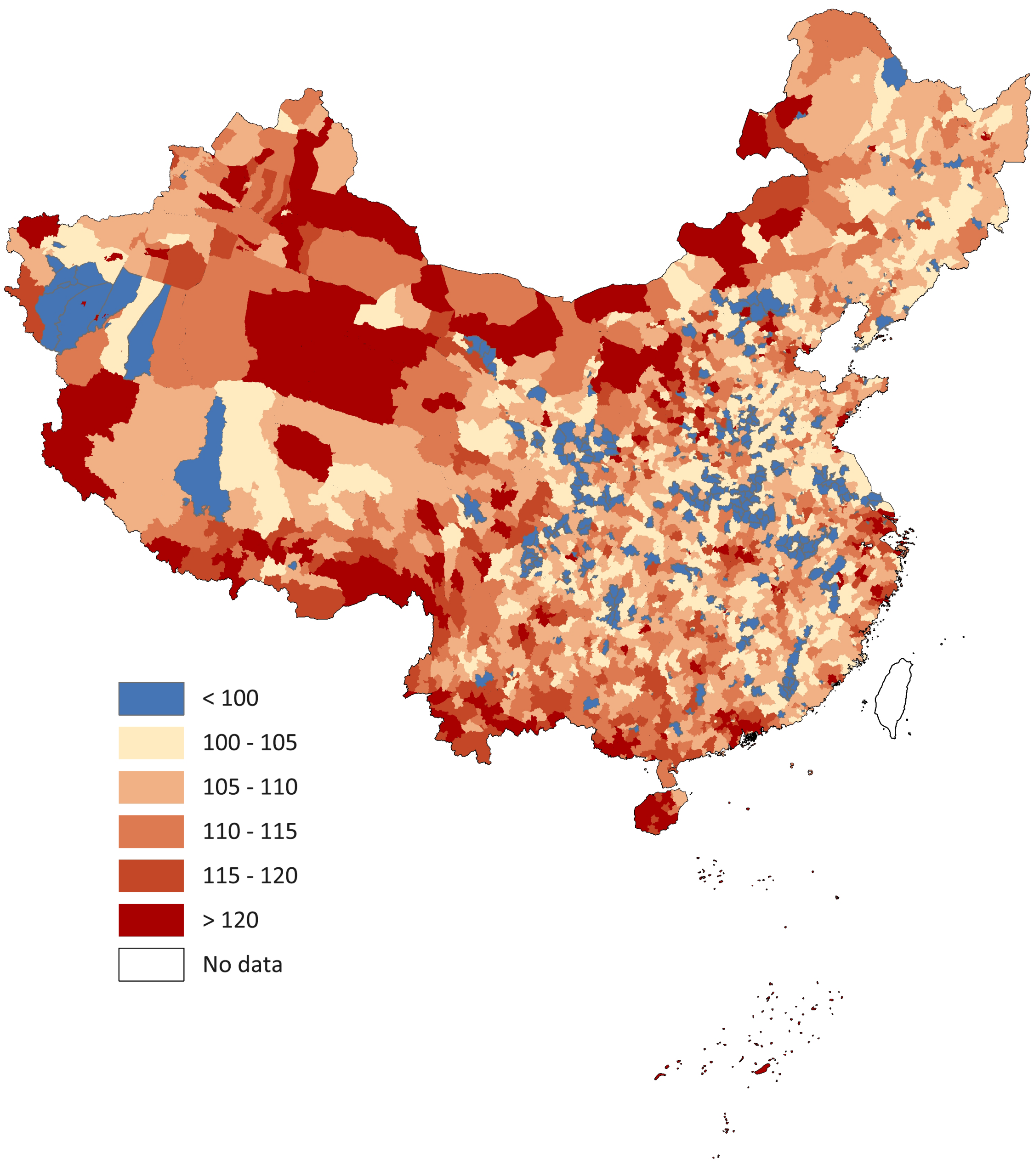

Figure A2: SR [15,50) (men per 100 women) by county, 2020

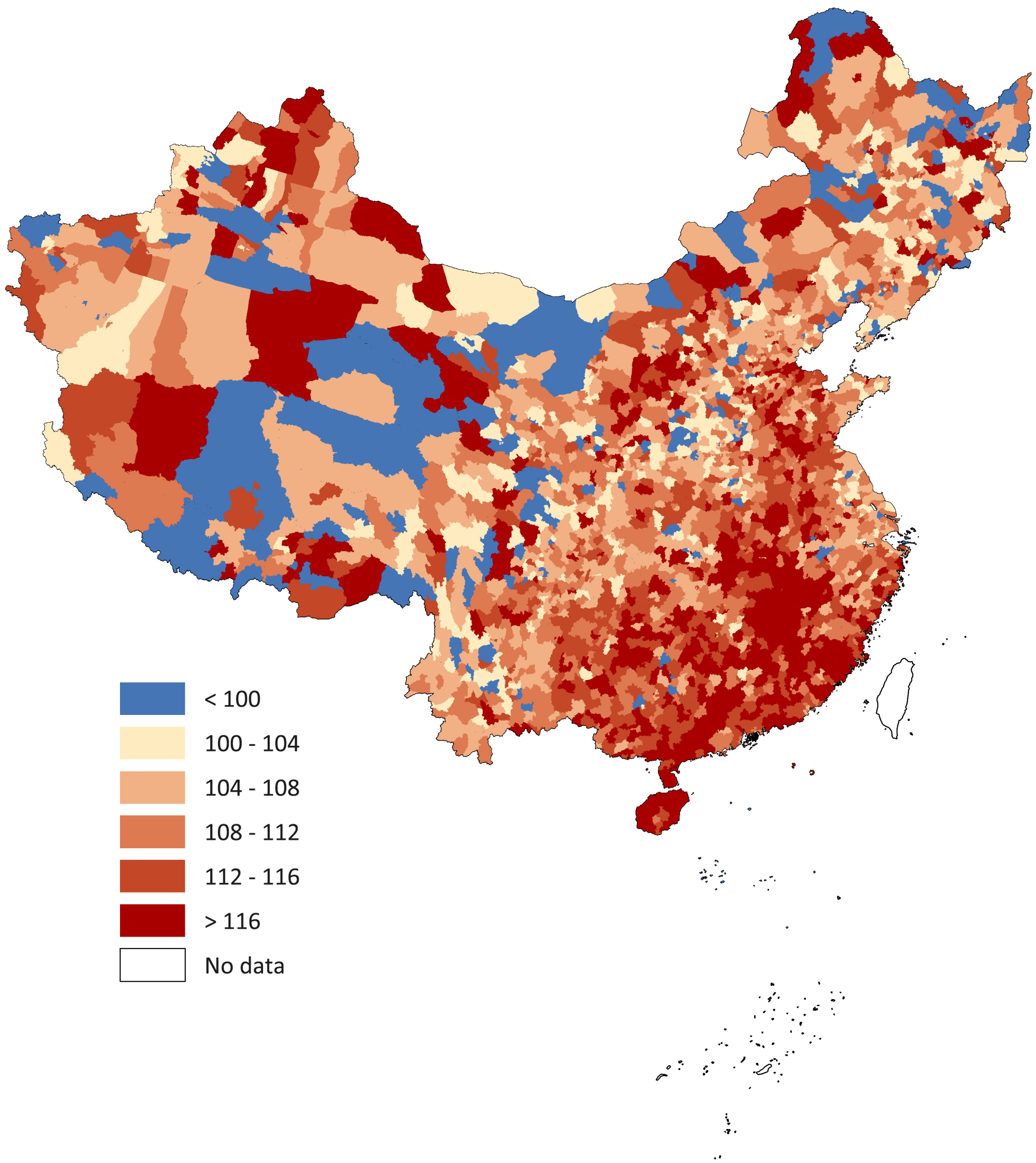

Figure A3: SR at birth (boys per 100 girls) by county, 2020

B Calculation of prefecture/county-level mortality and deaths, by gender and age group

To calculate the net out-migration ratio of each region, I performed the following calculations. First, I subtracted the total number of people aged 25-69 in a region in 2020 and the number of deaths among people in this age group in the region during 2010-2020 from the region's population aged 15-59 in 2010. I then divide this difference by the baseline population. This calculation compares the number of people of this age group who have out-migrated from the region during the decade to the sum of the number of people who have migrated to the region and the number of deaths. Additionally, I calculate the net out-migration ratio for males and females separately to examine the gender-specific impact of SR on migration.

To perform this calculation, I need the age-specific and gender-specific mortality rates for each region (prefecture and county) during the decade. Unfortunately, this data is not publicly available to the best of my knowledge. While some regions - which are few, do publish the number of local deaths by gender and age group after each Population Census (2010 and 2020), they do not publish these information in non-census years, which are the years between 2011 and 2019.

Therefore, I had to rely on other publicly available data. I used two sources: the *Tabulation on China Population Census by County* and the *China Population Census Yearbook* from each of the two censuses conducted in 2010 and 2020. The *Tabulation by County* provides total mortality rates for each county (cities and prefectures included) in each census year, while the *Statistical Yearbook* contains total mortality rates for each province, along with provincial-level death tolls by gender and age group. Here are the steps I followed:

1 I first calculated the total mortality rate and mortality rates by gender and age group for each province.

2 For each region (prefecture or county) in a province, I divided the region's total mortality rate by the province's total mortality rate. This ratio served as the weight to adjust the province's gender-specific and age-specific mortality rates, resulting in mortality rates by gender and age group specific to the region.

3 With the estimated mortality rate, I was able to gauge the number of deaths by gender and age group in each region for the years 2010 and 2020.

To illustrate with an example, consider age group $[g, g+5)$ of region r in province p in

2010. The estimated mortality rate (for both gender of this group) is calculated as:

$$Mortality_{[g,g+5),r,2010} = \frac{Mortality_{total,r,2010}}{Mortality_{total,p,2010}} \times Mortality_{[g,g+5),p,2010} \tag{B1}$$

where $Mortality_{total}$ denotes the total mortality rate of local population, which is publicly available in the *Tabulation by County*. Note that $Mortality_{[g,g+5),r,2010}$ represents the mortality rate of age group $[g, g+5)$ of region r only in year 2010. To calculate the total deaths among this cohort over a ten-year period for the net out-migration ratio, one can simply multiply this rate by 10, as done in Chen, Oliva and Zhang (2022). I opted for a similar but improved approach, where I took into account the mortality rate for this group ten years later. Using the same method, I also calculated the mortality rate for this age group in 2020 and took the average of the 2010 and 2020 mortality rates. This serves as the average mortality rate of this age group in the region during the decade. Hence, the calculation for $Death_{[15,50),r}$ in equation 1 is as follows:

$$Death_{[15,50),r} = \sum POP_{[g,g+5),r,2010} \times \frac{Mortality_{[g,g+5),r,2010} + Mortality_{[g+10,g+15),r,2020}}{2} \times 10 \tag{B2}$$

Here, $g \in \{15, 20, 25, 30, 35, 40, 45\}$.

Gender-specific mortality rates and deaths are calculated following the same method.

C An instrumental variable approach

Apart from the attempts to calculate an accurate and unbiased SR, I also employ an instrumental variable approach to the individual-level analyses, which utilizes exogenous variations that induce change in local SR.

Numerous researches have identified China's One-Child Policy since the late 1970s along with the introduction of the B-ultrasound since early 1980s as primary drivers for the surge in SR (e.g., Chen, Li and Meng, 2013; Chu, 2001; Yang and Chen, 2004). The OCP was a program implemented by the Chinese government in the late 1970s (and abolished only in recent years) to restrict most families to having only one offspring. The goal was to control the country's population growth. McElroy and Yang (2000) confirms the inhibitory impact of the OCP on the country's fertility rate, where many families are only allowed with one offspring. Consequently, families who prefer male children, a common cultural practice in China, often resort to prenatal sex selection to ensure their child is male (Hesketh and Xing, 2006; Chen, Li and Meng, 2013).

Meanwhile, the implementation of the OCP has been different among different regions, ethnic groups, and urban and rural areas, with much more leniency in policies for ethnic minorities compared to the Han ethnic group (Gu et al., 2007). Li, Yi and Zhang (2011) find that differences in the implementation of family planning policies between ethnic minority and Han groups explain 57% and 54% of the SR increase in the 1991-2000 and 2001-2005 cohorts, respectively.

Drawing from Cameron, Meng and Zhang (2019), I instrument for regional SR with the degree of stringency of the OCP, the share of the ethnic minorities, and an interaction between the two. The strictness of the OCP is proxied by the average number of live births among local women of childbearing age. The rationale underlying this set of instrumental variables is that the OCP exerts a direct influence on the local SR while its impact varies across different ethnic groups. Consequently, regions with differing proportions of ethnic minorities are expected to demonstrate heterogeneity in the effect of the policy on their SR.

The results are presented in the following tables. The sizes of these coefficients are highly similar across the models, with the coefficient estimated using instrumental variables slightly larger than that estimated using OLS by approximately 26.19% to 47.37%.

Table C4: Individual evidence: SRs and migration descisions

Dep. Var.	Migrant: Yes or No							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: SR based on hukou registration							
Model:	*OLS*				*Two-stage*			
Sex ratio	0.0072***	0.0082***	0.0042***	0.0042***	0.0069***	0.0107***	0.0054***	0.0053***
	(4.101)	(3.303)	(3.135)	(3.280)	(2.766)	(2.616)	(2.645)	(2.624)
Mean D.V.	0.0964	0.0964	0.0964	0.0964	0.0964	0.0964	0.0964	0.0964
Std. dev. D.V.	0.295	0.295	0.295	0.295	0.295	0.295	0.295	0.295
Obs.	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894
	Panel B: SR based on the timing of migration							
Model:	*OLS*				*Two-stage*			
Sex ratio	0.0070***	0.0082***	0.0056***	0.0046***	0.0083***	0.0115***	0.0063***	0.0060***
	(3.897)	(3.401)	(4.848)	(4.042)	(2.101)	(2.288)	(2.528)	(2.440)
Mean D.V.	0.0964	0.0964	0.0964	0.0964	0.0964	0.0964	0.0964	0.0964
Std. dev. D.V.	0.295	0.295	0.295	0.295	0.295	0.295	0.295	0.295
Obs.	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894	3,655,894
	Panel C: SR based on the timing of migration, w/o long-term migrants							
Model:	*OLS*				*Two-stage*			
Sex ratio	0.0056***	0.0060***	0.0040***	0.0038***	0.0068***	0.0094***	0.0058***	0.0056***
	(3.698)	(2.896)	(3.488)	(3.339)	(3.209)	(2.737)	(2.817)	(2.818)
Mean D.V.	0.0770	0.0770	0.0770	0.0770	0.0770	0.0770	0.0770	0.0770
Std. dev. D.V.	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267
Obs.	3,542,609	3,542,609	3,542,609	3,542,609	3,542,609	3,542,609	3,542,609	3,542,609
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects								
Prov.	No	Yes	Yes	absorbed	No	Yes	Yes	absorbed
Cohort	No	No	Yes	absorbed	No	No	Yes	absorbed
Prov. × Cohort	No	No	No	Yes	No	No	No	Yes
# Clusters	275	275	275	275	275	275	275	275

Notes: Regressions in columns (1)-(4) are estimated using OLS. Regressions in columns (5)-(8) are estimated using instrumental variables. Robust t-statistics in parentheses. Standard errors allow for clustering within home prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

Table C5: Heterogeneity: by gender and SES

Dep. Var.	Migrant: Yes or No				
	(1)	(2)	(3)	(4)	(5)
	Panel A: OLS				
		Low SES		*High SES*	
Sample:	*ALL*	*Rural hukou*	*High school & below*	*Urban hukou*	*College & above*
Sex ratio × Female	-0.0004***	-0.0003***	-0.0005***	-0.0001	0.0009***
	(-3.863)	(-2.911)	(-4.055)	(-0.410)	(2.720)
Sex ratio	0.0044***	0.0048***	0.0043***	0.0027***	0.0052***
	(3.356)	(3.339)	(3.393)	(2.809)	(3.220)
	Panel B: Two-stage				
		Low SES		*High SES*	
Sample:	*ALL*	*Rural hukou*	*High school & below*	*Urban hukou*	*College & above*
Sex ratio × Female	-0.0005***	-0.0005***	-0.0007***	0.0007***	0.0021***
	(-3.685)	(-3.124)	(-4.327)	(3.166)	(4.303)
Sex ratio	0.0056***	0.0057***	0.0056***	0.0023	0.0019
	(3.030)	(3.112)	(3.236)	(1.108)	(0.551)
Mean dep. var.	0.0964	0.105	0.0866	0.0732	0.179
Std. dev. dep. var.	0.295	0.307	0.281	0.261	0.384
Obs.	3,655,894	2,764,417	3,460,096	1,110,358	414,678
Controls	Yes	Yes	Yes	Yes	Yes
Fixed effects					
Pvc. × Cohort	Yes	Yes	Yes	Yes	Yes
# Clusters	277	277	277	277	277

Notes: Regression in Panel A are estimated using OLS. Regressions in Panel B are estimated using instrumental variables. Robust t-statistics in parentheses. Standard errors allow for clustering within home prefectures. * indicates significance at the 10 percent level, ** at the 5 percent level, *** at the 1 percent level.

D Data used in county-level checks

In the main paper, I calculate SR of each prefecture or city, and migrants are defined as those crossing prefectures. For further checks, I calculate SR of counties and define migrants as those crossing counties. County is the administrative division below prefecture.

To control for economic factors of each county, I gather data on per capita GDP from China's Statistical Yearbook. I primarily use two yearbooks, the *China Statistical Yearbook for Regional Economy* (CSYRE) 2010 and the *China County Statistical Yearbook* (CCoSY) 2010, to collect data for counties in 2009, and merge it with the 2010 Population Census. Counties that changed in terms of administration or names between 2009 - 2010 are adjusted manually. It should be noted that there are inconsistencies in the same indicator across these three data sources (the CSYRE, the CCoSY, and the Population Census), such as total local population. Additionally, economic indicators for some counties sometimes differ slightly between the two yearbooks. This may call into question the accuracy of data. Nevertheless, the vast majority of data from different sources is entirely consistent, with only a few instances where there are discrepancies. Even in those cases, the difference is minimal - typically less than one thousandth.

I rely on the census data for demographic-related variables and the CSYRE for economic-related variables, supplementing any missing information with data from the CCoSY. I am more inclined to refer to data presented in the CSYRE as it provides a more comprehensive coverage of economic indicators. Although the contents of the two yearbooks overlap significantly, they were compiled by two different departments under the Chinese National Bureau of Statistics in 2010. CSYRE is edited by the Department of Comprehensive Statistics of the National Economy, who is responsible for conducting monitoring, forecast and comprehensive analysis on the performance of the national economy and preparing macro-regulation proposals. On the other hand, the CCoSY is edited by the Organization of Rural Socio-Economic Survey, who collects, processes and provides statistical data on agricultural production, income and expenditure of rural households, rural economy, farming activities, etc. The former department mainly focuses on socioeconomic development, while the latter focuses on agriculture. Therefore, it is reasonable to assume that economic indicators, namely per capita GDP, calculated by the former are more credible.

Both the CSYRE and CCoSY mainly contain data for county-level cities, counties, autonomous counties, and other rural areas, while there is a lack of data for most municipal districts - simply put, the information in these yearbooks primarily covers rural rather than urban areas. To obtain data for municipal districts in urban areas, I consult Statistical Yearbooks from local governments at all levels nationwide. With this, I am able to acquire information on some municipal districts and combine it with county-level data from other yearbooks. This resulted in obtaining data on per capita GDP for a total of 2,048 out of 2,893 counties in 2010. For the rest of the counties lacking this information, I use the per capita GDP of the prefecture in which the county is located in 2009, collected from the China City Statistical Yearbook, as a proxy. Ultimately, I am able to collect data on per capita GDP for 2,790 counties (742 of which are proxied for), which I merge with data from the 2010 Population Census. I later perform a series of sensitivity checks excluding the counties with a proxied per capita GDP, and this does not affect the robustness of the results.