Prince and Cinderella: Sex Selection and Socioeconomic Status

Wanru Xiong
Office of Population Research
Princeton University

Abstract

This paper studies how local marriage market influences prospective parents' sex selection. I formulate an extended Roy model of sex selection. My model shows that when the sex ratio imbalance is severe in the local marriage market, prospective parents of higher socioeconomic status (SES) tend to choose boys; instead in the case of mild imbalance, parents of lower status indulge their son preference. Empirical tests using the fertility data of China in the past 15 years are congruent with the theoretical hypotheses. Specifically, a drop in the local marriage market sex ratio between 2000 and 2010 leads to a reversal in the general sex selection pattern by parent's SES. This paper also proposes a long-term self-correcting equilibrium mechanism on the sex ratio.

1 Introduction

Over the most important things in life we are granted with little autonomy. We do not choose to be born. We do not choose our parents. We may have a choice over the parent of our children, but whether to have a boy or girl has been beyond our choice until recent advances in technology allow prospective parents to have a say in the fetus's gender.

The other name of this kind of discretion is formally known as sex selective abortion (foeticide) or infanticide. This is not a piece of news in countries where son preference dominates (Das Gupta, 1987). Though the exact number of illegal female fetuses abortion and infanticide is beyond statistical measurement, rough estimations based on sex ratio suggest an astonishing figure. According to cross-country empirical studies, sex ratio at birth, defined in demography as the proportion of newborn males to females in a given population within a certain period, has stayed constant to be around 105 in human population without intended social and behavioral interventions (Johansson and Nygren, 1991). In China, however, the number never fell within the normal range since 1980, when the implementation of family planning policy met the introduction of ultrasound technology (Ebenstein, 2011). According to the census, the national average sex ratio at birth was around 110.0 in 1990, soared to 119.9 in 2000, kept high in 2005 to be 120.5 and showed no sign of reversing trend in 2010 with the ratio being 121.2 (National Bureau of Statistics of P.R.C.). Although prenatal sex screening and sex selective abortion are legally prohibited, they have been practiced behind the curtain.There is no way for sex ratio to be so skewed without sex selection in practice. Obviously, groups of parents are manipulating their babies' gender.

This paper answers the question about who select boys and why. To be specific, I focus on how local marriage market conditions influence sex selection decision made by prospective parents of different SES. The basic idea is that prospective parents make rational decision, if technology permits, about their fetus's gender in face of the family planning policy constraint, local marriage market condition and with their own SES in consideration. Marriage market condition matters because as a result of past sex selection, deficit of women in marriage market dooms the failure of some bachelors in marriage market. If parents care about whether their children could get married, or using a more formal term "reproduction prospect", they would take local marriage market condition into consideration. The effect of local marriage market pressure may be heterogenous because parents have different rearing ability. The general idea is driven by an equilibrium sense, that is, to find the causes of sex ratio imbalance in its own consequences. In a word, sex ratio could be endogenous in the long term.

This paper contains theoretical model and empirical evidence. In the theoretical part, I adapt the classical Roy Model (Roy, 1951) to fit in the context of sex selection. Similar enough to what happened in the case of international migration (Borjas, 1987), realized cases could be "positive hierarchical sorting" or "refugee sorting", in which parents who tend to choose son over girl are of high or low SES respectively. The key factor turns out to be the relative probability of getting married, which is related not only to the local marriage market sex ratio but also to the parents' SES and rearing ability. To be specific, when there is a substantial deficit of women in marriageable age, parents from the upper end in the society tend to select boys. In contrast, when marriage squeeze is mild, parents from the lower class indulge their son preference. In the following empirical part, I test the hypotheses and model implications. I first present aggregate level statistics based on the census data. Cross-tabulations outline a sketch of the reality, in which there is a quick and enigmatic reversal between 2000 and 2010. In city level, I find a negative correlation between sex ratio in local marriage market and sex ratio at birth. After that, I conduct an individual level logit regression analysis using China Family Panel Survey (CFPS) 2010 and 2012 micro level data. The results of individual level analysis are congruent with the aggregate level evidence. Together the empirical evidence and model yield consistent findings.

The rest of the paper is organized as follows. Section 2 gives a brief background about sex selection and marriage squeeze in China. Section 3 elaborates an extended Roy Model in the context of sex selection. Section 4 presents empirical evidence, including aggregate level statistics, city level and individual level regression analysis based on census and household survey data. The final section ends with discussion and concluding remarks.

2 Background

Sex selection has been drawing great attention. It contributes a part to the worldwide momentous "missing women" problem (Chen et al., 2012). Amartya Sen first coined this tragedy in 1990. In his article, Sen estimated that "a greater many more than 100 million women are 'missing' " in South Asia, West Asia and North Africa. China alone counted 50 million (Sen, 1990). Malnutrition and limited access to medical care are among other main causes of "missing girls". The deficit of women costs a high price for the society. The crux is that surplus bachelors lay on the other side of the same coin. There has been extensive research in demography, political science and economics on the consequences of sex ratio imbalance (Hudson and den Boer, 2004; Edlund et al., 2013; Das Gupta, Ebenstein and Sharygin, 2011; Poston and Glover, 2004).

Marriage squeeze is an important channel through which imbalance sex ratio leads to social problems. With male bachelors in excess supply, their prospects of marriage are overshadowed (Das Gupta et al., 2011). Even those who succeed in getting married do not win easily. In China, bride price keeps soaring in marriage market, putting an ever-heavier burden on the groom's side. In some areas, an apartment in town is considered a starting price for proposal. Therefore, saving for son's wedding is a common saving motive, especially in small towns where son's marriage becomes a family campaign (Wei and Zhang, 2009). The "missing girls" are gone, while the society and the surplus bachelors in particular are paying for the loss.

Investigation into the causes of "missing girls" is the first step towards any solution. In regard to macro-level causes, researchers have almost reached a consensus. The leading cause should be son preference culture (Coale and Banister, 1994; Almond, Edlund and Milligan, 2013). In traditional Chinese culture, the family line passes only through the sons. Therefore the desire for sons is not only related to labor concerns but also to family values. Another main factor is the family planning policy (Ebenstein, 2011; Johansson and Nygren, 1991; Zeng, 2007). The family planning policy in China after 1979 - commonly known as the "One Child Policy" - imposed a binding constraint on the number of childbearing based on parent's identity. When the number of children was restricted, parents with strong son preference turned to gender manipulation (Ebenstein, 2011). This kind of sex selection incentive could be explained by the theory of quality-quantity trade-off in family and household economics, when a boy is perceived as a higher quality child (Becker and Lewis, 1974). Furthermore, some researchers argued that the family planning policy did not play solo. They believed that it was the synchronization of one child policy along with the introduction of ultrasound technology that triggered the start of severe sex ratio imbalance (Ebenstein, 2010). Other policies such as social security reform (Das Gupta et al., 2003) and media campaign (Das Gupta et al., 2003) are helpful in addressing the imbalance sex ratio problem. A recently paper provide evidence from China that sex ratio in the marriage market has an effect on sex ratio at birth (Li et al., 2016).

However, "who abandoned the girls and why?" still remains a puzzle. There have been some micro-level studies trying to find reasons for parents' sex selection behavior in their personal characteristics. Following the biological approach, scientists found that some psychological and physiological factors might influence the probability of conceiving a boy (James, 1995). In the field of social science, researchers got divergent evidence. With respect to income, Gaulin and Robbins (1993) noticed that in North America, low-income mothers invested more in daughters. However, low-income parents have stronger son preference in China (Chu, 2001). Education matters (Chen, Li and Meng, 2012), but sometimes in a counterintuitive way. Contrary to common belief that the idea of gender equality is more accepted by well-educated parents, in India parents with higher education strive for sons (Bhat and Zavier, 2007). Social status of the parents in general is related with sex selection behavior, though evidence are inconsistent. Cronk (1989) conducted a field survey in African tribe, from where he came out with evidence of daughter preference among the underclass. In Asia however, there is strong discrimination against daughters in the bottom of the society (Banister, 2004).

The above contradictory empirical evidence each gains theoretical supports. Edlund (1999) tried to model endogenous sex selection choice. In her model, she assumed parents' son preference as well as reproducing concerns, and a strict hierarchical mating rule in marriage market. The model inferred that the upper class chose male offspring, which eventually led to permanent social stratification by sex. Her model was consistent with some of the above findings where parents with better SES favored sons. On the contrary, Ebenstein (2011) formulated and estimated a dynamic model of fertility choice in China under a game theory framework. His simulation yielded the opposite inference that low-educated peasants were responsible for the missing girls. In sum, researchers have not yet reached a consensus neither in theory nor in empirical evidence about micro-level factors of sex selection.

3 The Model

In this paper, the Roy model of sex selection with son preference and reproduction uncertainty is a theoretical attempt to answer the question about who conceived with a girl tends to conduct sex selection. Roy model (Roy, 1951) applies to sex selection in the sense of rational self-selection, as in its classical application of migration choice. The over arching rule is that given natural state (pregnant with a boy or a girl), son preference and sex selection cost, prospective parents make sex selection decision based on comparison between expected utility in girl state and boy state, where in both states rearing ability of their own and reproductive uncertainty derived from marriage market are taken into consideration.

3.1 Model settings

Consider sex selection as transition between two states, namely "boy state" (denoted as state 1) and "girl state" (state 0). A state is a set of outcomes with positive probabilities. Initial state is given by nature, in other words random conception. In keeping with the spirit of this paper, I focus on the transition from girl state 0 to boy state 1, which is also in consistence with the following son preference assumption. To explicitly incorporate son preference into the Roy Model, I denote the demeaned individual specific son preference as β_i, which is normally distributed in population:

$$\beta \sim \mathcal{N}(0, \sigma_\beta^2).$$

Larger value of β_i indicates stronger son preference. Notice that β_i is a demeaned value; therefore even a negative β_i could indicate son preference, though it is girl preference in a relative sense. In my model, this preference index appears symmetrically in the utility functions, that is, utility gain from having a boy corresponds to an equal loss from a girl. Giving up this symmetric preference assumption does not alter the essence of my model, as long as the preference component exists in the utility function.

Assume additively separable utility function. Utility of prospective parents in natural girl state is:

$$\begin{aligned} U_{0i} &= \mu_0 + \epsilon_{0i} \\ \epsilon_{0i} &= \alpha_{0i} + \beta_i \\ \alpha_0 &\sim \mathcal{N}(0, \sigma_\alpha^2). \end{aligned}$$

The mean utility in the population from having a girl is denoted as μ_0. Idiosyncratic utility component ϵ_{0i} is decomposed into two parts, ability α_{0i} and preference β_i. It is useful to think of α_0 as demeaned value of parent's ability in nurturing a girl. In general, parenting ability and gender preference need not be independent, for both of them could be functions of education and SES. In China, gender equality has been an integral part of education. Social value surveys show that son preference is more prevalent among farmers and migrant workers, who tend to receive poor education (Liao, Cao and Tao, 2012). Therefore, for conciseness, I assume the covariance $\sigma_{\alpha\beta}$ between α_0 and β are negative and identical across individuals, i.e.

$$\sigma_{\alpha\beta} < 0.$$

This assumption is not crucial to the following analysis.

The expected utility of prospective parents in natural girl state were they transit to boy state are given by:

$$\begin{aligned} U_{1i} &= \mu_1 + \epsilon_{1i} \\ \epsilon_{1i} &= \alpha_{1i} + \beta_i \\ \alpha_1 &\sim \mathcal{N}(0, \lambda^2\sigma_\alpha^2). \end{aligned}$$

The parameter μ_1 is the mean utility that prospective parents in natural girl state would gain in boy state if all of then were to transit to boy state.

The key specification of the model comes with λ, the uncertainty index in boy state. This uncertainty derives itself from marriage market, where with excess supply of single bachelors, some boys may not be lucky enough to get married thus discounting the expected utility gain from having a boy from the reproduction prospective. As long as women hold the advantage position in marriage market, the relative uncertainty in boy state is greater, implying $\lambda > 1$.

The transition from girl state to boy state, i.e. sex selection, is not free, both in terms of monetary cost and maternal health cost. Assume the prevailing cost of sex selection to be $\mathcal{C}$. The corresponding utility loss π is a constant across all prospective parents. For conciseness, I assume perfectly efficient sex selection technology, meaning one to one replacement. The probability of selection failure and its corresponding loss could be discounted as part of selection cost and utility deduction.

In the following discussions I hold two assumptions for simplicity.

Assumption 3.1. *In the population, $\mu_0 \approx \mu_1$. For representative parents, boy and girl are equally desirable.*

Assumption 3.2. *For each parent,* $\mathrm{corr}(\alpha_{0i}, \alpha_{1i}) = 1$. *Parenting ability for raising a boy or a girl is perfectly correlated.*

Given the above assumptions, the variance of ϵ_0 is:

$$\sigma_0^2 = \mathrm{E}(\epsilon_0)^2 = \sigma_\alpha^2 + \sigma_\beta^2 - 2\sigma_{\alpha\beta}.$$

The variance of ϵ_1 is:

$$\sigma_1^2 = \mathrm{E}(\epsilon_1)^2 = \lambda^2\sigma_\alpha^2 + \sigma_\beta^2 + 2\sigma_{\alpha\beta}.$$

The correlation between ϵ_0 and ϵ_1 is given by:

$$\rho = \frac{\sigma_{01}}{\sigma_0\sigma_1}$$
$$\sigma_{01} = \mathrm{cov}(\epsilon_0, \epsilon_1) = \lambda\sigma_\alpha^2 - \sigma_\beta^2 + (\lambda - 1)\sigma_{\alpha\beta}.$$

Prospective parents are aware of selection cost $\mathcal{C}$ (and utility equivalence π), μ_0 and μ_1, and their own natural state as well as α and β. Thus they have knowledge of ϵ_0 and ϵ_1. Based on this information, the sex selection rule becomes:

A prospective parent given natural girl state will do sex selection if and only if

$$U_0 < U_1 - \pi.$$

That is

$$\mu_0 - \mu_1 - \pi + (\epsilon_1 - \epsilon_0) > 0.$$

Define indicator variable $\mathcal{I} = 1$ if there is sex selection. Let $\nu \equiv \epsilon_1 - \epsilon_0$. The transition rate from girl state to boy state is:

$$\mathrm{P} = \Pr[\mu_0 - \mu_1 - \pi + (\epsilon_1 - \epsilon_0) > 0] = 1 - \Phi(z),$$

where $z = (\mu_1 - \mu_0 - \pi)/\sigma_\nu$ and Φ is the standard normal distribution function. This transition rate is exactly the counterpart of migration rate (Borjas, 1987). Obviously, the transition rate goes up as (a) the mean utility gap between average utility in boy state and girl state gets wider and (b) the cost of sex selection falls. That is

$$\frac{\partial P}{\partial \mu_0} < 0, \frac{\partial P}{\partial \mu_1} > 0, \frac{\partial P}{\partial \pi} < 0.$$

The conditional mean utilities $\mathrm{E}(U_0|\mathcal{I}=1)$ and $\mathrm{E}(U_1|\mathcal{I}=1)$ are given by

$$\begin{aligned}
\mathrm{E}(U_0|\mathcal{I}=1) &= \mu_0 + \mathrm{E}\big(\epsilon_0\big|\frac{\nu}{\sigma_\nu} > z\big) = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}(\rho - \frac{\sigma_0}{\sigma_1})\varphi \\
\mathrm{E}(U_1|\mathcal{I}=1) &= \mu_1 + \mathrm{E}\big(\epsilon_1\big|\frac{\nu}{\sigma_\nu} > z\big) = \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}(\frac{\sigma_1}{\sigma_0} - \rho)\varphi,
\end{aligned}$$

where $\varphi = \phi(z)$ and ϕ is the density of the standard normal. Here, self-selection bias emerges as a result of utility maximization behavior when initially $\mathrm{P} < 1$ so that at least part of the prospective parents decide not to transit their states, which is consistent with the reality that most parents just go and meet their destiny anyway. Whether the average utility of those who choose to conduct sex selection is higher than that of the parents naturally pregnant with a boy is depend on $\rho \lesseqgtr \sigma_1/\sigma_0$. On the other hand, whether the average utility of those who choose to conduct sex selection is higher than that of the parents naturally pregnant with a girl is depend on $\rho \lesseqgtr \sigma_0/\sigma_1$. The later condition characterizes the prospective parents who conduct sex selection in the nature girl state, with whom my paper concerns.

3.2 Cases and hypotheses

To explicitly explain the mechanism, I discuss the following three cases. I use the same terms of the cases as in Borjas 1987[5]. Frist, define

$$\begin{aligned}
Q_0 &= \mathrm{E}(\epsilon_0|\mathcal{I}=1) = \mathrm{E}(\alpha - \beta|\mathcal{I}=1) \\
Q_1 &= \mathrm{E}(\epsilon_1|\mathcal{I}=1) = \mathrm{E}(\alpha + \beta|\mathcal{I}=1),
\end{aligned}$$

where Q_0 and Q_1 are individual specific component of utility from having children. Notice that they could be decomposed into two parts with $\alpha_i (i = 0, 1)$ representing parenting ability as a function of parent's SES and β indicating personal gender preference.

3.2.1 Case I: Positive hierarchical sorting

Positive hierarchical sorting is the case where the prospective parents with the highest utility from children given girl state transit to boy state and gain greater utility than the average of those who naturally pregnant with a boy, i.e.

$$Q_0 > 0; Q_1 > 0.$$

This type of sorting occurs if and only if the following two conditions hold:

$$\frac{\sigma_1}{\sigma_0} > 1 \tag{1}$$

$$\rho > \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right). \tag{2}$$

For condition 1,

$$\frac{\sigma_1}{\sigma_0} > 1 \Leftrightarrow (\lambda^2 - 1)\sigma_\alpha^2 + 4\sigma_{\alpha\beta} > 0.$$

Since $\sigma_{\alpha\beta} < 0$ and $\lambda > 1$, the uncertainty in boy state is greater than girl state considering the marriage market condition. In the simple case where $\sigma_{\alpha\beta}$ is sufficiently close to zero, say the ability and preference are almost independent, the above inequality always holds in sex selection context.

For Condition 2,

$$\begin{aligned}
&\rho > \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right) \\
\overset{\frac{\sigma_1}{\sigma_0}>1}{\Longleftrightarrow} &\rho > \frac{\sigma_0}{\sigma_1} \\
\Longrightarrow &\lambda > \frac{\sigma_\alpha^2 + 2\sigma_\beta^2 - 3\sigma_{\alpha\beta}}{\sigma_\alpha^2 - \sigma_{\alpha\beta}} = 1 + 2\frac{\sigma_\beta^2 - \sigma_{\alpha\beta}}{\sigma_\alpha^2 - \sigma_{\alpha\beta}}.
\end{aligned}$$

When λ is large enough and $\sigma_\beta^2 - \sigma_{\alpha\beta}$ is sufficiently small, the second condition holds and positive hierarchical sorting takes place. In this case, among the prospective parents in girl state, those with ϵ_0 falling in the upper tail of the distribution (higher ability or mild son preference) would conduct sex selection.

3.2.2 Case II: Negative hierarchical sorting

Negative hierarchical sorting is the case when prospective parents negatively self- select from the natural girl state and their realized utility after transition is below the average of those in boy state. In this case, the realized outcomes become:

$$Q_0 < 0; Q_1 < 0.$$

In other words, prospective parents with lower parenting ability and stronger son preference tend to conduct sex selection and eventually end up with relatively lower utility in boy state, though with greater utility compared to their original utility in girl state. This happens when the following conditions are satisfied:

$$\frac{\sigma_1}{\sigma_0} < 1 \tag{3}$$

$$\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right). \tag{4}$$

As will be shown in Case III, the conditions 3 and Condition 4 can never hold simultaneously. Therefore in the sex selection context there is no negative hierarchically sorting as in the case of migration.

3.2.3 Case III: Refugee sorting

This case is called refugee sorting because the prospect parents who choose to transit come from the lower tail of the expected utility distribution in girl state and end up in the upper tail in boy state, i.e.

$$Q_0 < 0; Q_1 > 0.$$

This type of sorting occurs if and only if

$$\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right). \tag{5}$$

When $\sigma_1/\sigma_0 > 1$, Condition 5 becomes

$$\begin{aligned}
&\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right) \\
\overset{\frac{\sigma_1}{\sigma_0}>1}{\Longleftrightarrow} &\rho < \frac{\sigma_0}{\sigma_1} \\
\Longrightarrow &\lambda < \frac{\sigma_\alpha^2 + 2\sigma_\beta^2 - 3\sigma_{\alpha\beta}}{\sigma_\alpha^2 - \sigma_{\alpha\beta}} = 1 + 2\frac{\sigma_\beta^2 - \sigma_{\alpha\beta}}{\sigma_\alpha^2 - \sigma_{\alpha\beta}}.
\end{aligned}$$

When $\sigma_1/\sigma_0 < 1$, Condition 5 becomes

$$\begin{aligned}
&\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right) \\
\overset{\frac{\sigma_1}{\sigma_0}>1}{\Longleftrightarrow} &\rho < \frac{\sigma_1}{\sigma_0} \\
\Longrightarrow &(\lambda - \lambda^2)\sigma_\alpha^2 - 2\sigma_\beta^2 - (1+\lambda)\sigma_{\alpha\beta} < 0.
\end{aligned}$$

This inequality holds for any $\lambda > 1$, which means that as long as $\sigma_1/\sigma_0 < 1$, the case will be like refugee sorting. An important implication is that the possibility of negative selection is ruled out. Put it in the sex selection context, it is reasonable to predict that when the marriage market is instead in favor of men, those who get very low utility from a girl will have less concern to try their luck again.

3.2.4 Hypotheses

Which case best describes the reality remains an empirical question. Before going to empirical evidence, I propose the following hypotheses and predictions based on my model.

Hypothesis 1: In the simple case where $\sigma_{\alpha\beta} \approx 0$, ceteris paribus, when λ, say the uncertainty index in boy state is high, we will observe positive hierarchically sorting, where prospective parents with high ability and mild son preference choose to conduct sex selection and settle down in a desirable position. When λ drops below a critical value, there will be "refugee sorting" instead. In that case, prospective parents with strong son preference and weaker parenting skill transit to boy state, in which they are better off because they successfully indulge their strong son preference.

Hypothesis 2: When $|\sigma_{\alpha\beta}|$ ($\sigma_{\alpha\beta} < 0$) becomes sufficiently large, the case will be refugee sorting, ceteris paribus.

4 Empirical evidence

This section contains empirical evidence of sex selection in recent China, mainly focuses on the cohorts starting from 2000 to 2010, when sex selection technology such as diagnostic ultrasound had been almost universally accessible. I will show both summary tabulations and regression analysis. Aggregate level data outline a sketch of the general patterns of the relationship between 1) sex selection rate and mother's education; 2) the correlation between local marriage market and sex ratio at birth. Then, I use real fertility data from CFPS to do individual level regression analysis. After that, I discuss the link and gap between empirical evidence and theory implications.

4.1 Aggregated level statistics and regression analysis

4.1.1 Sex selection and social-economic status: sex ratio at birth by mother's education

This section uses the cross tabulations in census about mother's education and children's gender by parity. Public census summarized data of year 2000 and 2010 contain tabulations of detail fertility information during a year before the national cencus (1999.11.1 to 2000.10.31 and 2009.11.1 to 2010.10.31 respectively) (National Bureau of Statistics of P.R.C.). For clarity, I calculate a sex selection rate based on the counted sum number. Sex selection rate (denoted as t) is defined as the rate of realized sex selection in practice implied by the actual live births. In other words, it takes a certain proportion (the sex selection rate) of prospective parents to succeed in conducting sex selection to distort the natural sex ratio at birth to what is observed in reality. Similar as in the model, here I assume perfectly efficient sex selection technology, meaning one to one replacement. In practice, the risk of failure cannot be neglected, thus the actually attempted sex selection rate is higher than the realized sex selection rate defined here, because the latter is discounted by the inefficiency of sex selection technology. Sex selection rate gives a direct assessment of the phenomenon, although it represents the lower bound.

Define the actual number of male births as B and female births as G. If there is no sex selection, potential number of female births amount to G'. Set the natural sex ratio at birth to be 105 (Johansson and Nygren, 1991). To be consistent with the model, positive sex selection rate indicates transition from girl to boy, while negative the other way round. Keeping the total number of children constant, by nature we have

$$G' = (B + G) \cdot (100/(100 + 105)).$$

With replacement (selection), it becomes

$$(G + tG)/(B - tG') = 100/105.$$

Solving for sex selection rate, we get

$$t = (100B - 105G)/(100G + 100B).$$

Notice the sex selection rate conveys no more information than the number of live births and the assumption of natural sex ratio at birth. Therefore sex selection rate can be derived from an exact transform of sex ratio at birth (denote as SR):

$$t = (SR - 105)/(100 + SR).$$

Table 1 presents the sex selection rate by mother's education and parity. From Table 1, we can see that in 2000, sex selection rate strictly increases with mother's education, growing from a negative rate for the illiterate group to a positive one for those attend high school and above. The fact that as many as 7.65% of the illiterate mothers choose to abandon a baby boy for a girl in 2000 is quite astonishing. Common sense may be that low-educated mothers are more enchanted by traditional son preference. One possible explanation refers to physiological reasons, for example prevalent malnutrition of low-educated mothers, which may endanger male fetus more. However, that cannot explain the rates observed in second parity (Table 2) and 2010 in general. Keeping our focus on 2000 case, the sex selection rates at second birth show similar pattern, except that all groups select boys. In general, the pattern in 2000 is close to "positive hierarchical sorting" in the theoretical model.

However, things are different in 2010. At first parity, mothers with primary education are most obsessed with sons, followed by those with junior high school education. This fits the case of "refugee sorting", though not perfectly, where parents of lower SES indulge their son preference. The pattern changes at second parity, where mothers in the highest education tend most to abandon girls for boys, as in the case of 2000. In 2010, none group in the table select girls.

Comparing the cases in 2000 and 2010, we notice that in general, sex selection rate at first parity is higher in 2010 than in 2000, but the opposite for the second parity. A relevant fact is that 20% more children of second parity are born in 2010 compared with 2000. Together the facts may reflect some effects of recent family planning policy relaxation. Second, there is much less variation between education groups in 2010 than in 2000. It is a hint that sex selection decision may be more dominated by socioeconomic concerns in 2000.

In both years there are differences between urban, county and rural areas, although the patterns are inconsistent across years and education groups. As parents with different residential registration status (hukou) are subjected to different family policy regulation rules, simple cross-sectional comparison between urban and rural areas may not give much insight.

A preliminary finding from these aggregate level cross-tabulations is that with respect to first parity, the 2000 case is similar to "positive hierarchical sorting" while the 2010 case is close to "refugee sorting".

4.1.2 Self-correcting mechanism: sex ratio at birth and sex ratio at local marriage market

The model hypothesizes a self-correcting equilibrium mechanism in sex ratio balance with sex selection. The implication that highly skewed sex ratio in local marriage market is associated with more balanced sex ratio at birth is borne out by the data. Figure 1 plots the city level sex ratio at birth against local marriage market sex ratio for year 2000 and 2010 respectively, revealing a negative correlation.

The negative correlation between sex ratio at birth and that at local marriage market is not necessarily causal. Sex ratio at local marriage market is a result of both past local sex ratio at birth and migration, if we neglect mortality before marital age. One concern is that factors that affect migration also influence sex ratio at birth. For example, fast economic development attracts migrants while influences the fertility decisions by changing social norms.

To investigate whether this negative correlation is driven by self-correcting force, I do city level regression analysis to control for confounding variables. The model is a multiple regression with province fixed effects:

$$Sex\ Ratio\ at\ Birth = \alpha + \beta\ Sex\ Ratio\ at\ Marriage\ Market + \theta\ City + i.province + \epsilon$$

The unit of observation is prefecture-level city. Prefecture-level city is an administrative unit. There are over 300 prefecture-level cities in 34 provinces in mainland China. *City* represents a set of city-level economic and demographic control variables, including GDP per capita of the year (log), GDP growth, the proportion of the tertiary industry in the local economy, population density (log) and population growth rate. The sample is restricted to cities with complete information regarding to these control variables. Most excluded cities locate in the west, where the population density is very low. Summary statistics are shown in Table 3 and Table 4.

For sex ratio at marriage market, I use several ratios referring to different age groups for robustness checks and perform a placebo test using sex ratio of the total population. Province fixed effects tries to capture the unobserved factors such as son preference culture, family planning policy strictness and technology accessibility. Standard errors are clustered at province level to allow for within province correlation.

Regression results, shown in Table 5, support the self-correcting mechanism hypothesis. After controlling for the above economic and demographic variables, higher sex ratio at marriage market is associated with lower sex ratio at birth. Column (4) and (8) show a placebo test using sex ratio of the total population as explanatory variable. In contrast with that of local marriage market, this coefficient turns out to be positive, indicating a special role of marriage market sex ratio. However, I fail to rule out the possibility that observed high sex ratio urges the government to implement a stricter restriction on select selection, which may explain the negative correlation.

Among the four definitions of local marriage market sex ratio, the one covering males between 20-35 and females between 20-30 is the most significant (the ratio has been adjusted to fit the normal sex ratio scale). It is reasonable because the average age at first marriage is higher for males and the age gap in marriage mainly falls between 0-3 (husband wife) in China (Wei, Dong and Jiang, 2013). According to census 2010, 53% of males between 20-34 have got married, while only 38% in age group 20-29. For females, 53% got married in age group 20-29 (National Bureau of Statistics, P.R.C.). Other definitions have the expected signs, but are not always significant. The age gap in marriage actually plays an important role, which will be discussed later. It is interesting to see that the economic and demographic variables seem to have some explanatory power over sex ratio at birth, but that is beyond the scope of this paper.

This correlation between marriage market sex ratio and sex ratio at birth echoes the story that in China when there were significant surplus of males in marriage market around 1980s due to the Great Famine, sex ratio at birth decreased (Li et al., 2016) However, my finding suggests a more general and persistent pattern between marriage market sex ratio and sex ratio at birth from cross-sectional data.

4.2 Individual level regression analysis

In this subsection, I perform regression analysis using individual level real fertility data. The data come from China Family Panel Studies (CFPS). CFPS is a nationally representative, annual longitudinal survey of Chinese communities, families, and individuals launched in 2010 by the Institute of Social Science Survey of Peking University, China. In the 2010 baseline survey, the CFPS successfully interviewed almost 15,000 families and almost 30,000 individuals within these families, for an approximately response rate of 79%(CFPS, 2010). 2012 follow-up wave covers newborn family members.

I construct two samples, namely the 2000 cohort and the 2010 cohort. The 2010 cohort is defined as children under 2 in 2012. In order to get a counterpart cohort with equal data quality and representativeness, I construct the 2000 cohort sample using observations of children aged 7-10 in the 2010 survey. By doing this, I might suffer from sample attrition bias. If mortality rate is higher for boys in poorer families, the sample thus gives false evidence for sex selection and SES. Children within the same cohort are pooled together to form a cross-sectional data.

The samples are limited to Han children with non-agricultural hukou. In China, the family planning policy imposes different regulation rules on parents with different residential registration status. In most rural areas, parents who first have a girl are allowed to have a second child, while in urban, "One Child Policy" is a near universal rule. The classification of urban and rural is based on parents' hukou status, which is inherited by the child. Therefore, to avoid unnecessary complication caused by variation in policy constraints, I restrict my sample to the first-born children with non-agriculture hukou (urban) at birth. Furthermore, the Family Planning Policy is not applied to minority groups, so I exclude them from my sample.

The empirical model is an individual level regression with region fixed effects. Region refers to the east, middle and west classification in China. Region fixed effects is included to capture unobserved regional factors such as prevalence of son preference culture and technology. There is not enough variation within each province to allow for province level fixed effects because few cities are sampled within each province.

$$\begin{aligned}
\mathrm{P}(Boy) &= \alpha child + \beta_1 mother + \beta_2 father + \theta localmarriagemarket + \delta city + i.region + \epsilon \\
\mathrm{P}(Boy) &= \alpha child + \beta_1 mother + \beta_2 father + \theta localmarriagemarket \\
&\quad + \gamma(localmarriagemarket \times mother'seducation) + \delta city + i.region + \epsilon
\end{aligned}$$

$mathrmP(Boy)$ is the probability of having a baby boy, and *child* is a dummy about whether the child was born in urban area (by geographic definition). Mother's characteristics, denoted as *mother*, include mother's age at the birth of the child and its quadratic form. Similar variables are controlled for father, except for the age quadratic form. *City* is a set of variables controlling for city characteristics as before, including GDP per capita of the year (log), GDP growth, proportion of the tertiary industry in the local economy, population density (log) and population growth rate.

Parents' years of schooling are included in the model as a proxy for SES. A lot of studies have proved the positive correlation between education and SES (Fershtman, Murphy and Weiss, 1996). Other commonly used SES variables such as income and working status are endogenous, in that the information is collected after the child was born. Incentives like preventive saving motive alter family resources allocations (Wei and Zhang, 2009). Besides, parents' income is affected by life cycle factors, which are hard to adjust.

The variable of special interest to this paper is the local marriage market sex ratio. Here, local marriage market is defined in prefecture-level city level, as discussed in section 4.1.2. In each city there are several counties and districts. Most non- agricultural hukou people live in urban areas, where transportation and communication across districts and counties are convenient enough to form an aggregate prefecture-level city marriage market. Therefore, it is reasonable to define local marriage market in prefecture-level city given my sample.

In the analysis, sex ratio in marriage market at the time of birth is used as a proxy for local marriage market uncertainty. The more severe the problem of sex ratio imbalance is, the worse the condition for males. By using the local sex ratio in marriage market at the time of birth, I make an important yet controversial assumption that parents do not foresee the future marriage market condition for their child. All the information available to the parents comes from the present situation. This assumption actually rules out rational expectation of prospective parents, which undermines the equilibrium sense of this model. To be specific, the local marriage market sex ratio is the city level ratio of males between 20-35 to all females between 20-30 regardless of their marital status. Here I cover a wider cohort of males than females in marriage market, as discussed in section 4.1.2. This sex ratio is the most statistically significant in the regression of sex ratio at birth on sex ratio at local marriage market.

The interaction term between mother's years of schooling in the second regression model captures the heterogeneity in response of parents of different SES (education) when facing the same local marriage market condition.

The model is estimated using weighted logit and OLS (Linear Probability Model). The standard error is clustered in city level.

Table 7 and Table 8 report the results of models for 2000 and 2010 cohort sample respectively. There are two basic specifications for each cohort, with and without the interaction term between mother's education and local marriage market sex ratio.

I detect signs of sex selection in the regression results, although in natural state without intended sex selection, none of the above variables should be significant. Across the regressions in Table 7 for the 2000 cohort sample, local marriage market sex ratio enters with the expected negative sign, though not significant before adding the interaction term and city level control variables. It is reasonable if divergent response of different parents when facing the marriage market conditions cancel out. In column (4) and (8) where the interaction term between mother's schooling and local marriage market condition is specified, we get a significantly negative coefficient of local marriage market sex ratio. The coefficient of the interaction term is positive and statistically significant, which means mothers with different years of schooling response to local marriage market sex ratio in different ways. For example, according to column (4), given a local marriage market sex ratio of 113 (sample mean), the probability of having a first-born boy is 0.27 higher with 3 more years of schooling of the mother. This result is consistent with the aggregate level evidence in section 4.1. Both of them depict the pattern of "positive hierarchical sorting" in 2000 as described in the model.

The results of 2010 cohort sample give evidence to "refugee sorting" instead. In column (4) and (8) with the interaction specified, we get a negative coefficient of mother's education and a positive one of the interaction term. As in the case of 2000, mothers of higher education are more sensitive to local marriage market condition. What makes it different from the case in 2000 cohort is that the coefficient of mother's education is larger than that in 2000 (-0.819 compared with -0.700). The pure effect of schooling now out weights the effect of its interaction with local marriage market sex ratio. According to column (4), given a local marriage market sex ratio of 96 (sample mean), the probability of having a boy is 0.15 lower with 3 more years of education now. In this sense, the general pattern reverses to become "refugee sorting", although the basic individual sex selection incentives stay the same. As the model predicts, the general pattern depends on the local marriage market condition.

4.3 The shift

The most interesting finding of this paper is the enigmatic shift of sex selection pattern between 2000 and 2010 in China. What happened during this period that led to such a fundamental reversal?

Recall the hypothesis of the model that when λ, the index of risk in marriage market, drops below a critical value, "positive hierarchically sorting" case transits to "refugee sorting". Tracing this model implication, I investigate the changes in local marriage market condition by checking the record of sex ratio in marriage market. It turns out that sex ratio in local marriage market dropped significantly during this period. It may not sound plausible since the sex ratio at birth has been kept high for decades. However, once we consider the population structure as a consequence of the family planning policy starting from the late 1970 and the baby boom in the early 1960 in China, things become reasonable. The birth heaping in the early 1960 as a compensatory population growth left a lasting bump in the age structure, and generated another bump, though less significant, when the baby boomers reached their reproductive age. As we can see from the following population pyramids, the bumps play a role in the marriage market. In 2000, given the gender difference in marriageable age range, the population in the range covers more cohorts before family planning policy for men than for women. However, in 2010 the bump falls outside of the age range for marriage, releasing the pressure in the main marriage market. In sum, the imbalance sex ratio is attenuated by the fluctuation of cohort size, causing variation of sex ratio in the marriage market despite of near constant sex ratio at birth.

The transition of sex ratio in marriage market is clearer in city level. The following graph plots the sex ratio in urban marriage market of each prefecture level city in 2000 and 2010. The fact that most dots lay far below the 45-degree line indicates that for most cities, sex ratios in marriage market are higher in 2000 than in 2010. Using the terms in the model, there has been a significant drop of λ and that may be the reason why "positive hierarchically sorting" case in 2000 transited to "refugee sorting" in 2010.

5 Conclusions

The main conclusion of this paper is that prospective parents respond rationally to the local marriage market in their sex selection decision. The extended Roy model of sex selection with son preference and reproduction uncertainty shows that when the future marriage uncertainty for boys is high enough, parents with higher SES choose boys. This kind of "positive hierarchical sorting" is observed in 2000. According to the model, when the future marriage uncertainty for boys drops below a certain threshold, parents with lower parenting ability and strong desire for son indulge their preference. Interesting enough, this phenomenon described as "refugee sorting" turns out to be the case in 2010. Between 2000 and 2010, the sex ratio in local marriage market witnessed a significant drop, which reflected the vicissitude of the recent population history in China. Notice that in the next 10 years, the population structure in the marriage market will be similar to the one in 2000. If the theory holds, we are going to observe "positive hierarchically sorting" in China again.

This paper proposes a long-term self-correcting equilibrium mechanism in fertility decisions. Sex ratio at birth can be endogenous, in that people response to the local marriage market sex ratio when making sex selection decisions. However, this self-correcting equilibrium mechanism, if exists, runs at the cost of potential inequality, when boys tend to come from better-off family while girls tend to be born in the lower class. That is the story of prince and Cinderella.

References

[1] Almond, D., Edlund, L. and Milligan, K. (2013). Son preference and the persistence of culture: evidence from South and East Asian Immigrants to Canada. *Population and Development Review*, 39(1), 75-95.

[2] Banister, J. (2004). Shortage of girls in China today. *Journal of Population Research*, 21(1), 19-45.

[3] Becker, G. S. and Lewis, H. G. (1974). Interaction between quantity and quality of children. In Economics of the family: Marriage, children, and human capital (pp. 81-90). University of Chicago Press.

[4] Bhat, P. M. and Zavier, A. F. (2007). Factors influencing the use of prenatal diagnostic techniques and the sex ratio at birth in India. *Economic and Political Weekly*, 2292-2303.

[5] Borjas, George J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, Vol.77, No.4. 531-553.

[6] Chen, Y., Li, H. and Meng, L. (2013). Prenatal Sex Selection and Missing Girls in China: Evidence from the Diffusion of Diagnostic Ultrasound. *Journal of Human Resources*, 48(1), 36-70.

[7] China Family Panel Survey – http://www.isss.edu.cn/cfps/

[8] Chu, Junhong. (2001). Prenatal sex determination and sex?selective abortion in rural central China. *Population and Development Review*, 27(2), 259-281.

[9] Coale, A. J. and Banister, J. (1994). Five decades of missing females in China. *Demography*, 31(3), 459-479.

[10] Cronk, L. (1989). Low Socioeconomic Status and Female?Biased Parental Investment: The Mukogodo Example. *American Anthropologist*, 91(2), 414-429.

[11] Das Gupta, M. (1987). Selective discrimination against female children in rural Punjab, India. *Population and development review*, 77-100.

[12] Das Gupta, M., Ebenstein, A. and Sharygin, E. J. (2011). China's Bride Shortage and Upcoming Challenges for Elderly Men'. Development Research Group. The World Bank.

[13] Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, X., Chung, W. and Hwa-Ok, B. (2003). Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *The Journal of Development Studies*, 40(2), 153-187.

[14] Ebenstein, A. (2010). The "missing girls" of China and the unintended consequences of the one child policy. *Journal of Human Resources*, 45(1), 87-115.

[15] Ebenstein, A. (2011). Estimating a dynamic model of sex selection in China. *Demography,* 48(2), 783-811.

[16] Edlund, L. (1999). Son preference, sex ratios, and marriage patterns. *Journal of Political Economy*, 107(6), 1275-1304.

[17] Edlund, L., Li, H., Yi, J. and Zhang, J. (2013). Sex ratios and crime: evidence from China. *Review of Economics and Statistics*, 95(5), 1520-1534.

[18] Fan, C. C. and Huang, Y. (1998). Waves of rural brides: Female marriage migration in China. *Annals of the Association of American Geographers*, 88(2), 227-251.

[19] Fershtman, C., Murphy, K. M. and Weiss, Y. (1996). Social status, education, and growth. *Journal of Political Economy*, 108-132.

[20] Gaulin, S. J. and Robbins, C. J. (1991). Trivers?Willard effect in contemporary North American society. *American Journal of Physical Anthropology*, 85(1), 61-69.

[21] Hudson, V. M. and Den Boer, A. M. (2004). Bare branches: The security implications of Asia's surplus male population (pp. 275). Cambridge, MA: MIT Press.

[22] James, W. H. (1995). What stabilizes the sex ratio?. *Annals of Human Genetics*, 59(2), 243-249.

[23] Johansson, S. and Nygren, O. (1991). The missing girls of China: A new demographic account. *The Population and Development Review*, 35-51.

[24] Li, X., Chan, M. L., Spencer, B. G., and Yang, W. (2016). Does the marriage market sex ratio affect parental sex selection? Evidence from the Chinese census. *Journal of Population Economics*, 1-20.

[25] Liao, Q., Cao, G. and Tao, R. (2012). The fertility intention and gender preference of migrants and their determinants. *Population and Development*, 2-12.

[26] Poston, D. L. and Glover, K. S. (2005). Too many males: Marriage market implications of gender imbalances in China. *Genus*, 119-140.

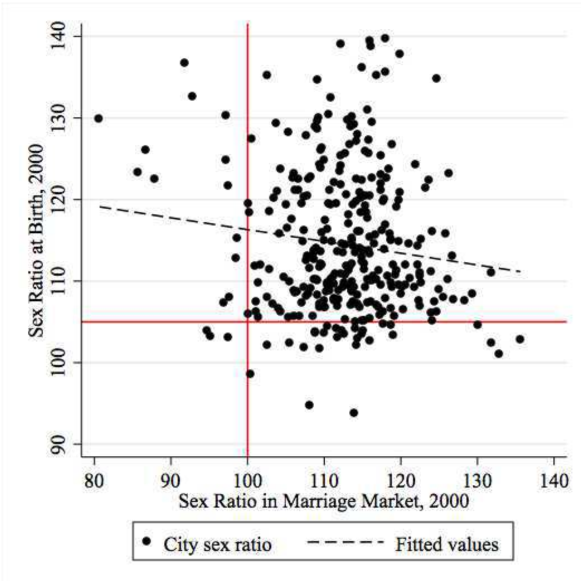
[27] Roy. A. D (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers, New Series*, Vol.3, No.2.135-146.

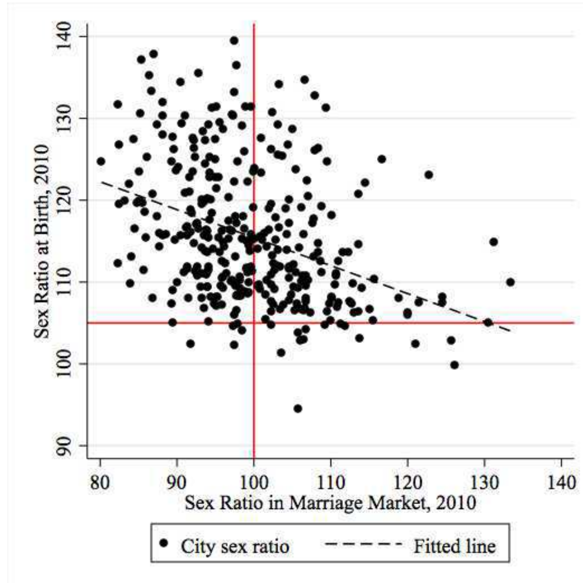
[28] Sen, A. (1990). More than 100 million women are missing. *The New York Review of Books.*

[29] Wei, S. J. and Zhang, X. (2009). The competitive saving motive: Evidence from rising sex ratios and savings rates in China (No. w15093). National Bureau of Economic Research.

[30] Zeng, Y., Ping, T., Baochang, G., Yi, X., Bohua, L. and Yongpiing, L. (1993). Causes and implications of the recent increase in the reported sex ratio at birth in China. *Population and Development Review*, 283-302.

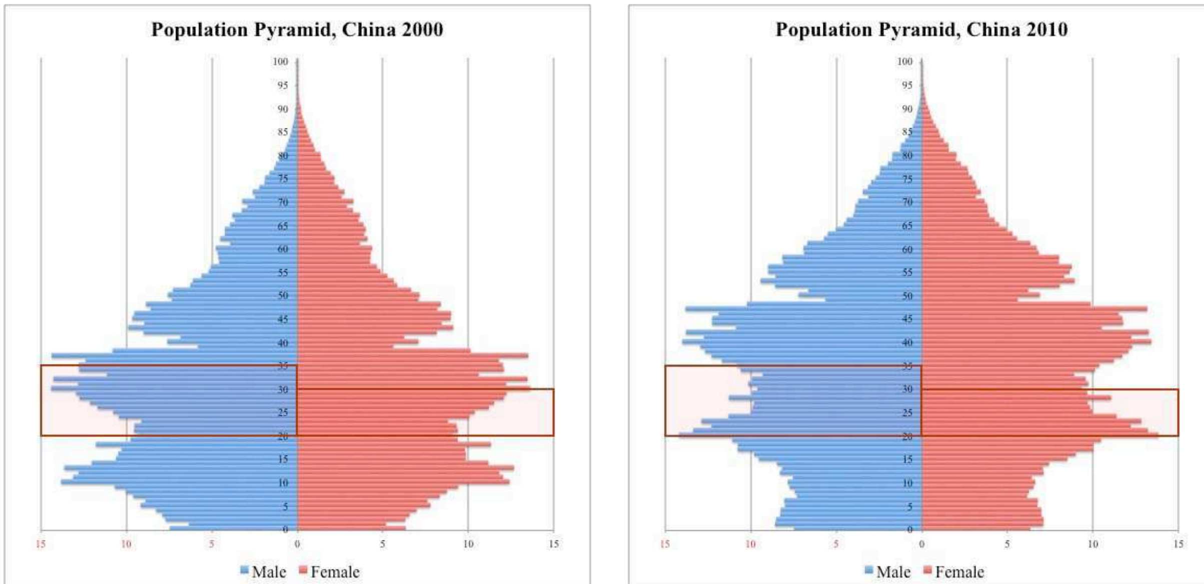
Figure 1: Sex ratio at birth and local marriage market sex ratio, prefecture-level city





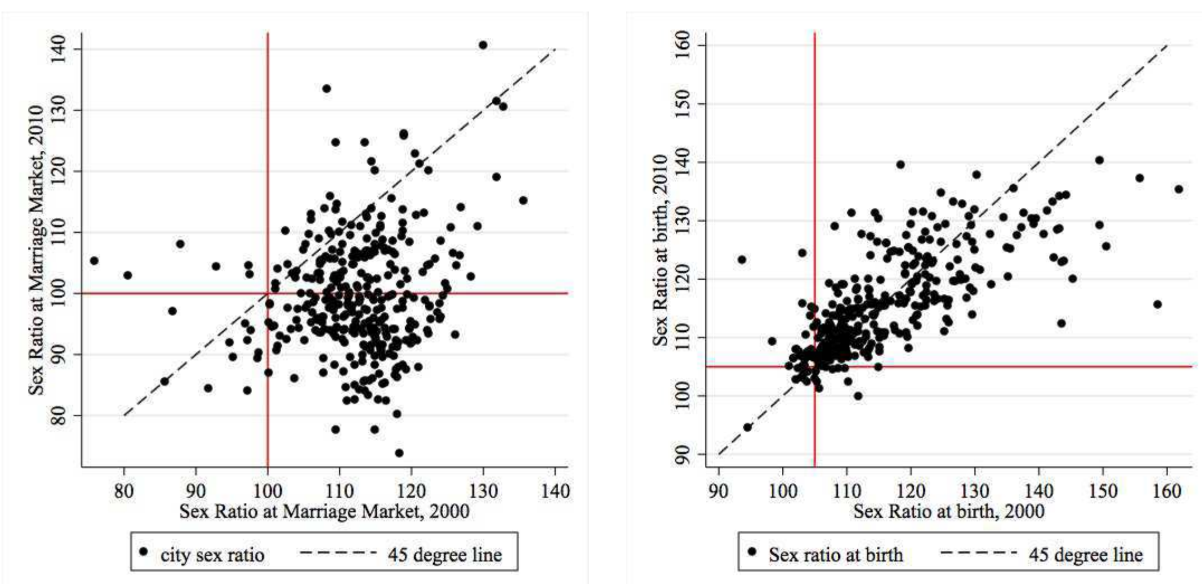
Source: Census data 2000/2010, National Bureau of Statistics, P.R.C.

Figure 2: Population pyramid of China, 2000 and 2010



Source: Census data 2000/2010, National Bureau of Statistics, P.R.C.

Figure 3: Sex ratio at marriage market and sex ratio at birth, city



Source: author's calculation based on census 2000/2010, National Bureau of Statistics, P.R.C.

Table 1: Sex selection rate by mother's education at first birth, China

Sex selection rate		2000 National Census		
Mother's education	National	Urban	Town	Rural
Illiterate	-7.65%	-5.00%	-7.05%	-7.85%
Primary school	-0.21%	0.23%	1.61%	-0.42%
Junior high school	1.46%	2.02%	2.29%	1.10%
High school or above	2.49%	2.06%	3.55%	2.43%
Sex selection rate		2010 National Census		
Mother's education	National	Urban	Town	Rural
Illiterate	3.09%	0.54%	2.63%	3.36%
Primary school	4.83%	6.10%	4.25%	4.75%
Junior high school	3.96%	5.44%	4.21%	3.42%
High school or above	2.89%	2.63%	3.87%	2.74%

Source: author's calculation based on census 2000/2010, National Bureau of Statistics, P.R.C.

Table 2: Sex selection rate by mother's education at second birth, China

Sex selection rate		2000 National Census		
Mother's education	National	Urban	Town	Rural
Illiterate	8.56%	11.13%	12.27%	8.21%
Primary school	17.10%	15.82%	17.09%	17.20%
Junior high school	21.87%	19.18%	21.50%	22.35%
High school or above	17.44%	14.18%	18.59%	19.09%
Sex selection rate		2010 National Census		
Mother's education	National	Urban	Town	Rural
Illiterate	10.73%	5.68%	6.60%	11.69%
Primary school	10.10%	11.90%	9.10%	10.06%
Junior high school	13.55%	14.93%	14.36%	12.95%
High school or above	14.71%	15.14%	14.08%	14.58%

Source: author's calculation based on census 2000/2010, National Bureau of Statistics, P.R.C.

Table 3: Summary statistics, prefecture-level city, China 2000

Year:2000 (obs=255)	Mean	S.D.	Min	Max
GDP per capita (Yuan)	15215	13750	2169	152099
GDP growth rate (%)	11.9	4.0	1.4	39.9
Proportion of tertiary industry (%)	41.0	9.8	7.3	75.0
Population density (people / sq km)	1150.4	1009.6	13.0	6416.0
Population growth rate (%)	7.3	5.3	-2.1	31.0
Local sex ratio				
At birth	118.4	11.8	102.1	170.6
(Male: 20-35; Female: 20-30)	112.2	7.5	75.9	135.6
(Male: 20-30; Female: 20-30)	103.3	6.4	76.5	120.1
(Male: 20-35; Female: 20-35)	103.8	5.9	80.3	121.0
(Male: 20-40; Female: 20-40)	104.2	5.6	82.4	121.5
(Male: All; Female: All)	106.4	4.0	89.9	124.4

Source: 2000 Census, National Bureau of Statistics, P.R.C.; CEInet Statistics Database.
Note: GDP per capital and population density data for year 2000 is not available for many cities in the sample. The missing values are replaced by their counterpart values in 2001.

Table 4: Summary statistics, prefecture-level city, China 2010

Year:2010 (obs=276)	Mean	S.D.	Min	Max
GDP per capita (2001, Yuan)	44865	28875	6222	249040
GDP growth rate (2002, %)	14.8	3.9	4.1	40.3
Proportion of tertiary industry (%)	41.5	11.1	9.8	78.7
Population density (people / sq km)	1015.9	1079.7	13.3	11449.3
Population growth rate (%)	5.6	6.9	-10.4	48.6
Local sex ratio				
At birth	116.7	8.3	102.3	140.3
(Male: 20-35; Female: 20-30)	97.9	8.9	73.7	133.5
(Male: 20-30; Female: 20-30)	100.8	7.6	80.1	136.3
(Male: 20-35; Female: 20-35)	101.5	7.2	82.0	137.9
(Male: 20-40; Female: 20-40)	102.2	6.9	84.4	139.3
(Male: All; Female: All)	105.1	4.2	89.7	132.6

Source: 2010 Census, National Bureau of Statistics, P.R.C.; CEInet Statistics Database.

Table 5: Regression of sex ratio at birth on local marriage market sex ratio, prefecture-level city, China 2000 / 2010

	Local sex ratio at birth, 2000				Local sex ratio at birth, 2010			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP per capita (log)	-4.037***	-3.170**	-2.805**	-3.191**	-0.659	-1.229	-1.095	-2.715
	(1.101)	(1.288)	(1.353)	(1.401)	(1.411)	(1.515)	(1.499)	(1.613)
GDP growth rate	-0.134	-0.111	-0.107	-0.083	0.009	0.003	-0.000	-0.025
	(0.125)	(0.124)	(0.121)	(0.185)	(0.096)	(0.102)	(0.102)	(0.098)
Population density (log)	0.206	0.257	0.134	0.620	-1.016	-0.407	-0.512	0.188
	(1.069)	(1.214)	(1.201)	(1.528)	(1.121)	(1.173)	(1.166)	(1.210)
Population growth rate	0.262	0.302	0.298	0.276*	0.147	0.177*	0.179*	0.135
	(-0.156)	(0.180)	(0.181)	(0.143)	(0.100)	(0.103)	(0.103)	(0.111)
Proportion of tertiary industry	-0.103**	-0.077*	-0.072	-0.069*	-0.111**	-0.088*	-0.081*	-0.101**
	(0.040)	(0.042)	(0.042)	(0.037)	(0.052)	(0.047)	(0.046)	(0.046)
Local sex ratio	-0.307**				-0.270***			
(Male: 20-35; Female: 20-30)	(0.143)				(0.080)			
Local sex ratio		-0.223				-0.143*		
(Male: 20-35; Female: 20-35)		(0.145)				(0.084)		
Local sex ratio			-0.336**				-0.168*	
(Male: 20-40; Female: 20-40)			(0.154)				(0.094)	
Local sex ratio				0.766***				0.284
(Male: All; Female: All)				(0.254)				(0.185)
Province Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	183.881***	162.343***	170.641***	56.919**	158.449***	147.878***	149.563***	116.125***
	(18.414)	(14.219)	(13.732)	(26.272)	(14.423)	(13.258)	(14.005)	(17.583)
Observations	255	255	255	255	276	276	276	276
R-square	0.476	0.462	0.470	0.489	0.462	0.434	0.436	0.437

Note: Robust standard errors in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 6: Summary statistics

	2010 cohort (obs=167)				2010 cohort (obs=356)			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Male	0.49	0.50	0	1	0.53	0.50	0	1
Urban	0.75	0.43	0	1	0.90	0.29	0	1
Mother's age at birth	26.97	4.40	18	37	26.26	4.21	17	39
Father's age at birth	29.11	4.32	21	41	28.51	4.89	18	48
Mother's schooling	10.10	5.35	0	18	10.28	3.77	0	16
Father's schooling	10.28	5.55	0	18	10.81	3.76	0	22
Local marriage market sex ratio	95.9	7.5	80.2	111.7	112.7	8.0	75.9	124.7
GDP per capita	54381	26920	6222	199657	20564	16614	2169	152099
GDP growth rate (%)	13.0	3.2	8.2	26.0	11.8	3.3	1.4	24
Population density	1689	1520	110	11449	1656.7	1126.9	132.0	4288.0
Population growth rate (%)	4.4	5.0	-4.3	14.9	6.3	5.5	-2.1	23.2
Proportion of tertiary industry (%)	45.5	11.7	14.0	75.5	42.0	10.4	7.3	63.5

Sample: CFPS 2010/2012, Han with non-agricultural hukou.
Note: (1) GDP per capita, GDP growth rate and population density information for 2000 is not available. Counterpart of year 2001, 2002 and 2001 is used as proxy for them respectively; (2) 151 (2010) and 296 (2000) observations with complete city level controls in the sample.

Table 7: Dependent variable: gender of the first-born child, cohort 2000 (age 7 to 10 in CFPS 2010 wave)

Sample: 2000 cohort	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Logit	Logit	Logit	Logit	OLS	OLS	OLS	OLS
Urban	0.456	0.392	0.472	0.428	0.109	0.093	0.113	0.101
	(0.613)	(0.510)	(0.616)	(0.502)	(0.148)	(0.121)	(0.148)	(0.119)
Father's age at child's birth	0.023	0.031	0.023	0.033	0.005	0.007	0.005	0.007
	(0.041)	(0.043)	(0.041)	(0.044)	(0.010)	(0.010)	(0.010)	(0.010)
Father's years of schooling	-0.070	-0.067	-0.068	-0.063	-0.017	-0.015	-0.016	-0.014
	(0.048)	(0.056)	(0.048)	(0.050)	(0.011)	(0.011)	(0.011)	(0.011)
Mother's age at child's birth	0.215	0.181	0.214	0.187	0.051	0.042	0.051	0.043
	(0.283)	(0.298)	(0.284)	(0.296)	(0.067)	(0.069)	(0.067)	(0.068)
(Age square)	-0.005	-0.005	-0.005	-0.005	-0.001	-0.001	-0.001	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)
Mother's years of schooling	0.062*	0.070**	-0.465	-0.700*	0.015*	0.016**	-0.105	-0.144
	(0.036)	(0.035)	(0.369)	(0.411)	(0.008)	(0.008)	(0.085)	(0.092)
Local marriage market sex ratio	-0.023	-0.028	-0.070**	-0.099**	-0.005	-0.006	-0.016**	-0.021**
(Male: 20-35; Female: 20-30)	(0.018)	(0.022)	(0.035)	(0.039)	(0.004)	(0.005)	(0.008)	(0.008)
Mother's years of schooling ×			0.005	0.007*			0.001	0.001*
Local marriage market sex ratio			(0.003)	(0.004)			(0.001)	(0.001)
GDP per capita (log)		0.006		-0.021		0.001		-0.004
		(0.206)		(0.201)		(0.049)		(0.048)
GDP growth rate		0.075*		0.082**		0.017*		0.018**
		(0.039)		(0.037)		(0.008)		(0.008)
Population density (log)		-0.113		-0.148		-0.025		-0.031
		(0.141)		(0.139)		(0.034)		(0.033)
Population growth rate		0.034		0.032		0.008		0.008
		(0.031)		(0.031)		(0.007)		(0.007)
Proportion of tertiary industry		-0.013		-0.015		-0.003		-0.003
		(0.012)		(0.012)		(0.003)		(0.003)
Region fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.350	0.774	4.990	9.137	0.416	0.652	1.620	2.355
	(4.300)	(4.804)	(6.066)	(6.699)	(1.026)	(1.155)	(1.399)	(1.522)
Observations	356	356	356	356	356	356	356	356
R-square					0.039	0.062	0.043	0.069

Note: Robust standard errors in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 8: Dependent variable: gender of the first-born child, cohort 2010 (age 0 to 2 in CFPS 2012 wave)

Sample: 2010 cohort	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Logit	Logit	Logit	Logit	OLS	OLS	OLS	OLS
Urban	0.141	0.399	0.102	0.387	0.032	0.090	0.024	0.085
	(0.481)	(0.502)	(0.452)	(0.478)	(0.116)	(0.121)	(0.106)	(0.111)
Father's age at child's birth	-0.108	-0.116	-0.111	-0.119	-0.024	-0.023	-0.024	-0.023
	(0.097)	(0.105)	(0.100)	(0.110)	(0.022)	(0.023)	(0.022)	(0.023)
Father's years of schooling	0.026	0.022	0.028	0.024	0.006	0.005	0.006	0.005
	(0.037)	(0.040)	(0.038)	(0.042)	(0.008)	(0.009)	(0.008)	(0.009)
Mother's age at child's birth	0.747	0.835	0.647	0.745	0.136	0.150	0.118	0.137
	(0.522)	(0.529)	(0.530)	(0.530)	(0.116)	(0.119)	(0.116)	(0.119)
(Age square)	-0.014	-0.016*	-0.012	-0.014	-0.003	-0.003	-0.002	-0.003
	(0.009)	(0.009)	(0.010)	(0.009)	(0.002)	(0.002)	(0.002)	(0.002)
Mother's years of schooling	-0.034	-0.022	-0.855*	-0.819*	-0.007	−0.004	-0.184*	-0.179*
	(0.038)	(0.039)	(0.494)	(0.477)	(0.009)	(0.009)	(0.107)	(0.103)
Local marriage market sex ratio	-0.005	-0.003	-0.095*	-0.085*	-0.001	-0.000	-0.020**	-0.019*
(Male: 20-35; Female: 20-30)	(0.025)	(0.032)	(0.049)	(0.048)	(0.006)	(0.007)	(0.010)	(0.010)
Mother's years of schooling ×			0.009*	0.008*			0.002	0.002*
Local marriage market sex ratio			(0.005)	(0.005)			(0.001)	(0.001)
GDP per capita (log)		-0.618		-0.718		-0.145		-0.160
		(0.507)		(0.526)		(0.107)		(0.106)
GDP growth rate		0.122		-0.103		0.023		0.020
		(0.104)		(0.093)		(0.021)		(0.019)
Population density (log)		0.034		0.106		0.015		0.027
		(0.352)		(0.343)		(0.077)		(0.074)
Population growth rate		0.020		0.009		0.004		0.002
		(0.048)		(0.048)		(0.010)		(0.010)
Proportion of tertiary industry		-0.003		-0.005		-0.001		-0.001
		(0.021)		(0.020)		(0.005)		(0.004)
Region fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-5.903	-2.358	4.039	7.757	-0.410	0.428	1.627	2.459
	(7.698)	(8.051)	(8.161)	(8.639)	(1.728)	(1.792)	(1.733)	(1.749)
Observations	157	157	157	157	157	157	157	157
R-square					0.108	0.142	0.132	0.165

Note: Robust standard errors in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.