# Individual vs. Social Motives in Identity Choice: Theory and Evidence from China[*]

Ruixue Jia[†]and Torsten Persson[‡]

May 23, 2015

### Abstract

This paper provides a framework to study how individual and social motives shape identity choice and applies it to the ethnic choice of children in ethnically mixed marriages. The model highlights the interaction of material benefits, identity costs, and social reputations. It is consistent with two motivating facts for ethnic choices in China, and also delivers a set of auxiliary predictions. The empirical tests on Chinese microdata find support for these predictions. In particular, social motives significantly crowd in changes in individual motives in some localities, and crowd out the same changes in other localities.

[†]School of International Relations and Pacific Studies, University of California San Diego, and CIFAR, rxjia@ucsd.edu.

[‡]Institute for International Economic Studies, Stockholm University, and CIFAR, torsten.persson@iies.su.se

# 1 Introduction

In this paper, we provide a theoretical framework to analyze how material benefits, intrinsic costs, and social norms interact in shaping the choice of ethnic identity. Specifically, we study how parents in ethnically mixed marriages choose the ethnicity of their children. Confronting the model predictions with microdata from Chinese censuses, we find robust empirical support that social motives strongly modify the effects of individual motives on choice. These results are certainly quite specific to ethnicity choices in China, but we believe that they also speak to a more general set of issues in the social sciences.

In their theoretical and empirical work, economists typically consider how *individual*, most often material, motives shape individual decisions. By contrast, other social scientists – like sociologists and social psychologists – mainly consider how *social* motives shape individual decisions. To caricature and quip: most economists still think about how individual decisions drive social outcomes, while most sociologists still think about how social outcomes drive individual decisions. So what, the reader may ask – the different approaches may both be valuable and reflect an effective division of labor in the social sciences. Perhaps that is so, but there is also a risk that important issues fall in the cracks between different disciplines.[1]

One such issue is the interaction between individual and social motives. It is easy to think about individual economic, political or social choices in which both individual and social motives are important: e.g., tax compliance, political participation, and fertility decisions. Suppose the government intervenes to encourage a certain choice, by modifying some individual motive that it can influence. Do the social motives help or hinder that intervention? Put differently, are the stronger individual motives crowded in or crowded out by the social motives? We do not know too much about this general issue, at least not empirically.

This paper studies a specific instance of the general issue: how do social attitudes and government policy interventions jointly shape ethnic identification? This answer to this question could have important implications. For instance, conflicts could be exacerbated by institutions that motivate individuals to socially identify with specific ethnic groups (Horowitz, 2000).

Ethnic identification is the broad topic of seminal research by Bisin and Verdier (2000). Motivated by a large sociological literature, these authors aim to theoretically understand why cultural convergence is so slow, even in the US. They model the persistent propensity of ethnic and religious minorities to marry within their own kin and socialize their children in the same mold. Bisin, Topa and Verdier (2001) formulate

---

[1]Of course, there is a growing literature in economics on individual choices and social interactions. See e.g., Brock and Durlauf (2001) and Blume et al (2011) for discussions of the general issues of the economics and econometrics of social spillovers in that literature. There is also a related (and older) literature in sociology, called economic sociology. See Smelser and Swedberg (2005) for an exhaustive survey. To the best of our knowledge none of these litteratures have addressed the general issue we fouce on here, namely the how individual and social motives interact to shape individual choices.

a related model of cultural transmission, which they structurally estimate on US data for different religious groups.

In much of the literature, the desire to adopt a certain ethnicity or religion has intrinsic motives with social and psychological roots, such as a desire for social recognition or self esteem. Yet, history is ripe with examples of groups that gradually or suddenly change their identity to reap individual material benefits. For instance, Bates (1974) discusses how economic and political change drove emerging ethnic groups to compete for the spoils of patronage in post-colonial Africa. The case studies in Vail (1989) describe how people in different parts of southern Africa came to identify with vaguely defined ethnic groups in colonial times, but how these became major interest groups in post-colonial times to lay claims to government resources. Botticini and Eckstein (2007) demonstrate how material incentives played an important historical role in individual transitions between Judaism and Christianity. Cassan (2013) shows how higher-caste groups in Punjab, at the turn of the past century, adopted a lower-caste identity, in order to take advantage of a large land-distribution program.

Existing research thus emphasizes social motives as well as individual motives. But it is likely that persistence or change of ethnic or religious identities reflect the interplay of the two types of motives: a tradeoff between material benefits and the costs shaped by existing social self-images or norms. Which way social motives tilt that tradeoff is far from clear, however, so we do not want to decide on this issue *a priori*. Indeed, recent theoretical research by Benabou and Tirole (2011) shows that higher individual incentives to make a certain choice may be either crowded out or crowded in by concerns for social reputations, and how this is driven by the choices of other members each individual's peer group.

China is an interesting testing ground when it comes to government policies and family choices in the realm of ethnicity. A multiethnic society with 55 officially recognized ethnicities beyond the dominant Han (about 91.5% of the population), the country is still relatively homogenous despite some ethnic tensions with occasional riots in Tibet and Xinjiang. Meanwhile, the population shares of different ethnicities are quite dispersed across China's regions: the combined share of all minorities ranges from 0.3% in Jiangxi province to 94% in Tibet. Also, the national and provincial governments have made policy interventions that remind of affirmative action for US minorities. Moreover, mixed ethnic couples are free to choose whichever of their two ethnicities for their children at birth and we can observe these choices directly in the data, Section 2 describes and discusses these issues further.

Two facts on the ethnicity of children stand out from the data.[2] One is:

**F1** *The propensity to choose minority identity for children is much higher in mixed couples with a minority man and a Han woman than in those with a Han man and a minority woman.*

---

[2]We use the 1982, 1990, 2000 censuses and the 2005 mini-census – see Section 4 for more detail on sources.

On average, the probability of having minority children in minority-man and Han-man mixed marriages are 94 percent and 44 percent, respectively. Figure 1 plots this probability over time, by five-year birth cohorts, for the two types of mixed marriages. The figure illustrates another fact:

**F2** *The share of minority children in mixed marriages are clearly increasing over time in mixed couples with a Han man, especially after 1980.*

Average minority identity among the children of such couples is 38 percent in cohorts born before 1980 but 49 percent in cohorts born after 1980. Differently, we observe little change in mixed couples with a minority man: 95 percent have minority children in cohorts born before 1980 against 93 percent after 1980.

### [**Figure 1 about here**]

Against this background, our paper studies ethnic choices in China theoretically and empirically. Theoretically, we set up a model for the choice of ethnicity for children that is consistent with F1-F2. Building on Benabou and Tirole (2011), we study the interplay between individual and social motives. Mixed couples thus make decisions on their children's ethnicity based on three interacting factors: material benefits (tied to policies favoring minorities), identity costs (of having a child with a different ethnicity), and social norms (as reflected in choices of other mixed couples in the same peer group). Having shown that the model implies facts F1-F2, we derive a set of new predictions that can be empirically tested with Chinese microdata.

The most important new prediction is that the effect of higher material benefits should be larger in regions and peer groups where the initial share of minority children is smaller, because individual motives driven by material benefits are crowded in rather than crowded out by social motives. A variant of this prediction concerns the effects of benefits in peer groups with different quartiles of shares of minority children. We find strong and robust empirical support that the predicted effects are statistically as well as quantitatively significant. To the best of our knowledge, no earlier empirical work – on this topic or others – arrives at such results regarding the interaction between individual and social motives.

The model has two additional auxiliary predictions, which we take as sanity checks on the model that delivers our main prediction. One concerns a heterogenous effect of material benefits across different minorities, and the other concerns the interaction effects between materials benefits and intrinsic identity costs. We explore information on differential ethnic policies, gender of children, as well as religion and find that these predictions too are borne out by the data.

Of course, it is important to consider alternative explanations for the patterns in the data. We discuss three alternatives – changing bargaining power of women, endogenous mixed marriages, and a kind of censoring – in some detail. While we find that some of these may contribute to the changing ethnicity of children to mixed couples, none

of these explanations change our main empirical results on the interaction between individual and social motives.

Generally, our study contributes to a better understanding of the interplay between material incentives and social motives when identity is a choice. Thereby, it provides a complementary perspective to the studies of ethnicity choice discussed earlier in this section. Our our paper also adds to the few existing studies of ethnicity in China done by sociologists. Guo and Li (2008) document a pattern similar to F1. Relying on the 1-percent sample of the 2000 census, they find that the average probability of having a minority child is more than one half, and argue that this raises the minority population share over time. To the best of our knowledge, no existing research has systematically analyzed ethnic decisions in China from a rational-choice perspective. Our paper tries to fill the gap.

However, the main variables we explore to test our theory and alternatives to it – such as the rollout of one-child policy and sex ratios – have been widely used in other contexts. For instance, Ebenstein (2010) examines how the one-child policy has affected sex ratios in China, Wei and Zhang (2011) show that higher male-female ratios related to the one-child policy might explain a large part of increased Chinese saving rates, whereas Edlund et al. (2013) document that higher sex ratios has led to more crimes.

Our study also makes a more general contribution. The method we suggest to estimating the interaction of individual and social motives can also be applied to other individual choices in the economic, political or social arena. For instance, Besley, Jensen and Persson (2014) apply a related model to derive predictions so as to empirically study the evasion from local property taxes in the UK.

The next section describes the institutional background to our study. Section 3 formulates our model and spells out its predictions. Section 4 discusses which data can be used to test these predictions. Section 5 confronts the main predictions on the interaction between individual and social motives with data. Section 6 tests two additional predictions of the model. Section 7 discusses alternative explanations for the patterns in the data and whether these explanations drive our main results. Finally, Section 8 provides a brief conclusion. To save space, some additional estimation results are relegated to a Web Appendix.

# 2   Background

China is composed of 56 ethnic groups, namely the dominant Han plus 55 minorities. As of 2000, the combined population of minority groups stood at about 106 million, circa 8.5% of the total population on the mainland. The 55 minority groups vary widely in size. With a population of more than 15 million (in 2000), the Zhuang is the largest minority group and the Lhoha, with only 2,965, the smallest. Minority groups also vary greatly in culture, spoken language and religious practice – 53 minority groups speak languages of their own, 23 have their own written language, 10 groups are

predominantly Muslim, and seven follow Tibetan Buddhism. Some minority groups, like the Uighurs, look physically very different from Han Chinese while other groups look broadly similar to the Han.

## 2.1  Costs and Benefits of Having a Minority Child

There is no legal barrier for mixed marriages between any two ethnic groups. At the birth of their child, a mixed married couple can choose either ethnicity for their children. Along with the name and the birth date, one's ethnic identity appears in almost every context, including the birth certificate and all the forms which have to be filled out at school. As a result, the chosen ethnicity can be thought of as public information to peers. Choosing minority identity brings both benefits and costs for the child, and hence indirectly for the parents.[3]

The benefit side comes from the ethnic policies of the governments. Since the beginning of the People's Republic of China (1949-), the government has employed different policies to the benefit of ethnic minorities. Such policies exist in three areas:

(i) *Family planning.* When the family-planning policy started in the 1960s, minorities were more favorably treated than the Han majority. Over time, there has also some regional variation in the treatment of different minorities. As detailed in Section 4, family planning policies became much more stringent in the years around 1980 with the implementation of the one-child policy.

(ii) *Entrance to higher education.* Since the restoration of entrance exams to colleges in 1977, minorities have enjoyed additional points in the exams that decide upon the entry to different levels of education, especially college and high school. These benefits also vary by province.

(iii) *Employment.* The national ethnic policy states that minorities should have favorable treatment in employment. However, explicit quotas for minority employment are rare. As minorities are often discriminated in employment, it is unclear that this policy would make people tend to choose minority identity for children.

The cost side of having minority children has two aspects. First, there might be some discrimination against a minority in the labor market, even for people with similar educational background. However, this cost may be less critical since being a minority increases the chance of receiving higher education due to the ethnic policies. At the birth of a child, these benefits are likely to dominate the potential discrimination costs in the labor market.

The main cost of having a minority child is the identity cost, especially for Han fathers. In a patriarchal society such as China, children are expected to follow the

---

[3]According to government regulation, couples with the same ethnicity cannot choose any other ethnicity for their children. Regarding switches later in life, children from mixed marriages can apply to change their ethnicities given at birth before the age of 20. However, the applications have to be made by the parents for those younger than 18. Since these applications are costly and approval is uncertain, the impact of policy interventions on switches later in life should be much less important than the ethnicity choices by parents at the birth of their children.

ethnicity and family name of the father.[4] Additionally, the identity cost are likely to be affected by social norms, as expressed by the choices of a relevant peer group. Therefore, a Han man will face a trade-off between material benefits (on behalf of the child) and social status. The problem for a minority man is rather different, as having a minority child is the main expectation of society.

Based on these considerations, our model incorporates three different motives: individual material benefits, intrinsic benefits or costs, and social reputations. We will also build in the asymmetry for Han and minority men. Before presenting the model, we highlight some anecdotal evidence that illustrate these motives.

## 2.2 Some Anecdotal Evidence

There is almost no existing research on ethnic choices for children in China. However, one can find numerous discussions online among parents.

One example of a suggestive discussion [in our own translation from Chinese] appears at the website babytree.com:[5]

> *Anonymous* asked: "If the father is a Han and the mother is a minority, could the child be a minority?"
> *Biruoxinian* answered: "The child usually follows the father's ethnicity. It is also fine if you insist on following the mother's."
> *Xiximami12* answered: "You can follow the mother's, because there are ethnic favors toward minorities."
> *Sankouzhijiatu* answered: "The child should follow the father's ethnicity. Only the children of a live-in husband will follow the mother's."

These answers suggest that the prosocial norm is that children follow the fathers' ethnicity. However, benefits due to ethnic policies are considered as a motive in the other direction. In addition, a Han man with a minority child is considered to be of lower status – as the wife usually go to live in the husband's family, only lower-status men (with children of minority ethnicity) will consider becoming live-in husbands.

Another illustrative discussion [also in our own translation from Chinese] can be found at the website jzb.com (the meaning of jzb in Chinese is parents' helper):[6]

> *Zhongermen* said: "I went to register the birth of my child a while ago. I am a Han man and my wife is a minority. I told the police that I want my child to be a Han. The police kindly suggested that I should choose minority for the child. She said that one score lower implies an extra playground of competitors in the high-school entrance exam and that I should

---

[4]The link between family name and ethnicity is not very close for most of China's ethnic minorities. Therefore, it is difficult to build an empirical strategy upon family names as a source of variation.

[5]http://www.babytree.com/ask/detail/3690549

[6]http://jzb.com/bbs/thread-335421-1-1.html?action=printable

be responsible for my child's future. But I insisted on choosing Han in the end. I hope that my child's future will reply on his own ability, not ethnic favors."

*fh2315* remarked: "Choosing minority is not a big deal if the minority does not practice religion."

*hello_friend* remarked: "You will regret at the time of the high-school exam! You'd better be more realistic!"

*claetitia* remarked: "Well, if you despise the ethnic favor for extra scores, minorities can at least have more children!"

*donna2276* remarked: "First, I admire you. Many people would feel ashamed of their choices. Second, I would like to point out your mistake. Many children fail the exam due to one score less. Do you want your kid to fail?"

Once again, these arguments reflect the tradeoff between material individual benefits and intrinsic or social costs when choosing ethnic identity for children.

## 2.3 Patterns of Mixed Marriages

To be sure, entering into a mixed marriage is also a subject of choice. Among the married couples that appear in all four censuses, the probability for a minority man to marry a Han woman is 17% and that for a minority woman to marry a Han man is 18%. But the difference across genders is much less striking than the corresponding difference in black-white marriages in the US, where 6% of black men marry white women while 2.9% of black women marry white men around the year 2000 (Fryer, 2007).

Table A1 in the appendix presents the patterns of four types of marriages, as well as the information on the education and age differences between the husband and the wife. Compared with couples of the same ethnicity, the education difference between the parties in mixed couples is slightly lower, suggesting a bit more assortative matching in the dimension of education. The age difference between husband and wife does not differ substantively across marriage types.

The probability of mixed marriage has changed over time and also varies across regions. For instance, the probability to marry a Han man for minority women born in the 1940s (and hence married in the 1960s) was 15%, whereas it went up to 21% for minority women born in the 1970s (and hence married in the 1990s). This increase is likely to be correlated with the ethnic policies favoring minority children – no specific policy favors mixed marriages as such, but the benefits for the children affect the "continuation value" for mixed marriages. Huang and Zhou (2015) argue that the one-child policy has raised the probability of mixed marriages in China.

Our approach focuses on the choice of ethnicity for the children, given an earlier choice to enter into a mixed marriage. This is reasonable because, at least in China, the marriage choice has generally been made before having a child. The tradeoffs between material benefits and social costs in the web discussions cited above concern couples

who already had their child. In our empirical analysis to follow, we will still check that our results are robust to the cohort-specific frequency of mixed marriages. We will also present the results for a subsample where the couple married before the ethnic policies and their children were born after the policies. Endogenous mixed marriage is less of a concern for this subsample.

# 3  The Model

We use a framework similar to the one in Benabou and Tirole (2011) to model the ethnicity choice for children as a choice that involves individual (material and intrinsic payoffs), as well as social payoffs (social norms and culture). But we extend their setup to encompass two different groups. As the main purpose of the model is to set the stage for our empirical work, we include only prospective determinants of ethnicity choices that can be measured – or proxied – with some degree of confidence. These include material benefits for minority children, cultural differences across ethnicities, and (the reflections of) social norms. As further discussed in Section 4, we can gauge these determinants at the regional level and some at the individual level.

   Our model is certainly highly stylized. However, it is not only consistent with facts F1 and F2, but it also yields several additional and testable predictions. In particular, it provides a new prediction on the interaction between material benefits and social motives, the main issue we focus on in this paper. We think about the other two predictions more as suggesting additional sanity checks of our model.

## 3.1  Setup

Consider a region – a prefecture, to be concrete – with a continuum of households in a given cohort. There are two ethnicities $J \in \{H, M\}$, where $H$ denotes Han and $M$ Minority. Households have children that yield the same basic benefit $v$ for everyone. Each household has a single binary decision to make: whether to choose minority status for their children, $m = 1$, or not, $m = 0$.[7] In line with the social situation in China, we assume that this choice primarily reflects the husband's preferences. We focus on the decisions by mixed couples $(H, M)$ or $(M, H)$, where the first entry is the ethnicity of the man. Non-mixed couples are not considered, as they always choose their joint ethnicity for their children (they have no other choice). The framework considers individual – material and intrinsic – motives as well as social motives and, not the least, the interaction between the two.

---

[7]We focus on the ethnicity of children, given the number of children. One might argue that the number of children would be another outcome to study. However, Guo and Li (2008) report that the number of children in mixed marriages does not seem to be significantly different from that in Han-Han marriages, while the number is significantly higher in Minority-Minority marriages. As we find a similar pattern in our own data, we leave the analysis of the number of children to future research on marriage choices.

**Han-Minority mixed couples**   Suppose first that the man is Han and the woman is minority. Then, the preference function of the couple is

$$v + (b - e(H) - \varepsilon)m + \mu E(\varepsilon \mid m) , \tag{1}$$

where $b$ is the net *material individual* benefit of having minority children, which is controlled by the regional government. This parameter could differ across regions or time, due to different policies favoring minority children (such as these children being allowed to have more children conditional on marrying a minority, or advantages in college entrance exams). Further $e(H) + \varepsilon$, is the *intrinsic individual* cost of having a minority child (different from the Han man's own ethnicity). Its first component is the average stigma perceived individually by the household when the ethnicity of their child does not coincide with that of the Han man. By definition, this component is common and deterministic to everyone in the same peer group (region), but it could differ across these groups; it could also differ across ethnicities depending on "cultural or linguistic distance". The second component $\varepsilon$ captures the variation in intrinsic cost and is the main source of heterogeneity in the model. We assume that $\varepsilon$ is distributed across couples with mean $E(\varepsilon) = 0$, c.d.f. $G(\varepsilon)$, and continuous, differentiable, and single-peaked p.d.f. $g(\varepsilon)$.

According to these individual motives alone, households with high values of $\varepsilon$ would have a Han child, while those with a low value would have a minority child. In line with the discussion in Section 2, we assume that having a Han child is the prosocial choice, according to the prevailing social norm in the peer group. The the final term in (1) is a way to model the strength of this social motive. Specifically, it captures the household's *social reputation*, or self image – how society views the mixed couple, or the couple views itself – given the ethnicity decision that it makes. Taken literally, the model thus assumes that the choices of $m$ are perfectly observable by everybody in the peer group. In reality, observability is indeed realistic since the ethnic choice follows the child through life, as discussed in Section 2.[8]

Because households with high values of $\varepsilon$ make the prosocial choice, we assume that the household's social reputation is given by its "expected type" $E(\varepsilon \mid m)$, the conditional mean of $\varepsilon$ of those couples in the relevant *peer group*, who make the same choice as the couple does. Parameter $\mu$, is the weight on this social reputation relative to the individual payoff. Below, we often label this concern as the household's social motive.

**A cutoff rule**   It is useful to define the difference

$$\Delta = E(\varepsilon \mid m = 0) - E(\varepsilon \mid m = 1) . \tag{2}$$

The value of $\Delta$ is the *gain* in social reputation to the couple within its peer group from conforming to a norm of giving its child the same ethnicity as the Han man rather than

---

[8]The model can easiliy be modified to allow for stochastic observation – in that case, one part of parameter $\mu$ reflects the probability that $m$ is observed.

minority ethnicity. In the language of Benabou and Tirole (2011), the first term is the social "honor" when the child has Han identity. This will be the choice of couples with sufficiently high $\varepsilon$. The term deducted from this honor is the social "stigma" when the child has minority identity. This will be the choice of those with sufficiently low values of $\varepsilon$.

With this notation, it follows from (1) and (2) that the mixed couple will be indifferent about the child's identity when

$$b - e(H) - \varepsilon_H^* = \mu\Delta(\varepsilon_H^*) \ . \tag{3}$$

This equality implicitly defines a cutoff value of $\varepsilon$, below which couples have minority children and above which it will have Han children. For the marginal couple, the net individual benefit of having a minority child (the LHS) is equal to the gain in social reputation of having a Han child (the RHS). By (1), $\varepsilon_H^*$ is a function of $b, e$ and $\mu$. Based on this cutoff rule, the equilibrium gain in social reputation becomes

$$\Delta(\varepsilon_H^*) = E(\varepsilon \mid \varepsilon > \varepsilon_H^*) - E(\varepsilon \mid \varepsilon < \varepsilon_H^*) > 0 \ . \tag{4}$$

By definition of truncated means (of a mean-zero variable), the first term is always positive and the second term is always negative. Hence, $\Delta(\varepsilon_H^*)$ is always positive. Note also that for the whole peer group, social reputation is like a zero-sum game: under a veil of ignorance about $\varepsilon$, the ex ante expected value of $\mu E(\varepsilon \mid m)$ is zero (since the unconditional mean of $\varepsilon$ is zero).

The properties of the equilibrium and its comparative statics crucially reflect the sign of $\frac{d\Delta(\varepsilon_H^*)}{d\varepsilon}$, i.e., the derivative of the gain in social reputation from Han children. Suppose $\varepsilon_H^*$ goes up, such that more Han-minority couples have minority children. Then, both the honor and the stigma terms in (4) go up, so the question is which goes up by more. By the results in Jewitt (2004), the single peak of $g$ implies that $\Delta$ has a unique interior minimum. Moreover, for low values of $\varepsilon_H^*$, when few mixed couples (with Han men) have minority kids, the stigma goes up faster than the honor, such that $\frac{d\Delta(\varepsilon_H^*)}{d\varepsilon^*} < 0$. Therefore, the gain in social reputation from conforming with the norm of having a Han child goes down, so more couples have a minority child. It follows that ethnicity choices for children are strategic complements in the part of the distribution where $\frac{d\Delta(\varepsilon_H^*)}{d\varepsilon^*} < 0$. By an analogous but converse argument, for high values of $\varepsilon_H^*$ when many Han-minority couples have minority kids, we have $\frac{d\Delta(\varepsilon_H^*)}{d\varepsilon^*} > 0$ which makes ethnicity choices strategic substitutes.

For some of results below, we also assume that the second derivative of $\Delta(\varepsilon_H^*)$ is everywhere positive $\frac{d^2\Delta(\varepsilon_H^*)}{d\varepsilon^{*2}} > 0$. Similarly to Benabou and Tirole (2011), we assume that $1 + \mu\frac{d\Delta(\varepsilon_M^*(b,e,\mu))}{d\varepsilon^*} > 0$ – this guarantees that $\mu$ is not large enough to create multiple equilibria (when ethnic choices are strategic complements).

**Minority-Han mixed couples**   In a $M, H$ mixed couple, where the man is minority rather than Han, the preference function analogous to (1) can be written:

$$v + mb - (1 - m)(e(M) + \varepsilon) + \mu E(\varepsilon \mid m) \ , \tag{5}$$

where $e(M)$ and $\varepsilon$ now represent the average and idiosyncratic intrinsic cost of having a Han child, different from the minority man's own ethnicity.

We specifically assume that the distribution function $G$ for $\varepsilon$ and the weight on social reputation $\mu$ are exactly the same in the two types of families in the same locality. This is a strong assumption, although one can think of arguments why $\mu$, say, could be either higher or lower among minorities than majorities – the former may be more eager to fit in or more eager to preserve their identities. We do not pursue this issue further, however. The main argument is measurement: since proxies for $\mu$ and the distributions of $\varepsilon$ would be very hard to find in available data, theoretical predictions would be empirically empty.

The mixed couple will have a Han child when $b + \mu E(\varepsilon \mid m = 1) > -(e(M) + \varepsilon) + \mu E(\varepsilon \mid m = 0)$. Defining the gain in social reputation in an analogous way as before – i.e., $\Delta$ is the difference between the honor of having a minority child, $\mu E(\varepsilon \mid m = 1)$ and the stigma of having a Han child, $\mu E(\varepsilon \mid m = 0)$ – we can write the indifference condition for having a minority child as

$$-b - e(M) - \varepsilon_M^* = \mu \Delta(\varepsilon_M^*) \ . \tag{6}$$

Thus, minority-Han households with $\varepsilon$ larger (smaller) than $\varepsilon_M^*$ will have minority (Han) children. Because $\Delta$ is always positive, it follows that $\varepsilon_M^* < 0$.

## 3.2 Consistency with F1 and F2

**Comparison across mixed marriages ($\varepsilon_M^*$ versus $\varepsilon_H^*$) – Fact F1**   Having formulated the child stage of the model, we show that its implications for ethnicity choices are consistent with facts F1 and F2 noted in the introduction.

In terms of the model, F1 requires that (in the majority of prefectures) $G(\varepsilon_H^*) < 1 - G(\varepsilon_M^*)$. Suppose first that the distribution of $\varepsilon$ is symmetric. By symmetry, $G(-\varepsilon_M^*) = 1 - G(\varepsilon_M^*)$. It follows from (6) and (3) that $\varepsilon_H^* < -\varepsilon_M^*$. Since $G$ is increasing, a symmetric distribution indeed implies F1. If the distribution of $\varepsilon$ has positive skew, with more mass in its right tail, we have $G(\varepsilon_H^*) < G(-\varepsilon_M^*) < 1 - G(\varepsilon_M^*)$. A sufficient condition for the model to be consistent with fact F1 is thus that the distribution of $\varepsilon$ does not have strong enough negative skew.

The intuition is straightforward: on average, minority men experience both extrinsic material benefits $b$ and intrinsic benefits $e(M)$ of a minority child, so – compared to Han men – more of them will choose minority identity for the children.

**The effect of material benefits ($b$) – Fact F2**   How does a Han-minority family react to an increase in material benefits, $b$? Consider the probability of seeing minority kids in a population of such couples. This probability (share) is just $G(\varepsilon_H^*(b, e, \mu))$, given that the cutoff value $\varepsilon_H^*$ itself is a function of the benefits and costs of having minority kids. Using the definition $b - e - \varepsilon_H^* = \mu \Delta(\varepsilon_H^*)$, we can calculate how the

probability (share) of minority kids shifts in response to a higher net benefit:

$$\frac{\partial G(\varepsilon_H^*(b,e,\mu))}{\partial b} = g(\varepsilon_H^*(b,e,\mu)) \frac{1}{1 + \mu \frac{d\Delta(\varepsilon_H^*(b,e,\mu))}{d\varepsilon^*}} > 0 \; . \tag{7}$$

Similarly, the effect of material incentives $b$ on the probability (share) of minority kids for a minority-Han family is:

$$\frac{\partial (1 - G(\varepsilon_M^*(b,e,\mu)))}{\partial b} = g(\varepsilon_M^*(b,e,\mu)) \frac{1}{1 + \mu \frac{d\Delta(\varepsilon_M^*(b,e,\mu))}{d\varepsilon^*}} > 0 \; . \tag{8}$$

Thus, higher material benefits raises the probability of having a minority child in both types of families, everything else equal. But we can say more. First, comparing the two expressions, $g(\varepsilon_M^*(b,e,\mu))$ is smaller than $g(\varepsilon_H^*(b,e,\mu))$, because minority-man couples having Han children is more of a tail event than Han-man couples having minority children. Second, the social-reputation derivatives fulfill $\frac{d\Delta(\varepsilon_M^*(b,e,\mu))}{d\varepsilon^*} > \frac{d\Delta(\varepsilon_H^*(b,e,\mu))}{d\varepsilon^*}$ – ceteris paribus, fewer Han-man mixed couples than minority-man mixed couples have minority children. Comparing the marginal Han man to the marginal minority man, having a minority child is thus a strategic complement rather than a strategic substitute – or, if it is a strategic complement (substitute) for both, complementarity (substitutability) is larger (smaller). Compared to mixed couples with minority men, concerns for social reputation are thus more likely to crowd in rather than crowd out material incentives in mixed couples with Han men – or more likely to crowd them in more (crowd them out less). For these two reasons, the effect of material incentives is always higher for Han-minority families.

This prediction, *together* with the fact that benefits for minority children have gone up over time (see Section 3 for a discussion of the benefits), makes the model consistent with fact F2: a more pronounced trend over time to have minority kids in mixed marriages with Han men than in those with minority men.

Having established consistency between our model and facts F1 and F2, we turn our interest to some auxiliary predictions from the model. These are the ones we will test empirically.

## 3.3 Main Prediction: Material benefits ($b$) and social motives ($\frac{d\Delta}{d\varepsilon^*}$)

Our most important prediction concerns the strength of the interaction between individual (material) motives and social motives. We focus on the effects on mixed households with Han men. From (7), material benefits are crowded in by social reputation – i.e., the social multiplier $\frac{1}{1 + \mu \frac{d\Delta(\varepsilon_H^*(b,e,\mu))}{d\varepsilon^*}}$ is larger than 1 – when few people have minority kids and their ethnicity choices are strategic complements (i.e., when $\frac{d\Delta(\varepsilon_H^*(b,e,\mu))}{d\varepsilon^*} < 0$). Instead, benefits are crowded out – leading to a social multiplier smaller than 1 – when many people have minority kids ($\frac{d\Delta(\varepsilon_H^*(b,e,\mu))}{d\varepsilon^*} > 0$).

But the effect in (7) of a change in benefits also includes the density $g(\varepsilon_H^*)$ at the cutpoint. When considering this channel, we impose the same condition as when ensuring consistency with fact F1 – i.e., the distribution of $\varepsilon$ is has (weakly) positive skew[9]. Specifically, we assume that the median $\varepsilon_{50}$ (and the mean) of the distribution lies (weakly) to the right of the mode. Suppose we compare two localities with cutpoints at percentiles equidistant from – and not too far from – the median, i.e., $\varepsilon_{50+n}^*$ and $\varepsilon_{50-n}^*$. Because of the positive skew, we have $g(\varepsilon_{50-n}^*) \geq g(\varepsilon_{50+n}^*)$. The larger effects of material benefits due to the higher social multiplier at $\varepsilon_{50-n}^*$ compared to $\varepsilon_{50+n}^*$ is thus reinforced by a higher density. We can now repeat this comparison for every other twin percentile cutpoints above and below the median. Therefore, if the cutpoints in the localities we observe in the data are continuously distributed along the support of $\varepsilon$, we may conclude that the average effect of material benefits in regions with cutpoints $\varepsilon_H^*$ below the median must be higher than the average effect in regions with cutpoints above the median.

Of course, we do not observe the cutpoints $\varepsilon_H^*$ of different localities directly but only the shares of households who get minority children $G(\varepsilon_H^*)$. However, the cutpoints and shares are one-to-one. Based on the argument above, we can therefore state:

**P1** *Suppose all peer groups in a province experience the same increase in benefits, due to a provincial policy. Then, peer groups with the share of minority children in male-Han mixed marriages below a cutoff share close to the median display a larger positive effect on the probability of having minority kids relative to prefectures above the cutoff share.*

In the data, we will evaluate prediction P1 by difference in differences, comparing prefectures and cohorts above and below alternative shares of minority children in the neighborhood of the median.

Prediction P1 relies on a comparison of the effects above and below a threshold share. An alternative way to get at the interaction of individual and social motives is to consider the comparative statics for different parts of the distribution of shares $G(\varepsilon^*)$ observed in the data, say different quartiles of shares corresponding to different quartiles of cutpoints $\varepsilon^*$. Let $\varepsilon_q^*$, $q = 1, 2, 3, 4$ denote cutpoints located at the middle of the quartiles of the $\varepsilon$ distribution. The (weak) positive skew of the distribution implies that $g(\varepsilon_4^*) \leq g(\varepsilon_1^*)$ and $g(\varepsilon_4^*) \leq g(\varepsilon_3^*) \leq g(\varepsilon_2^*)$. Moreover, under the assumption that $\frac{d^2\Delta}{d\varepsilon^{*2}} > 0$, the first derivatives of the social multiplier are monotonically ordered as: $\frac{d\Delta(\varepsilon_1^*)}{d\varepsilon^*} < \frac{d\Delta(\varepsilon_2^*)}{d\varepsilon^*} < 0 < \frac{d\Delta(\varepsilon_3^*)}{d\varepsilon^*} < \frac{d\Delta(\varepsilon_4^*)}{d\varepsilon^*}$. Using these facts in (7), we obtain an alternative testable prediction:

**P1'** *Suppose all peer groups in a province experience the same increase in benefits, due to a provincial policy. Then, the effect on the probability of having minority children is (i) larger in the first, second and third quartile than in the fourth quartile*

---

[9]As in the discussion of consistency with F1, this assumption can be weakened to say that the distribution of $\varepsilon$ does not have too much negative skew.

*of the share distribution, (ii) larger in the second than in the third quartile, (iii) ambiguous when we compare the first and second quartiles.*

P1 and P1' about the interaction of individual and social motives are the main new predictions of the model.

## 3.4 Additional Predictions

**Heterogeneity in material effects ($b$)** In this subsection, we formulate two additional auxiliary predictions, which do have some independent interest but also serve as "sanity checks" of the underlying model. One of these predictions is straightforward:

**P2** *Minority groups that enjoy a smaller increase in material benefits are less likely respond by choosing minority for their children than minority groups that enjoy a larger increase.*

**Material benefits ($b$) and intrinsic costs ($e$)** The other additional prediction is a bit more involved. We have analyzed how higher benefits of minority children shape the probability that mixed couples choose minority identity for their children. Do the (average) intrinsic costs $e(H)$ of having minority children systematically alter the effect of material benefits $b$ for Han-minority couples? In the model, this is the interaction effect of $b$ and $e$ on $G(\varepsilon_H^*(b, e, \mu))$. Given (7), this interaction effect can be written:

$$\frac{\partial G(\varepsilon_H^*(b, e, \mu))}{\partial b \partial e} = \left( \frac{dg}{d\varepsilon_H^*} - \frac{\mu \frac{d^2 \Delta}{d\varepsilon^{*2}}}{1 + \mu \frac{d\Delta(\varepsilon_H^*)}{d\varepsilon^*}} \right) \frac{1}{1 + \mu \frac{d\Delta(\varepsilon_H^*)}{d\varepsilon^*}} \cdot \frac{\partial \varepsilon_H^*}{\partial e(H)} \ .$$

The first multiplicative term on the right-hand side itself includes two effects which both depend on the cutoff value $\varepsilon_H^*$. The sign of the first effect depends on the change in the density $\frac{dg(\varepsilon_H^*(b,e,\mu))}{d\varepsilon^*}$ , which is positive before the single peak of $g$ and negative thereafter. The sign of the second effect is negative as the second derivative of the gain in social reputation $\frac{d^2 \Delta}{d\varepsilon^{*2}}$ is positive; thus the social multiplier goes down as the cutoff increases. As for the second multiplicative term $\frac{\partial \varepsilon_H^*}{\partial e}$, we know that it is negative – i.e., with higher intrinsic costs, fewer couples have minority kids. Putting these results together, we have:

**P3** *The interaction effect of $b$ and $e$ on the share of Han-minority couples having minority children is more likely to be negative if the share of minority children in the peer group of these couples is small.*

# 4 Data and Measurement

This section discusses how to measure the variables and parameters in the model. We also provide more information on the background for each variable. Outcome variables

and some control variables are measured at the individual level, whereas the individual and social incentives are measured at the prefecture, residency, education-group, or ethnicity level.

**Linking of datasets**  We draw on two sources of data. The first involves three of China's censuses: the 1-percent samples of the 1982, 1990 and 2000 censuses. Our second source is the 20-percent sample of the 2005 population survey, also known as the mini-census (covering about 1 percent of the population). These data provide demographic information and some information on socioeconomic status for altogether about 36 million people. One drawback of the data is that they give the location of the household at the time of the respective census (or mini-census), rather than at the time of childbirth. Therefore, our results could be biased by migration. We deal with this prospective problem in a couple of ways below.

As in the model, we are interested in the husband-wife-children structure of households. The husband or wife data draws on the information about the gender of the head of household. In some cases, parents or parents-in-law of the household head or the spouse cohabit with them. We drop this relatively small part of the sample, as the censuses do not distinguish parents from parents-in-law in the 1982 and 1990 censuses.

The administrative units we focus on are the areas defined by four-digit census codes: prefectures or cities. As some areas change names and codes over time, we unify the boundaries based on the year 2000 information to end up with 346 prefectures and cities. Since more than 330 of these are prefectures, we refer to all units by this label.

**Measuring ethnicity outcomes ($G(\varepsilon^*)$ in the model)**  We can directly identify children in the 2000 census and the 2005 mini-census. The 1982 and 1990 censuses do not distinguish between children and children-in-law. To identify children in the earlier data sets, we limit ourselves to unmarried children who still live with their parents. The results we report below are robust to using the 2000 census and the 2005 mini-census only.

As shown by the summary statistics in Table 1, 44 percent of the children in Han-minority families are minorities whereas 94 percent of the children in minority-Han families are minorities. This is fact F1 in the introduction. The high share of minority children in minority-Han is not only associated with little covariation over time, but also with little variation over space (see Figure 2b below). This is why our analysis focuses on the children in Han-minority families.

[**Table 1 about here**]

In the Web Appendix, we show that F1 is true not only at the aggregate but also at the individual level – the domain of the model – when we allow for prefecture fixed effects, year fixed effects, and (linear) provincial trends.[10]

---

[10]See Table A1. Column (1) in this table compares the probability of minority for a child in

In our analysis to follow, we take the mixed marriages as given and focus on the choice of ethnicity for the children. It is possible that some regions are more open to mixed marriages as well as to minority identity for the children. To take this into consideration, we always control for prefecture fixed effects and non-parametric provincial trends in our econometric specifications. We also discuss this possibility among the alternative explanations for our results considered in Section 7, where we also provide some additional estimates to address the prospective problem.

**Measuring material benefits ($b$ in the model)**   We measure material benefits of minority children in alternative ways. Since ethnic policies are a bundle of regulations (see the background in Section 2), it is not straightforward to quantify regional variation over time in these policies. To check that our results are robust, we try four ways:

**1. The rollout of one-child policies.** Some policies like family-planning gave favorable treatment to minorities already in the 1960s. But these policies became more generous and salient in the 1980s, when family planning was switched more strictly to a one-child policy. However, minority-minority couples were still allowed to have two or more children: by giving them minority status, parents could thus create an option value in their children's future family choices.

To measure the rollout of one-child policies, we employ the rollout timing for 27 provinces used in Edlund et al. (2013).[11] As explained in that paper and earlier work on family planning policies (e.g., Peng, 1996), the one-child policy is an umbrella term for a raft of policies. Although officially launched in 1979, it covers programs with an earlier start date. Edlund et al. (2013) focus on the following three programs: (i) family-planning science and technology-research institutes, (ii) family-planning education centers, and (iii) family-planning associations. Since all these programs indicate the salience of one-child policy, we consider the first year that any of them was present as the first year of one-child policy in the province. This starting date ranges from 1976 (in Jiangsu) to 1984 (in Guangxi).

This measure has the advantage of being staggered across provinces. It has the disadvantage of not capturing other benefits, such as those in education.

**2. Fines for violating one-child policies.** The fines charged by provinces for an extra child (expressed in terms of average annual income) measures not only the existence but also the intensity of one-child policies. This time-varying measure has been used by Ebenstein (2010) and Wei and Zhang (2011) to analyze the consequences of one-child policies. Since the fine information is only available after 1980, we employ it to conduct a robustness check in the subsample of children born after 1980. Naturally, this measure suffers from a similar disadvantage as the first one.

---

minority-Han families and that for a child in Han-minority families. Similar to the aggregate pattern, the difference is around 50%. Columns (2) and (3) present the results after including prefecture fixed effects and year fixed effects. Column (4) further allows for provincial (linear) trends. The estimates are very similar to those in column (1).

[11] Beijing, Shanghai, Tianjin and Chongqing are not included. We thank Lena Edlund for providing this data.

**3. Pre- and Post-1980.** To obtain a broader measure that also captures the benefits of minorities in higher education, we use a dummy for cohorts of parents that give birth to children after 1980. The increasing benefits to having minority children over time, together with the theoretical result in Section 3 that the effects of benefits are larger in mixed marriages with a Han man, makes the model consistent with fact F2.

In the Web Appendix (Table A3), we show that the increasing propensity for such couples to have minority children also holds up at the individual level, when we control for cohort and prefecture fixed effects, as well as linear provincial trends and use children born before 1970 as a comparison group. A higher probability of minority children is visible already from the 1970s, which is consistent with some ethnic policies being implemented already before 1980. However, the increase after 1980 is more striking. What we explore below is this increase at the intensive margin.

The drawback of using this measure is that it may be confounded with time trends. In the analysis to follow, we always include non-parametric provincial trends (province-by-cohort fixed effects) to take care of potentially variable time trends in policy. We also vary the threshold year of comparison to check whether the pre- and post-1980 differences are indeed salient.

**4. Heterogeneous benefits.** The third measure we explore exploits heterogeneity in the beneficiaries of pro-minority policies. Specifically, most of the preferential policies are limited to minorities with a population smaller than 10 million. In particular, the No. 7 Document by the Chinese Communist Party Central Committee in 1984 requires that minority groups with a population size over 10 million are treated in the same way as the Han in family-planning policies.

As there was more than 13 million members of the Zhuang minority already in the 1982 census, this group enjoyed fewer benefits than other minority groups. Therefore, we will compare the Zhuang minority with other minority groups. As Table 1 shows, the probability of having a Zhuang wife among Han men with minority wives is around 17 percent. To check whether a heterogeneous impact is driven by population size, we also control for minority population size by ethnic group in our analysis. Some scholars argue that Zhuang is an ethnic group very much integrated with Han (Kaup, 2000). This feature would imply a prediction opposite to the impact of lower benefits though: if those identity costs go down over time, this would make a Han-Zhuang family more likely to choose minority status for their child.

**Measuring peer groups for social motives (related to $\frac{d\Delta}{d\varepsilon^*}$ in the model)** Following the discussion about crowding in or crowding out in Section 3 (the sign of $\frac{d\Delta}{d\varepsilon^*}$), we measure social motives by the shares of minority children in mixed marriages. In order not to run into the reflection problem discovered and discussed by Manski (1993), we want to treat the social motives for a particular cohort as predetermined by previous choices in the relevant peer group. Because we cannot observe the relevant peer group directly and our data derive from a sample of the population, we define the peer group

relevant for the social motives in a number of different ways in the hope of avoiding biased estimates.[12]

**1. Previous cohort in the same prefecture.** Given the dramatic economic development in the past few decades, social motives may have changed fairly quickly. We define the peer group relevant for the prevailing social motive as the birth cohort one decade before the Han husband in the same prefecture. For example, the Han-men mixed households in the prefecture that have children in the 1980s become the peer group for Han-men mixed household that have children in the 1990s, and so on.

**2 A-C. Previous cohort in the same prefecture plus residence, education, or wife's ethnicity**. The first measure only uses ethnicity of the husband, birth cohort, and prefecture to define the peer group. But we also consider a number of finer peer groups. (**A**) The first refinement is to condition also on urban or rural residency and define the peer group at the prefecture-ethnicity-cohort-residency level. Specifically, we base the distinction between urban and rural on the husband's legal residency (*Hukou*). A disadvantage of this measure is that it implies smaller groups, due to the disaggregation itself and the fact that rural/urban information is only available in the 2000 and 2005 censuses. Hence, the number of observations in each cell becomes smaller than that for the previous measure. (**B**) One may also plausibly argue that peer groups may be formed by people with different levels of education. A second refinement is therefore to condition on education of the father. Specifically, we base the distinction on whether he has an education corresponding to completed high-school or above. This way, we define the peer group at the prefecture-ethnicity-cohort-education level. (**C**) Yet another possibility is that the relevant peer group for a Han man and minority woman of a certain ethnicity is limited to other couples where the wife has the same ethnicity. We consider this possibility as well by defining the peer group at the prefecture-(male)ethnicity-cohort-(female)ethnicity level.

**3. 1970s cohort in the same prefecture.** We also exploit the variation across prefectures in the birth cohort of the 1970s, i.e., before the dramatic changes in ethnic policies. This is a simpler measure than the one in **1** and treats the social motives as predetermined over the period of changing policies. This serves as yet another check for the peer group relevant for the social motives emphasized in our model.

Figure 2 plots the distribution across prefectures of the shares that have a minority child in the two types of mixed families for the children born in the 1970s cohort. It shows a great deal of variation across prefectures for male-Han mixed families. In terms of the model, this dispersion reflects the joint distribution of parameters $b, e(H), \mu$ leading to different cutoffs $\varepsilon^*$ and the mapping from these cutoffs into shares via distribution $G$.

However, for male-minority mixed families, most prefectures are concentrated at the right end, leaving little variations across prefectures. As stated before, we therefore

---

[12]The prospective econometric problems of estimating the influence of unobserved peer groups in a sample from the population appear related to the biases due to measurment error when estimating peer effects for members of partially sampled networks (Chandrasekhar and Ellis, 2011).

focus on the effect of social motives for Han-minority families.

In addition, the pattern for the Han-minority families in Figure 2 also suggests that the likelihood for the sons to be a minority is lower than that for the daughters. This is consistent with the fact that the identity costs for parents are higher than the sons. Below we will explore this variation to proxy $e(H)$.

[**Figure 2 about here**]

Figure 3 maps the spatial distribution across China of the ethnicity choices (based on the 1970s cohort) by male-Han mixed families. It suggests that the social motives vary considerably across prefectures, and that this variation is not strongly geographically clustered. For instance, province fixed effects only explain about a third of the variation across prefectures.

For Han-minority families, our model predicts a strategic complementarity $\frac{d\Delta}{d\varepsilon^*} < 0$ for low values of the cutoff $\varepsilon^*$ (when the share of mixed couples having minority kids is small) and a strategic substitutability $\frac{d\Delta}{d\varepsilon^*} > 0$ for high values of $\varepsilon^*$ (when a large share of mixed couples have minority kids). In theory, if the distribution of $\varepsilon$ were symmetric, the sign would flip at a critical cutoff of $\varepsilon^*_{50} = 0$, corresponding to a share of minority kids at 0.5. Allowing for a non-symmetric distribution, we will check how the estimates behave as we vary the assumption about the critical share of minority kids in the neighborhood of 0.5, in consistency with the model predictions.

[**Figure 3 about here**]

**Measuring intrinsic costs ($e(H)$ in the model)**  A first measure of intrinsic (individual) cost $e$ is whether the child is a son or a daughter. Consistent with Confucian values, the intrinsic costs of having a child with different ethnicity than the father are higher for a son than a daughter. Consequently, we examine whether the impact of benefits on ethnic choices is smaller for sons.

A second measure of intrinsic costs is whether the spouse belongs to a religious minority group. It is conceivable that minority identity is more costly for a Han man if it requires his child to practice religion (recall the discussion in Section 2). To clarify, this measure is available at the ethnic group level, not the individual level. Out of the 55 minority groups, 18 practice Islam or Tibetan Buddhism. We define a wife as religious if she belongs to one of these 18 minority groups.

Naturally, men who marry religious women constitute a selected sample, but our question concerns how a religious wife shapes the effect of material benefits on ethnic choice for children, rather than the effect of a religious wife itself. The summary statistics in Table 1 show that the share of male-Han mixed families with a religious wife is about 19 percent.

**Taking migration into account** The variation across prefectures and provinces discussed in this section is based on residency at census time. However, this residency may be different than birth place, due to migration. Only the 2000 census includes information whether an individual's birth place coincides with her current residency (the 1982 and 1990 censuses spells out whether one lived in the same county five years ago, and the 2005 mini-census only has information on whether one lived in the same province one year ago). Based on the 2000 census, over 85 percent of individuals were born in the same county as their current residency, while 94 percent were born in the same province. Given that prefecture is the administrative level above county, these facts suggest that migration is unlikely to make a major difference for our main results. Moreover, Frijters, Gregory and Meng (2013) document that rural-urban migration did not take off until 1997.

Nevertheless, we conduct robustness checks by omitting the most recent census from 2005 from the sample, and by excluding individuals whose birth county and residency county are different. This should minimize the potential impact of migration.

# 5   Individual vs. Social Motives

The most important new predictions from our model are P1-P1' on the interactions between individual and social motives. To the best of our knowledge, no similar predictions have been studied in the existing literature. This section confronts these predictions with data.

## 5.1   Testing Prediction P1

Our first prediction about the interactions between individual and social motives says that the effect of higher material benefits should be larger in peer groups where the initial share of minority children is smaller, because individual motives driven by material benefits are crowded in rather than crowded out by prevailing social motives. To test this, we ask whether $\beta_b$ is positive in the difference-in-difference specification:

$$
\begin{aligned}
MinChild_{i,p,t} \;=\; & \beta_b Post_{r,t} \times I(\leq \mathrm{X})_{p,t-10} + I(\leq \mathrm{X})_{p,t-10} + Post_{r,t} + birth_t \\
& + ethn_g + pref_p + prov_r \times birth_t + \varepsilon_{i,p,t} \;,
\end{aligned}
\tag{9}
$$

where $MinChild_{i,p,t}$ is a dummy indicating whether child $i$ (with minority mother of ethnicity group $g$), in prefecture $p$ (belonging to province $r$), and birth cohort $t$ is a minority. In the main specification, we use a dummy $Post_{r,t}$ to measure whether province $r$ has established family-planning institutions. We use another dummy $Post1980_t$ as an alternative to check for robustness of the baseline results.

$I(\leq \mathrm{X})_{p,t-10}$ is an indicator for whether the peer group – according to Definition 1 in Section 4, i.e., by parents with children in the past birth cohort in the same prefecture – has a share of minority children smaller than some critical value X between 0 and 1. Thus, the parameter of interest $\beta_b$ measures the difference in the effect of material

benefits post the introduction of the one-child policy (or post 1980) between prefectures below and above the assumed cutoff. To identify this parameter, we include the time-varying share definition itself.[13] We also include cohort fixed effects (see below), which absorb any direct effects of post policy (post-1980) cohorts.

To allow for an effect of prefecture characteristics that are time-invariant or change slowly over time – such as the attitudes towards mixed marriages – we control for prefecture fixed effects ($pref_p$). To hold constant factors that affect the ethnicity choices by different cohorts across China (including the average effects of post-policy or post-1980 benefits), we also include birth-cohort fixed effects ($birth_t$, for every ten years). To control for ethnicity-specific factors that are time-invariant or change slowly over time, we include ethnicity fixed effects ($ethn_g$). For example, some ethnic groups may have stronger preference of having a child within the same ethnicity. Since we focus on the children Han-minority families, ethnicity fixed effects refer to the ethnicity of the wife. Finally, we include non-parametric province-specific trends ($prov_r \times birth_t$) to control as flexibly as possible for different evolutions across provinces, such as different growth rates or different provincial policies.[14] When estimating this specification, we always cluster the standard errors at the prefecture level.

Since P1 refers to cutoffs close to the median, Table 2 presents the results for the cutoff range between 0.4 and 0.6. Column (1) shows that the average effect of $Post_{r,t}$ is around 0.08. Columns (2)-(4) show that the estimated effect of material incentives is indeed significantly larger when the share of minority children is smaller than the cutoff value. The estimated interaction effects are quantitatively large, at least on the order of the average effect in column (1). This is consistent with Prediction P1 that benefits have a larger effect in peer groups where few mixed households have minority children, because they are crowded in by a strategic complementarity (giving a social multiplier above 1), rather than crowded out by a strategic substitutability (giving a social multiplier below 1). For completeness, Figure 4 visualizes the corresponding interaction estimates and their 95 percent confidence intervals for all cutoffs between 0.1 and 0.9.

[**Table 2 about here**]

[**Figure 4 about here**]

Similarly, columns (5)-(8) present the results replacing $Post_{r,t}$ with $Post1980_t$. The estimates of the interaction effects are slightly larger than those in columns (2)-(4). To check whether our choice of 1980 as the average cutoff year for the step-up of ethnic policies, Figure 5 illustrates the estimated coefficients from our baseline specification, when we use all years from 1975 to 1985 as alternative cutoff years. As the figure

---

[13]The reason is that the minority birth share in some prefectures switch from falling short of to surpassing $X$ across different birth cohorts. Thus, the effect of the minority birth share itself is not (completeley) absorbed by the prefecture fixed effects (see below).

[14]The results are also robust to including prefecture-specific linear trends ($pref_p \times t$).

shows, the estimate of our main parameter of interest has a relatively sharp peak for our pre-chosen cutoff year.

[**Figure 5 about here**]

**Robustness checks** We conduct three additional robustness checks for the findings in Table 2. First, we use the 1970s cohort to calculate the shares of minority children in a prefecture and employ these shares as a measure of predetermined social motives. The results are presented in the Web Appendix (Table A4). The results are slightly larger than those in Table 2. This is consistent with the fact that our baseline measure $I(\leq X)_{p,t-10}$ takes changing social motives norms into account.

Second, to deal with the concern that peer groups are mismeasured due to migration, we re-estimate the baseline results, dropping all data after the 2000 census as well as individuals whose birth county and residency county are different in the 2000 census. The results in the Web Appendix (Table A5) entail coefficients similar to those in Table 2.

Finally, instead of using only the rollout of one-child policies, we rely on the fine information to measure the intensity of one-child policies. The results are presented in the Web Appendix (Table A6). They show that our baseline results hold up well within the subsample of children born after 1980.

## 5.2 Peer Groups

The notion of a peer group plays a key role in our model. The empirical estimates we have shown so far rely on the assumption that the previous cohort of Han-minority couples in the same prefecture makes up the relevant peer group for ethnicity decisions. It is important to consider other alternatives, however, since peer groups are not observable. In particular, one may argue that our definition is too wide and that couples are more influenced by people that live under similar conditions, have the same education, or exactly the same ethnicity composition. This subsection considers these three possibilities.

**Subdivision by residency** We begin with Table 3A, which present results for, respectively, urban-resident and rural-resident members of the same ethnicity-prefecture-cohort – i.e., when using peer-group definition 2A. The average effect in columns (1) and (5) show that the change in material benefits had a larger effect on the ethnicity choices of urban couples (about 10% of the mean) than on those of rural couples (4% of the mean). This makes sense, since the benefits in terms of family planning and college entrance are likely more salient for urban couples.[15] Although based on a considerably smaller sample, the estimates of the interaction between individual and social motives deliver a similar message as the prefecture-cohort-level results in Table 2.

---

[15]Fewer exceptions from the one-child policy are allowed in urban areas, and more children born in urban areas go on to college.

[**Table 3A about here**]

**Subdivision by education**  In another attempt to vary the definition of the peer group, we further subdivide each cohort of mixed couples with Han men in the prefecture by the educational background of these men. In particular, we split the sample, according to definition 2B, into those with less than a high-school education, and those with high-school or more.

Thе results are presented in Table 3B. They show that the baseline findings in Table 2 are not driven by one particular educational group. The interaction effect between material benefits and social norms appears to be larger for the smaller group with more education than high school. A possible interpretation of this, in terms of the model, would be that households in this group put a higher weight $\mu$ on their social reputation.

[**Table 3B about here**]

**Subdivision by wife's ethnicity**   In the estimates presented so far, we have assumed that all mixed couples with a Han man and a minority wife, no matter which minority, form the basis of the relevant peer group. But one may argue that the peer group is specific to each specific minority group of the wife. To check whether this produces different results, we consider definition 2C above, where each cohort in a prefecture could thus, in theory, make up 55 different peer groups (in practice, the average number is much smaller due to the regional dispersion of minorities). The estimation results, presented in Table 3C, do not differ dramatically from the baseline estimates in Table 2.

[**Table 3C about here**]

## 5.3   Testing Prediction P1'

We now consider the prediction P1', about the interaction between individual and social motives, based on the behavior in different quartiles. To do that we replace $I(\leq X)_{p,t-10}$ in (9) with three indicators for the share of minority children in the previous cohort being in one of the three first quartiles: $I(0\text{-}0.25)_{p,t-10}$, $I(0.25\text{-}0.50)_{p,t-10}$ and $I(0.50\text{-}0.75)_{p,t-10}$. We thus leave the fourth quartile as the reference group.

Table 4 shows the results, when the peer group is always defined as the mixed couples in the previous (ten years older) cohort in the same prefecture (Definition 1). Column (1) presents the interaction effects with prefecture fixed effects. Columns (2) and (3) add ethnic fixed effects (referring to the ethnicity of the minority wife), cohort fixed effects and province-specific nonparametric trends. Column (4) excludes migrants. Columns (5)-(8) reproduce the specifications in columns (1)-(4) but replace $Post_{r,t}$ with $Post1980_t$.

Consistent with prediction P1', the effect is generally larger for the first, second and third quartile. In addition, the point estimates for the second quartile are indeed

significantly higher than that for the third quartile (with a $p$-value smaller than 0.05 in all specifications including trends). These effects are again large: the difference in effects of higher material benefits, say, in the first vs. the fourth quartile is on the order of the average effect estimated in column 1 of Table 2. This corresponds to the theoretical prediction of a social multiplier above 1 in the first quartile – due to crowding in – and a social multiplier below 1 in the fourth quartile – due to crowding out.

[**Table 4 about here**]

In summary, the data are clearly consistent with predictions P1 and P1'. The results reported in this section constitute solid evidence that peer-group dependent social motives help shape the effect of individual material benefits on individual ethnicity choices, with crowding in same places and crowding out in others. The interaction between individual and social motives appears to be not only statistically significant but also quantitatively significant.

# 6 Additional Results

## 6.1 Heterogeneous Material Benefits – Testing P2

Auxiliary prediction P2 says that the effect of higher benefits should be smaller for mixed households where the man is Han and the wife is Zhuang rather than some other minority, simply because the Zhuang experienced a smaller increase in minority benefits. To test this, we check whether $\beta_z < 0$ in the specification:

$$
\begin{aligned}
MinChild_{i,p,t} \;=\;\; & \beta_z Post_{r,t} \times ZhuangWife_i + \gamma ZhuangWife_i + Post_{r,t} \\
& + ethn_g + pref_p + birth_t + prov_r \times birth_t + \varepsilon_{i,p,t} \;.
\end{aligned}
$$

In addition to the same controls as in the baseline, we also control for the share of minority population for each ethnic group and its interaction with the post-policy indicator. This enables us to tell apart the impact of fewer benefits from any direct impact of population size.

The estimates are displayed in Table 5. Column (1) reports the results with prefecture fixed effects. Columns (2) and (3) add cohort fixed effects and non-parametric province-specific trends. Column (4) further controls for ethnicity fixed effects and column (5) excludes migrants. In all specifications, having a Zhuang wife indeed decreases the effect of material benefits, in line with prediction P2. The magnitude of the interaction effect $\beta_z$ is close to that of $Post_{r,t}$ in Table 2, implying that the Han-minority households with a Zhuang wife did not significantly respond to the policies. This finding does not reflect the effect of a larger population size. If anything, the impact of material benefits on ethnicity is higher if the minority wife belongs to a larger ethnic group.

Similar to the robustness checks in the previous section, columns (6)-(8) present the results using $Post1980_{p,t}$ (controlling for all fixed effects and trends) instead of $Post_{r,t}$. The results are close to those in columns (1)-(5).

[**Table 5 about here**]

Altogether, the econometric estimates are consistent with prediction P2. These findings work as a sanity check for our model.

## 6.2   Material benefits and intrinsic costs – Testing P3

Our second additional prediction involves the interaction effect of material benefits and intrinsic costs due to cultural distance on the choice of a minority child. Given that the average share of minority children for Han-Minority households is quite small (around 0.4), our model predicts that this interaction effect is negative. As discussed in Section 4, we proxy the intrinsic costs by dummy variables indicating whether the child is a son and whether the minority wife is religious. Thus, we estimate:

$$
\begin{aligned}
MinChild_{i,p,t} \;=\; & \beta_s Post_{r,t} \times Son_i + \delta Son_i + Post_{r,t} + ethn_g \\
& + pref_p + birth_t + prov_r \times birth_t + \varepsilon_{i,p,t} \;,
\end{aligned}
$$

and

$$
\begin{aligned}
MinChild_{i,p,t} \;=\; & \beta_l Post_{r,t} \times ReligiousWife_i + \delta ReligiousWife_i + Post_{r,t} \\
& + ethn_g + pref_p + birth_t + prov_r \times birth_t + \varepsilon_{i,p,t} \;,
\end{aligned}
$$

expecting to find negative values of $\beta_s$ and $\beta_l$.

Table 6 presents the results when we compare sons and daughters in specifications otherwise very similar to those in Table 5, except that we do not need to include population size. The results show that having a son cuts the effect of material benefits, consistent with the predicted negative interaction. The estimate of $\beta_s$ is -0.017, which amounts to about 20% of the average effect of the ethnic policies.

Conceptually, a sharper specification would be to examine the choices within a household before and after the implementation of ethnic policies. However, this would require both the presence of multiple children (despite the one-child policy) with birth years before and after the policy. Unfortunately, we do not have enough such observations and thus lack the statistical power to conduct a within-household analysis.

[**Table 6 about here**]

Table 7 presents the results on the effect on Han men with religious minority wives. The specifications are similar to those in the Table 6, except that here we do control for population shares of the wives' minority groups. Consistent with the model, having a religious wife also cuts the effect of material benefits. The estimate of $\beta_s$ is -0.03, about 40% of the average effect of ethnic policies.

26

In sum, we find that the estimates are consistent with prediction P3. Thus our model also passes the second of our sanity checks.

# 7 Alternative Explanations

Our model is consistent with motivating facts F1-F2 in the Chinese micro data under certain assumptions (not too much negative skew in the $\varepsilon$ distribution). With the same assumptions, three auxiliary predictions from the model are also borne out by the data. Together, these results show that our model provides a plausible framework to understand the interaction of individual motives – entailed in material benefits and intrinsic costs – and social motives.

Can our findings be explained by other mechanisms? In this section, we discuss three possible alternatives and argue that they are unlikely to be the main drivers of our results.

## 7.1 Bargaining Power

Bargaining is an alternative mechanism to the material benefits of having a minority child that might explain facts F1 and F2 in Figure 1. Specifically, minority women's bargaining power may have gone up over time leading to a higher number of Han-minority couples choosing minority status for their children. This mechanism may have become more powerful after 1980, due to social and economic factors, like unbalanced and increasing sex ratios – a larger number of men relative to women – among the Han.

Even though this mechanism might be consistent with Figure 1, increasing bargaining power is a less plausible explanation for some of our findings in the tests of the model predictions P1-P3. In particular, it cannot explain the findings in Table 5 that the impact on the incidence of minority children after 1980 is smaller for a wife of Zhuang ethnicity than a wife of other minorities. If the culprit is an increasing sex ratio among Han men, why would the Zhuang women's bargaining power become lower over time than the bargaining power of women from other minorities? If anything, their bargaining power may be higher because the Zhuang has a larger population than other minorities and (if anything) better integrated with the Han than other minorities. Similarly, the findings in Table 6 of a smaller effect on ethnic choices in couples with Han men and religious minority women would require a smaller increase in the bargaining power of religious wives than of non-religious wives.

How would women's bargaining power have to change in order to explain our main findings on the interaction between individual and social motives in Tables 2-4? For this to be the case, minority women's bargaining power not only has to go up over time, but also has to increase faster in the peer groups where the share of minority children is low. In this subsection, we check this prospective explanation for three plausible roots of bargaining power.

**Education differences**    The education gap between the husband and the wife is a natural candidate to measure variations in their relative bargaining power, as the party with higher education plausibly has more bargaining power. To gauge that gap, we unify the categorical education levels across censuses into four groups: 1 indicates completion of less than primary school, 2 completion of primary school, 3 completion of secondary school (High school), and 4 completion of collage or above. On average, the education difference between the husband and the wife is 0.23, meaning that the average woman marries a man with more education.

Before including this measure in our analysis, we note how it varies with time and with the share of minority children. Panel (a) in Table A7 of the Web Appendix presents a matrix of education differences by children's birth cohort and by the share of minority children born in the 1970s in the prefecture. Two patterns appear in the data. First, the education gap decreases over time, consistent with the interpretation that bargaining power of minority women went up over time. However, this change in the education gap is not systematically larger where the share of minority children was small. Therefore, bargaining power associated with educational differences is unlikely to explain our main finding on individual-social interactions.

As a further check, we add the education difference – and its interaction with the share indicator $I(\leq X)$ – to specifications similar to those underlying Table 2. The results are presented in columns (1)-(3) in Table 8. There seems to be, at best, a weak effect of a minority wife's bargaining power vs. her Han husband on the probability that their child gets minority identity. More importantly, this measure of bargaining power does not seem to drive the interaction between individual and social motives, as the estimated interaction coefficients are very similar to those in Table 2.

[**Table 8 about here**]

**Age differences**    Another candidate proxy for bargaining power is the age difference between the husband and the wife. On the one hand, a younger woman may have more bargaining power than an older woman in a marriage with the same man. On the other hand, an older man may have more bargaining power than a younger man in a marriage with the same woman (due to higher income or assets).

In the data, the average age difference between the husband and the wife is 2.6 years. Panel (b) in Table A7 shows the change in age differences by birth cohort and by the share of minority children. Once again, the age gap decreases over time, consistent with the interpretation of increasing bargaining power of minority women. However, the change over time is more pronounced in the prefectures with a high share of minority children. If this factor played an important role, it would thus imply the opposite of Prediction 1 of our model.

Similar to the analysis for education differences, columns (4)-(6) in Table 8 present the results after including the age difference between husband and wife and its interaction with the share indicator $I(\leq X)$. Older Han men seem to be less willing to

concede minority status on their children. However, the estimates for the strength of individual-social interactions are very close to those in Table 2.

**Sex ratios**   A third candidate to measure bargaining power is the (male to female) sex ratio in the father's birth cohort of Han men within the same prefecture. It is natural to assume that a higher such ratio increases the bargaining power of the wife. However, the timing of unbalanced sex ratios does not square well with the patterns in the data. Higher sex ratios at birth indeed became salient after the 1980s. But the children born in the 1980s had parents born in the 1960s when sex ratios were roughly balanced. This can also be seen in Panel (c) of Table A7: the change in sex ratios for the fathers is not very important in the cohorts we consider.

Nevertheless, we add to our specification the sex ratio and its interaction with the share indicator $I(\leq X)$. The results are presented in columns (7)-(9) of Table 8. The relation between sex ratios and the probability of having a minority child goes in the expected direction. But including this measure of bargaining power cannot explain the interaction of individual and social motives.

Taken together, the results in this subsection do not speak against the hypothesis that bargaining power may play a role for understanding the patterns of ethnicity decisions we see in the data. But they do show that bargaining power cannot explain our main findings.

## 7.2   Endogenous Mixed Marriages

Another concern may be that some unobserved regional factors affect the incidence of mixed marriages as well as the ethnic choices for children, in such a way that our findings in Tables 2-4 just reflect a proxy for those omitted variables rather than an interaction between individual and social motives. It is true that in a broader context mixed marriages are endogenous. In fact, we are currently doing additional research on the incidence of mixed marriages. We leave the question about the drivers of mixed marriages for an accompanying paper, but it is still important to examine whether this concern matters for our main findings. We conduct three checks on this concern.

**Correlation patterns**   As a first check, we investigate the correlation between the probability of mixed marriages and the probability of mixed couples having minority children in the prefecture-cohort data. We distinguish two types of marriages. Columns (1)-(3) in Table A8 present the results for the HM marriages. The two first columns suggest a significant and positive correlation between the probability of mixed marriages and the probability of mixed couples having minority children. However, as the third column shows, this correlation becomes insignificant once we include the kind of provincial non-parametric trends $prov_r \times birth_t$ that always enter the baseline specifications in Tables 2-4. Thus, the correlation between mixed marriages and ethnicity of children primarily reflects differences in attitudes across province-chorts. Columns (4)-(6) show that the correlation is generally weaker for the MH marriages.

Nevertheless, we re-estimate our baseline specification in Table 2, while controlling for the share of mixed marriages and its interaction with the share indicator in the father's cohort in the same prefecture. We present the estimates from these specifications in Table 9. The incidence of mixed marriages is weakly positively – and sometimes significantly – correlated with the incidence of minority children in these marriages. However, including the incidence of mixed marriages only very marginally alters the estimates of the central interaction effects in Tables 2.

[**Table 9 about here**]

**A specific subsample**  Further, we focus on a subsample where endogenous marriages is less of a problem. In particular, we look at couples married before the introduction of ethnic policies, but with children born after this introduction. This restriction plus the fact that marriage-year information is available only in the 2000 and 2005 censuses, means that we have a much smaller sample. The results are presented in columns (5)-(8) of Table 9. Since the subsample is drawn from more recent censuses, most of the children in the sample were born after 1970. This explains why the average effect of the policies in columns (5) and (7) is smaller than that in the whole sample. Moreover, the interaction effect with the social norms is less precisely estimated in columns (6) and (8). Nevertheless, the pattern of results is similar to those in the baseline results. Moreover, the interaction effects become larger and more precisely estimated if we use cutoffs of higher values than 0.5 (estimates not shown).

The bottom line from this subsection is that endogenous mixed marriages are unlikely to explain our main results.

## 7.3  Censoring

A final concern is that our main result might have a mechanical explanation, due to a kind of upward censoring. Specifically, could not our finding of a larger policy effect in prefectures where the share of minority children is small be driven by the simple fact that there is little room to respond when this share is large and approaching one? To check for this possibility, we restrict the estimation sample to prefecture-cohorts with a share of minority children between 0.3 and 0.7. In this interval, there should be enough room for mixed households in every prefecture-cohort to respond without hitting a constraint. As shown in Table 10, the estimates from the restricted sample are similar to the baseline estimates from the full sample in Table 2. In other words, upward censoring does not drive our main findings on individual-social interactions.

[**Table 10 about here**]

# 8   Conclusion

We provide a framework to analyze the ethnicity of children in interethnic marriages. Drawing on earlier work by Benabou and Tirole (2011), we present a model which is consistent with two motivating facts for China. The model also delivers a set of auxiliary predictions. The empirical tests we carry out on Chinese census data generally support these predictions. Most importantly the results suggest that changes in individual motives triggered by policy interventions are sometimes crowded in and sometimes crowded out by social motives, precisely in the way that theory predicts.

Our empirical results speak to two issues that have rarely been studied in the data. One is specific to China, namely the ethnicity of children in mixed marriages. In future work, we hope to extend our empirical analysis to predictions from a model of directed marriage search, asking which individuals end up in mixed couples in the first place. Then, the ethnic choices for children analyzed in this paper would help determine the continuation value from the marriage stage.

The second issue is more general, namely how individual (material and intrinsic) motives and social motives interact in shaping individual choices. Our methodology for empirically investigating how social concerns modify the effect of individual incentives can plausibly be applied to other economic, political or social choices – e.g., in tax evasion, political participation, or fertility – which may reflect an interplay between individual and social motives.

# References

[1] Bates, Robert (1974), "Ethnic Competition and Modernization in Contemporary Africa", *Comparative Political Studies* 6, 457-484.

[2] Benabou, Roland and Jean Tirole (2011), "Laws and Norms", NBER Working Paper, No 17579.

[3] Besley, Timothy, Anders Jensen, and Torsten Persson (2014), "Norms, Enforcement and Tax Evasion", mimeo, IIES and LSE.

[4] Blume, Lawrence, William Brock, Steven Durlauf, and Yannis Ioannidis (2011), "Identification of Social Interactions", in J. Benhabib, Jess, Alberto Bisin, and Matthew Jackson (eds.), *Handbook of Social Economics*, Volume 1B, North Holland.

[5] Bisin, Alberto and Thierry Verdier (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits", *Quarterly Journal of Economics* 115, 955-988.

[6] Bisin, Alberto, Giorgo Topa, and Thierry Verdier (2001), "Religious Intermarriage and Socialization in the United States", *Journal of Political Economy* 112, 615-664.

[7] Botticini, Maristella and Zvi Eckstein (2007), "From Farmers to Merchants, Conversions, and Diaspora: Human Capital and Jewish History," *Journal of the European Economic Association* 5, 885-926.

[8] Brock, William and Steven Durlauf (2001), "Discrete Choice with Social Interactions", *Review of Economic Studies* 68, 235-260.

[9] Chandrasekhar, Arun and Randall Lewis (2011), "Econometrics of Sampled Networks", mimeo, MIT.

[10] Cassan, Guilhem (2013), "Identity-Based Policies and Identity Manipulation: Evidence from Colonial Punjab", *American Economic Journal: Economic Policy*, forthcoming.

[11] Edlund, Lena, Hongbin Li, Junjian Yi, and Junsen Zhang (2013), "Sex Ratios and Crime: Evidence from China", *Review of Economics and Statistics*, forthcoming.

[12] Ebenstein Avraham (2010), "The 'Missing Girls' of China and the Unintended Consequences of the One Child Policy", *Journal of Human Resources* 45, 87-115.

[13] Frijters, Paul, Robert Gregory, and Xin Meng (2013), "The Role of Rural Migrants in the Chinese Urban Economy", in Dustmann, Christian (ed.), *Migration– Economic Change, Social Challenge*, Oxford University Press.

[14] Fryer, Roland (2007), "Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century", *Journal of Economic Perspectives* 21, 71-90.

[15] Guo, Zhigang and Rui Li (2008), "Cong Renkoupuchashuju kan zujitonghun fufu de hunling, shengyushu jiqi zinu de minzu xuanze" (Marriage Age, Number of Children Ever Born, and the Ethnic Identification of Children of Inter-Ethnic Marriage: Evidence from China Population Census in 2000), Shehuixue Jianjiu (*Research on Sociology*) 5, 98-116.

[16] Huang, Wei and Yi Zhou (2015), "Impact of One-Child Policy on Interethnic Marriage in China", mimeo, Harvard University.

[17] Horowitz, Donald (2000), *Ethnic Groups in Conflict*, University of California Press.

[18] Jewitt, Ian (2004), "Notes on the Shape of Distributions", mimeo, Oxford University.

[19] Kaup, Katherine Palmer (2000), *Creating the Zhuang: Ethnic Politics in China*, Lynne Rienner Publishers.

[20] Manski, Charles (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies* 60, 531-542.

[21] Peng, Peiyun (ed.) (1996), *Zhong Guo Ji Hua Sheng Yu Quan Shu* (*Encyclopedia of China's Family Planning Program*), China Population Press.

[22] Smelser, Neil and Richard Swedberg (eds.) (2005), *The Handbook of Economic Sociology*, 2nd edition, Princeton University Press.

[23] Vail, Leroy (1989), *The Creation of Tribalism in Southern Africa*, University of California Press.

[24] Wei, Shang-jin and Xiaobo Zhang (2011), "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China," *Journal of Political Economy* 119, 511-564.

# Figures and Tables

**Figure 1  Share of Minority Children by Type of Mixed Marriage and Birth Cohort**
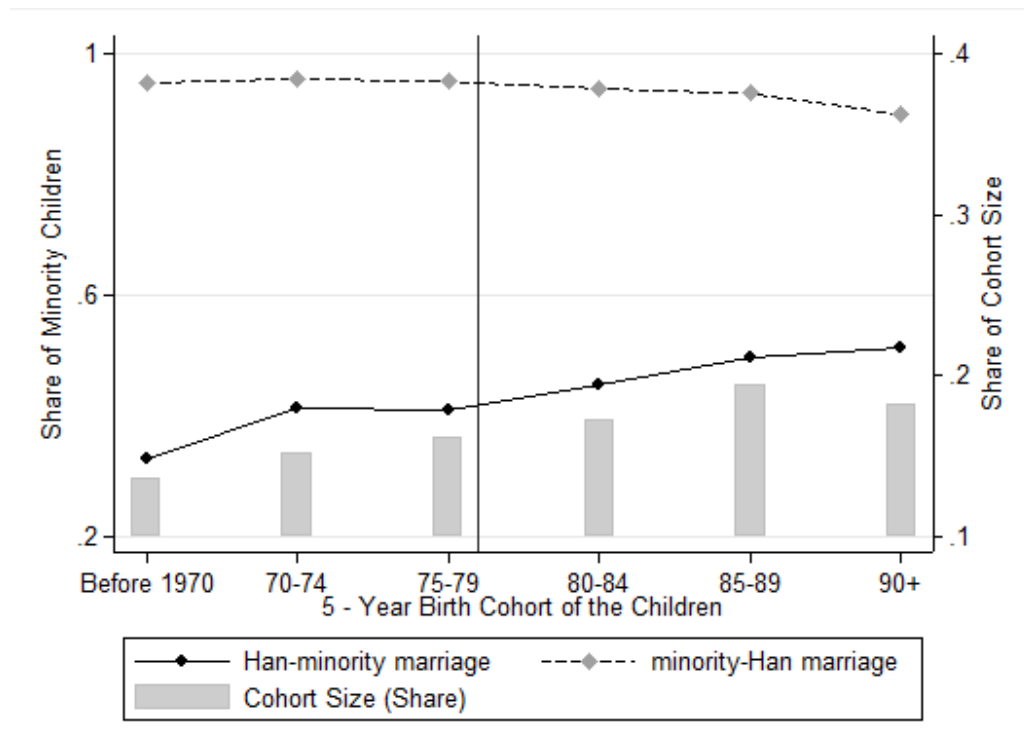


**Notes**: This figure displays the share of minority children in mixed marriages by cohorts. It shows that (F1) the children are more likely to be a minority in mixed marriages with a male-minority and (F2) there is an increasing trend of minority children in mixed marriages with a male-Han, especially after the 1980s.

**Figure 2  Distribution of Minority Shares across Prefectures (for 1970s Cohort)**

(a) HM-Families



(b) MH-Familily
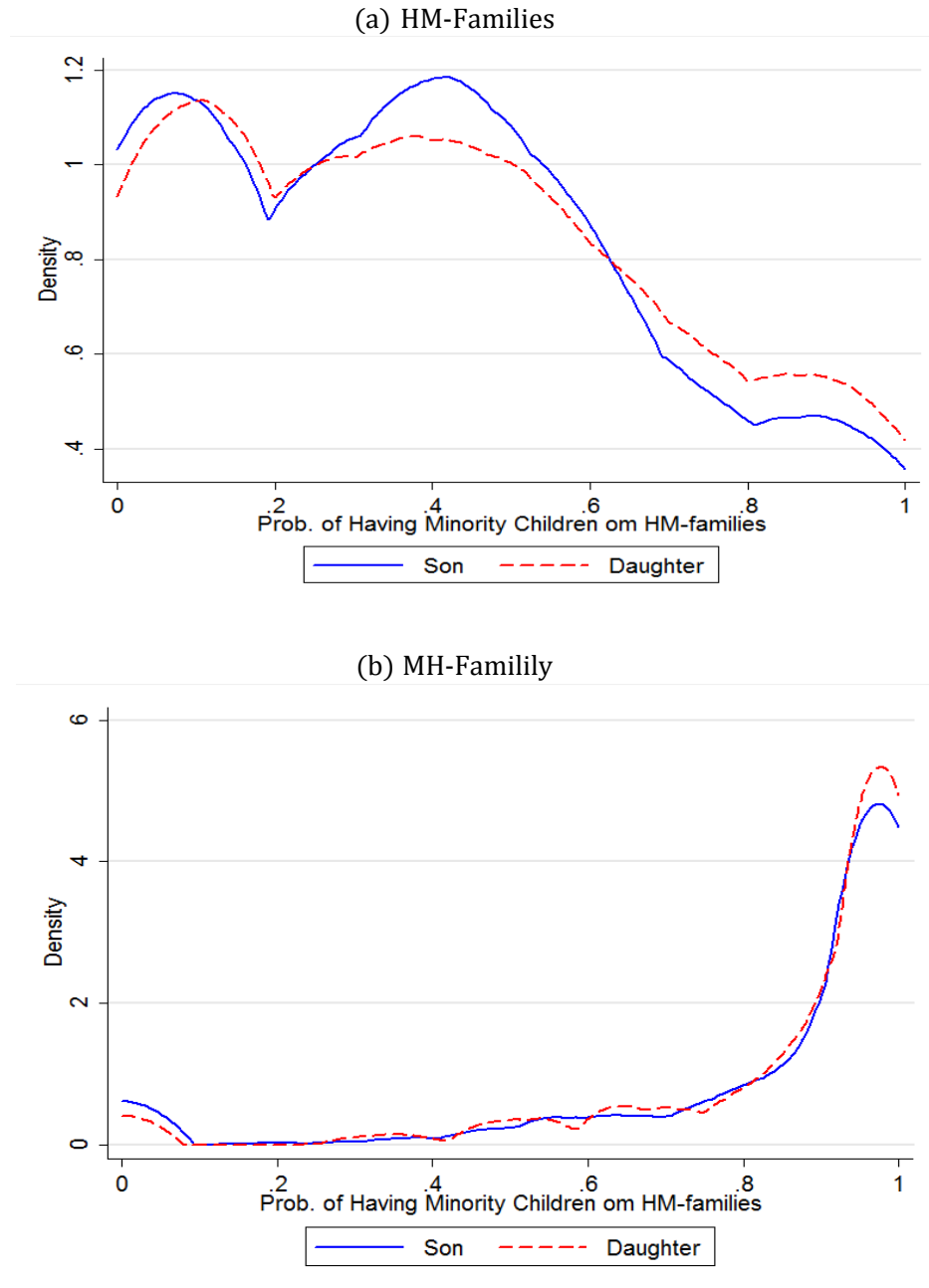


**Notes**: The two figures plot the distribution of shares of minority children across prefectures in the 1970s cohort. Panel (a) shows that this share varies a great deal across regions for HM-marriages, whereas panel (b) shows that MH-couples in most prefectures have almost exclusively minority children.

**Figure 3  Spatial Distribution of Share of Minority Children (for 1970s Cohort)**
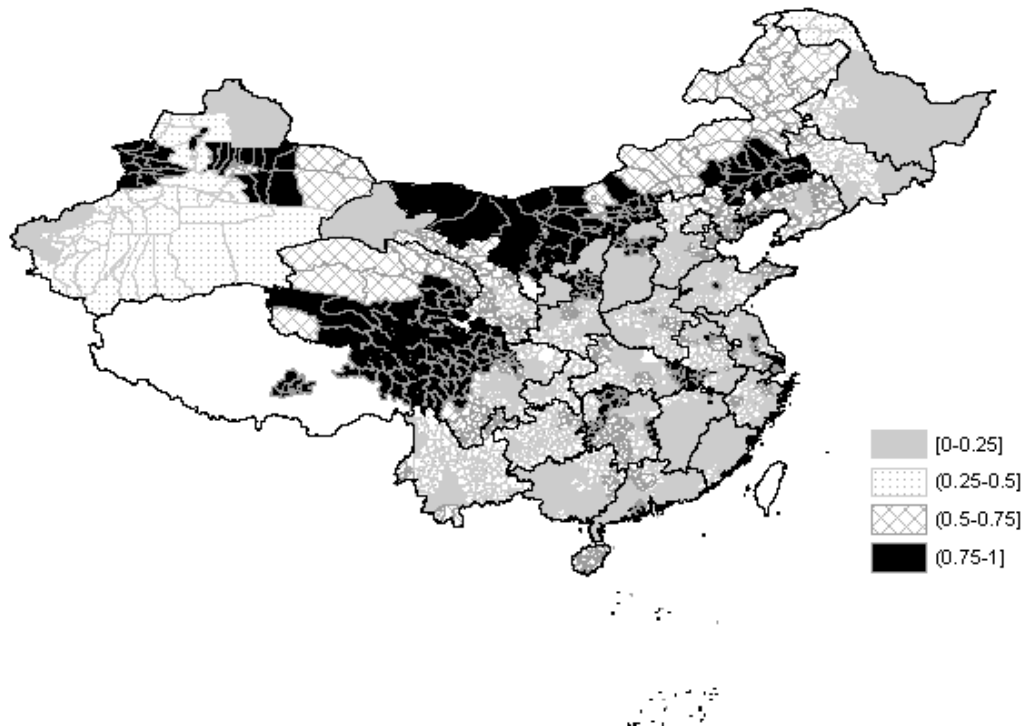


**Notes**: This figure maps the average share of minority children in mixed marriages with a male-Han. The share is calculated based on the 1970s cohorts. A set of province fixed effects explains only about 35% of the variations across prefectures.

## Figure 4 The Effects of Individual Benefits (Post Policy) * Social Motives
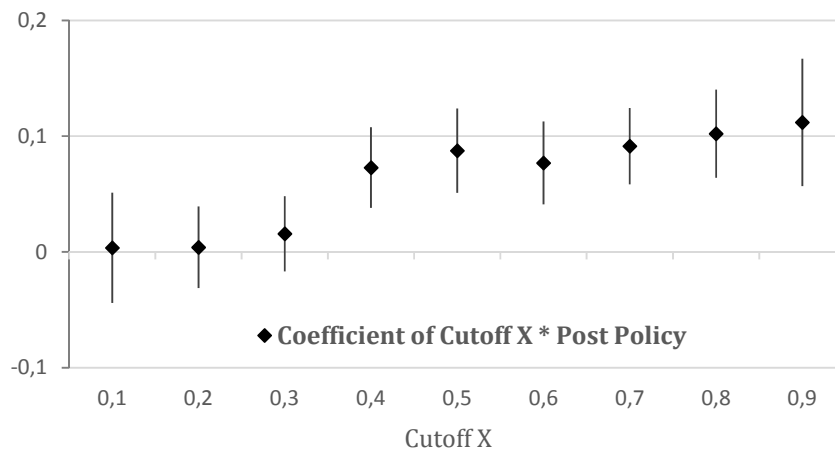


**Notes**: This figure plots the results for prediction P1 using different cutoff values for the share of minority children, ranging from 0.1 to 0.9. The line through each dot indicates the 95% confidence interval. Post Policy is defined based on the rollout of one-child policy.

## Figure 5 Social Motives*Post Year (1975, 1976,....,1985)
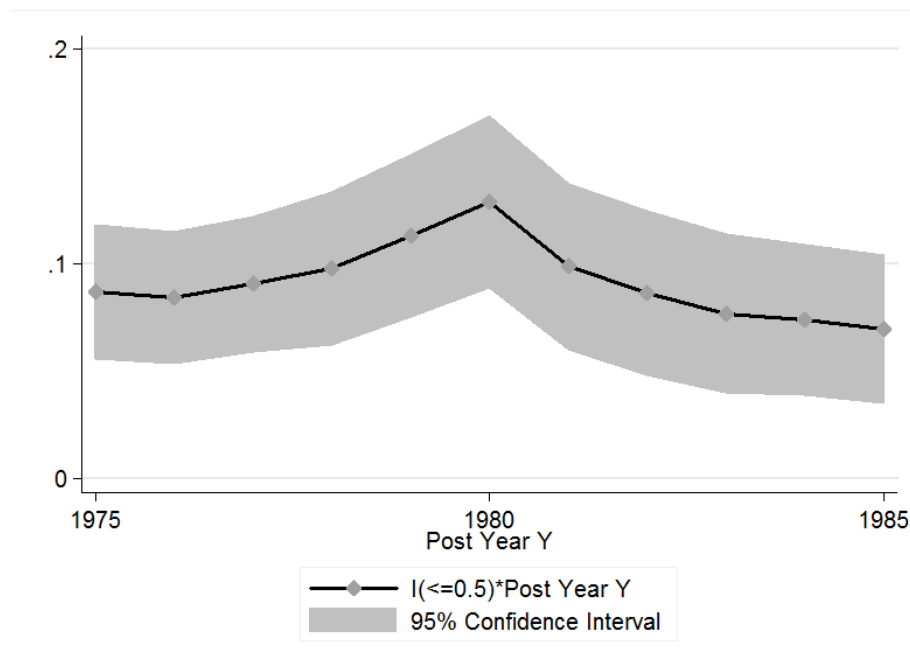


**Notes**: This figure presents the the estimates of the individual-social interaction (using 0.5 as the cutoff share), for each possible cutoff year from 1975 to 1985. It shows that the estimated effect is the largest, when we compare the effects before and after 1980.

## Table 1  Summary Statistics by Subsample

| Variable | All Obs. | Mean (s.d.) | Before 1980 Obs. | Mean (s.d.) | After 1980 Obs. | Mean (s.d.) |
|---|---|---|---|---|---|---|
| Child Minority (MH-family) | 129,447 | 0.94 (0.24) | 63,136 | 0.95 (0.21) | 66,311 | 0.93 (0.26) |
| Child Minority (HM-family) | 143,770 | 0.44 (0.50) | 60,055 | 0.38 (0.49) | 83,715 | 0.49 (0.50) |
| Zhuang Wife (HM-family) | 143,770 | 0.17 (0.37) | 60,055 | 0.17 (0.37) | 83,715 | 0.17 (0.38) |
| Son (HM-family) | 143,770 | 0.54 (0.50) | 60,055 | 0.55 (0.50) | 83,715 | 0.53 (0.50) |
| Religious Wife (HM-family) | 143,770 | 0.19 (0.39) | 60,055 | 0.18 (0.38) | 83,715 | 0.20 (0.40) |
| Post Policy (HM-families) | 140,296 | 0.51 (0.50) | | | | |
| Post 1980 (HM-families) | 143,770 | 0.58 (0.49) | | | | |

**Notes**: Consistent with the model, the children in MH-families are very likely to be minorities (the mean is 94%). As this leaves little variation for us to explore, our analysis focuses on HM-families.

## Table 2  Results I for P1: Individual and Social Motives

| | (1) | (2) X=0.4 | (3) X=0.5 | (4) X=0.6 | (5) | (6) X=0.4 | (7) X=0.5 | (8) X=0.6 |
|---|---|---|---|---|---|---|---|---|
| Mean (before 1980) | 0.41 | | | | 0.41 | | | |
| I($\leq$X)*Post Policy | | 0.073*** | 0.087*** | 0.077*** | | | | |
| | | (0.018) | (0.019) | (0.018) | | | | |
| Post Policy | 0.079*** | -0.000 | -0.016 | -0.017 | | | | |
| | (0.011) | (0.013) | (0.012) | (0.014) | | | | |
| I($\leq$X)*Post 1980 | | | | | | 0.120*** | 0.129*** | 0.128*** |
| | | | | | | (0.018) | (0.020) | (0.025) |
| Post 1980 | | | | | 0.075*** | | | |
| | | | | | (0.010) | | | |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Province FE*Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Observations | 122,581 | 122,581 | 122,581 | 122,581 | 125,654 | 125,654 | 125,654 | 125,654 |
| R-squared | 0.291 | 0.297 | 0.297 | 0.297 | 0.285 | 0.293 | 0.293 | 0.292 |

**Notes**: Columns (1)-(4) present the results using the measure Post Policy, while columns (5)-(8) use Post 1980 as a robustness check. Figure 5 visualizes the results for alternative cutoff years for higher minority benefits. The data come from three censuses and a mini census from 1982-2005. Cohorts born before 1960 are dropped as they cannot be matched with information on the share of minority children in the previous cohort. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

## Table 3A  Results II for P1: Peer Groups by Residency

| | (1) Urban | (2) X=0.4 Urban | (3) X=0.5 Urban | (4) X=0.6 Urban | (5) Rural | (6) X=0.4 Rural | (7) X=0.5 Rural | (8) X=0.6 Rural |
|---|---|---|---|---|---|---|---|---|
| Mean (before 1980) | 0.60 | | | | 0.49 | | | |
| I(≤X)*Post Policy | | 0.061** | 0.055* | 0.072*** | | 0.047** | 0.060*** | 0.074*** |
| | | (0.028) | (0.029) | (0.027) | | (0.023) | (0.022) | (0.023) |
| Post Policy | 0.061*** | 0.025 | 0.018 | 0.006 | 0.020* | -0.002 | -0.015 | -0.030* |
| | (0.012) | (0.016) | (0.018) | (0.018) | (0.010) | (0.015) | (0.016) | (0.016) |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Province FE*Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Observations | 14,036 | 14,036 | 14,036 | 14,036 | 42,271 | 42,271 | 42,271 | 42,271 |
| R-squared | 0.249 | 0.254 | 0.255 | 0.256 | 0.338 | 0.338 | 0.338 | 0.338 |

**Notes**: Columns (1)-(4) present the results when the peer group is made up by rural couples (classified based on the husband's Hukou) in the previous cohort in the same prefecture, while columns (5)-(8) present the results for peer groups of urban couples. The data come from the censuses 2000 and 2005, where Hukou information was present. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

## Table 3B  Results III for P1: Peer Groups by Education

| | (1) Above HS | (2) X=0.4 Above HS | (3) X=0.5 Above HS | (4) X=0.6 Above HS | (5) Below HS | (6) X=0.4 Below HS | (7) X=0.5 Below HS | (8) X=0.6 Below HS |
|---|---|---|---|---|---|---|---|---|
| Mean (before 1980) | 0.53 | | | | 0.40 | | | |
| I($\leq$X)*Post Policy | | 0.103*** | 0.117*** | 0.118*** | | 0.065*** | 0.056*** | 0.071*** |
| | | (0.029) | (0.029) | (0.027) | | (0.019) | (0.018) | (0.018) |
| Post Policy | 0.055*** | -0.002 | -0.011 | -0.022 | 0.074*** | -0.004 | -0.003 | -0.021 |
| | (0.015) | (0.016) | (0.017) | (0.016) | (0.011) | (0.014) | (0.012) | (0.014) |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y2 | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Province FE*Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Observations | 20,208 | 20,208 | 20,208 | 20,208 | 101,748 | 101,748 | 101,748 | 101,748 |
| R-squared | 0.294 | 0.302 | 0.303 | 0.303 | 0.294 | 0.300 | 0.300 | 0.300 |

Notes: Columns (1)-(4) present the results when the peer group is made up by couples in which the husband has not completed high school in the previous cohort in the same prefecture, while columns (5)-(8) present the results for couples in which the husband's education is high school or above.  Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

Table 3C Results IV for P1: Peer Groups by Wife's Ethnicity

| | (1) | (2) X=0.4 | (3) X=0.5 | (4) X=0.6 | (5) | (6) X=0.4 | (7) X=0.5 | (8) X=0.6 |
|---|---|---|---|---|---|---|---|---|
| Mean (before 1980) | 0.41 | | | | 0.41 | | | |
| I(≤X)*Post Policy | | 0.062*** | 0.066*** | 0.084*** | | | | |
| | | (0.016) | (0.018) | (0.017) | | | | |
| Post Policy | 0.083*** | 0.008 | 0.002 | -0.017 | | | | |
| | (0.011) | (0.011) | (0.013) | (0.012) | | | | |
| I(≤X)*Post 1980 | | | | | | 0.071*** | 0.078*** | 0.111*** |
| | | | | | | (0.017) | (0.019) | (0.020) |
| Post 1980 | | | | | 0.079*** | | | |
| | | | | | (0.010) | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Province FE*Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Observations | 117,310 | 117,310 | 117,310 | 117,310 | 120,211 | 120,211 | 120,211 | 120,211 |
| R-squared | 0.291 | 0.303 | 0.302 | 0.300 | 0.285 | 0.297 | 0.297 | 0.295 |

**Notes**: Columns (1)-(4) present the results when the peer group is made up by couples in the previous cohort in the same prefecture with a Han husband and a wife of one specific minority (out of 55) and the policy shift in each province is measured by the roll out of the one-child policy. Columns (5)-(8) present the results for the same definition of the peer group, but the policy shift is instead measured by a post-1980 dummy. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

Table 4  Results for P1':  Interaction Effects by Quartiles

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| I(0-0.25)*Post Policy | 0.068*** | 0.070*** | 0.109*** | 0.100*** | | | | |
| | (0.018) | (0.018) | (0.023) | (0.023) | | | | |
| I(0.25-0.5)*Post Policy | 0.092*** | 0.095*** | 0.124*** | 0.120*** | | | | |
| | (0.025) | (0.026) | (0.024) | (0.024) | | | | |
| I(0.5-0.75)*Post Policy | 0.048** | 0.049** | 0.050** | 0.056** | | | | |
| | (0.022) | (0.023) | (0.023) | (0.023) | | | | |
| Post Policy | -0.003 | -0.017 | -0.043** | -0.041** | | | | |
| | (0.014) | (0.014) | (0.017) | (0.017) | | | | |
| I(0-0.25)*Post 1980 | | | | | 0.064*** | 0.064*** | 0.178*** | 0.174*** |
| | | | | | (0.021) | (0.021) | (0.033) | (0.034) |
| I(0.25-0.5)*Post 1980 | | | | | 0.080*** | 0.086*** | 0.174*** | 0.171*** |
| | | | | | (0.026) | (0.026) | (0.031) | (0.032) |
| I(0.5-0.75)*Post 1980 | | | | | 0.038 | 0.039 | 0.073*** | 0.077*** |
| | | | | | (0.029) | (0.029) | (0.028) | (0.029) |
| Post 1980 | | | | | -0.005 | | | |
| | | | | | (0.019) | | | |
| *p*-value | 0.092 | 0.100 | 0.004 | 0.011 | 0.106 | 0.000 | 0.000 | 0.000 |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | | Y | Y | Y | | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Prov. FE*Cohort FE | | | Y | Y | | | Y | Y |
| Exclude Migrants | | | | Y | | | | Y |
| Observations | 122,581 | 122,581 | 122,581 | 113,889 | 125,654 | 125,654 | 125,654 | 116,373 |
| R-squared | 0.281 | 0.294 | 0.298 | 0.308 | 0.275 | 0.289 | 0.293 | 0.304 |

**Notes**: This table contains tests of Prediction P1' on the relative effects in different qurtiles, when the policy shift is measured by the roll out of the one-child policy (columns (1)-(4)) and by the 1980-dummy (columns (5)-(8)).  According to the model, the interaction effects estimated for the first three quartiles should be larger than that for the fourth quartile. Further, the effect for the second quartile should be larger than that for the third quartile – the *p*-values refer to tests for a difference between the effects in the second and third quartile. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Zhuang Wife * Post Policy | -0.097*** | -0.088*** | -0.073*** | -0.080*** | -0.088*** | | | |
| | (0.033) | (0.031) | (0.026) | (0.024) | (0.024) | | | |
| Post Policy | 0.067*** | 0.021 | 0.000 | 0.003 | 0.003 | | | |
| | (0.022) | (0.022) | (0.019) | (0.019) | (0.019) | | | |
| Zhuang Wife * Post 1980 | | | | | | -0.082** | -0.097** | -0.097** |
| | | | | | | (0.039) | (0.037) | (0.038) |
| Zhuang Wife | -0.076 | -0.081* | -0.087* | | | -0.079* | | |
| | (0.049) | (0.048) | (0.045) | | | (0.046) | | |
| Minority Share (Wife)*Post Policy | 0.063* | 0.061* | 0.074*** | 0.073*** | 0.075*** | | | |
| | (0.032) | (0.031) | (0.025) | (0.024) | (0.025) | | | |
| Minority Share (Wife) | -0.071** | -0.069** | -0.081*** | | | -0.086*** | | |
| | (0.035) | (0.034) | (0.031) | | | (0.032) | | |
| Minority Share (Wife)*Post 1980 | | | | | | 0.077*** | 0.085*** | 0.086*** |
| | | | | | | (0.027) | (0.027) | (0.028) |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | Y | Y | Y | Y |
| Province FE*Cohort FE | | | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | | | | Y | Y | | Y | Y |
| Exclude Migrants | | | | | Y | | | Y |
| Observations | 140,296 | 140,296 | 140,296 | 140,296 | 131,352 | 143,770 | 143,770 | 134,200 |
| R-squared | 0.278 | 0.280 | 0.288 | 0.296 | 0.305 | 0.283 | 0.292 | 0.301 |

**Notes**: This table shows that the effect of the policy change is smaller for families with a Zhuang wife, consistent with the fact that the Zhuang minority enjoys fewer benefits. It also shows that this finding is not driven by the size of minorities.  Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

## Table 6  Results I for P3:  Heterogenous Effects of Intrinsic Costs (Sons vs. Daughters)

|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Son*Post Policy | -0.016*** | -0.019*** | -0.018*** | -0.019*** | -0.017*** | | | |
|  | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | | | |
| Post Policy | 0.107*** | 0.059*** | 0.051*** | 0.050*** | 0.049*** | | | |
|  | (0.013) | (0.009) | (0.009) | (0.009) | (0.009) | | | |
| Son*Post 1980 | | | | | | -0.018*** | -0.019*** | -0.016*** |
|  | | | | | | (0.005) | (0.005) | (0.005) |
| Son | -0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 |
|  | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | Y | Y | Y | Y |
| Province FE*Cohort FE | | | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | | | | Y | Y | | Y | Y |
| Exclude Migrants | | | | | Y | | | Y |
| Observations | 140,296 | 140,296 | 140,296 | 140,296 | 131,352 | 143,770 | 143,770 | 134,200 |
| R-squared | 0.273 | 0.276 | 0.284 | 0.296 | 0.305 | 0.279 | 0.292 | 0.301 |

**Notes**: This table shows that the effect of the policy change is smaller for sons, consistent with the interpretation that it is more costly for a Han man to have a son (than a daughter) with a different ethnicity.  Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

### Table 7  Results II for P3:  Heterogenous Effects of Intrinsic Costs (Religious vs. Non-religious Wives)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Religious Wife*Post Policy | -0.062*** | -0.061*** | -0.037*** | -0.032** | -0.034*** | | | |
| | (0.017) | (0.017) | (0.013) | (0.013) | (0.013) | | | |
| Post Policy | 0.122*** | 0.071*** | 0.033* | 0.033* | 0.035* | | | |
| | (0.022) | (0.020) | (0.018) | (0.018) | (0.018) | | | |
| Religious Wife*Post 1980 | | | | | | -0.039** | -0.028* | -0.031** |
| | | | | | | (0.015) | (0.015) | (0.015) |
| Religious Wife | 0.087*** | 0.087*** | 0.069*** | | | 0.076*** | | |
| | (0.019) | (0.019) | (0.017) | | | (0.018) | | |
| Minority Share (Wife)*Post Policy | -0.018 | -0.014 | 0.018 | 0.018 | 0.015 | | | |
| | (0.021) | (0.020) | (0.018) | (0.018) | (0.019) | | | |
| Minority Share (Wife)*Post 1980 | | | | | | 0.023 | 0.036 | 0.036 |
| | | | | | | (0.021) | (0.023) | (0.023) |
| Minority Share (Wife) | -0.081*** | -0.082*** | -0.104*** | | | -0.108*** | | |
| | (0.027) | (0.027) | (0.025) | | | (0.024) | | |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | Y | Y | Y | Y |
| Province FE*Cohort FE | | | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | | | | Y | Y | | Y | Y |
| Exclude Migrants | | | | | Y | | | Y |
| Observations | 140,296 | 140,296 | 140,296 | 140,296 | 131,352 | 143,770 | 143,770 | 134,200 |
| R-squared | 0.277 | 0.280 | 0.287 | 0.296 | 0.305 | 0.282 | 0.292 | 0.301 |

**Notes**: This table shows that the effect of the policy change is smaller when the wife belongs to a minority group with Islamic or Tibetan religion, consistent with the hypothesis that it is more costly for a Han man to have a child of a minority ethnicity, when that ethnicity practices a different religion.  Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

## Table 8 Alternative Explanations I: Bargaining Power

| | (1) X=0.4 | (2) X=0.5 | (3) X=0.6 | (4) X=0.4 | (5) X=0.5 | (6) X=0.6 | (7) X=0.4 | (8) X=0.5 | (9) X=0.6 |
|---|---|---|---|---|---|---|---|---|---|
| I(≤X)*Post Policy | 0.072*** | 0.087*** | 0.076*** | 0.075*** | 0.090*** | 0.080*** | 0.071*** | 0.085*** | 0.076*** |
| | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.017) | (0.018) | -0.018 |
| Post Policy | -0.000 | -0.016 | -0.016 | -0.002 | -0.018 | -0.020 | 0.003 | -0.013 | -0.014 |
| | (0.013) | (0.012) | (0.014) | (0.013) | (0.012) | (0.015) | (0.012) | (0.012) | (0.014) |
| I(≤X)* (Husband – Wife Edu) | -0.011* | -0.012* | -0.017** | | | | | | |
| | (0.006) | (0.007) | (0.007) | | | | | | |
| Husband – Wife Edu | -0.003 | -0.001 | 0.003 | | | | | | |
| | (0.004) | (0.005) | (0.005) | | | | | | |
| I(≤X)* (Husband – Wife Age) | | | | 0.005*** | 0.004*** | 0.005*** | | | |
| | | | | (0.001) | (0.001) | (0.001) | | | |
| Husband – Wife Age | | | | -0.004*** | -0.004*** | -0.005*** | | | |
| | | | | (0.001) | (0.001) | (0.001) | | | |
| I(≤X)*Sex Ratio | | | | | | | -0.022 | -0.079 | 0.011 |
| | | | | | | | (0.078) | (0.076) | (0.086) |
| Sex Ratio | | | | | | | 0.134* | 0.174** | 0.118 |
| | | | | | | | (0.069) | (0.072) | (0.072) |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Province FE*Cohort FE | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 122,581 | 122,581 | 122,581 | 122,581 | 122,581 | 122,581 | 122,581 | 122,581 | 122,581 |
| R-squared | 0.298 | 0.298 | 0.297 | 0.298 | 0.298 | 0.297 | 0.298 | 0.298 | 0.297 |

**Notes**: This table shows that our main results cannot be explained by changes in the bargaining power of minority women, when this is proxied by education differences, age differences and sex ratios. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

## Table 9  Alternative Explanations II:  Endogeneous Mixed Marriages

| | (1) | (2) X=0.5 | (3) | (4) X=0.5 | (5) | (6) X=0.5 | (7) | (8) X=0.5 |
|---|---|---|---|---|---|---|---|---|
| I($\leq$X)*Post Policy | | 0.083*** (0.016) | | | | 0.011 (0.020) | | |
| Post Policy | 0.061*** (0.009) | -0.018 (0.011) | | | 0.026*** (0.009) | 0.021*** (0.011) | | |
| I($\leq$X)*Post 1980 | | | | 0.121*** (0.019) | | | | 0.054* (0.031) |
| Post 1980 | | | 0.057*** (0.009) | | | | 0.015* (0.008) | |
| I($\leq$X)*HM Share | | 0.002 (0.002) | | 0.002 (0.002) | | | | |
| HM Share | 0.006** (0.003) | 0.002 (0.002) | 0.007*** (0.003) | 0.002 (0.002) | | | | |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | | Y | | Y | | Y |
| Province FE*Cohort FE | | Y | | Y | | Y | | Y |
| Observations | 122,581 | 122,581 | 125,654 | 125,654 | 17,779 | 17,779 | 20,104 | 20,104 |
| R-squared | 0.292 | 0.298 | 0.287 | 0.293 | 0.305 | 0.307 | 0.284 | 0.287 |

**Notes**: Endogenous mixed marriages are unlikely to explain our findings, as we explore decisions that are made once couples have entered into marriage. Columns (1)-(4) show that the main results indeed hold when we control for the incidence of mixed marriages in the prefecure and cohort. Columns (5)-(8) focus on the small subsample where the couple married before the policy and the child were born after the policy. The information on marriage year is only available for censuses 2000 and 2005.  Standard errors are clustered at the prefecture level. Significance: ***, 1%, *, 5%, *, 10%.

Table 10  Alternative Explanations III:  Censoring (Shares between 0.3 and 0.7 only)

| | (1) | (2) X=0.4 | (3) X=0.5 | (4) X=0.6 | (5) | (6) X=0.4 | (7) X=0.5 | (8) X=0.6 |
|---|---|---|---|---|---|---|---|---|
| I(≤X)*Post Policy | | 0.069*** | 0.105*** | 0.081** | | | | |
| | | (0.026) | (0.029) | (0.040) | | | | |
| Post Policy | 0.111*** | 0.043** | 0.003 | -0.002 | | | | |
| | (0.017) | (0.018) | (0.020) | (0.039) | | | | |
| I(≤X)*Post 1980 | | | | | | 0.103*** | 0.127*** | 0.124** |
| | | | | | | (0.024) | (0.030) | (0.050) |
| Post 1980 | | | | | 0.108*** | | | |
| | | | | | (0.017) | | | |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Province FE*Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Observations | 55,145 | 55,145 | 55,145 | 55,145 | 55,145 | 56,684 | 56,684 | 56,684 |
| R-squared | 0.096 | 0.104 | 0.104 | 0.103 | 0.094 | 0.103 | 0.103 | 0.102 |

**Notes**: This table shows the baseline results on a sample restricted such that the share of minority children lies between 0.3 and 0.7. It  shows that the room to change should not be a critical concern. Standard errors are clustered at the prefecture level. Significance: ***, 1%, *, 5%, *, 10%.

# Web Appendix

Table A1 describes mixed-marriage patterns in China.

Tables A2-A3 present individual-level checks on facts F1 and F2 depicted in Figure 1.

Tables A4-A6 show additional robustness checks on the test of the main prediction P1.

Table A7 describes the changes in bargaining power of wives across time and shares of minority children.

Table A8 presents correlations between mixed marriages and the probability of having a minority child.

### Table A1: Differences across Marriages

|  | HH | MM | HM | MH | |
|---|---|---|---|---|---|
| #Couples | 6436486 | 417089 | 90704 | 81570 | |
| Share in total marriages | 91.60% | 5.90% | 1.30% | 1.20% | |
|  |  |  |  |  | |
| Share of HM marriage for a minority woman |  |  |  |  | 1.3/(1.3+5.9)=18% |
| Share of MH marriage for a minority man |  |  |  |  | 1.2/(1.2+5.9)=17% |
|  |  |  |  |  | |
| Husband Edu-Wife Edu | 0.27 | 0.26 | 0.21 | 0.23 | |
| Husband Age-Wife Age | 2.41 | 2.72 | 2.8 | 2.48 | |

**Notes:** This table describes the marriage patterns among all the married couples in the four censuses.

## Table A2  Fact F1:  HM-Families versus MH-Families
### Dependent Variable: Minority Child = 0/1

|  | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| MH-Families | 0.493*** | 0.471*** | 0.473*** | 0.473*** |
|  | (0.029) | (0.028) | (0.028) | (0.028) |
|  |  |  |  |  |
| Prefecture FE |  | Y | Y | Y |
| Cohort FE |  |  | Y | Y |
| Province FE*Cohort |  |  |  | Y |
| Observations | 273,217 | 273,217 | 273,217 | 273,217 |
| R-squared | 0.279 | 0.380 | 0.382 | 0.385 |

**Notes**: This table shows that fact F1 in Figure 1 also holds at the individual level. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

### Table A3  Fact F2:  Ethnicity of Children by Cohorts
#### Dependent Variable: Minority Child = 0/1

|  | (1) HM | (2) HM | (3) HM | (4) MH | (5) MH | (6) MH |
|---|---|---|---|---|---|---|
| Cohort 70-74 | 0.083*** | 0.070*** | 0.061*** | 0.006* | 0.004 | 0.018** |
|  | (0.011) | (0.011) | (0.015) | (0.003) | (0.003) | (0.008) |
| Cohort 75-79 | 0.080*** | 0.087*** | 0.069** | 0.002 | 0.002 | 0.028* |
|  | (0.012) | (0.013) | (0.027) | (0.004) | (0.004) | (0.016) |
| Cohort 80-84 | 0.122*** | 0.118*** | 0.090** | -0.010** | -0.010** | 0.028 |
|  | (0.017) | (0.015) | (0.039) | (0.005) | (0.005) | (0.024) |
| Cohort 85-89 | 0.166*** | 0.158*** | 0.123** | -0.018*** | -0.016*** | 0.035 |
|  | (0.020) | (0.019) | (0.051) | (0.006) | (0.005) | (0.031) |
| Cohort 90+ | 0.183*** | 0.179*** | 0.133** | -0.053*** | -0.042*** | 0.020 |
|  | (0.025) | (0.021) | (0.064) | (0.007) | (0.006) | (0.039) |
|  |  |  |  |  |  |  |
| Prefecture FE | Y | Y | Y | Y | Y | Y |
| Cohort FE |  | Y | Y |  | Y | Y |
| Province FE*Cohort |  |  | Y |  |  | Y |
| Observations | 143,770 | 143,770 | 143,770 | 129,447 | 129,447 | 129,447 |
| R-squared | 0.014 | 0.271 | 0.278 | 0.007 | 0.075 | 0.080 |

**Notes**: This table shows that fact F2 in Figure 1 also holds at the individual level. The probability of minority children in HM-families is increasing over time, but the increase becomes more salient after 1980s. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

**Table A4: Results V for P1: Using the 1970s Cohort as the Peer Group**

|  | (1) | (2) X=0.4 | (3) X=0.5 | (4) X=0.6 | (5) | (6) X=0.4 | (7) X=0.5 | (8) X=0.6 |
|---|---|---|---|---|---|---|---|---|
| I($\leq$X)*Post Policy |  | 0.060*** | 0.089*** | 0.088*** |  |  |  |  |
|  |  | (0.018) | (0.020) | (0.017) |  |  |  |  |
| Post Policy | 0.100*** | 0.008 | -0.015 | -0.023* |  |  |  |  |
|  | (0.012) | (0.012) | (0.013) | (0.014) |  |  |  |  |
| I($\leq$X)*Post 1980 |  |  |  |  |  | 0.072*** | 0.109*** | 0.107*** |
|  |  |  |  |  |  | (0.021) | (0.023) | (0.019) |
| Post 1980 |  |  |  |  | 0.098*** |  |  |  |
|  |  |  |  |  | (0.011) |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE |  | Y | Y | Y |  | Y | Y | Y |
| Province FE*Cohort FE |  | Y | Y | Y |  | Y | Y | Y |
| Observations | 140,296 | 140,296 | 140,296 | 140,296 | 143,770 | 143,770 | 143,770 | 143,770 |
| R-squared | 0.286 | 0.296 | 0.297 | 0.297 | 0.281 | 0.292 | 0.293 | 0.292 |

**Notes**: This table shows that the results in Table 2 are robust to using the 1970s cohort in the same prefecture as the peer group. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

| | (1) | (2) X=0.4 | (3) X=0.5 | (4) X=0.6 | (5) | (6) X=0.4 | (7) X=0.5 | (8) X=0.6 |
|---|---|---|---|---|---|---|---|---|
| I(≤X)*Post Policy | | 0.068*** | 0.080*** | 0.074*** | | | | |
| | | (0.017) | (0.018) | (0.018) | | | | |
| Post Policy | 0.080*** | 0.002 | -0.012 | -0.015 | | | | |
| | (0.011) | (0.013) | (0.012) | (0.014) | | | | |
| I(≤X)*Post 1980 | | | | | | 0.118*** | 0.122*** | 0.128*** |
| | | | | | | (0.017) | (0.020) | (0.026) |
| Post 1980 | | | | | 0.075*** | | | |
| | | | | | (0.010) | | | |
| | | | | | | | | |
| Prefecture FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Ethnicity FE (Wife) | Y | Y | Y | Y | Y | Y | Y | Y |
| Excluding Migrants | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Province FE*Cohort FE | | Y | Y | Y | | Y | Y | Y |
| Observations | 113,889 | 113,889 | 113,889 | 113,889 | 116,373 | 116,373 | 116,373 | 116,373 |
| R-squared | 0.301 | 0.308 | 0.308 | 0.307 | 0.296 | 0.304 | 0.304 | 0.303 |

**Notes**: This table shows that the results in Table 2 are robust to dropping migrants from the sample. People born before 1960 are dropped as they cannot be matched with a previous cohort. Standard errors are clustered at the prefecture level. Significance: ***, 1%, **, 5%, *, 10%.

### Table A6: Results VII for P1: Using Information on Fines

|  | (1) | (2) 0.4 | (3) 0.5 | (4) 0.6 |
|---|---|---|---|---|
| I(≤X)*Fines |  | 0.054*** | 0.053*** | 0.029*** |
|  |  | (0.011) | (0.013) | (0.011) |
| Fines | 0.027*** | -0.019*** | -0.025*** | -0.019** |
|  | (0.007) | (0.006) | (0.008) | (0.008) |
|  |  |  |  |  |
| Prefecture FE | Y | Y | Y | Y |
| Ethnicity FE | Y | Y | Y | Y |
| Cohort FE |  | Y | Y | Y |
| Province FE*Cohort FE |  | Y | Y | Y |
| Observations | 82,727 | 82,727 | 82,727 | 82,727 |
| R-squared | 0.302 | 0.306 | 0.306 | 0.306 |

**Notes:** This table presents results using the level of fines to measure the importance of one-child policy in the subsample of mixed couples whose children are born after 1980. The same data has been used by Ebenstein (2010) and Wei and Zhang (2011).

**Table A7: Changes in Bargaining Power by Time and Share of Minority Children**

| | Cohort 1970, HM family | |
|---|---|---|
| | Share of minority children | |
| | <0.5 | >=0.5 |
| *(a) Husband's Education – Wife's Education* | | |
| Children born before 1980 | 0.32 | 0.28 |
| | (0.59) | (0.59) |
| Children born after 1980 | 0.18 | 0.17 |
| | (0.56) | (0.60) |
| | | |
| *(b) Husband's Age – Wife's Age* | | |
| Children born before 1980 | 3.06 | 3.01 |
| | (4.00) | 4.01 |
| Children born after 1980 | 2.87 | 2.34 |
| | (4.23) | (3.83) |
| | | |
| *(c) Sex Ratio in Father's Cohort* | | |
| Children born before 1980 | 1.04 | 1.05 |
| | (0.10) | (0.10) |
| Children born after 1980 | 1.09 | 1.05 |
| | (0.09) | (0.09) |

**Notes:** This table presents changes in three measures of bargaining power. It shows that bargaining power of women (measured by education and age differences between the husband and the wife) increases over time. However, there is no evidence that this increase is more pronounces for prefectures with a smaller share of minority children. The sex ratios faced by the fathers changed very little because the fathers of children born in the 1980s are generally married in the 1960s. Standard deviations are reported in paraphesis.

## Table A8: Correlations between Mixed Marriage and Minority Identity for the Children

(Prefecture-cohort rather than individual data; D.V.: share of minority children)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
|  | HM | HM | HM | MH | MH | MH |
| HM marriage share for a Han man | 0.004*** | 0.004*** | 0.002 |  |  |  |
|  | (0.002) | (0.002) | (0.003) |  |  |  |
| MH marriage share of a minority man |  |  |  | 0.031 | 0.031 | 0.081 |
|  |  |  |  | (0.046) | (0.046) | (0.054) |
|  |  |  |  |  |  |  |
| Prefecture FE |  | Y | Y |  | Y | Y |
| Province FE*Cohort FE |  |  | Y |  |  | Y |
| Observations | 1,633 | 1,633 | 1,633 | 1,527 | 1,527 | 1,527 |
| R-squared | 0.696 | 0.696 | 0.738 | 0.531 | 0.531 | 0.591 |

**Notes**: The tale displays the correlations between the shares of mixed marriages and the shares of minority children in the same type of marriage.