北京大学中国经济研究中心
China Center for Ecnomic Research
讨论稿系列
Working Paper Series

E2025005 2025-03-11

Rationed Fertility: Treatment Effect Heterogeneity in the Child Quantity–Quality Tradeoff*

Rufei Guo† Junjian Yi‡ Junsen Zhang§ Ning Zhang¶

February 25, 2025

Abstract

We develop a generalized theory of rationed fertility to analyze treatment effect heterogeneity in the child quantity–quality tradeoff. An exogenous increase in fertility can be either desired or undesired. We derive a positive rationing income effect on child quality for desired fertility increases, but a negative rationing income effect for undesired fertility increases. We propose an econometric framework to identify treatment effects of desired and undesired fertility increases, and estimate a structural model to gauge the quantitative importance of the novel rationing income effect. Our study highlights the importance of distinguishing between desired and undesired changes when evaluating social programs.

*We wish to thank the editor and three anonymous referees for providing extremely helpful comments on earlier versions of the paper. We appreciate valuable discussions with Yi Chen, Pierre-André Chiappori, David Figlio, Naijia Guo, Haoran He, Juan Pantano, Mark Rosenzweig, Shangjin Wei, Liqiu Zhao, and Minghai Zhou. We are also thankful to seminar participants at the 2016 Annual Meeting of the China Labor Economists Forum, the 2017 China Meeting of the Econometric Society, the 2017 Annual Conference of the Asian Bureau of Finance and Economic Research, the 2019 International Symposium on Contemporary Labor Economics, the 2022 Kobe University Workshop on Trade and the Labor Market, the 2023 Wuhan University-Ohio State University Joint Workshop, the 2023 Quarterly Meeting of the China Labor Economists Forum, Beijing Normal University, Peking University, Wuhan University, and Zhejiang University of Finance and Economics. Rufei Guo acknowledges partial financial support from the National Natural Science Foundation of China (Grant No. 72373113). Junjian Yi acknowledges partial financial support from the National Natural Science Foundation of China (Grant No. 72342031). Junsen Zhang acknowledges partial financial support from the Key Program of the National Natural Science Foundation of China (Grant No. 72034006). This article was edited by James J. Heckman.

†Center for Health Economics and Management, Economics and Management School, Wuhan University. Email: rufei_guo@whu.edu.cn.

‡National School of Development, Peking University; Institute for Global Health and Development, Peking University. Email: junjian@nsd.pku.edu.cn.

§School of Economics, Zhejiang University. Email: jszhang@cuhk.edu.hk. Corresponding author.

¶Department of Economics, Chinese University of Hong Kong. Email: ning.zhang@cuhk.edu.hk.

1 Introduction

Since Becker (1960), the neoclassical paradigm considers fertility to be a human choice. Not infrequently, fertility is "rationed"—that is, fertility is determined largely by external forces outside of an individual's own control. The concept of rationed fertility is traced back at least as far as Malthus (1798). In the modern world, fertility is rationed on many occasions. Coercive fertility control once prevailed in China and India (Panandiker and Umashankar, 1994; Zhang, 2017), and it still exists around the world (WHO, 2014).[1] Rationed fertility also includes infertility and unwanted fertility, two general scenarios that constrain millions of people (Pantano, 2016).[2] Twinning is also a form of fertility rationing. Rationed fertility is a real-world phenomenon and has profound economic implications.

Rationed fertility, such as twinning, coercive fertility control, or infertility shock, has been widely used as natural or quasi-natural experiments in empirical studies on fertility. For example, economists are particularly interested in estimating the causal effect of fertility on child quality, i.e., the quantity–quality (QQ) effect, based on Becker and Lewis's (1973) theory of the child QQ trade-off, which predicts that the average quality of children is lower in larger families than in smaller ones.[3] However, estimates of the QQ effect based on rationed fertility and other exogeneous varations in fertility fall into a wide range. For example, Rosenzweig and Wolpin (1980), Hanushek (1992), Rosenzweig and Zhang (2009), and Bagger et al. (2021) find negative QQ effects; by contrast, Black, Devereux, and Salvanes (2005) and Angrist, Lavy, and Schlosser (2010) show that the effect is insignificant or positive. Mogstad and Wiswall (2016) and Brinch, Mogstad, and Wiswall (2017) find heterogeneous QQ effects among different fertility levels or different propensities for procreating. Despite a substantial body of empirical research, theoretical progress in understanding the QQ effect remains slow (Guo, Yi, and Zhang, 2022).

[1]China's "One-child" policy is coercive, as the control methods include mandated abortion, insertion of intrauterine device (IUD), tubal ligation (female sterilization), and vasectomy (male sterilization). In 1979–2000, China had 234 million cases of IUD insertion, 218 million abortion, 91 million tubal ligation, and 25 million vasectomy (National Health and Family Planning Committee, 2013). For members of the China Communist Party or employees of the public sector (including governments, universities, hospital, state-owned enterprises, etc.), violation of the fertility policy would lead to demolition or even dismissal. In India, a forced sterilization campaign was launched after 1975, leading to 8.1 million cases of sterilization in one year only (Panandiker and Umashankar, 1994). Recent decades continue to witness forced sterilization toward certain population groups, such as ethnic and sex minorities, and people with disability and illness. The World Health Organization, in collaboration with six prominent international organizations, issued a 36-page statement on "Eliminating forced, coercive and otherwise involuntary sterilization" (WHO, 2014).

[2]Infertility is the failure of pregnancy after frequent and unprotected sex for at least a year. Treatment for infertility, such as in vitro fertilization, is expensive and generally inaccessible to people in developing countries. More than 10% of the population suffer from infertility (Petraglia, Serour, and Chapron, 2013). Unwanted fertility, which is resulted from imperfect fertility control, is also widespread. In the US, more than 15% of births are unwanted; the proportion of unwanted children rises by birth order, and reaches at a startling 37% for the third births (Lin, Pantano, and Sun, 2020).

[3]We use child quantity, family size, and fertility interchangeably in this paper.

To analyze heterogeneous QQ effects, we develop a generalized theory of rationed fertility that distinguishes between desired and undesired changes. This distinction provides a new perspective to understand treatment effects and program evaluation in general. In particular, we use rationing theory to extend the QQ model, and find a new rationing income effect that generates differential QQ effects between desired and undesired fertility increases. In the original Becker–Lewis setup, child quantity and quality enter the household budget constraint in a multiplicative manner, so an increase in child quantity raises the cost of child quality (Becker and Lewis, 1973). The Becker–Lewis model generates a negative *correlation* between child quantity and quality. One is unable to directly derive the comparative statics of child quantity on child quality in the Becker–Lewis setup, since both child quantity and quality are choice variables. Rosenzweig and Wolpin (1980) appear to be the first to conduct a comparative static analysis of the effect of child quantity on child quality. In the Rosenzweig–Wolpin setup, fertility is rationed at the *optimal* level; the QQ effect consists of a price effect and a substitution effect. The price effect is always negative, as implied in the original Becker–Lewis setup. The sign of the substitution effect is determined by parental preference over child quantity and quality.

We generalize the comparative static analysis of the child quantity–quality tradeoff in the Rosenzweig–Wolpin setup, and evaluate the QQ effects when fertility is rationed at *any possible* level. In addition to the price and substitution effects in Rosenzweig and Wolpin (1980), we find a rationing income effect that appears for the first time in the literature. The sign of the rationing income effect is determined by the type of fertility change. A desired fertility change, which moves fertility toward the optimal level, has a positive rationing income effect. By contrast, an undesired fertility change, which moves fertility away from the optimal level, has a negative rationing income effect. The rationing income effect implies differential QQ effects between desired and undesired fertility changes.

We investigate the empirical content of our rationed fertility theory in three steps. First, we propose an econometric framework to separately identity QQ effects for desired and undesired fertility increases, respectively. Second, we separately estimate these QQ effects, exploiting the natural experiment of twin births and China's birth-control policy. Third, based on the estimated QQ effects, we gauge the quantitative importance of the rationing income effect by estimating a structural model that allows unobserved heterogeneity in both resources and preference. We find that the rationing income effect explain the difference in the total QQ effects between desired and undesired fertility increase.

Our econometric framework exploits both twinning and a birth-control policy to shift fertility. For brevity, we assume a pool of mothers with either two or three children. With neither a birth-control policy nor twinning at the second birth, mothers achieve their optimal fertility levels. Specifically, type-A mothers have two children, and type-B mothers have three children. With the birth-control policy, which targets two children per family, type-A mothers still have two children. By contrast, the policy rations the fertility level of type-B mothers, who consequently achieve two children under the policy.

With this simple setup, we identify the QQ effects for desired and undesired fertility increases. Without the policy, twinning shifts the realized fertility of type-A mothers from the optimal level of two to three, which is an undesired fertility increase. In this case, we identify the QQ effects for undesired fertility increases induced by twinning. With the policy, twinning also shifts the realized fertility of type-A mothers. At the same time, twinning helps type-B mothers circumvent the policy and achieve their optimal level of three children, which represents a desired fertility increase. Twinning under the policy enables us to identify the QQ effects for a combination of undesired and desired fertility increases. By comparing QQ effects without and with the policy, we identify the effects for desired fertility increases.

Our econometric analysis is closely related to the literature on multivalued treatments. Heckman, Hohmann, Smith, and Khoo (2000) show that the sign and magnitude of a treatment effect depends on treated individuals' alternatives if they do not take the treatment. For example, when evaluating the effects of the Job Training Partnership Act (JTPA), the alternatives for the treatment can either be participating in other training programs or no training at all. Heckman et al. (2000) show that the effects of the JTPA are larger for people who otherwise take no training than for people who otherwise take other training programs. Heckman and Vytlacil (2007a), Heckman, Urzua, and Vytlacil (2008), and Heckman and Urzúa (2010) find that the treatment effect is a weighted average of the effects for people induced to the treatment from different alternatives. Heckman and Pinto (2018) and Lee and Salanié (2018) further develop theoretical frameworks of multivalued treatments. The empirical literature on multivalued treatments focuses on isolating and identifying the treatment effects for people induced to the treatment from different alternatives (Kline and Walters, 2016; Kirkeboen, Leuven, and Mogstad, 2016; Hull, 2018; Mountjoy, 2020). Building on Hull (2018), we take twinning as the instrumental variable (IV) for child quantity, and use the birth-control policy as a "stratifying variable." Although our observed treatment is binary—mothers have either two or three children—it represents two types of fertility increases: desired or undesired. In the absence of the policy, twinning mainly induces undesired fertility increases; under the policy, twinning induces a mix of desired and undesired fertility increases.

We then use Chinese data to estimate QQ effects for desired and undesired fertility increases. We use fines for unauthorized births compiled by Ebenstein (2010) to measure the intensity of the coercive "One-child" policy, which varies across provinces and birth cohorts. We measure child quality by middle school attendance, because the middle school attendance rate changed dramatically across our sample period, but the rates for primary school and high school attendance remained stable. The "One-child" policy in rural China restricts births at the third or higher parities, but not at the second parity. We restrict our sample to rural mothers with at least two children, and examine the effects of twinning at the second birth on fertility and child quality, without and with the policy. We find that the policy magnifies the estimated effect of twinning on fertility. At the same time, the policy dampens or even reverses the negative effect of twinning on child quality. The estimates deliver a negative QQ effect for undesired fertility increases, and a positive QQ effect for desired fertility increases. The difference in QQ effects is consistent with the theoretical prediction based on the rationing income effect.

Finally, we develop a structural model that allows for unobserved heterogeneity in both preferences and resource constraints to quantify the rationing income effect in accounting for the difference in QQ effects between desired and undesired fertility increases. We structurally estimate the model by matching the model-predicted QQ effects and other key choices to our empirical estimates. Based on our structural model, decomposition analysis shows that the rationing income effect is the dominant force explaining the difference in QQ effects between type-A and type-B mothers. Specifically, the sum of price and substitution effects accounts for 20.9% of the difference in QQ effects between type-A and type-B mothers, whereas the rationing income effect accounts for 79.1% of the difference.

Our theory of rationed fertility provides a synthesis for three strands of literature on the heterogeneous QQ effects.[4] The first explores treatment effect heterogeneity for "wanted" versus "unwanted" fertility changes. Unwanted births occur when parents do not want more children, possibly as a result of imperfect fertility control. An unwanted fertility increase, as a deviation from parents' optimal choice, represents a case of undesired fertility increase in our theory. Early studies find that reductions in unwanted births, as induced by access to legalized abortion or contraceptives, improve educational and labor market outcomes of the affected cohorts in the US (Gruber, Levine, and Staiger, 1999; Ananat et al., 2009; Ananat and Hungerman, 2012; Pantano, 2016). These studies, however, do not directly test the QQ effects. More

[4] Guo, Yi, and Zhang (2022) provides a comprehensive review of the literature on the QQ tradeoff. The heterogeneous QQ effects along other dimensions, such as birth order, is beyond the scope of this paper.

recently, Sun (2019) exploits county-level roll-out of abortion clinics and finds a negative QQ effect for unwanted fertility increases. Similarly, exploiting county-level roll-out of family planning programs in the US, Bailey, Malkova, and McLaren (2019) detect a negative QQ effect. Their findings are consistent with our theory, as an unwanted fertility increase represents an undesired fertility increase, which induces a negative rationing income effect.

The closest work to ours is Lin et al. (2019), who use the framework of unordered monotonicity to study heterogeneous QQ effects (Heckman and Pinto, 2018). They identify the distinct effects of "planned" versus "unplanned" expansions in family size on the outcomes of older siblings. Our analysis differs from theirs in two respects. First, our primary objective is to test the generalized theory of rationed fertility, while they contribute to the QQ literature by distinguishing the effects of planned versus unplanned children using a framework with multiple treatments. Second, we use the twin instrument and China's birth-control policy, while they use the same-sex instrument, along with exogenous variation in opportunities to avoid unplanned births and exploit data on pregnancy intention to distinguish between planned and unplanned births.

Second, Black, Devereux, and Salvanes (2010) explore treatment effect heterogeneity for "expected" and "unexpected" fertility increases. They consider that a fertility increase induced by the same-sex instrument is expected, while a fertility increase induced by twinning is unexpected. Using administrative data in Norway, they find that the effect of the unexpected fertility increase on child IQ is negative, while the effect of the expected fertility increase is negligible. The finding is consistent with our theory: The fertility increase induced by twinning is less desired than those induced by child gender.

Third, Schultz (2007) classifies population policies in developing countries into two types: "voluntary" and "mandatory" population policies. Schultz (2007, p. 3265) writes, "...voluntary versus mandatory population policies, which might lead to the same decline in fertility, could theoretically have a different effect on other family outcomes and on the distribution of welfare losses and gains." Our model of rationed fertility provides a theoretical basis for this intuitive insight. On the one hand, a voluntary policy, such as the distribution of contraception tools and knowledge, helps parents reduce fertility to their desired level, and generates a positive rationing income effect. The empirical evidence consistently shows that fertility reductions induced by voluntary policies improve child quality (Rosenzweig and Schultz, 1987; Joshi and Schultz, 2013; Dang and Halsey Rogers, 2016). Macroeconomic studies have also demonstrated that voluntary policies targeting the reduction of unwanted births are effective in promoting human capital accumulation, particularly among disadvantaged socioeconomic groups (Cavalcanti, Kocharkov, and Santos, 2021; Seshadri and Zhou, 2022). On the other hand, a mandatory policy forces parents to deviate from their desired fertility level, and generates a negative rationing income effect. The empirical evidence is mixed for the effect induced by mandatory policies. For example, Qian (2009) shows that fertility reductions induced by China's coercive "One-child" policy do not increase school enrollment for boys, and even reduce school enrollment for girls; Liu (2014) find that the same fertility reductions improve child health, but have no effect on child education. Our theory implies that, compared with mandatory population policies, voluntary population policies can better promote human capital development.

Although our study focuses on the child QQ tradeoff, the distinction between desired and undesired changes in rationing theory has broader implications, especially for the literature on treatment effects and policy evaluation. First, rationing theory complements the generalized Roy model, which gives rise to the choice-theoretical framework for the literature on treatment effects and program evaluation (Heckman and Vytlacil, 2007a,b). In the Roy model, individuals make choices by comparing the costs and benefits of different options. To identify the treatment effects, researchers use IVs, which shift the costs or benefits, and induce individuals to choose different treatment statuses. Many instruments are based on changes in public policies and laws that are compulsory or mandatory, which induce drastic changes in the costs and benefits of individual choices. Rationing theory is particularly suitable to the analysis of compulsory policies.

Consider a compulsory education law that subsidizes and mandates high school attendance. To illustrate, we assume two types of children who did not attend high school before the law. One type includes gifted children who did not attend because their families were too poor to support them. The other type includes less gifted children, who did not attend because the return to high school is less than the cost, and not because of the borrowing constraint. If the law mandates that both types of children attend, the increase in education for the first type can be considered to be desired, and for the second type to be undesired. Rationing theory predicts a positive rationing income effect on child outcomes for the first type and a negative effect for the second type.

Second, the differential treatment effects between desired and undesired changes also have major implications for program evaluation based on randomized controlled trials (RCTs). Participation in most RCTs is voluntary; subjects choose to comply, report, and stay in the experiment based on cost-and-benefit calculations (Heckman, 2020). The IV estimate using randomization as an instrument captures the treatment effect for compliers. Compliers in RCTs, based on revealed preference theory, experience desired changes in the treatment status, which generates a positive income effect on their outcomes, as implied by rationing theory.

We should be cautious when extrapolating IV estimates based on RCTs to contexts that involve compulsory or mandatory policies, which induce both desired and undesired changes.

The remaining parts of this study is organized as follows. Section 2 derives a generalized theory of rationed fertility. Section 3 develops an econometric method to identify QQ effects for desired and undesired fertility changes. Section 4 exploits twin births and China's birth-control policy to estimate these QQ effects, based on which Section 5 quantify the rationing income effect using a structural model. Section 6 concludes.

2 A Theory of Rationed Fertility

We extend Becker and Lewis (1973) and Rosenzweig and Wolpin (1980) to derive the generalized comparative statics of the effect of rationed fertility on child quality. In contrast to Rosenzweig and Wolpin (1980), real-world constraints require that comparative statics must be evaluated in non-optimal fertility levels. We find a novel rationing income effect, which predicts differential QQ effects for different types of fertility increases.

2.1 Becker–Lewis Setup

We consider the model setup in Becker and Lewis (1973). Parents maximize utility by choosing fertility or child quantity (n), child quality (q), and a composite consumption good (s),

$$\begin{aligned} \max_{n,q,s} \quad & U(n, q, s), \\ \text{subject to} \quad & \pi_n n + \pi_{nq} nq + p_s s \leq y, \end{aligned} \tag{P1}$$

where y is the monetary income of the family, and p_s is the price of the composite good. The price of child quantity, i.e., the cost of an additional child, is $p_n = \pi_n + \pi_{nq} q$, where π_n represents the "fixed cost" of an additional child, including the cost of giving birth and any necessities to keep the child alive; $\pi_{nq} q$ represents the costs that increase in child quality, such as tuition fees and healthcare expenditure. Similarly, the price of child quality is $p_q = \pi_{nq} n$, including tuition fees and healthcare expenditure that increase by child quantity. Solving P1 gives the optimal child quantity (n^o), child quality (q^o), and composite good (s^o).

Because n enters the price of q, a decline in n caused by external forces, would reduce p_q and raise the equilibrium demand for q, generating a negative correlation between n and q (Becker and Lewis, 1973). This insight was built into macroeconomic models to show that technological progress simultaneously reduces n and enhances q, shifting the economy from Malthusian stagnation to demographic transition and modern growth (Becker and Barro, 1988; Galor and Weil, 2000). By contrast, the microeconomic literature focuses on exploring quasi-natural or natural experiments to estimate the effects of fertility on child quality. The estimated effects seem appealing, because they may inform population control policies that target decreasing fertility. However, these estimates do not have an economic interpretation in the Becker–Lewis setup, in which fertility is a choice variable. We are unable to derive the comparative statics of n on q in the Becker–Lewis setup. Rosenzweig and Wolpin (1980) is among the first to theoretically derive the effects of fertility on child quality.

Before presenting the Rosenzweig–Wolpin setup, we define the shadow prices of child quantity (v_n) and quality (v_q) as:

$$v_n = \frac{\partial U}{\partial n} / \lambda,$$
$$v_q = \frac{\partial U}{\partial q} / \lambda,$$

where λ is the Lagrangian multiplier of P1, representing the marginal utility of 1 dollar; thus, v_n and v_q are the marginal utilities of child quantity and quality measured in dollars. For brevity, we call v_n and v_q the "returns" of child quantity and quality. In equilibrium, $n = n^o$, $q = q^o$, and $s = s^o$, and parents achieve maximal utility by equalizing the returns and prices of all choice variables: $v_n = p_n$, $v_q = p_q$, and $v_s = p_s$.

2.2 Rosenzweig–Wolpin Setup

To derive the comparative statics of child quantity on child quality, Rosenzweig and Wolpin (1980) build on the rationing theory of Tobin and Houthakker (1950) to treat child quantity as a parameter. The problem is

$$\begin{aligned} \max_{q,s} \quad & U(\bar{n}, q, s), \\ \text{subject to} \quad & \pi_n \bar{n} + \pi_{nq} \bar{n} q + p_s s \leq y. \end{aligned} \tag{P2}$$

P2 differs from P1, because child quantity ($\bar{n}$) is no longer a choice variable. P2 also differs from the standard utility maximization problem in the textbook, because the rationed child quantity, as part of the price of child quality, directly enters parental utility. Rosenzweig and Wolpin (1980) solve P2 and derive the comparative statics of $\bar{n}$ on q when rationed child quantity ($\bar{n}$) equals optimal child quantity (n^o) in P1:

$$\frac{\partial q}{\partial \bar{n}} = \pi_{nq} \frac{\partial \tilde{q^{*c}}}{\partial p_q^*} + \frac{\partial \tilde{q^{*c}}}{\partial \bar{n}}, \tag{1}$$

where $\tilde{q^{*c}}$ is a Hicksian demand for child quality;[5] and p_q^* is the price of child quality when child quantity is rationed, such that $p_q^* = \pi_{nq}\bar{n}$.[6]

2.3 Generalized Comparative Statics of the Child Quantity–quality Tradeoff

We generalize Rosenzweig and Wolpin (1980) by solving P2 and deriving the comparative statics of rationed fertility ($\bar{n}$) on child quality at any possible values of $\bar{n}$. Real-world constraints, such as population policies, infertility, or unwanted fertility, require evaluations of the QQ effects at non-optimal fertility levels. We find a rationing income effect, which has yet been identified in the literature. The rationing income effect explains heterogeneous effects of fertility on child quality for different types of fertility increases.

Solving P2, the effect of rationed fertility on child quality is

$$\underbrace{\frac{\partial q}{\partial \bar{n}}}_{QQ} = \underbrace{\pi_{nq} \frac{\partial \tilde{q^{*c}}}{\partial p_q^*}}_{QQ^P\text{: price effect}} \underbrace{+ (1 - \alpha_\Delta \cdot \epsilon_{n^* \cdot y}) \frac{\partial \tilde{q^{*c}}}{\partial \bar{n}}}_{QQ^S\text{: substitution effect}} + \underbrace{(v_n - p_n)\frac{\partial q^*}{\partial y}}_{QQ^I\text{: rationing income effect}}, \tag{2}$$

where q^* is the Walrasian demand for child quality;[7] α_Δ is the share of compensating income change out of total monetary income—that is, $\alpha_\Delta = \frac{(v_n - p_n)\bar{n}}{y}$; and $\epsilon_{n^* \cdot y}$ is the income elasticity of child quantity.[8] Note that the result in Rosenzweig and Wolpin (1980) is a special case of our Eq. (2). When $\bar{n} = n^o$ and $v_n = p_n$, Eq. (2) collapses to Eq. (1), the result in Rosenzweig and Wolpin (1980).

Our generalized Eq. (2) can be interpreted as an extended version of the standard Slutsky equation (Mas-Colell, Whinston, and Green, 1995). Eq. (2) shows that the QQ effect consists of three terms: $QQ = QQ^P + QQ^S + QQ^I$. The first term, QQ^P, is similar to the Hicksian substitution effect in the Slutsky equation, because $\bar{n}$ is part of the price of child quality in P2. A one-unit increase in $\bar{n}$ raises p_q^* by π_{nq} and reduces q via the Hicksian substitution effect. Thus, QQ^P, is negative. We call QQ^P the price effect, as $\bar{n}$ affects q through the change in the price of q. The negative price effect is a direct implication of the Becker–Lewis

[5] The Hicksian demand is derived from $\min_{q,s}\{p_n^*\bar{n} + p_q^* q + p_s^* s | U(\bar{n}, q, s) \geq u\}$, where $p_q^* = \pi_{nq}\bar{n}$, $p_n^* = \pi_n$, $p_s^* = \pi_s$.

[6] Eq. (1) is the same as Eq. (18) in Rosenzweig and Wolpin (1980, p. 231) with different notations.

[7] The Walrasian demand is derived from $\max_{n,q,s}\{U(n, q, s) | p_n^* n + p_q^* q + p_s^* s \leq y\}$.

[8] Appendix A1 presents the full derivation. Using the rationing theory of Neary and Roberts (1980) and duality theorem, we solve P2 and derive Eq. (2) by linking P1 and P2 with models with non-interactive budget constraints.

setup, where child quantity and quality interactively enter the budget constraint. The price effect appears in both Eqs. (1) and (2).

The second term, QQ^S, arises because $\bar{n}$, as part of the price of child quality, directly enters the utility function. QQ^S is not present in the Slutsky equation, because a price does not enter a standard utility function (Mas-Colell, Whinston, and Green, 1995). Since the sign of QQ^S is determined by whether child quantity and quality are net substitutes or complements, we call it the substitution effect. The first component, $1 - \alpha_\Delta \cdot \epsilon_{n^* \cdot y}$, is positive, because we expect the absolute value of α_Δ to be small (Appendix A1). The second component can be further decomposed as

$$\frac{\partial \widetilde{q^{*c}}}{\partial \bar{n}} = \frac{\partial q^{*c}}{\partial p_n^*} \left(\frac{\partial n^{*c}}{\partial p_n^*} \right)^{-1},$$

where $q^{*c}(p^*, u)$ and $n^{*c}(p^*, u)$ are the Hicksian demand functions for child quantity and quality.[9] The own-price effect, $\frac{\partial n^{*c}}{\partial p_n^*}$, is negative. The cross-price effect, $\frac{\partial q^{*c}}{\partial p_n^*}$, is positive (negative) if q and n are net substitutes (complements). Rosenzweig and Wolpin (1980) first observe the substitution effect as shown in Eq. (2), but they only consider the case of $\bar{n} = n^o$, $v_n = p_n$, and $\alpha_\Delta = 0$.

The third term, QQ^I, which consists of $v_n \frac{\partial q^*}{\partial y}$ and $-p_n \frac{\partial q^*}{\partial y}$, generalizes the income effect in the standard Slutsky equation. When $\bar{n}$ rises by one unit, a compensatory income increase of p_n is required to make the original (q, s) bundle affordable. A one-unit increase in $\bar{n}$ reduces the monetary income spent on other goods by p_n, and tends to reduce q if q is a normal good. In this sense, the component $-p_n \frac{\partial q^*}{\partial y}$ is similar to the income effect in the standard Slutsky equation. The component $v_n \frac{\partial q^*}{\partial y}$, which is absent from the standard Slutsky equation, arises because $\bar{n}$ enters parental utility.

To understand the generalized income effect, we define "social income" in the Beckerian sense as follows (Becker, 1991):

$$W = v_n n + v_q q + v_s s,$$

where v_n, v_q, and v_s are the marginal utilities of n, q, and s measured in dollars; that is, the returns of n, q, and s, respectively. In P1, parents achieve the highest social income ($W = W^o$) by equalizing the returns and costs of n, q, and s: $v_n = p_n$, $v_q = p_q$, and $v_s = p_s$. In P2, while parents can still equalize the returns and costs of q and s ($v_q = p_q$ and $v_s = p_s$), they cannot choose n, and in general cannot equalize v_n and p_n.

[9] The problem is $\min_{n,q,s}\{p_n^* n + p_q^* q + p_s^* s | U(n, q, s) \geq u\}$.

Consider a one-unit increase in $\bar{n}$ that is financed by a reduction in s.[10] The one-unit increase in $\bar{n}$ costs p_n, and thus reduces s by $\frac{p_n}{p_s}$. The change in social income is

$$\frac{\partial W}{\partial \bar{n}} = v_n - \frac{p_n}{p_s} v_s = v_n - p_n,$$

where the second equality holds because parents always equalize the return and cost of s ($v_s = p_s$). Parents achieve the highest social income (W^o) at the optimal fertility level ($\bar{n} = n^o$ and $v_n = p_n$). When fertility is rationed below the optimal fertility level ($\bar{n} < n^o$), parents prefer more children ($v_n > p_n$), and the social income is below W^o. A marginal increase in fertility moves it closer to the optimal level, raises social income ($\frac{\partial W}{\partial \bar{n}} > 0$), and induces a positive income effect. In contrast, when fertility is rationed above the optimal level ($\bar{n} > n^o$), parents prefer fewer children ($v_n < p_n$), and social income is also below W^o. A marginal increase in fertility moves fertility further away from the optimal level, reduces social income ($\frac{\partial W}{\partial \bar{n}} < 0$), and induces a negative income effect.

QQ^I, which we call a rationing income effect, appears for the first time in the literature. Rosenzweig and Wolpin (1980) only consider the case of $\bar{n} = n^o$ and $v_n = p_n$, so QQ^I does not appear in Eq. (1). The sign of the rationing income effect is determined by the type of fertility change. We define two types of fertility changes:

1. Desired fertility change: A change of fertility toward the optimal level in P1 (n^o).

2. Undesired fertility change: A change of fertility away from the optimal level in P1 (n^o).

As illustrated in Figure 1a, desired fertility changes include the fertility increase when $\bar{n} < n^o$ and the fertility reduction when $\bar{n} > n^o$; undesired fertility changes include the fertility increase when $\bar{n} \geq n^o$ and the fertility reduction when $\bar{n} \leq n^o$. Desired fertility changes increase the social income of the family, inducing a positive rationing income effect. By contrast, undesired fertility changes reduce the social income of the family, inducing a negative rationing income effect.

The theory does not predict the sign of the overall QQ effect, which is a summation of three components. The price effect QQ^P is negative. The substitution effect QQ^S depends on parental preference. The rationing income effect QQ^I is determined by the type of fertility change.

[10] Appendix A3.6 examines the impact of fertility increases on parental consumption. Based on a unique twins survey in China, we find that twining reduces clothing expenditure of both parents, and also reduces mothers' expenditure on cosmetics.

Mogstad and Wiswall (2016) suggest that most empirical studies interpret the overall QQ effect based on the negative price effect. By contrast, Mogstad and Wiswall (2016) emphasize the role of the substitution effect in interpreting heterogeneous QQ effects. They also note that the substitution effect hardly exhibits predictable variations a priori, because the substitution effect hinges on parental preference, about which we have little information. The new rationing income effect exhibits predictable variations by the type of fertility increase: The rationing income effect is positive for desired fertility increases and negative for undesired increases.

2.4 A Roadmap to Empirical Analyses

We conduct three analyses to explore the empirical content of our rationed fertility theory in the following sections. First, we devise an econometric method to separately identify QQ effects for desired and undesired fertility increases. Conceptually, we have two cases to distinguish between desired and undesired fertility increases. Figure 1a shows the first case. For a representative mother, we observe a desired fertility increase from $n^o - 1$ to n^o, and an undesired fertility increase from n^o to $n^o + 1$. In this case, desired and undesired fertility increases are observed at different birth parities for the same mother. Figure 1b shows the second case. In this case, there are two types of mothers, type-A and type-B. For all mothers, we observe fertility increases at the same birth parity, from $\bar{n}$ to $\bar{n} + 1$. For type-A mothers, where $n^o_A = \bar{n}$, we observe an undesired fertility increase; for type-B mothers, where $n^o_B = \bar{n} + 1$, we observe a desired fertility increase. In this second case, although desired and undesired fertility increases are observed at the same birth parity, type-A and type-B mothers differ in optimal fertility, $n^o_A \neq n^o_B$. Our identification strategy is based on the second case.

Second, we separately estimate QQ effects for desired and undesired fertility increases in the second case, leveraging the natural experiment of twin births and China's birth-control policy. The difference in QQ effects between desired and undesired fertility increases in this case, however, does not unambiguously inform the rationing income effect, because the two types of mothers have different levels of optimal fertility, which reflect differences in preferences and constraints, and thus differential price and substitution effects between desired and undesired fertility increases.[11]

To quantify the rationing income effect, in the third analysis, we structurally estimate our rationed fer-

[11] The difference in QQ effects between desired and undesired fertility increases in the first case also does not unambiguously inform the rationing income effect, because of the nonlinearity of the model, and price and substitution effects do not necessarily stay the same at different birth parities.

tility model that incorporates unobserved heterogeneity in both preferences and resource constraints across mothers. Based on structural estimates, simulation results show that the rationing income effect explains the difference in QQ effects between desired and undesired fertility increases.

3 Identification of QQ Effects for Desired and Undesired Fertility Increases

In this section, we devise an empirical strategy to separate desired and undesired fertility increases, and estimate QQ effects for the two types of fertility increases. Because of the rationing income effect, the effect on child quality of a desired increase in fertility should be less negative than that of an undesired increase.

3.1 Distinguishing between the Desired and Undesired Fertility Increases

Our econometric framework builds on the literature of multivalued treatments (Kline and Walters, 2016; Kirkeboen, Leuven, and Mogstad, 2016; Hull, 2018; Mountjoy, 2020). Particularly, we adapt the analysis of Hull (2018) with one IV and one stratifying variable. For purposes of illustration, we assume a pool of three types of mothers ($i \in A, B, C$), as shown in Table 1. The shares of the three types are P_A, P_B, and P_C, which are unobservable to researchers. Without twinning at the second birth ($Z_i = 0$) and the birth-control policy ($X_i = 0$), the three types of mothers can achieve their optimal fertility levels in P1, which are two, three, and three (column (1)). By Hull (2018), twinning is the IV that imposes an exogenous increase of $\bar{n}$, while the policy is the stratifying variable.

We consider the differences between twinning-induced fertility increases without and with the policy. Without the policy ($X_i = 0$), twinning at the second birth shifts the realized fertility of type-A mothers from two to three, but does not affect the realized fertility of type-B and type-C mothers (column (2)). In this case, type-A mothers are "compliers" of twinning; type-B and type-C mothers are "always-takers" of twinning.[12]

We assume that the birth-control policy targets two children per family. Because type-A mothers' optimal fertility is two, the policy does not affect the realized fertility of type-A mothers (column (3)). The policy rations fertility of type-B mothers, reducing type-B mothers' fertility from three to two. Type-C mothers realize three children, regardless of the policy.

Under the policy ($X_i = 1$), twinning shifts realized fertility from two to three for both type-A and type-

[12]In the terminology of Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996), compliers are individuals whose treatment status is affected by the instrument, and always-takers are those who are treated irrespective of whether the instrument is switched on or off.

B mothers (column (4)). In this case, both type-A and type-B mothers are compliers of twinning; type-C mothers remain always takers of twinning.

In this simple scenario, we observe two types of fertility changes. Without the policy, twinning shifts fertility of type-A mothers from the optimal two to an undesired three (the arrow between columns (1) and (2) in Table 1). Thus, the twinning-induced fertility increase for type-A mothers is undesired.[13] By contrast, because the policy rations the fertility level of type-B mothers at two, twinning helps type-B mothers break the rationing and achieve the optimal fertility of three, representing a desired fertility increase (the arrow between columns (3) and (4) in Table 1).[14]

To distinguish between the two types of fertility increases, we use the following regression:

$$D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_i \cdot X_i + \alpha_3 X_i + v_i, \tag{3}$$

where D_i is an indicator that is equal to one if mother i has three children, and zero otherwise; and v_i represents idiosyncratic fertility preference shocks.[15] The coefficient α_1 captures the proportion of mothers who increase fertility conditional on $X_i = 0$. When $X_i = 0$, twinning shifts fertility for type-A mothers only, $\alpha_1 = P_A$. The coefficient $\alpha_1 + \alpha_2$ captures the proportion of mothers who increase fertility conditional on $X_i = 1$. When $X_i = 1$, twinning shifts fertility for both type-A and type-B mothers; $\alpha_1 + \alpha_2 = P_A + P_B$ and $\alpha_2 = P_B$. As shown in Table 1, the twining-induced fertility increase from two to three for type-A mothers is undesired, and for type-B mothers is desired. Consequently, α_1 and α_2 capture undesired and desired fertility increases, respectively. The coefficient α_3 captures the proportion of mothers who decrease their fertility because of the policy—that is, the share of type-B mothers, conditional on $Z_i = 0$. So, $\alpha_3 = -P_B$. A correctly specified Eq. (3) should yield $\alpha_2 + \alpha_3 = P_B - P_B = 0$, which serves as a specification test for the empirical implementation.

[13]The notion of twinning-induced "undesired" fertility change is well noted in the literature. Black, Devereux, and Salvanes (2010) consider twinning to be an "unexpected" or "unplanned" shock to fertility. Similarly, Mogstad and Wiswall (2016, p. 173) conclude that "twin births increase the number of siblings beyond the desired family size." Since a compulsory birth-control policy is absent in the Norwegian data used by the two studies, they do not consider twinning-induced "desired" fertility increases.

[14]Two more changes are present in Table 1. Given the policy ($X_i = 1$), twinning shifts the fertility of type-A mothers from the optimal two to an undesired three, which represents an undesired increase. Given non-twinning ($Z_i = 0$), the policy rations fertility of type-B mothers and reduces fertility from three to two, which represents an undesired decrease.

[15]Section A2.6 presents the model with continuous X_i. Interpretation of the coefficients is similar under discrete or continuous X_i.

3.2 Identifying QQ Effects

We next estimate QQ effects for the two types of fertility changes by considering the following regression:

$$Y_i = \rho_0 + \rho_1 Z_i + \rho_2 Z_i \cdot X_i + \rho_3 X_i + \varepsilon_i. \tag{4}$$

where Y_i is the quality of the child of mother i, and ε_i is the idiosyncratic shocks to child quality. The coefficient ρ_1 captures the effect of twinning on middle school attendance conditional on $X_i = 0$; $\rho_1 + \rho_2$ captures the effect of twinning on middle school attendance conditional on $X_i = 1$. So, ρ_2 captures the change in the effect of twinning on middle school attendance when the birth-control policy switches on (X_i changes from 0 to 1).

To interpret the coefficients in Eq. (4), we adopt the potential-outcome framework. The indicator for three children, D_i, is the realized treatment status of mother i ($D_i = 1$ if mother i has three children, and $D_i = 0$ otherwise). Denote D_{zi} as the potential treatment status of mother i when $Z_i = z$ ($Z_i = 1$ if mother i has second-born twins, and $Z_i = 0$ otherwise). We have $D_i = D_{0i} + (D_{1i} - D_{0i}) \cdot Z_i$. As Y_i is the realized outcome, we further denote $Y_i(d, z)$ as the potential outcome when $D_i = d$ and $Z_i = z$.

We make three standard assumptions on the independence, exclusion, and monotonicity of twinning (Z_i).

Assumption 1 *Conditional on policy exposure X_i, twinning status Z_i is independent of the potential outcomes.* $\{Y_i(d, z), D_{zi}\}_{\forall z \in \{0,1\}, \forall d \in \{0,1\}} \perp Z_i\big|_{X_i}$.

Assumption 2 *Conditional on policy exposure X_i, twinning status Z_i affects Y_i only via D_i.* $Y_i(d, 0) = Y_i(d, 1) = Y_i(d) = Y_{di}\big|_{X_i}, \forall d \in \{0, 1\}$

Assumption 3 *Twinning status Z_i monotonically shifts D_i for everyone.* $D_{1i} \geq D_{oi}, \forall i$.

We discuss potential concerns when twinning may violate Assumptions 1 and 2 in Section A3.2. The monotonicity assumption automatically holds in our setting, because mothers with twins at the second birth have at least three children. We also require the "relevancy" condition, $\mathbb{E}[D_{1i} - D_{0i}|X_i] > 0$, which is automatically satisfied for the twin instrument.

We make two additional assumptions on the policy (X_i).

Assumption 4 *The policy X_i does not change the mothers' type.* $\mathbf{Pr}(i \in S|X_i = x) = \mathbf{Pr}(i \in S) = P_S, \forall x \in \{0, 1\}, \forall S = A, B, C$.

Assumption 5 *The policy X_i can be excluded from the average treatment effect of D_i on Y_i for each type of mother.* $\mathbb{E}[Y_{1i} - Y_{0i}|i \in S, X_i = x] = \mathbb{E}[Y_{1i} - Y_{0i}|i \in S], \forall x \in \{0, 1\}, \forall S = A, B, C$.

Assumption 4 holds by definition. Assumption 5, which is similar to assumption A4 in Hull (2018), states that the policy does not change the average treatment effect for each type of mothers.

Under Assumptions 1–5,[16] we have

$$\rho_1 = P_A \cdot \mathbb{E}[Y_{1i} - Y_{0i}|i \in A] = P_A \cdot \beta_A,$$
$$\rho_2 = P_B \cdot \mathbb{E}[Y_{1i} - Y_{0i}|i \in B] = P_B \cdot \beta_B,$$

where $\beta_A = \mathbb{E}[Y_{1i} - Y_{0i}|i \in A]$ is the QQ effect for type-A mothers who experience undesired fertility increases; $\beta_B = \mathbb{E}[Y_{1i} - Y_{0i}|i \in B]$ is the QQ effect for type-B mothers who experience desired fertility increases; and $\rho_3 = -P_B \cdot \beta_B + \{\mathbb{E}[Y_{0i}|X_i = 1] - \mathbb{E}[Y_{0i}|X_i = 0]\}$. Note that we allow heterogeneity in the QQ effects within each type of mothers, that is, we allow $Y_{1i} - Y_{0i} \neq Y_{1j} - Y_{0j}, \forall i, j \in S$, where $S = A, B, C$. Because $P_A = \alpha_1 > 0$ and $P_B = \alpha_2 > 0$, we have

$$\beta_A = \mathbb{E}[Y_{1i} - Y_{0i}|i \in A] = \frac{\rho_1}{\alpha_1}, \tag{5}$$
$$\beta_B = \mathbb{E}[Y_{1i} - Y_{0i}|i \in B] = \frac{\rho_2}{\alpha_2}. \tag{6}$$

QQ effects β_A and β_B share the signs of ρ_1 and ρ_2, respectively. Appendix A2 details the proof.

In an IV estimation framework, Eq. (3) is regarded as the "first-stage" regression, and Eq. (4) is regarded as the "reduced-form" regression. The "second-stage" regression can be specified as follows:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 D_i \cdot X_i + \gamma_3 X_i + \epsilon_i,$$

where D_i and $D_i \cdot X_i$ are instrumented by Z_i and $Z_i \cdot X_i$. The coefficient γ_1 captures the effect on child quality of twinning-induced fertility increases for type-A mothers without the policy ($X_i = 0$)—that is, $\gamma_1 = \beta_A$.

[16] Assumptions 1, 2, and 5 are valid only conditional on covariates discussed in the empirical analysis. For brevity, we relegate discussions on covariates to Appendix A3.1.

And $\gamma_1 + \gamma_2$ captures the weighted average effects for both type-A and type-B mothers under the policy ($X_i = 1$)—that is, $\gamma_1 + \gamma_2 = \frac{P_A}{P_A+P_B}\beta_A + \frac{P_B}{P_A+P_B}\beta_B$. We have $\gamma_2 = \frac{P_B}{P_A+P_B}(\beta_B - \beta_A)$. Appendix A2.5 contains the detailed proof. Because the interpretation of γ_2 is not straightforward, our empirical analyses focus on Eqs. (3) and (4).

For heuristic purpose, we have assumed X_i to be a dichotomous policy. Appendix A2.6 considers continuous policy intensity, and proves that the interpretation of the estimation coefficients when X_i is continuous is similar to that when X_i is dichotomous. An explicit modelling of continuous policy also enables the estimation of β_B for a given policy intensity. That is, we are able to examine the change in β_B under different levels of X_i, estimating β_{B_x} for $X_i = x$, which we will discuss in Section 4.4. Further details of the empirical implementation, such as the handling of covariates, are discussed in Sections 4.2 and A3.1.

Our observed treatment is binary, so it would be natural to use the framework for the marginal treatment effect (Heckman and Vytlacil, 1999, 2005, 2007b; Heckman, 2010). We should be able to identify a parametric function of the marginal treatment effect if the birth-control policy (X_i) satisfies the independence, exclusion, relevance, and monotonicity assumptions (Brinch, Mogstad, and Wiswall, 2017).[17] However, we do not use the framework for the marginal treatment effect. Instead, we develop the current identification framework for three reasons. First, our Assumption 5 on the birth-control policy is weaker than the exclusion assumption of an IV. We only require the birth-control policy to be a stratifying variable rather than an IV. Second, we separately identify the average treatment effect for type-A and type-B mothers (β_A and β_B) without any functional form assumptions on the marginal treatment effect. Finally, and most importantly, our theory presented in Section 2 has explicit predictions for the difference in QQ effects between type-A and type-B mothers. The primary objective of our empirical analysis is not to identify heterogeneous treatment effects per se, but to test how the theory predicts differential treatment effects between desired and undesired fertility increases.

4 Estimation of QQ Effects for Desired and Undesired Fertility Increases

In this section, we first describe the background of China's "One-child" policy. We then introduce the China population census, present the empirical specification, and describe the variables used in the estimation.

[17] If twinning serves as the only IV, we are unable to identify the function of the marginal treatment effect, because there are no never-takers with twinning.

4.1 Background

The "One-child" policy started rolling out across China in 1979, and was officially written into the Constitution in 1982 (Zhang, 2017). The policy did not restrict each family to one child in its literal sense. Due to widespread opposition and implementation problems in rural areas, the central government issued Central Document No. 7 in April 1984, which allowed rural families to have a second child. In general, urban couples employed by the government can only have one child, but most rural couples can have at least two children, especially if the first child is a girl. Minorities, or non-Han Chinese, are exempted from the policy.

Figures 2a plots completed fertility for rural Han mothers born in 1940–1960, for whom different proportions of fecund years fell under the policy. The proportion of rural mothers who had at least two children (solid line) remained stable and stayed above 80% across all cohorts. The proportion of rural mothers who had at least three children (dashed line) dropped dramatically, from 96% for the 1940 cohort to 31% for the 1960 cohort. The policy in rural China restricted births at third and higher parities, but not at the second parity. Figure 2b plots completed fertility for urban Han mothers. Different from Figure 2a, the proportions of both urban mothers who have at least two children (solid line) and at least three children (dashed line) dropped sharply across birth cohorts. The policy in urban China restricted births at the second and higher parities.

Our empirical analysis focuses on rural China. In urban China, the policy restricts the second birth. Under the policy, the variation in fertility should come from twinning at the first birth. When estimating the QQ effect using twinning at the first birth, we must compare the child quality of first-born twins with that of first-born singletons. This is less than satisfactory because of the innate differences between twins and singletons, as pointed out by Rosenzweig and Zhang (2009). In rural China, the policy restricts third birth. Therefore, the variation in fertility should come from twinning at the second birth under the policy. We compare first-born singleton children in rural families with or without second-born twins.

4.2 Data and Empirical Specification

Our primary data source is the 1% samples of the 1982 and 1990 China population censuses.[18] Both censuses were conducted by the National Bureau of Statistics. The 1982 sample covers 10,039,191 individuals from 2,428,658 households, and the 1990 sample covers 11,835,947 individuals from 3,152,818 households.

[18]We extract the data sets from the Minnesota Population Center (2014).

The census includes information at both household and individual levels. Variables at the individual level include age, gender, ethnicity, educational attainment, and marital status. Women aged 16–55 report their fertility history.

We construct the working sample using the following steps. (i) Restrict to rural Han mothers born in 1940–1960 with at least two children. When the policy launched in 1979, the youngest 1960 cohort was aged 19, and their fecund ages are fully covered by the policy; the eldest 1940 cohort was aged 39, and their fecund ages are not much covered by the policy. (ii) Restrict to mothers whose oldest child is no older than 17 years, and whose children all reside in her household. As the census does not enumerate children who have left home in the household, this restriction minimizes sample selection, as discussed by Li, Zhang, and Zhu (2008) and Huang, Lei, and Zhao (2016). (iii) Drop mothers in Tibet, because the population-control policy in Tibet differs from that in the rest of China. (iv) Drop mothers with first-born twins. (v) Restrict to mothers whose first child is aged 13–17. When the first child is at least 13, most mothers should have their third birth if they desire. Our interested measure of child quality is middle school attendance, as described below. The normal age for middle school attendance is from 13 to 15. With these (i)–(v) restrictions, we obtain a sample of 264,013 mothers that will be used in the analysis below.

We estimate the fertility equation, Eq. (3) in Section 3.1, as follows,

$$D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_i \cdot X_i + \alpha_3 X_i + \mathbf{C_i^D} \tilde{\alpha}_4 + \epsilon_i; \tag{7}$$

and estimate the child-quality equation, Eq. (4),

$$Y_i = \rho_0 + \rho_1 Z_i + \rho_2 Z_i \cdot X_i + \rho_3 X_i + \mathbf{C_i^Y} \tilde{\rho}_4 + \varepsilon_i, \tag{8}$$

where Y_i is an indicator variable on whether the first-born child has ever attended middle school.

There are three differences between Eqs. (3)–(4) and Eqs. (7)–(8). First, following Brinch, Mogstad, and Wiswall (2017), D_i is an indicator variable on whether mother i has three or more children. Second, we use fines for unauthorized births to measure the intensity of the coercive "One-child" policy, as discussed in Section 4.3. The policy intensity is a continuous variable; by contrast, X_i is a dichotomous variable in Eqs. (3)–(4). In Appendix A2.6, we prove that the interpretation of the estimation coefficients when X_i is continuous is similar to that when X_i is dichotomous. Third, the vector $\mathbf{C_i^D}$ or $\mathbf{C_i^Y}$ includes an estimated index based on the set of control variables $\mathbf{C_i}$, and its interactions with Z_i, X_i, and $Z_i \cdot X_i$, the inclusion of which is to ensure that Assumptions 1, 2, and 5 in Section 3.2 are plausible. We discuss details of the covariates in Appendix A3.1. Bootstrapped standard errors are clustered by province and maternal education.

The identification and interpretation of the coefficients $(\alpha_1, \alpha_2, \rho_1, \rho_2)$ are carefully developed in Section 3. Specifically, α_1 captures the proportion of type-A mothers who experienced undesired fertility increases; and α_2 captures the proportion of type-B mothers who experienced desired fertility increases. Coefficient ρ_1 captures the effect of twinning on middle school attendance conditional on $X_i = 0$; and ρ_2 captures the change in the effect of twinning on middle school attendance when the birth-control policy switches on (X_i changes from 0 to 1). Eqs. (5) and (6) imply $\frac{\rho_1}{\alpha_1} = \beta_A$ and $\frac{\rho_2}{\alpha_2} = \beta_B$. Estimate of β_A identifies QQ_U, the QQ effect for undesired fertility increases; and β_B identifies QQ_D, the QQ effect for desired fertility increases. By our generalized theory of rationed fertility, if the rationing income effect drives the QQ effect, we expect $\beta_B - \beta_A > 0$.

4.3 Variable Construction and Summary Statistics

Child Quality (Y_i)

Child quality is measured by the first-born child's middle school attendance. Figure 3a shows that the middle school attendance rate increases from 60% for the 1965 cohort to 80% for the 1980 cohort (dashed line). By contrast, the primary school attendance rate (solid line) remains stable at a high level of 96% across all cohorts, and the high school attendance rate (dot-dashed line) remains below 10% across all cohorts. Thus, we choose children's middle school attendance to measure child quality.

Figure 3b shows the proportion of children attending primary, middle, and high schools for children aged 6–17. Approximately 95% of children over age 10 have attended primary school (solid line). Children start to attend middle school at age 13, and the attendance rate increases to over 65% by age 17 (dashed line). Only a few children are attending high school by age 17 (dot-dashed line). Table 2 shows that 51% of children in our sample have ever attended middle school.

Child Quantity (D_i)

Table 2 shows that 68% of mothers without twins have at least three children. All mothers with second-born twins have at least three children.

Twin (Z_i)

Our sample includes 262,956 mothers without twins and 1,057 mothers with second-born twins. The twinning rate is 0.402%, which is similar to that reported in previous studies using Chinese censuses (Li, Zhang, and Zhu, 2008; Huang, Lei, and Zhao, 2016).

Policy (X_i)

The "One-child" policy is an umbrella term for a package of birth-control programs, including forced sterilization, mandated abortion, fines on unauthorized births, job dismissal in the case of policy violation, bonuses for policy compliance, etc. We use fines for unauthorized births to measure the *overall enforcement strength* of the policy. Ebenstein (2010) first compiled fines as multiples of local household annual income to inform the strictness of birth-control policy across provinces and over years. Our interpretation of the fines data follows Ebenstein (2010). Although fines were not the main way of violation punishment especially from 1979 to 2000, fines highly correlate with the various coercive birth-control measures. As the variable of fines has been widely used in the literature (Ebenstein, 2010; Huang, Lei, and Sun, 2021; García, 2024), the use of fines also facilitates comparisons with existing studies. We plot the over-time variation of fines by province in Appendix Figure A1. Huang (2017) discusses the fines data in detail.[19]

We construct a measure of policy intensity for the third birth as the empirical counterpart of X_i. This measure is the weighted average fines for 10 years after a mother's second birth. We denote $Prob(s)$ as the probability of a third birth in year s after the second birth ($s = 1, 2, \ldots, 10$). Figure A3 depicts $Prob(s)$ based on the empirical distribution of birth spacing between the second and third births of mothers born in 1930–1939 in the 1982 census. As spacing between births can be affected by female wages, fines, and other factors (Heckman and Walker, 1990a,b), we use mothers born in 1930–1939 who gave birth before the "One-child" policy. For mothers with three or more children, 95% of the mothers' third birth was within 6 years after the birth of the second child. For mother i whose second birth is in year t and province p, the policy intensity is defined as

$$X_i = \sum_{s=1}^{10} Prob(s) \cdot fines_{t+s,p}, \tag{9}$$

where $fines_{t+s,p}$ represents the fines on unauthorized births as multiples of average household annual income in province p and year $t + s$.

[19]The "One-child" policy is not the only birth-control policy in China. During the early1970s, China implemented the "Later, Longer, and Fewer" (LLF) campaign, which encouraged couples to marry and give birth at older ages, to have longer spacing between births, and to have fewer children (Wang, 2016; Chen and Huang, 2020; Chen and Fang, 2021). We focus on the "One-child" policy, which better reflect coercive fertility control. Our results remain robust after controlling for the intensity of LLF campaign. See Appendix Section A3.4 for detailed discussions.

4.4 Estimation Results

Baseline Estimates

Column (1) of Table 3 reports our baseline estimates. Column (1) in Panel A reports estimates of the fertility equation (Eq. (7)). The estimated coefficient on twinning (α_1) is positive and statistically significant. Without the policy ($X_i = 0$), twinning increases the proportion of mothers with at least three children by 21 percentage points. The estimate of α_1 represents the share of mothers who have experienced undesired fertility increases by twinning.

The estimated coefficient on the policy (α_3) is negative and statistically significant. Without twinning ($Z_i = 0$), the proportion of mothers with at least three children decreases by 12 percentage points when X_i increase by one. X_i are measured by multiples of local household annual income, and its standard deviation is 0.48. The estimate of α_3 represents the share of mothers who have experienced undesired fertility reductions under the policy.

The estimated coefficient on the interaction term (α_2) is positive and statistically significant. Compared with $X_i = 0$, when $X_i = 1$ twinning additionally increases the proportion of mothers with at least three children by 12 percentage points. The estimate of α_2 represents the share of mothers who have experienced desired fertility increases. Their fertility should have been reduced by the policy, but twining helps them break the policy and keep their fertility at a more desired level. As predicted in Section 3, $\hat{\alpha}_2 + \hat{\alpha}_3 = 0$, suggesting a correct empirical specification.

Column (1) in Panel B of Table 3 reports estimates of the child-quality equation (Eq. (8)). The estimated coefficient before twinning (ρ_1) is negative and statistically significant, which implies a negative QQ effect for undesired fertility increases. The estimated coefficient on the interaction term (ρ_2) is positive and statistically significant ($p = 0.051$), which implies a positive QQ effect for desired fertility increases. Based on Eqs. (5) and (6), the calculated QQ effect for undesired fertility increases (β_A) is -0.157 (= -0.033/0.212), and that for desired fertility increases (β_B) is 0.412 (= 0.050/0.121). Consistent with our theory of rationed fertility, the difference ($\beta_B - \beta_A$) is positive and statistically significant.

Our estimates suggest that an undesired fertility increase lowers the probability of attending middle school by 15.7 percentage points, while a desired fertility increase raises the probability by 41.2 percentage points. One reason for the large effect size lies in the measure of child quality. We are examining whether a child between ages 13–17 have attended middle school or not, an indicator on a sensitive margin of educational investment. To put the effect size in perspective, we convert the effect size into the commonly used scale of completed schooling years. Specifically, we calculate the correlation between the probability of middle school attendance in 13–17 and completed schooling years in adulthood, exploiting repeated observations of the same cohort in different ages.[20] A ten-percentage-point increase in the probability of middle school attendance in 13–17 is associated with 0.369 more year of completed schooling in adulthood. Under this conversion scale, an undesired fertility increase results in 0.579 (= 1.57×0.369) less years of completed schooling, while a desired fertility increase causes 1.52 (= 4.12×0.369) more years of completed schooling.

Robustness

We present a series of robustness analyses in Appendix Sections A3.2–A3.4. Appendix A3.2 discusses the independence and exclusion assumptions on twin instrument (Assumptions 1 and 2). Appendix A3.3 addresses concerns on the policy intensity as a valid stratifying variable (Assumption 5). Appendix A3.4 analyzes gender difference in the QQ effects, controls on an alternative birth-control policy, and QQ effects at higher birth parities. Our baseline estimates remain robust.

Because the policy intensity is a continuous stratifying variable, we examine heterogeneity in β_B by policy intensity. As the policy intensity increases, the measure of the policy intensity better reflects the coercive nature of the birth control, which twinning breaks for type-B mothers. The more stringent is the coercive policy, the more positive is the twinning-generated rationing income effect. Therefore, we expect a more positive β_B in case of a higher policy intensity. Appendix A2.6 derives the identification result under continuous policy intensity, and Appendix A3.5 presents the estimation of heterogeneous treatment effects by policy intensity. Consistent with our theory of rationed fertility, we find that the estimated β_B generally increases with the policy intensity.

Placebo Tests

Our identification strategy is based on the multivalued-treatment framework, in which twinning serves as an IV, and the policy a stratifying variable. To differentiate the QQ effects between desired and undesired

[20] We use a two-way fixed-effect regression, $\ln(S_{pc}) = \delta + \gamma Y_{pc} + \epsilon_{pc}$, where Y_{pc} is the middle school attendance rate in 13–17 of cohort c in province p, and S_{pc} is the average completed schooling years of cohort c in province p. We calculate Y_{pc} when our sample cohorts are 13–17 years old, using the same data source as in our main analysis (1982 and 1990 China population censuses). We obtain S_{pc} from the 2000 China population census, when our sample cohorts were at least 23 years old. The estimate of γ is 0.424, suggesting that a ten-percentage-point increase in the probability of middle school attendance is associated with a 4.24% increase in completed schooling years. Because the mean of completed schooling years is 8.7 years, the 4.24% increase represents 0.369 (= 0.0424×8.7) year of completed schooling.

fertility increases, it is crucial that the stratifying variable reflects the coercive nature of the policy. We conduct placebo tests, using the indicator of first-born son or maternal schooling years as the stratifying variable. These two variables shift fertility, reflecting differences in parental preferences and constraints. The average fertility level is lower for mothers with first-born sons or more schooling years. Different from the policy, these two variables do not result in rationing income effects by affecting fertility through coercion. If the significant difference in our QQ estimates between desired and undesired fertility increases is driven by variations in price and substitution effects, which reflect preferences and constraints, we would expect to observe similar significant difference in QQ estimates when using either of these variables as a stratifying variable.

Columns (2) and (3) of Table 3 report the result of our placebo tests. We use first-born son and maternal schooling years as the stratifying variables in columns (2) and (3), respectively. Panels A and B, respectively, report the regression results for the fertility equation (Eq. (7)) and the child-quality equation (Eq. (8)). Contrary to the expectation, the result show that the estimates of β_A and β_B are not statistically significant. This suggests that our estimated differential QQ effects due to the coercive policy are unlikely to be driven by these observable dimensions of preferences and constraints.

5 Quantifying the Rationing Income Effect

Our results from reduced-form estimation and placebo tests provide suggestive evidence of the rationing income effect. However, the reduced-form estimate captures a combination of price, substitution, and rationing income effects (Eq. (2)). To address this, we structurally estimate our rationing fertility model in this section, which enables us to separately quantify the three components of the QQ effects for both desired and undesired fertility increases. We find that the rationing income effect is the driving force behind the difference in the QQ estimates between these two types of fertility increases.

5.1 A Parametric Model

Our structural model takes into account both heterogeneity in household resources and preferences regarding fertility choices. This allows us to quantitatively decompose the overall QQ effect into price, substitution, and rationing income effects while considering the unobservable differences in preferences and resources among different types of mothers. To estimate our structural model, we apply an indirect inference approach that matches the model-predicted comparative statics and moments with the corresponding empirical estimates.

Specifically, we begin by solving a parametric version of the unrestricted household problem P1, which is augmented with unobserved heterogeneity in both income and preference. The utility maximization problem for household i is:

$$\begin{aligned} \max_{n_i,q_i,s_i} \quad & U_i(n_i,q_i,s_i) = U_1^{\theta} s_i^{1-\theta}, \\ \text{subject to} \quad & (\pi_n + \tau y_i)n_i + \pi_{nq} n_i q_i + p_s s_i \leq y_i, \\ & U_1 = (\alpha n_i^{\rho} + (1-\alpha) h_i^{\rho}))^{\frac{1}{\rho}}, \\ & h_i = q_i^{\gamma}, \\ & \theta = \theta_0 + \epsilon_i. \end{aligned} \tag{P1'}$$

where y_i captures heterogeneity in household resources (income), and ϵ_i captures heterogeneity in household preferences for children.[21] We assume y_i and ϵ_i follows a joint normal distribution, with a correlation of $\sigma_{y\epsilon}$, i.e.,

$$\begin{pmatrix} y_i \\ \epsilon_i \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_y \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{y\epsilon} \\ \sigma_{y\epsilon} & \sigma_\epsilon^2 \end{pmatrix} \right].$$

Solving P1' gives optimal choices n_i^o, q_i^o, and s_i^o. Consistent with the Sections 3 and 4, we define type-A mothers as those whose optimal number of children $n_i^o = 2$; and type-B mothers as those whose optimal number of children is $n_i^o = 3$. We define type-C mothers as those whose preference for the number of children ϵ_i is greater than a certain threshold $\bar{\epsilon}$. That is, type-C mothers are those with extreme preferences for the number of children, and their fertility decisions will not be affected by rationing policies or twinning. Thus, type-B mothers have $\epsilon_i \leq \bar{\epsilon}$.

After solving P1' and classifying type of mothers for each household i in the model, we next turn to the

[21] The utility function and the human capital production function forms follow Mogstad and Wiswall (2016). We follow De La Croix and Doepke (2003) and references therein to include labor market costs τy_i into the price of children, reflecting the time input of raising children.

maximization problems when fertility is rationed, i.e.,

$$\begin{aligned}
\max_{q_i, s_i} \quad & U_i(\bar{n}, q_i, s_i) = U_1^{\theta} s_i^{1-\theta}, \\
\text{subject to} \quad & (\pi_n + \tau y_i)\bar{n} + \pi_{nq}\bar{n}q_i + p_s s_i \leq y_i, \\
& U_1 = (\alpha \bar{n}^{\rho} + (1-\alpha)h_i^{\rho}))^{\frac{1}{\rho}}, \\
& h_i = q_i^{\gamma}, \\
& \theta = \theta_0 + \epsilon_i.
\end{aligned} \tag{P2'}$$

We solve P2' for type-A and type-B mothers when $\bar{n} = 2$ and $\bar{n} = 3$, respectively. We can calculate the QQ effects for type-A and type-B mothers,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\left(q^A_{i|\bar{n}=3} - q^A_{i|\bar{n}=2}\right)}{\left(q^A_{i|\bar{n}=3} + q^A_{i|\bar{n}=2}\right)/2} = \beta_A,$$
$$\frac{1}{n}\sum_{i=1}^{n} \frac{\left(q^B_{i|\bar{n}=3} - q^B_{i|\bar{n}=2}\right)}{\left(q^B_{i|\bar{n}=3} + q^B_{i|\bar{n}=2}\right)/2} = \beta_B,$$

where $q^j_{i|\bar{n}=3} - q^j_{i|\bar{n}=2}$ ($j \in [A, B]$) represents the child quality difference when $\bar{n} = 2$ and when $\bar{n} = 3$ for household i, and $\left(q^j_{i|\bar{n}=3} + q^j_{i|\bar{n}=2}\right)/2$ ($j \in [A, B]$) is the average child quality when $\bar{n} = 2$ and when $\bar{n} = 3$. These are the theoretical and simulated QQ effects for type-A and type-B mothers derived from our model when introducing fertility rationing. Correspondingly in our empirical analysis, we estimate the QQ effects, denoted as β_A and β_B, for type-A and type-B mothers, respectively. These estimates are derived when the number of children is shifted from two to three (from $\bar{n} = 2$ to $\bar{n} = 3$) due to twinning at the second birth. We thus map the model-implied QQ effects into their empirical estimates by different types of mothers.

5.2 Structural Estimation

We normalize $p_s = 1$ and calibrate $\tau = 0.075$ from existing literature (De La Croix and Doepke, 2003). We calibrate $\gamma = 0.424$, the parameter of human capital production, based on a two-way fixed-effect regression of completed schooling years on middle-school attendance, as discussed in Section 4.4. The parameters regarding logarithm income process (μ_y, σ_y) are estimated to be (9.713, 0.497) using China Health and Nutrition Survey (CHNS, 1991 and 1993), and the set of parameters $\Theta = (\alpha, \rho, \theta_0, \sigma_\epsilon, \sigma_{y\epsilon}, \pi_n, \pi_{nq}, \bar{\epsilon})$ are estimated within the model.

We estimate the set of parameters Θ using indirect inference, by matching the model-predicted effects to the empirical estimates. Given a specific set of parameters Θ, our model is capable of generating the optimal choices for each household on number and quality of children. Here we provide a heuristic argument on how we identify each of the parameter in the model. First, we match the model predicted QQ effects for type-A and type-B mothers to the reduced-form estimates β_A and β_B to pin down the parameters of α and ρ. Second, to estimate the price parameters (π_n, π_{nq}), we match the model simulated expenditure on children by the corresponding category to the empirical estimates of household expenditure on children from the Chinese Household Income Project (CHIP, 1988 and 1995). Third, the correlation between household income and a dummy variable of whether the household has three or more children ($corr(y_i, I_{n_i \geq 3})$), the correlation between household income and quality of children (measured by middle school attendance of the first-born child) ($corr(y_i, q_i)$), and the share of type-A and type-B mothers (α_A, α_B) jointly help pin down $(\theta_0, \sigma_\epsilon, \sigma_{y\epsilon}, \bar{\epsilon})$, which are the parameters governing the preferences for number and quality of children. We have 8 parameters and 8 moments, rendering the model just identified.[22]

We minimize the distance between the model simulated statistics with the empirical estimates, i.e.,

$$f(\Theta) = [d - s(\Theta)]W[d - s(\Theta)]',$$

where d is the vector of empirical estimates and $s(\Theta)$ is the vector of simulated statistics. W is a weighting matrix with $\frac{1}{d^2}$ on the diagonal and zero elsewhere. The estimated result of the parameter is presented in Table 4. For example, the CES substitution parameter ρ is estimated to be -7.92, which is similar to the value -9 used in Mogstad and Wiswall (2016). Our estimate of ρ indicates that the elasticity of substitution between child quantity and quality is 0.11, suggesting strong complementarity between the two. The correlation coefficient between household income and preference is estimated to be -0.01, suggesting a negative relationship between household income and preference for children.

5.3 Decomposition

With the estimated parameters, we use Eq. (2) to quantitatively decompose the overall QQ effects into the price, substitution, and rationing income effect for type-A and type-B mothers, respectively. The de-

[22] See more details on the calculation and value of the data moments in Appendix A3.7.

composition analysis based on estimated parameters enables us to gauge the quantitative importance of the rationing income effect. The result is presented in Table 5, where the rationed fertility level increases from two to three for both type-A and type-B mothers. For type-A mothers, optimal fertility is two ($n^o_A = 2$), the fertility increase is undesired. For type-B mothers, optimal fertility is three ($n^o_B = 3$), the fertility increase is desired. The price effects are similar for type-A and type-B mothers, and the substitution effects are of the same order of magnitude for type-A and type-B mothers. It is the rationing income effect that significantly drives the different observed QQ effects for type-A and type-B mothers. In particular, our decomposition indicates that price and substitution effects account for 3.5% and 17.4% of the difference in QQ effects for type-A and type-B mothers, whereas the rationing income effect accounts for 79.1% of the difference.

We re-estimate the structural model based on different pre-set values of τ and γ to assess the sensitivity of our decomposition results. Appendix Table A12 explores values of τ ranging from 0.05 to 0.10 and values of γ ranging from 0.30 to 0.55, centered on the baseline calibration parameter values of $\tau = 0.075$ and $\gamma = 0.424$. Across these sensitivity checks, the rationing income effect consistently accounts for 76.6%–91.7% of the difference in QQ effects for type-A and type-B mothers.

The above decomposition quantifies the rationing income effect at a certain birth parity across mother types, the case that we explore in the reduced-form analysis. The estimated structural model also enables us to quantify the rationing income effect at different birth parities for the same mother. In Figure 4, we plot the decomposition of the overall QQ effects for a representative type-B mother (with mean income and preference for fertility) when evaluated at different levels of rationed fertility. We see that rationing income effect plays a dominant role in driving the overall QQ effects when rationed fertility level varies from one to five. Consistent with our theory of rationed fertility, the magnitude of rationing income effect increases when fertility shifts away from the unrestricted optimum of three children. Altogether, both decomposition results, either across mother types or across birth parities, consistently highlight the role of the rationing income effect in accounting for the observed QQ effect.

6 Conclusion

We develop a generalized theory of rationed fertility to analyze treatment effect heterogeneity in the child quantity–quality tradeoff. Although fertility is a choice, it is rationed by external factors. An exogenous increase in fertility can either be desired—that is, a move toward optimal fertility; or undesired—that is, a move away from optimal fertility. While a desired fertility increase generates a positive rationing income effect on child quality, an undesired fertility increase has a negative rationing income effect.

Our empirical analysis infers the rationing income effect from differential treatment effects between desired and undesired fertility increases. We explore the natural experiment of twin births and China's "One-child" policy. The estimated QQ effect for undesired fertility increases is negative; by contrast, the estimated QQ effect for desired fertility increases is positive. We quantify the rationing income effect in accounting for the overall QQ effects using a structural model with unobserved heterogeneity. We find that the rationing income effect explains a large proportion of the difference in QQ effects between desired and undesired fertility increases.

Data Availability

Code replicating the tables and figures in this article can be found in Guo et al. (2025) in the Harvard Dataverse, DATASET URL.

References

Ananat, Elizabeth Oltmans, Jonathan Gruber, Phillip B. Levine, and Douglas Staiger. 2009. "Abortion and selection." *Review of Economics and Statistics* 91 (1):124–136.

Ananat, Elizabeth Oltmans and Daniel M Hungerman. 2012. "The Power of the Pill for the Next Generation: Oral Contraception's Effects on Fertility, Abortion, and Maternal & Child Characteristics." *The Review of Economics and Statistics* 94 (1):37–51.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434):444.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics* 28 (4):773–824.

Bagger, Jesper, Javier A Birchenall, Hani Mansour, and Sergio Urzúa. 2021. "Education, Birth Order and Family Size." *The Economic Journal* 131 (633):33–69.

Bailey, Martha J., Olga Malkova, and Zoe M. McLaren. 2019. "Does access to family planning increase children's opportunities?" *Journal of Human Resources* 54 (4):825–856.

Becker, Gary S. 1960. "An Economic Analysis of Fertility." In *Demographic and Economic Change in Developed Countries*. Columbia University Press, 209–240.

———. 1991. *A Treatise on the Family*. Cambridge, Massachusetts: Harvard University Press, enlarged e ed.

Becker, Gary S. and Robert J. Barro. 1988. “A Reformulation of the Economic Theory of Fertility.” *Quarterly Journal of Economics* 103 (1):1–25.

Becker, Gary S. and H. Gregg Lewis. 1973. “On the Interaction between the Quantity and Quality of Children.” *Journal of Political Economy* 81 (2):S279–S288.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. “The More the Merrier? The Effect of Family Size and Birth Order on Children’s Education.” *Quarterly Journal of Economics* 120 (2):669–700.

———. 2010. “Small Family, Smart Family? Family Size and the IQ Scores of Young Men.” *Journal of Human Resources* 45 (1):33–58.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. 2017. “Beyond LATE with a Discrete Instrument.” *Journal of Political Economy* 125 (4):985–1039.

Cavalcanti, Tiago, Georgi Kocharkov, and Cezar Santos. 2021. “Family Planning and Development: Aggregate Effects of Contraceptive Use.” *The Economic Journal* 131 (634):624–657.

Chen, Yi and Hanming Fang. 2021. “The long-term consequences of China’s “Later, Longer, Fewer” campaign in old age.” *Journal of Development Economics* 151 (March):102664.

Chen, Yi and Yingfei Huang. 2020. “The power of the government: China’s Family Planning Leading Group and the fertility decline of the 1970s.” *Demographic Research* 42:985–1038.

Dang, Hai Anh H. and F. Halsey Rogers. 2016. “The decision to invest in child quality over quantity: Household size and household investment in education in vietnam.” *World Bank Economic Review* 30 (1):104–142.

De La Croix, David and Matthias Doepke. 2003. “Inequality and growth: why differential fertility matters.” *American Economic Review* 93 (4):1091–1113.

Ebenstein, Avraham. 2010. “The ‘Missing Girls’ of China and the Unintended Consequences of the One Child Policy.” *Journal of Human Resources* 45 (1):87–115.

Galor, Oded and David N. Weil. 2000. “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond.” *American Economic Review* 90 (4):806–828.

García, Jorge Luis. 2024. “Pricing Children, Curbing Daughters: Fertility and the Sex-Ratio During China’s One-Child Policy.” *Journal of Human Resources* 59 (5):1319–1352.

Gruber, Jonathan, Phillip Levine, and Douglas Staiger. 1999. “Abortion legalization and child living circumstances: Who is the “marginal child”?” *Quarterly Journal of Economics* 114 (1):263–291.

Guo, Rufei, Junjian Yi, and Junsen Zhang. 2022. “The Child Quantity-quality Trade-off.” In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmermann. Springer, Cham, 1–23.

Guo, Rufei, Junjian Yi, Junsen Zhang, and Ning Zhang. 2025. “Replication Data for: Rationed Fertility: Treatment Effect Heterogeneity in the Child Quantity-Quality Tradeoff.” *Harvard Dataverse* URL `https:`.

Hanushek, Eric A. 1992. “The Trade-off between Child Quantity and Quality.” *Journal of Political Economy* 100 (1):84.

Heckman, James J. 2010. “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy.” *Journal of Economic Literature* 48 (2):356–98.

Heckman, James J. 2020. “Randomization and Social Policy Evaluation Revisited.” *NBER Working Paper* .

Heckman, James J., N. Hohmann, J. Smith, and M. Khoo. 2000. “Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment.” *The Quarterly Journal of Economics* 115 (2):651–694.

Heckman, James J. and Rodrigo Pinto. 2018. “Unordered Monotonicity.” *Econometrica* 86 (1):1–35.

Heckman, James J. and Sergio Urzúa. 2010. “Comparing IV with Structural Models: What Simple IV Can and Cannot Identify.” *Journal of Econometrics* 156 (1):27–37.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2008. “Instrumental Variables in Models with Multiple Outcomes.” *Annales d’Économie et de Statistique* 91/92:151–174.

Heckman, James J. and Edward Vytlacil. 2005. “Structural Equations, Treatment Effects, and Econometric Policy Evaluation.” *Econometrica* 73 (3):669–738.

Heckman, James J. and Edward J. Vytlacil. 1999. “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (8):4730–4734.

———. 2007a. “Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation.” In *Handbook of Econometrics*, vol. 6, chap. 70. Elsevier, 4779–4874.

———. 2007b. “Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments.” In *Handbook of Econometrics*, vol. 6, chap. 71. Elsevier, 4875–5143.

Heckman, James J. and James R. Walker. 1990a. “The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data.” *Econometrica* 58 (6):1411.

———. 1990b. “The Third Birth in Sweden.” *Journal of Population Economics* 3 (4):1812.

Huang, Wei. 2017. “How Does the One Child Policy Impact Social and Economic Outcomes?” *IZA World of Labor* .

Huang, Wei, Xiaoyan Lei, and Ang Sun. 2021. “Fertility Restrictions and Life Cycle Outcomes: Evidence from the One-Child Policy in China.” *The Review of Economics and Statistics* 103 (4):694–710.

Huang, Wei, Xiaoyan Lei, and Yaohui Zhao. 2016. “One-Child Policy and the Rise of Man-Made Twins.” *Review of Economics and Statistics* 98 (3):467–476.

Hull, Peter. 2018. “IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons.” *Woring Paper* .

Imbens, Guido W and Joshua D Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62 (2):467.

Joshi, Shareen and T. Paul Schultz. 2013. “Family Planning and Women’s and Children’s Health: Long-Term Consequences of an Outreach Program in Matlab, Bangladesh.” *Demography* 50 (1):149–180.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. “Field of Study, Earnings, and Self-Selection.” *The Quarterly Journal of Economics* 131 (3):1057–1111.

Kline, Patrick and Christopher R Walters. 2016. “Evaluating Public Programs with Close Substitutes: The Case of Head Start.” *The Quarterly Journal of Economics* 131 (4):1795–1848.

Lee, Sokbae and Bernard Salanié. 2018. “Identifying Effects of Multivalued Treatments.” *Econometrica* 86 (6):1939–1963.

Li, Hongbin, Junsen Zhang, and Yi Zhu. 2008. “The Quantity-Quality Trade-Off of Children in a Developing Country: Identification Using Chinese Twins.” *Demography* 45 (1):223–243.

Lin, Wanchuan, Juan Pantano, Rodrigo Pinto, and Shuqiao Sun. 2019. “Identification of Quantity-Quality Trade-Off with Imperfect Fertility Control.” *Working Paper* .

Lin, Wanchuan, Juan Pantano, and Shuqiao Sun. 2020. “Birth order and unwanted fertility.” *Journal of Population Economics* 33 (2):413–440.

Liu, Haoming. 2014. “The Quality-Quantity Trade-off: Evidence from the Relaxation of China’s One-child Policy.” *Journal of Population Economics* 27 (2):565–602.

Malthus, Thomas Robert. 1798. *An Essay on the Principle of Population*. London: J. Johnson.

Mas-Colell, Andreu, Michael Dennis Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford University Press.

Minnesota Population Center. 2014. *Integrated Public Use Microdata Series, International: Version 6.3 [Machine-readable database]*. Minneapolis: University of Minnesota.

Mogstad, Magne and Matthew Wiswall. 2016. “Testing the Quantity-Quality Model of Fertility: Estimation using Unrestricted Family Size Models.” *Quantitative Economics* 7:157–192.

Mountjoy, Jack. 2020. “Community Colleges and Upward Mobility.” *American Economic Review* :1–83.

National Health and Family Planning Committee. 2013. *China Health and Family Planning Yearbook*. Beijing: China Union Medical College Press.

Neary, J. P. and K. W. S. Roberts. 1980. “The Theory of Household Behaviour under Rationing.” *European Economic Review* 13 (1):25–42.

Panandiker, V. A.P. and P. K. Umashankar. 1994. “Fertility control and politics in India.” *Population & Development Review* 20:89–104.

Pantano, Juan. 2016. “Unwanted Fertility, Contraceptive Technology and Crime: Exploiting a Natural Experiment in Access to The Pill.” *Working Paper* .

Petraglia, Felice, Gamal I. Serour, and Charles Chapron. 2013. “The changing prevalence of infertility.” *International Journal of Gynecology and Obstetrics* 123 (SUPPL. 2):4–8.

Qian, Nancy. 2009. “Quantity-Quality and the One Child Policy:The Only-Child Disadvantage in School Enrollment in Rural China.” *NBER Working Paper* .

Rosenzweig, Mark R. and T. Paul Schultz. 1987. “Fertility and Investments in Human Capital: Estimates of the Consequence of Imperfect Fertility Control in Malaysia.” *Journal of Econometrics* 36 (1-2):163–184.

Rosenzweig, Mark R. and Kenneth I. Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica* 48 (1):227.

Rosenzweig, Mark R. and Junsen Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-child" Policy." *Review of Economic Studies* 76 (3):1149–1174.

Schultz, T. Paul. 2007. "Population Policies, Fertility, Women's Human Capital, and Child Quality." In *Handbook of Development Economics*, vol. 4, edited by T. Paul Schult and John Strauss, chap. 52. North Holland, 3249–3303.

Seshadri, Ananth and Anson Zhou. 2022. "Intergenerational mobility begins before birth." *Journal of Monetary Economics* 129:1–20.

Sun, Shuqiao. 2019. "Less is More: How Family Size in Childhood Affects Long-Run Human Capital and Economic Opportunity." *Working Paper* .

Tobin, James and H. S. Houthakker. 1950. "The Effects of Rationing on Demand Elasticities." *Review of Economic Studies* 18 (3):140–153.

Wang, Fei. 2016. "Using New Measures to Reassess the Fertility Effects of China's Family Planning Policies Fei Wang." *Working Paper* :1–40.

WHO. 2014. "Eliminating forced, coercive and otherwise involuntary sterilization." Tech. rep., World Health Organization.

Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and Its Effects on Family Outcomes." *Journal of Economic Perspectives* 31 (1):141–160.

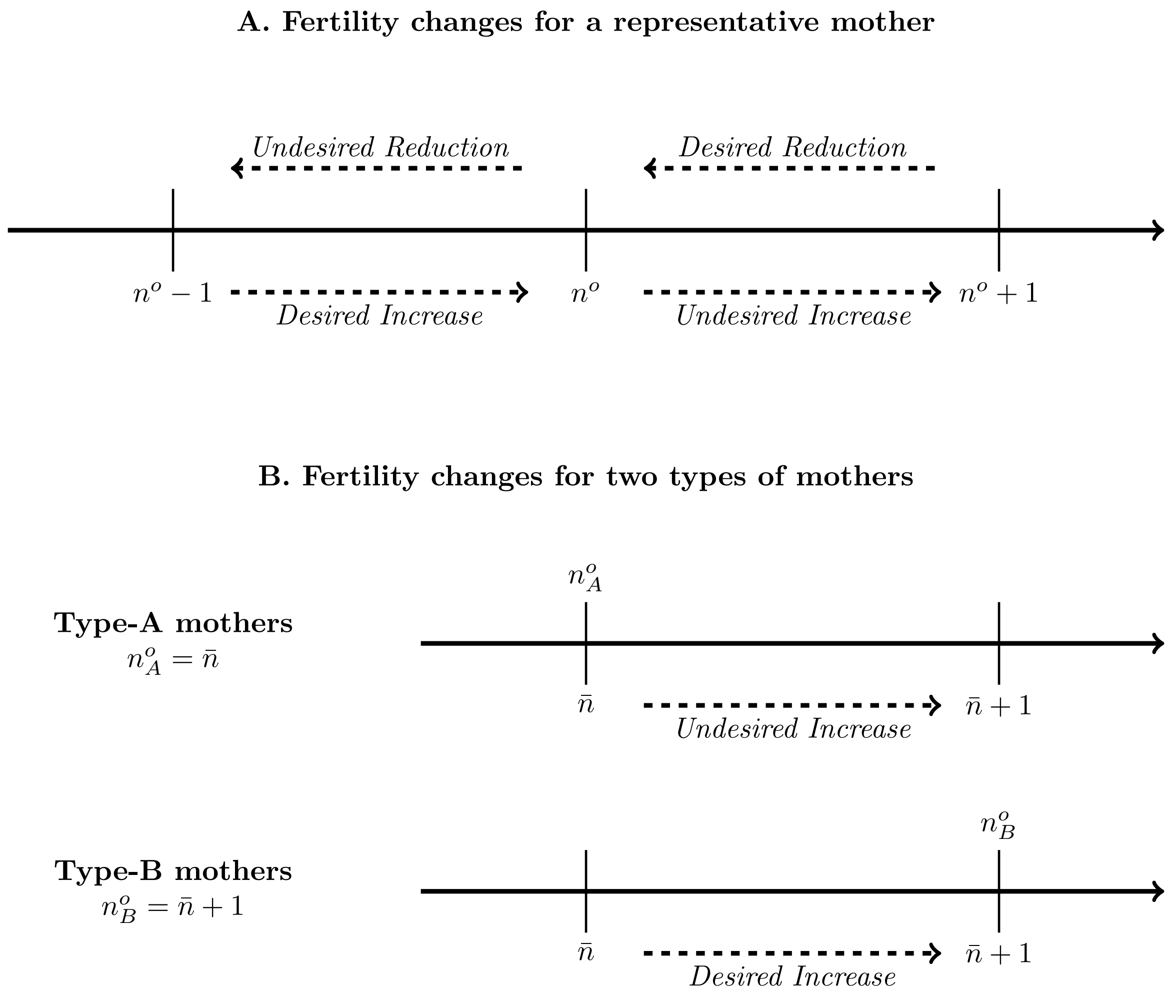

Figure 1: Undesired versus desired fertility changes

Notes: In sub-figure A, n^o is the optimal fertility level in the solution of the utility maximization problem P1 when fertility is not rationed. In sub-figure B, $n^o_A = \bar{n}$ is the optimal fertility for type-A mothers, and $n^o_B = \bar{n} + 1$ is the optimal fertility for type-B mothers.

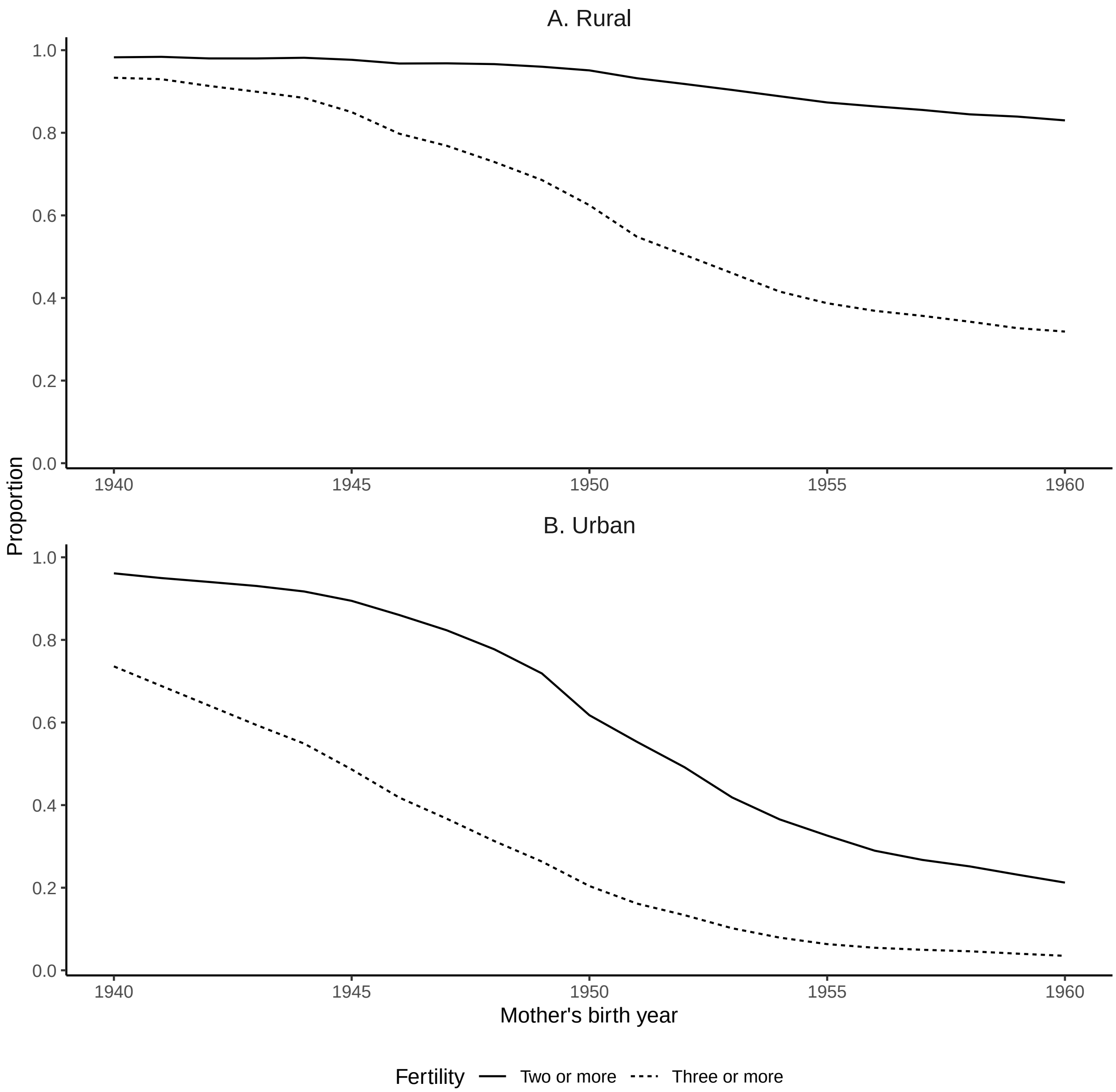

Figure 2: Trends in fertility

Notes: This figure shows the probability of a mother having at least two, three, or four children by the mother's birth year. Sub-figure A includes rural mothers, and sub-figure B includes urban mothers. The data source is the 1990, 2000, 2005, and 2010 waves of the China population census. We use Han mothers who were at least 40 years old in the census year.

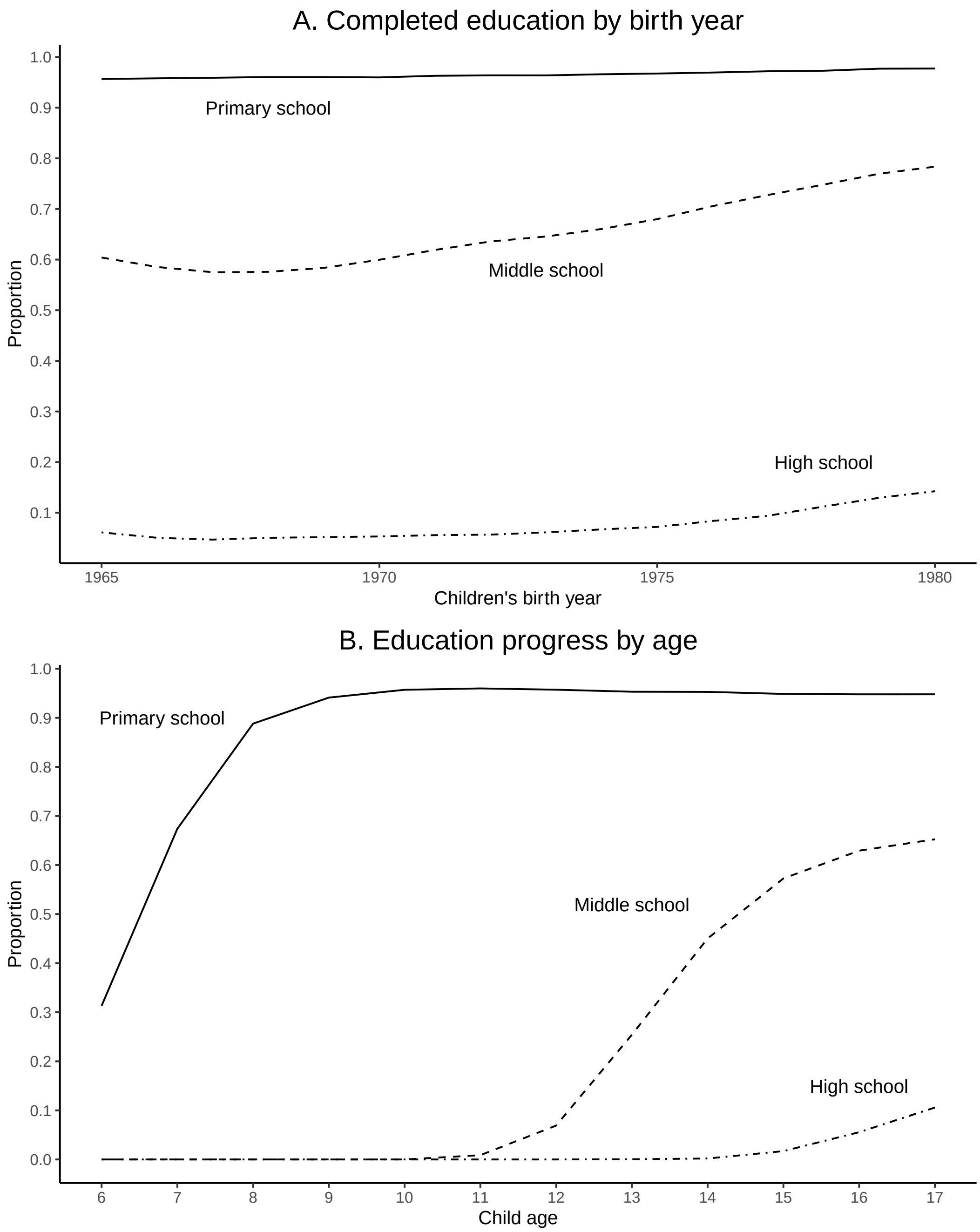

Figure 3: Children's education attainments

Notes: Sub-figure A shows the proportion of children completing primary, middle, or high school by children's birth year. The sample includes rural people born in 1965–1980 in the 2000 wave of the China population census. Sub-figure B shows the proportion of children attending primary, middle, or high school for children aged 6–17. The sample includes first-born children of mothers born in 1940–1960 in the 1982 and 1990 waves of the China population census.

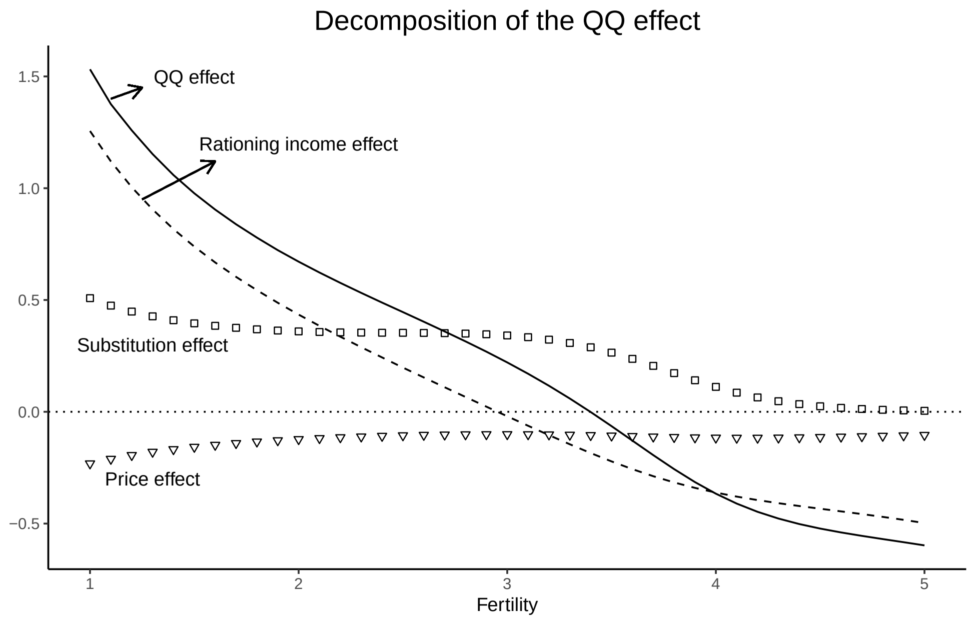

Figure 4: Decomposition results for a representative type-B mother

Notes: The figure shows decomposition results for a representative type-B mother when evaluated at different levels of rationed fertility. The x-axis is the rationed fertility level. The optimal fertility for a type-B mother is three children ($n_B^o = 3$). The y-axis shows the effect of rationed fertility on child quality (solid line) and its three components: price effect (line in triangles), substitution effect (line in squares), and rationing income effect (dashed line).

Table 1: Realized fertility for three types of mothers

	Fertility						Proportion
Policy	$X_i = 0$			$X_i = 1$			
Twinning	$Z_i = 0$		$Z_i = 1$	$Z_i = 0$		$Z_i = 1$	
	(1)		(2)	(3)		(4)	(5)
Type-A	2	$\xrightarrow{\alpha_1}$	3	2		3	P_A
Type-B	3		3	2	$\xrightarrow{\alpha_2}$	3	P_B
Type-C	3		3	3		3	P_C

Notes: The variable X_i is an indicator variable that equals one if mother i is under the birth-control policy, and zero otherwise. The variable Z_i is an indicator variable that equals one if mother i has second-born twins, and zero otherwise. The table plots realized fertility for three types of mothers under each combination of X_i and Z_i.

Table 2: Summary statistics

Sample	Full		Without twins		With twins	
	Mean (1)	S.D. (2)	Mean (3)	S.D. (4)	Mean (5)	S.D. (6)
Child quality						
Middle school attendance	0.51	0.50	0.51	0.50	0.49	0.50
Child quantity						
Three or more children	0.69	0.46	0.68	0.46	1.00	0.00
Birth-control policy						
Policy intensity	0.54	0.48	0.54	0.48	0.61	0.50
Twinning status						
Twin at the second birth	0.40%	0.06				
First-born child characteristics						
Male	0.50	0.50	0.50	0.50	0.49	0.50
Age	14.90	1.43	14.90	1.43	14.72	1.42
Maternal characteristics						
Maternal schooling years	3.85	3.38	3.85	3.38	4.11	3.43
Maternal age	36.99	2.76	36.99	2.76	37.12	2.54
Maternal age at the 1st birth	22.07	2.53	22.07	2.53	22.39	2.41
Maternal age at the 2nd birth	24.92	2.94	24.91	2.94	25.71	2.96
Observations	264,013		262,956		1,057	

Notes: This table shows summary statistics of mothers born in 1940–1960 and their first-born children in the 1982 and 1990 waves of the China population census. Columns (1) and (2) show the mean and standard deviation of each variable for the full sample, columns (3) and (4) for mothers without twins, and columns (5) and (6) for mothers with second-born twins.

Table 3: Baseline estimates and placebo tests

	Baseline estimates	Placebo tests	
Stratifying variable	Policy intensity (1)	First-born son (2)	Maternal schooling years (3)
	Panel A. Dependent variable: three or more children		
Twin (α_1)	0.212***	0.243***	0.312***
	(0.021)	(0.019)	(0.018)
Stratifying variable × Twin (α_2)	0.121***	0.098***	0.009***
	(0.011)	(0.005)	(0.001)
Stratifying variable	-0.121***	-0.098***	-0.009***
	(0.011)	(0.005)	(0.001)
F-statistic	136.09	184.52	143.72
R-squared	0.32	0.32	0.30
	Panel B. Dependent variable: middle school attendance		
Twin (ρ_1)	-0.033**	-0.003	-0.005
	(0.016)	(0.018)	(0.014)
Stratifying variable × Twin (ρ_2)	0.050**	0.001	-0.001
	(0.024)	(0.034)	(0.004)
Stratifying variable	0.026***	0.118***	0.027***
	(0.006)	(0.009)	(0.001)
R-squared	0.18	0.18	0.16
Observations	264,013	264,013	264,013
$\beta_A = \frac{\rho_1}{\alpha_1}$	-0.157**	-0.011	-0.017
	(0.078)	(0.073)	(0.043)
$\beta_B = \frac{\rho_2}{\alpha_2}$	0.412*	0.005	-0.139
	(0.211)	(0.352)	(0.410)
$\beta_B - \beta_A$	0.569**	0.016	-0.122
	(0.270)	(0.413)	(0.391)

Notes: Column (1) presents the baseline estimates, and columns (2) and (3) present the placebo tests. In column (1), the stratifying variable is the policy intensity, i.e., the expected fines for third-born children. In column (2), the stratifying variable is an indicator variable on whether the first-born child is a boy. In column (3), the stratifying variable is the schooling years of the mother. We subtract the median (six years) from maternal years of schooling before taking interactions. In all columns, the sample includes mothers born in 1940–1960 with at least two children in the 1982 and 1990 waves of the China population census. Control variables include maternal age at the second birth, maternal education, the child's gender, age, and age squared, fixed effects for province, maternal birth year, and census wave, as well as province-specific linear trends. Dependent variables are whether the mother has three or more children (Panel A) and whether the first-born child of the mother has ever attended middle school (Panel B). "Twin" is an indicator variable on whether the mother has second-born twins. We use block bootstrap with 100 repetitions to obtain standard errors clustered by province and maternal education (in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Model parameters estimates

Parameters	Description	Value	SE
α	Share of children quantity in utility function	0.612	0.018
ρ	CES substitution parameter	-7.920	0.285
θ_0	Share of children component in utility function	0.928	0.019
σ_ϵ	Std. of preference heterogeneity	0.300	0.012
$\sigma_{y\epsilon}$	Correlation coefficient between household income and preference	-0.010	0.001
π_n	Fixed cost of an additional child	1.467	0.037
π_{nq}	Quality-related cost of children	0.111	0.003
$\bar{\epsilon}$	Preference cutoff for type-C mothers	-0.140	0.004

Notes: This table reports the values and standard errors of estimated model parameters. Standard errors are bootstrapped.

Table 5: Decomposition results for type-A and type-B mothers

	Reduced-form estimates	Model simulation based on structural estimation			
	QQ effect	QQ effect	Price effect QQ^P	Substitution effect QQ^S	Rationing income effect QQ^I
Type-A mothers	-0.157	-0.162	-0.127	0.250	-0.286
Type-B mothers	0.412	0.421	-0.107	0.352	0.176
Difference (B − A)	0.569	0.583	0.020	0.102	0.462
% explained		100%	3.5%	17.4%	79.1%

Notes: This table reports the decomposition of overall QQ effects into price, substitution, and rationing income effects. The decomposition is based on Eq. (2). The rationed fertility level increases from two to three, which represents an undesired (a desired) increase for type-A (type-B) mothers. The first two rows shows the decomposition for type-A and type-B mothers, respectively. The second last row shows the difference in QQ effects and the three components between type-B and type-A mothers. The last row shows percent of the difference explained by each of the three components.

Online Appendix for
"Rationed Fertility: Treatment Effect Heterogeneity in the Child Quantity–Quality Tradeoff"

Rufei Guo* Junjian Yi† Junsen Zhang‡ Ning Zhang§

January 27, 2025

*Center for Health Economics and Management, Economics and Management School, Wuhan University. Email: rufei_guo@whu.edu.cn.

†National School of Development, Peking University; Institute for Global Health and Development, Peking University. Email: junjian@nsd.pku.edu.cn.

‡School of Economics, Zhejiang University. Email: jszhang@cuhk.edu.hk. Corresponding author.

§Department of Economics, Chinese University of Hong Kong. Email: ning.zhang@cuhk.edu.hk.

Contents

A1 Full Derivation of the Theory

In this section, we build on Rosenzweig and Wolpin's (1980) model using rationing theory in Neary and Roberts (1980), and formulate a general theory of *exogenous* fertility change from a truly rationed level. We first present and solve a model of rationed fertility. We then derive the effect of rationed fertility on child quality, and follow up with a simulation analysis.

A1.1 Becker–Lewis Setup

We first consider the three-commodity interactive (q^2) model, as in Becker and Lewis (1973) and Rosenzweig and Wolpin (1980). Parents maximize utility by choosing fertility or child quantity (n), child quality (q), and a composite consumption good (s),

$$\begin{aligned} \max_{n,q,s} \quad & U(n, q, s), \\ \text{subject to} \quad & \pi_n n + \pi_q q + \pi_{nq} nq + p_s s \leq y, \end{aligned} \tag{P1}$$

where y is the monetary income of the family. In the main text, we follow the recent literature to set $\pi_q = 0$. Becker and Lewis (1973, p. S283) use π_q to represent the costs of "public goods" or "family goods," such as "some aspects of training in the home and the 'handing down' of some clothing." The presence of π_q does not affect our theoretical analysis. We keep π_q in the Appendix for comparability with Becker and Lewis (1973).

The price of child quality, i.e., the cost of increasing q by one unit, is $p_q = \pi_q + \pi_{nq} n$, where $\pi_{nq} n$ represents the costs of "private goods," such as tuition fees and health care expenditure, which increases in n. Similarly, the price of child quantity, i.e., the cost of an additional child, is $p_q = \pi_n + \pi_{nq} q$, where π_n represents the "fixed cost" of an additional child, including the cost of giving birth and any necessities to keep the child alive; $\pi_{nq} q$ represents the cost of private goods, which increases in child quality; p_s is the price of the composite good. Solving P1 gives the optimal child quantity (n^o), child quality (q^o), and composite good (s^o).

We define the shadow prices of child quantity (v_n) and quality (v_q), respectively,

$$v_n = \frac{\partial U}{\partial n} / \lambda,$$
$$v_q = \frac{\partial U}{\partial q} / \lambda,$$

where λ is the Lagrangian multiplier of P1, representing the marginal utility of 1 dollar. So v_n and v_q are the marginal utilities of child quantity and quality measured by dollars. We then define the family "total income" in the Beckerian sense, as follows (Becker, 1991):

$$W = v_n n + v_q q + v_s s.$$

When $n = n^o$, $q = q^o$, and $s = s^o$, then $v_n = p_n$, $v_q = p_q$, and $v_s = p_s$; the family achieves the highest income, W^o.

To simplify notation, let $x = (n, q, s)$, $\pi = (\pi_{nq}, \pi_n, \pi_q, p_s)$, $p = (p_n, p_q, p_s)$, and $v = (v_n, v_q, v_s)$. Note that the shadow prices v are the monetary-equivalent utility values of n, q, and s, and the "actual" prices p are the monetary costs of n, q, and s. When there is no restriction on the choice of x, in equilibrium the shadow prices must equal the actual prices ($p = v$).[1]

If $\pi_{nq} > 0$, child quality enters the price of child quantity, and vice versa (Becker and Lewis, 1973). If $\pi_{nq} = 0$, the q^2 model is reduced to the standard non-interactive (q^1) model.

We defined the indirect utility function and uncompensated demand functions as

$$V(\pi, y) = \max_{n,q,s} \{U(n, q, s) | \pi_{nq} nq + \pi_n n + \pi_q q + p_s s \leq y\},$$

$$x(\pi, y) = \arg\max_{n,q,s} \{U(n, q, s) | \pi_{nq} nq + \pi_n n + \pi_q q + p_s s \leq y\},$$

where $x(\pi, y) = (n(\pi, y), q(\pi, y), s(\pi, y))$.

[1] Becker and Lewis (1973) and Rosenzweig and Wolpin (1980) conduct the comparative static analysis only at the unrestricted optimal point, so they do not distinguish p from v. To derive the implication of rationed fertility in the case of p not equal to v, a comparative static analysis off the unrestricted optimal point is necessary. Hence we call p the "actual" price of x to distinguish from the shadow price v of x.

The dual problem of the utility maximization problem in the q^2 model is the expenditure minimization problem,

$$\min_{n,q,s} \quad \pi_{nq}nq + \pi_n n + \pi_q q + p_s s,$$
$$\text{s.t} \quad U(n, q, s) \geq u.$$

We define the expenditure function and compensated demand functions as

$$e(\pi, u) = \min_{n,q,s}\{\pi_{nq}nq + \pi_n n + \pi_q q + p_s s | U(n, q, s) \geq u\},$$
$$x^c(\pi, u) = \arg\min_{n,q,s}\{\pi_{nq}nq + \pi_n n + \pi_q q + p_s s | U(n, q, s) \geq u\},$$

where $x^c(\pi, u) = (n^c(\pi, u), q^c(\pi, u), s^c(\pi, u))$. Throughout the paper, we use superscript "c" to denote compensated demand functions (in the Hicksian sense) in the expenditure minimization problem.

By the duality theorem, $x(\pi, y) = x^c(\pi, V(\pi, y))$. Specifically, we define the optimal fertility level $n^o \equiv n(\pi, y) = n^c(\pi, V(\pi, y))$.

A1.2 Rationed Fertility

We build on the rationing theory of Tobin and Houthakker (1950) and Neary and Roberts (1980) to extend Rosenzweig and Wolpin (1980), and derive a generalized comparative static analysis of an exogenous fertility change. As in Becker and Lewis (1973), both n and q are choice variables, so a direct comparative static analysis of the effect of n on q is not feasible. Rather, we focus primarily on the effect of rationed fertility.

To begin with, we fix n at $n = \bar{n}$. We call the q^2 (or q^1) model with fixed fertility ($n = \bar{n}$) the restricted q^2 (or q^1) model. The utility maximization problem in the restricted q^2 model is

$$\max_{q,s} \quad U(\bar{n}, q, s), \qquad \text{subject to} \quad \pi_n\bar{n} + \pi_q q + \pi_{nq}\bar{n}q + p_s s \leq y, \tag{P2}$$

where $\bar{n}$ is the number of children mandated by some exogenous force. Note that the restriction on or rationing of fertility makes $\bar{n}$ a parameter rather than a choice variable. We define the indirect utility function and uncompensated demand functions in the restricted q^2 model as

$$\tilde{V}(\pi, y, \bar{n}) = \max_{q,s}\{U(\bar{n}, q, s)|\pi_n \bar{n} + (\pi_{nq}\bar{n} + \pi_q)q + p_s s \leq y\},$$

$$\tilde{x}(\pi, y, \bar{n}) = \arg\max_{q,s}\{U(\bar{n}, q, s)|\pi_n \bar{n} + (\pi_{nq}\bar{n} + \pi_q)q + p_s s \leq y\},$$

where a tilde "∼" denotes functions in the restricted model.

The expenditure minimization problem in the restricted q^2 model is

$$\min_{q,s} \quad \pi_n \bar{n} + (\pi_{nq}\bar{n} + \pi_q)q + p_s s,$$

$$\text{subject to} \quad U(\bar{n}, q, s) \geq u.$$

The corresponding expenditure function and compensated demand functions are

$$\tilde{e}(\pi, u, \bar{n}) = \min_{q,s}\{\pi_n \bar{n} + (\pi_{nq}\bar{n} + \pi_q)q + p_s s|U(\bar{n}, q, s) \geq u\},$$

$$\tilde{x^c}(\pi, u, \bar{n}) = \arg\min_{q,s}\{\pi_n \bar{n} + (\pi_{nq}\bar{n} + \pi_q)q + p_s s|U(\bar{n}, q, s) \geq u\}.$$

To derive testable implications, we carry out the analysis in four steps. First, we link the restricted q^2 model and the unrestricted q^2 model. Second, we define two types of fertility changes. Third, we link the restricted q^2 model and the restricted q^1 model. Finally, we decompose the derivatives of q and s with respect to $\bar{n}$ in the restricted q^2 model into the derivatives of q, s, and n with respect to prices in the unrestricted q^1 model.[2]

In the restricted q^2 model, the shadow price of child quantity does not equal p_n if $\bar{n} \neq n^o$. Specifically, if $\bar{n} < n^o$, parents prefer more children, so the shadow price of child quantity is higher than p_n; if $\bar{n} > n^o$, parents prefer fewer children, so the shadow price is lower than p_n. We adjust

[2]The decomposition of derivatives in restricted models into derivatives in standard models helps deliver testable implications. This approach was pioneered by Becker and Lewis (1973), who decompose income and price elasticities of q and n in the q^2 model, which they call "observed" elasticities, into income and price elasticities of q and n in the q^1 model, which they call "true" elasticities.

the fixed price π_n (as a component of p_n) to equate p_n with the shadow price of n, inducing parents to choose $n = \bar{n}$ in the *unrestricted* q^2 model. The *supporting fixed price* $\bar{\pi}_n$ is defined by

$$\bar{n} = n^c(\pi_{-n}, \bar{\pi}_n, u), \tag{A1.1}$$

where $\pi_{-n} = (\pi_{nq}, \pi_q, p_s)$, and $n^c(\pi_{-n}, \bar{\pi}_n, u)$ is the compensated demand function of n in the unrestricted q^2 model (P2). The supporting fixed price $\bar{\pi}_n$ causes parents to choose $n = \bar{n}$ in the expenditure minimization problem in the unrestricted q^2 model. In the following analysis, functions derived from the unrestricted q^2 model are always evaluated at $(\pi_{-n}, \bar{\pi}_n, u)$.

Because the first-order conditions of q and s in the expenditure minimization problem in the restricted q^2 model and those in the unrestricted q^2 model (when $\pi = \bar{\pi}_n$) are the same, we have

$$\tilde{q}^c(\pi, u, \bar{n}) = q^c(\pi_{-n}, \bar{\pi}_n, u), \tag{A1.2}$$

$$\tilde{s}^c(\pi, u, \bar{n}) = s^c(\pi_{-n}, \bar{\pi}_n, u). \tag{A1.3}$$

By Eqs. (A1.1), (A1.2), and (A1.3), we connect the expenditure functions in the restricted q^2 model and those in the unrestricted q^2 model,

$$\tilde{e}(\pi, u, \bar{n}) = e(\pi_{-n}, \bar{\pi}_n, u) + (\pi_n - \bar{\pi}_n)\bar{n}. \tag{A1.4}$$

Differentiating Eq. (A1.4) with respect to $\bar{n}$, we have

$$\begin{aligned} \frac{\partial \tilde{e}}{\partial \bar{n}} &= \left(\frac{\partial e}{\partial \pi_n} - \bar{n} \right) \frac{\partial \bar{\pi}_n}{\partial \bar{n}} + (\pi_n - \bar{\pi}_n) \\ &= \pi_n - \bar{\pi}_n. \end{aligned} \tag{A1.5}$$

The second equality in Eq. (A1.5) holds because (i) $\frac{\partial e}{\partial \pi_n} = n^c$ by the envelope theorem (Shephard's lemma) and (ii) $n^c = \bar{n}$ by Eq. (A1.1).

Similarly, the indirect utility function in the restricted q^2 model is connected with that in the unrestricted q^2 model,

$$\tilde{V}(\pi, y, \bar{n}) = V(\pi_{-n}, \bar{\pi}_n, y + (\bar{\pi}_n - \pi_n)\bar{n}). \tag{A1.6}$$

Differentiating Eq. (A1.6) with respect to $\bar{n}$, we have

$$\begin{aligned}\frac{\partial \tilde{V}}{\partial \bar{n}} &= \left(\bar{n} + \frac{\partial V/\partial \bar{\pi}_n}{\partial V/\partial y}\right)\frac{\partial \bar{\pi}_n}{\partial \bar{n}}\frac{\partial V}{\partial y} + (\bar{\pi}_n - \pi_n)\frac{\partial V}{\partial y} \\ &= (\bar{\pi}_n - \pi_n)\frac{\partial V}{\partial y}.\end{aligned} \tag{A1.7}$$

The second equality in Eq. (A1.7) holds because (i) $\frac{\partial V/\partial \bar{\pi}_n}{\partial V/\partial y} = -n(\pi_{-n}, \bar{\pi}_n, y + (\bar{\pi}_n - \pi_n)\bar{n})$ by Roy's identity; (ii) $n(\pi_{-n}, \bar{\pi}_n, y + (\bar{\pi}_n - \pi_n)\bar{n}) = n^c(\pi_{-n}, \bar{\pi}_n, u)$ by the duality theorem; and (iii) $n^c = \bar{n}$ by Eq. (A1.1). We assume that the shadow price of income is always larger than zero, namely, $\frac{\partial V}{\partial y} > 0$.

A1.3 Desired and Undesired Fertility Changes

Based on Eqs. (A1.5) and (A1.7), we define desired and undesired fertility changes.

Definition A1 *When $\bar{n} < n^o$, $\bar{\pi}_n - \pi_n > 0$ and $\frac{\partial \tilde{V}}{\partial \bar{n}} > 0$, an increase in $\bar{n}$ is a **desired** fertility increase; a decrease in $\bar{n}$ is an **undesired** fertility decrease.*

Definition A2 *When $\bar{n} > n^o$, $\bar{\pi}_n - \pi_n < 0$ and $\frac{\partial \tilde{V}}{\partial \bar{n}} < 0$, an increase in $\bar{n}$ is an **undesired** fertility increase; a decrease in $\bar{n}$ is a **desired** fertility decrease.*

When fertility is rationed below the unrestricted optimum ($\bar{n} < n^o$), parents prefer more children. The shadow price of child quantity ($v_n = \bar{\pi}_n + \pi_{nq}\tilde{q}^c$) is higher than the actual price ($p_n = \pi_n + \pi_{nq}\tilde{q}^c$). As such, $\bar{\pi}_n - \pi_n = (\bar{\pi}_n + \pi_{nq}\tilde{q}^c) - (\pi_n + \pi_{nq}\tilde{q}^c) = v_n - p_n > 0$. By Eqs. (A1.5) and (A1.7), $\partial \tilde{e}/\partial \bar{n} < 0$ and $\partial \tilde{V}/\partial \bar{n} > 0$. An increase in n reduces the minimal expenditure to achieve a given utility level or raises the maximal utility attainable at a given income level. In this case, we call the fertility increase a desired one. Similarly, when fertility is rationed above the unrestricted optimum ($\bar{n} > n^o$), parents prefer fewer children. The shadow price of child quantity ($\bar{\pi}_n + \pi_{nq}\tilde{q}^c$) is lower than the actual price ($\pi_n + \pi_{nq}\tilde{q}^c$). As such, $\bar{\pi}_n - \pi_n = (\bar{\pi}_n + \pi_{nq}\tilde{q}^c) - (\pi_n + \pi_{nq}\tilde{q}^c) < 0$. By

Eqs. (A1.5) and (A1.7), $\partial\tilde{e}/\partial\bar{n} > 0$ and $\partial\tilde{V}/\partial\bar{n} < 0$. An increase in n either raises the minimal expenditure necessary to achieve a given utility level or reduces the maximal utility attainable at a given income level. In this case, we call the fertility increase an undesired one.

Undesired versus desired fertility changes are illustrated in Figure 1. Desired fertility changes move fertility toward the unrestricted optimal fertility level (n^0): Fertility increases at $\bar{n} < n^o$ and reductions at $\bar{n} > n^o$ are desired. In contrast, undesired fertility changes move fertility away from n^0: Fertility increases at $\bar{n} > n^o$ and reductions at $\bar{n} < n^o$ are undesired.

A1.4 Duality and Decomposition

By the duality theorem, we link uncompensated and compensated demand functions:

$$\tilde{q}(\pi, \tilde{e}(\pi, u, \bar{n}), \bar{n}) = \tilde{q}^c(\pi, u, \bar{n}), \tag{A1.8}$$

$$\tilde{s}(\pi, \tilde{e}(\pi, u, \bar{n}), \bar{n}) = \tilde{s}^c(\pi, u, \bar{n}). \tag{A1.9}$$

Differentiating Eqs. (A1.8) and (A1.9) with respect to $\bar{n}$ and invoking Eq. (A1.5), we have

$$\frac{\partial\tilde{q}}{\partial\bar{n}} = \frac{\partial\tilde{q}^c}{\partial\bar{n}} + (\bar{\pi}_n - \pi_n)\frac{\partial\tilde{q}}{\partial y}, \tag{A1.10}$$

$$\frac{\partial\tilde{s}}{\partial\bar{n}} = \frac{\partial\tilde{s}^c}{\partial\bar{n}} + (\bar{\pi}_n - \pi_n)\frac{\partial\tilde{s}}{\partial y}. \tag{A1.11}$$

Following Becker and Lewis (1973) and Rosenzweig and Wolpin (1980), we further decompose $\frac{\partial\tilde{q}}{\partial\bar{n}}$ and $\frac{\partial\tilde{s}}{\partial\bar{n}}$ into derivatives in the q^1 model. The expenditure minimization problem in the restricted q^1 model is

$$\begin{aligned} \min_{q,s} \quad & p_n^*\bar{n} + p_q^*q + p_s^*s, \\ \text{subject to} \quad & U(\bar{n}, q, s) \geq u, \end{aligned} \tag{P3}$$

where p_n^*, p_q^*, and p_s^* denote the actual prices in the non-interactive (q^1) model. We use the superscript "$*$" to denote functions in the q^1 model. The corresponding expenditure function and compensated demand functions are

$$\tilde{e^*}(p^*, u, \bar{n}) = \min_{q,s}\{p_n^* \bar{n} + p_q^* q + p_s^* s | U(\bar{n}, q, s) \geq u\},$$

$$\tilde{x^{*c}}(p^*, u, \bar{n}) = \arg\min_{q,s}\{p_n^* \bar{n} + p_q^* q + p_s^* s | U(\bar{n}, q, s) \geq u\},$$

where $p^* = (p_n^*, p_q^*, p_s^*)$ for notational brevity.

If $p_q^* = \pi_{nq}\bar{n} + \pi_q$, $p_n^* = \pi_n$, and $p_s^* = p_s$, the expenditure minimization problem in the restricted q^2 model (P2) is equivalent to that in the restricted q^1 model (P3). Hence the compensated demands of q and s in the two models are equal:

$$\tilde{q}^c(\pi, u, \bar{n}) = \tilde{q^{*c}}(\pi^*, u, \bar{n}), \tag{A1.12}$$

$$\tilde{s}^c(\pi, u, \bar{n}) = \tilde{s^{*c}}(\pi^*, u, \bar{n}). \tag{A1.13}$$

Differentiating Eqs. (A1.12) and (A1.13) with respect to $\bar{n}$,

$$\frac{\partial \tilde{q}^c}{\partial \bar{n}} = \pi_{nq}\frac{\partial \tilde{q^{*c}}}{\partial p_q^*} + \frac{\partial \tilde{q^{*c}}}{\partial \bar{n}}, \tag{A1.14}$$

$$\frac{\partial \tilde{s}^c}{\partial \bar{n}} = \pi_{nq}\frac{\partial \tilde{s^{*c}}}{\partial p_q^*} + \frac{\partial \tilde{s^{*c}}}{\partial \bar{n}}. \tag{A1.15}$$

If we further consider the utility maximization problem in the restricted q^1 model, we have

$$\frac{\partial \tilde{q}}{\partial \bar{n}} = \pi_{nq}\frac{\partial \tilde{q^*}}{\partial p_q^*} + \frac{\partial \tilde{q^*}}{\partial \bar{n}}, \tag{A1.16}$$

$$\frac{\partial \tilde{s}}{\partial \bar{n}} = \pi_{nq}\frac{\partial \tilde{s^*}}{\partial p_q^*} + \frac{\partial \tilde{s^*}}{\partial \bar{n}}, \tag{A1.17}$$

$$\frac{\partial \tilde{q}}{\partial y} = \frac{\partial \tilde{q^*}}{\partial y}, \tag{A1.18}$$

$$\frac{\partial \tilde{s}}{\partial y} = \frac{\partial \tilde{s^*}}{\partial y}, \tag{A1.19}$$

where $\tilde{q^*}$ and $\tilde{s^*}$ are uncompensated demand functions of q and s in the restricted q^1 model when

$p_q^* = \pi_{nq}\bar{n} + \pi_q$, $p_n^* = \pi_n$, $p_s^* = p_s$. Substituting Eqs. (A1.14) and (A1.18) into Eq. (A1.10), and Eqs. (A1.15) and (A1.19) into Eq. (A1.11), we have

$$\frac{\partial \tilde{q}}{\partial \bar{n}} = \pi_{nq} \frac{\partial \tilde{q^{*c}}}{\partial p_q^*} + \frac{\partial \tilde{q^{*c}}}{\partial \bar{n}} + (v_n - p_n) \frac{\partial \tilde{q^*}}{\partial y}, \tag{A1.20}$$

$$\frac{\partial \tilde{s}}{\partial \bar{n}} = \pi_{nq} \frac{\partial \tilde{s^{*c}}}{\partial p_q^*} + \frac{\partial \tilde{s^{*c}}}{\partial \bar{n}} + (v_n - p_n) \frac{\partial \tilde{s^*}}{\partial y}. \tag{A1.21}$$

Applying the Neary and Roberts's (1980, pp. 32-34) Eqs. (19), (24), and (29), we decompose derivatives of q and s with respect to $\bar{n}$ in the restricted q^1 model to derivatives in the unrestricted q^1 model,

$$\frac{\partial \tilde{q^{*c}}}{\partial \bar{n}} = \frac{\partial q^{*c}}{\partial p_n^*} \left(\frac{\partial n^{*c}}{\partial p_n^*} \right)^{-1}, \tag{A1.22}$$

$$\frac{\partial \tilde{q^{*c}}}{\partial p_q^*} = \frac{\partial q^{*c}}{\partial p_q^*} - \left(\frac{\partial n^{*c}}{\partial p_q^*} \right)^2 \left(\frac{\partial n^{*c}}{\partial p_n^*} \right)^{-1}, \tag{A1.23}$$

$$\frac{\partial \tilde{q^*}}{\partial y} = \frac{\partial q^*}{\partial y} - \frac{\partial \tilde{q^{*c}}}{\partial \bar{n}} \frac{\partial n^*}{\partial y}. \tag{A1.24}$$

As for the comparative static analysis with respect to s, we have

$$\frac{\partial \tilde{s^{*c}}}{\partial \bar{n}} = \frac{\partial s^{*c}}{\partial p_n^*} \left(\frac{\partial n^{*c}}{\partial p_n^*} \right)^{-1}, \tag{A1.25}$$

$$\frac{\partial \tilde{s^{*c}}}{\partial p_q^*} = \frac{\partial s^{*c}}{\partial p_q^*} - \frac{\partial n^{*c}}{\partial p_s^*} \frac{\partial n^{*c}}{\partial p_q^*} \left(\frac{\partial n^{*c}}{\partial p_n^*} \right)^{-1}, \tag{A1.26}$$

$$\frac{\partial \tilde{s^*}}{\partial y} = \frac{\partial s^*}{\partial y} - \frac{\partial \tilde{s^{*c}}}{\partial \bar{n}} \frac{\partial n^*}{\partial y}. \tag{A1.27}$$

Substituting Eq. (A1.24) into Eq. (A1.20), and Eq. (A1.27) into Eq. (A1.21), we have

$$\frac{\partial \tilde{q}}{\partial \bar{n}} = \pi_{nq} \frac{\partial \tilde{q^{*c}}}{\partial p_q^*} + (1 - \alpha_\Delta \epsilon_{n^* \cdot y}) \frac{\partial \tilde{q^{*c}}}{\partial \bar{n}} + (v_n - p_n) \frac{\partial q^*}{\partial y}, \tag{A1.28}$$

$$\frac{\partial \tilde{s}}{\partial \bar{n}} = \pi_{nq} \frac{\partial \tilde{s^{*c}}}{\partial p_q^*} + (1 - \alpha_\Delta \epsilon_{n^* \cdot y}) \frac{\partial \tilde{s^{*c}}}{\partial \bar{n}} + (v_n - p_n) \frac{\partial s^*}{\partial y}, \tag{A1.29}$$

where $\alpha_\Delta = \frac{(\bar{\pi}_n - \pi_n)\bar{n}}{y}$ and $\epsilon_{n^* \cdot y} = \frac{\partial n^*}{\partial y}\frac{y}{n^*}$. Note that α_Δ is the share of compensating income change,[3] $(\bar{\pi}_n - \pi_n)\bar{n}$, out of total monetary income y. $\alpha_\Delta > 0$ if child quantity is rationed below the unrestricted optimum ($\bar{n} < n^o$), and $\alpha_\Delta < 0$ if child quantity is rationed above ($\bar{n} > n^o$). The income elasticity of child quantity is given by $\epsilon_{n^* \cdot y}$. If $\bar{n} = n^o$, $\bar{\pi}_n = \pi_n$. Eqs. (A1.28) and (A1.29) are reduced to Eqs. (18) and (19) in Rosenzweig and Wolpin (1980, p. 231).[4] Importantly, Eq. (A1.28) enables us to derive the effect of $\bar{n}$, i.e., rationed fertility, on q, even though a comparative static analysis of the effect of n is not possible. Eq. (A1.28) is Eq. (2) in the main text.

A1.5 The Production Function of Child Quality

The model setup in our text is identical to Becker and Lewis (1973) and Rosenzweig and Wolpin (1980). Specifically, our model setup is,

$$\begin{aligned} \max_{n,q,s} \quad & U(n, q, s), \\ \text{subject to} \quad & \pi_n n + \pi_q q + \pi_{nq} nq + p_s s \leq y, \end{aligned} \tag{P1}$$

where q is child quality, and n is child quantity.

We modify our model to accommodate a nonlinear child quality production function, and find that our theoretical results remain qualitatively the same. Specifically, Mogstad and Wiswall (2016) use e to denote expenditure per child, and write child quality q as a monotonically increasing function of child expenditure e; the production output q then enters the utility function. Following

[3]Note that $(\bar{\pi}_n - \pi_n)\bar{n}$ appears in Eq. (A1.6). When π_n is adjusted to the supporting fixed price $\bar{\pi}_n$, income y should be adjusted to $y + (\bar{\pi}_n - \pi_n)\bar{n}$ to induce the unrestricted household to choose $\bar{n}$.

[4]Rosenzweig and Wolpin (1980) are the first to decompose derivatives of q and s with respect to $\bar{n}$ in the restricted q^2 model into derivatives of q, s, and n with respect to prices in the unrestricted q^1 model. They note that the method of decomposing derivatives in a restricted model into derivatives in an unrestricted model is analogous to the one used in rationing theory by Tobin and Houthakker (1950), which enables them to conduct a comparative static analysis only at the unrestricted optimal fertility level ($\bar{n} = n^o$). As a generalization of Tobin and Houthakker's (1950) rationing theory, Neary and Roberts (1980) apply duality techniques to evaluate functions both at and off the unrestricted optimum. We adopt their duality techniques to extend Rosenzweig and Wolpin's (1980) analysis. In this manner, we are able to evaluate derivatives of q and s with respect to $\bar{n}$ both at and off the unrestricted optimal fertility level (i.e., $\bar{n} = n^o$ and $\bar{n} \neq n^o$), respectively.

Mogstad and Wiswall (2016), our model setup becomes,[5]

$$\begin{aligned} \max_{n,e,s} \quad & U(n,q,s), \\ \text{subject to} \quad & q = q(e), q'(e) > 0, \\ & \pi_n n + \pi_e e + \pi_{ne} ne + p_s s \leq y. \end{aligned} \tag{P1*}$$

In P1*, we allow the quality function, $q(e)$, to be nonlinear. For example, if we assume that the quality function is concave ($q''(e) < 0$), then the cost function is convex ($e''(q) > 0$).

With this new setup, our decomposition result of Eq. (2) becomes:

$$\underbrace{\frac{\partial q}{\partial \bar{n}}}_{QQ} = \frac{\mathbf{d}q}{\mathbf{d}e}\frac{\partial e}{\partial \bar{n}} = \underbrace{\pi_{ne}\frac{\mathbf{d}q}{\mathbf{d}e}\frac{\partial \widetilde{e^{*c}}}{\partial p_e^*}}_{QQ^P:\text{ price effect}} + \underbrace{(1 - \alpha_\Delta \cdot \epsilon_{n^* \cdot y})\frac{\mathbf{d}q}{\mathbf{d}e}\frac{\partial \widetilde{e^{*c}}}{\partial \bar{n}}}_{QQ^S:\text{ substitution effect}} + \underbrace{(v_n - p_n)\frac{\mathbf{d}q}{\mathbf{d}e}\frac{\partial e^*}{\partial y}}_{QQ^I:\text{ rationing income effect}}, \tag{2*}$$

where $\frac{\mathbf{d}q}{\mathbf{d}e} = q'(e) > 0$. The decomposition result presented in Eq. (2*) is qualitatively the same as that in Eq. (2) in the paper.

A1.6 Theoretical Extensions on Parental Responses

We extend the model in Section 2 to incorporate parental time allocation. Parents care about children's average human capital (quality), the number of children, their own consumption, and leisure. Parents' utility function is $u = u(h, n, c, l)$, where h is the average human capital of children, c is parental consumption, and l is parental leisure. To enhance a child's human capital h, parents can either increase child expenditure q, or allocate more time to home tutorials t, i.e. $h = h(q, t)$. We use T to denote total parental time. Hence $d = T - t - l$ represents parental labor supply.

[5] Mogstad and Wiswall (2016) further assume $\pi_e = 0$ and $\pi_{ne} = 1$, which are not needed in our decomposition analysis.

The parents' utility maximization problem is

$$\begin{aligned} \max_{q,t,c,l} \quad & u = u(h, c, l, n), \\ \text{subject to} \quad & h = h(q, t), \\ & c + nq + \pi_n n + \pi_q q + w(t + l) = wT + y, \end{aligned} \tag{A1.30}$$

where $U(q, t, c, l, n) = u(h(q, t), c, l, n)$ is the reduced-form utility function. Here, q no longer represents child quality; instead, q represents parental investment in each child.[6] Meanwhile, c, t, and l are components of composite good s in the three-commodity model of Section 2.

Differentiating the budget constraint with respect to $\bar{n}$, and multiplying both sides by $\frac{1}{wT+y}$, we have

$$\epsilon_{cn} b_c + b_n + \epsilon_{qn} b_q + \epsilon_{(t+l)n} b_{t+l} = 0, \tag{A1.31}$$

where $\epsilon_{kn} = \frac{\partial k}{\partial n} \frac{1}{k}$, $b_k = \frac{p_k k}{wT+y}$, $\forall k = c, n, q, t + l$. Note that $p_c = 1$, $p_n = q + \pi_n$, $p_q = n + \pi_q$, and $p_{t+l} = w$; ϵ_{kn} is the semi-elasticity of k with respect to $\bar{n}$, $\forall k = c, q, t + l$; and b_k is the budget share of k, $\forall k = c, n, q, t + l$.[7]

If $\epsilon_{qn} = 0$, then $\epsilon_{cn} b_c + \epsilon_{(t+l)n} b_{t+l} = -b_n < 0$. Then either $\epsilon_{cn} < 0$ or $\epsilon_{(t+l)n} < 0$.[8] The theory implies that in response to an exogenous fertility increase, parents will either reduce self-consumption or increase labor supply, or both, in order to maintain expenditure per child.

[6] $\pi_{nq} = 1$ by construction. Child human capital h can also be multidimensional. Heckman (2007) emphasizes the importance of noncognitive skills as a form of human capital.

[7] $\epsilon_{(t+l)n} = \frac{b_t}{b_{t+l}} \epsilon_{tn} + \frac{b_l}{b_{t+l}} \epsilon_{ln}$. Note that $\epsilon_{ky} = \frac{\partial k}{\partial y} \frac{y}{k}$ is the income elasticity of k.

[8] $\epsilon_{(t+l)n} < 0$ means $\epsilon_{dn} > 0$ since $d = T - t - l$.

A2 Full Derivation of the Econometrics

A2.1 Setup

For purposes of illustration, we assume a pool of three types of mothers ($i \in A, B, C$), as shown in Table 1. The shares of the three types of mothers are P_A, P_B, and P_C (column (5)), which are unobservable to researchers. Without twinning at the second birth ($Z_i = 0$) and the birth-control policy ($X_i = 0$), the three types of mothers can achieve their optimal fertility levels in P1, which are two, three, and three (column (1)).

We consider the differences between twinning-induced fertility increases without and with the policy. Without the policy ($X_i = 0$), twinning at the second birth shifts the realized fertility of type-A mothers from two to three, but does not affect the realized fertility of type-B and type-C mothers (column (2)). In this case, type-A mothers are "compliers" of twinning; type-B and type-C mothers are "always-takers" of twinning.[9]

We assume that the birth-control policy targets two children per family. Because type-A mothers' optimal fertility is two, the policy does not affect the realized fertility of type-A mothers (column (3)). The policy rations fertility of type-B mothers, reducing type-B mothers' fertility from three to two. Type-C mothers do not comply with the policy, and realize three children regardless of the policy.

Under the policy ($X_i = 1$), twinning shifts the realized fertility from two to three for both type-A and type-B mothers (column (4)). In this case, both type-A and type-B mothers are compliers of twinning; type-C mothers remain always takers of twinning.

In this simple scenario, we observe two types of fertility changes. Without the policy, the twinning shifts fertility of type-A mothers from the optimal two to an undesired three (the arrow between columns (1) and (2) in Table 1). Black, Devereux, and Salvanes (2010) consider twinning to be an "unexpected" and "unplanned" shock to fertility. Similarly, Mogstad and Wiswall (2016,

[9] In the terminology of Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996), compliers are individuals whose treatment status is affected by the instrument, and always-takers are individuals who are treated irrespective of whether the instrument is switched on or off.

p. 173) conclude that "twin births increase the number of siblings beyond the desired family size." Thus, the twinning-induced fertility increase for type-A mothers is undesired. By contrast, because the policy rations the fertility level of type-B mothers at two, twinning helps type-B mothers circumvent the rationing and achieve the optimal fertility of three, which represents a desired fertility increase (the arrow between columns (3) and (4) in Table 1).[10]

A2.2 Assumptions

We adopt the framework from the literature on the treatment effect. The indicator of three children, D_i, is the realized treatment status of mother i ($D_i = 0, 1$). Denote D_{zi} as the potential treatment status of mother i when $Z_i = z$ ($z = 0, 1$). We have $D_i = D_{0i} + (D_{1i} - D_{0i}) \cdot Z_i$. As Y_i is the realized outcome, we further denote $Y_i(d, z)$ as the potential outcome when $D_i = d$ and $Z_i = z$.

We make the three standard assumptions on the independence, exclusion, and monotonicity of twinning (Z_i).

Assumption 1 *Conditional on policy exposure X_i, twinning status Z_i is independent of the potential outcomes.* $\{Y_i(d, z), D_{zi}\}_{\forall z \in \{0,1\}, \forall d \in \{0,1\}} \perp Z_i\big|_{X_i}$.

Assumption 2 *Conditional on policy exposure X_i, twinning status Z_i affects Y_i only via D_i.* $Y_i(d, 0) = Y_i(d, 1) = Y_i(d) = Y_{di}\big|_{X_i}, \forall d \in \{0, 1\}$

Assumption 3 *Twinning status Z_i monotonically shifts D_i for everyone.* $D_{1i} \geq D_{oi}, \forall i$.

The monotonicity assumption automatically holds in our setting, because mothers with twins at the second birth have at least three children. We also require the "relevancy" condition, $\mathbb{E}[D_{1i} - D_{0i}|X_i] > 0$, which automatically satisfies for the twin instrument.

We make two additional assumptions on the policy (X_i).

[10] Two more changes are present in Table 1. Given the policy ($X_i = 1$), twinning shifts the fertility of type-A mothers from the optimal two to an undesired three, which also represents an undesired increase. Given non-twinning ($Z_i = 0$), the policy rations fertility of type-B mothers, and reduces the fertility from three to two, which represents an undesired decrease.

Assumption 4 *The policy X_i does not change mothers' types.* $\mathbf{Pr}(i \in S | X_i = x) = \mathbf{Pr}(i \in S) = P_S, \forall x \in \{0, 1\}, \forall S = A, B, C.$

Assumption 5 *The policy X_i can be excluded from the average treatment effect of D_i on Y_i for each type of mother.* $\mathbb{E}[Y_{1i} - Y_{0i} | i \in S, X_i = x] = \mathbb{E}[Y_{1i} - Y_{0i} | i \in S], \forall x \in \{0, 1\}, \forall S = A, B, C.$

Assumption 4 holds by definition. Assumption 5, which is similar to assumption A4 in Hull (2018), states that the policy does not change the average treatment effect for each type of mother.

A2.3 Fertility Equation

To distinguish between the two types of fertility increases, we consider the fertility equation

$$D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_i \cdot X_i + \alpha_3 X_i + v_i,$$

where D_i is an indicator equal to one if mother i has three children, and zero otherwise; v_i represents idiosyncratic fertility preference shocks.

For mothers who are not exposed to the birth-control policy ($X_i = 0$), the effect of twinning on fertility is

$$\begin{aligned}
&\mathbb{E}[D_i | Z_i = 1, X_i = 0] - \mathbb{E}[D_i | Z_i = 0, X_i = 0] \\
=&\mathbb{E}[D_{1i} - D_{0i} | X_i = 0] \\
=&\mathbf{Pr}(D_{1i} > D_{0i} | X_i = 0) \\
=&\mathbf{Pr}(i \in A | X_i = 0) \\
=&\mathbf{Pr}(i \in A) \\
=&P_A,
\end{aligned}$$

where the first equality uses A1, the second equality uses A3, and the third equality uses A4. The compliers are type-A mothers, who desire two children and experience undesired fertility increases because of twinning. We have, $\alpha_1 = P_A$.

For mothers exposed to the birth-control policy ($X_i = 1$), the effect of twinning on fertility is

$$
\begin{aligned}
&\mathbb{E}[D_i|Z_i = 1, X_i = 1] - \mathbb{E}[D_i|Z_i = 0, X_i = 1] \\
=&\mathbb{E}[D_{1i} - D_{0i}|X_i = 1] \\
=&\mathbf{Pr}(D_{1i} > D_{0i}|X_i = 1) \\
=&\mathbf{Pr}(D_{0i} = 0|X_i = 1) \\
=&\mathbf{Pr}(i \in A|X_i = 1) + \mathbf{Pr}(i \in B|X_i = 1) \\
=&\mathbf{Pr}(i \in A) + \mathbf{Pr}(i \in B) \\
=&P_A + P_B,
\end{aligned}
$$

where the first equality uses A1, the second equality uses A3, and the fourth equality uses A4. When $X_i = 1$, twinning induces an undesired fertility increase with probability P_A, and a desired fertility increase with probability P_B. We have, $\alpha_1 + \alpha_2 = P_A + P_B$, and $\alpha_2 = P_B$.

The effect of the birth-control policy on the fertility of mothers without twins is

$$
\begin{aligned}
&\mathbb{E}[D_i|Z_i = 0, X_i = 1] - \mathbb{E}[D_i|Z_i = 0, X_i = 0] \\
=&\mathbb{E}[D_{0i}|X_i = 1] - \mathbb{E}[D_{0i}|X_i = 0] \\
=&\mathbf{Pr}(i \in C|X_i = 1) - \{\mathbf{Pr}(i \in C|X_i = 0) + \mathbf{Pr}(i \in B|X_i = 0)\} \\
=& - \mathbf{Pr}(i \in B) \\
=& - P_B,
\end{aligned}
$$

where the first equality uses A1 and the third equality uses A4. We have, $\alpha_3 = -P_B$.

A2.4 Child-quality Equation

We then estimate QQ effects for the two types of fertility changes by considering the child-quality equation:

$$Y_i = \rho_0 + \rho_1 Z_i + \rho_2 Z_i \cdot X_i + \rho_3 X_i + \varepsilon_i.$$

where Y_i is the quality of the child of mother i and ε_i is the idiosyncratic shocks to child quality.

For mothers who are not exposed to the birth-control policy ($X_i = 0$), the effect of twinning on child quality is

$$\begin{aligned}
&\mathbb{E}[Y_i|Z_i = 1, X_i = 0] - \mathbb{E}[Y_i|Z_i = 0, X_i = 0] \\
=&\mathbb{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}|X_i = 0] - \mathbb{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}|X_i = 0] \\
=&\mathbb{E}[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})|X_i = 0] \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0, X_i = 0] \cdot \mathbf{Pr}(D_{1i} - D_{0i} > 0|X_i = 0) \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|i \in A, X_i = 0] \cdot \mathbf{Pr}(i \in A|X_i = 0) \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|i \in A] \cdot \mathbf{Pr}(i \in A) \\
=& P_A \cdot \beta_A,
\end{aligned}$$

where the first equality uses A1 and A2, the third equality uses A3, and the fifth equality uses A4 and A5. Here $\beta_A = \mathbb{E}[Y_{1i} - Y_{0i}|i \in A]$ is the average treatment effect for type-A mothers who experience undesired fertility increases. We have $\rho_1 = P_A \cdot \beta_A$.

For mothers exposed to the birth-control policy ($X_i = 1$), the effect of twinning on child quality is

$$
\begin{aligned}
&\mathbb{E}[Y_i|Z_i = 1, X_i = 1] - \mathbb{E}[Y_i|Z_i = 0, X_i = 1] \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0, X_i = 1] \cdot \mathbf{Pr}(D_{1i} - D_{0i} > 0|X_i = 1) \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|D_{0i} = 0, X_i = 1] \cdot \mathbf{Pr}(D_{0i} = 0|X_i = 1) \\
=&\{\mathbb{E}[Y_{1i} - Y_{0i}|i \in A, X_i = 1] \cdot \mathbf{Pr}(i \in A|D_{0i} = 0, X_i = 1) \\
&+ \mathbb{E}[Y_{1i} - Y_{0i}|i \in B, X_i = 1] \cdot \mathbf{Pr}(i \in B|D_{0i} = 0, X_i = 1)\} \cdot \mathbf{Pr}(D_{0i} = 0|X_i = 1) \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|i \in A, X_i = 1] \cdot \mathbf{Pr}(i \in A|X_i = 1) \\
&+ \mathbb{E}[Y_{1i} - Y_{0i}|i \in B, X_i = 1] \cdot \mathbf{Pr}(i \in B|X_i = 1) \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|i \in A] \cdot \mathbf{Pr}(i \in A) + \mathbb{E}[Y_{1i} - Y_{0i}|i \in B] \cdot \mathbf{Pr}(i \in B) \\
=& P_A \cdot \beta_A + P_B \cdot \beta_B,
\end{aligned}
$$

where the first equality uses A1-A3, and the second-last equality uses A4 and A5. Coeffficient $\beta_B = \mathbb{E}[Y_{1i} - Y_{0i}|i \in B]$ is the average treatment effect for type-B mothers, who experience desired fertility increases. We have $\rho_1 + \rho_2 = P_A \cdot \beta_A + P_B \cdot \beta_B$, and $\rho_2 = P_B \cdot \beta_B$.

The effect of the birth-control policy on the child quality of mothers without twins is

$$
\begin{aligned}
&\mathbb{E}[Y_i|Z_i = 0, X_i = 1] - \mathbb{E}[Y_i|Z_i = 0, X_i = 0] \\
=&\,\mathbb{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}|X_i = 1] - \mathbb{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}|X_i = 0] \\
=&\,\mathbb{E}[Y_{0i}|D_{0i}, X_i = 1] \cdot \mathbf{Pr}(D_{0i} = 0|X_i = 1) + \mathbb{E}[Y_{1i}|D_{0i}, X_i = 1] \cdot \mathbf{Pr}(D_{0i} = 1|X_i = 1) \\
&- \mathbb{E}[Y_{0i}|D_{0i}, X_i = 0] \cdot \mathbf{Pr}(D_{0i} = 0|X_i = 0) - \mathbb{E}[Y_{1i}|D_{0i}, X_i = 0] \cdot \mathbf{Pr}(D_{0i} = 1|X_i = 0) \\
=&\,\mathbb{E}[Y_{0i}|i \in A \cup B, X_i = 1] \cdot (P_A + P_B) + \mathbb{E}[Y_{1i}|i \in C, X_i = 1] \cdot P_C \\
&- \mathbb{E}[Y_{0i}|i \in A, X_i = 0] \cdot P_A - \mathbb{E}[Y_{1i}|i \in B \cup C, X_i = 0] \cdot (P_B + P_C) \\
=&\{\mathbb{E}[Y_{0i}|i \in A, X_i = 1] - \mathbb{E}[Y_{0i}|i \in A, X_i = 0]\} \cdot P_A \\
&+ \{\mathbb{E}[Y_{0i}|i \in B, X_i = 1] - \mathbb{E}[Y_{0i}|i \in B, X_i = 0]\} \cdot P_B \\
&+ \{\mathbb{E}[Y_{1i}|i \in C, X_i = 1] - \mathbb{E}[Y_{1i}|i \in C, X_i = 0]\} \cdot P_C \\
&- \mathbb{E}[Y_{1i} - Y_{0i}|i \in B, X_i = 0] \cdot P_B \\
=&\{\mathbb{E}[Y_{0i}|X_i = 1] - \mathbb{E}[Y_{0i}|X_i = 0]\} - P_B \cdot \beta_B,
\end{aligned}
$$

where the first equality uses A1 and the last equality uses A5. The term $\mathbb{E}[Y_{0i}|X_i = 1] - \mathbb{E}[Y_{0i}|X_i = 0]$ is the difference in the potential outcome in child quality by the policy—that is, a "selection effect." We have $\rho_3 = -P_B \cdot \beta_B + \{\mathbb{E}[Y_{0i}|X_i = 1] - \mathbb{E}[Y_{0i}|X_i = 0]\}$.

A2.5 A Second Stage

Consider a "second-stage" regression,

$$
Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 D_i \cdot X_i + \gamma_3 X_i + \varepsilon_i,
$$

where we use Z_i and $Z_i \cdot X_i$ as the instrumental variables for D_i and $D_i \cdot X_i$. The coefficient γ_1 is the effect of D_i on Y_i when $X_i = 0$,

$$
\begin{aligned}
\gamma_1 =& \frac{\mathbb{E}[Y_i|Z_i = 1, X_i = 0] - \mathbb{E}[Y_i|Z_i = 0, X_i = 0]}{\mathbb{E}[D_i|Z_i = 1, X_i = 0] - \mathbb{E}[D_i|Z_i = 0, X_i = 0]} \\
=& \frac{\beta_A \cdot P_A}{P_A} \\
=& \beta_A.
\end{aligned}
$$

The coefficient γ_1 identifies the quantity-quality effect for type-A mothers, i.e., $\mathbb{E}[Y_{1i} - Y_{0i}|i \in A]$. The coefficient $\gamma_1 + \gamma_2$ is the effect of D_i on Y_i when $X_i = 1$,

$$
\begin{aligned}
\gamma_1 + \gamma_2 =& \frac{\mathbb{E}[Y_i|Z_i = 1, X_i = 1] - \mathbb{E}[Y_i|Z_i = 0, X_i = 1]}{\mathbb{E}[D_i|Z_i = 1, X_i = 1] - \mathbb{E}[D_i|Z_i = 0, X_i = 1]} \\
=& \frac{P_A \cdot \beta_A + P_B \cdot \beta_B}{P_A + P_B} \\
=& \frac{P_A}{P_A + P_B}\beta_A + \frac{P_B}{P_A + P_B}\beta_B.
\end{aligned}
$$

We can also show that γ_3 equals negative γ_2 plus a selection effect,

$$
\gamma_3 = -\gamma_2 + \{\mathbb{E}[Y_{0i}|X_i = 1] - \mathbb{E}[Y_{0i}|X_i = 0]\}.
$$

The sign of γ_2 informs the size of β_A versus β_B:

$$
\gamma_2 = \frac{P_B}{P_A + P_B}(\beta_B - \beta_A).
$$

If $\gamma_2 > 0$, then $\beta_B > \beta_A$.

A2.6 Continuous Policy

In the previous derivations, we have assumed that the policy exposure is a dummy variable. Relaxing the setting to a continuous policy does not change the model implications. Consider a continuous policy that ranges from zero to $\bar{X}$, $X_i \in [0, \bar{X}]$. The definitions of type-A and type-C mothers do not change: Type-A mothers always desire and realize two children, and type-C mothers always desire and realize three children. Suppose N_i is realized fertility, and N_i^o is optimal fertility in P1:

$$N_i = N_i^o = 2\big|_{Z_i=0,} \quad \forall i \in A,$$

$$N_i = N_i^o = 3\big|_{Z_i=0,} \quad \forall i \in C.$$

The definition for type-B mothers becomes subtler. A type-B mother, who desired three children, will have two children only when the policy is strong enough.

$$\begin{aligned}
&\forall i \in B, \exists \theta_i \in [0, \bar{X}), \\
&N_i = N_i^o = 3\big|_{Z_i=0, X_i \le \theta_i}, \\
&N_i = 2 < 3 = N_i^o\big|_{Z_i=0, X_i > \theta_i},
\end{aligned}$$

where θ_i is the minimal policy strength that reduces a type-B mother's fertility from three to two. Let $F_B(\cdot)$ denote the cumulative distribution function of θ_i among type-B mothers. At policy intensity x, $F_B(x) \equiv \mathbf{Pr}(\theta_i \le x | i \in B, X_i = x) = \mathbf{Pr}(D_{0i} = 0 | i \in B, X_i = x)$. We use type-$B_x$ mothers to denote type-B mothers with $\theta_i \le x$. In other words, $i \in B_x$ is equivalent to $(i \in B, \theta_i \le x)$. We have $P_{B_x} = \mathbf{Pr}(i \in B_x) = \mathbf{Pr}(i \in B, \theta_i \le x) = F_B(x)\mathbf{Pr}(i \in B)$.

Assumptions 4 and 5 should be modified as:

Assumption 4' *A mother's X_i does not change her type.* $\mathbf{Pr}(i \in S | X_i = x) = \mathbf{Pr}(i \in S) = P_S, \forall x \in [0, \bar{X}], \forall S = A, C$. *Also* $\mathbf{Pr}(i \in B, \theta_i \le x | X_i = x) = \mathbf{Pr}(i \in B, \theta_i \le x) = P_{B_x}$.

Assumption 5' *The X_i can be excluded from the average treatment effect of D_i on Y_i for each type of mother.* $\mathbb{E}[Y_{1i} - Y_{0i} | i \in S, X_i = x] = \mathbb{E}[Y_{1i} - Y_{0i} | i \in S], \forall x \in [0, 1], \forall S = A, C$. *Also* $\mathbb{E}[Y_{1i} - Y_{0i} | i \in B, \theta_i \le x, X_i = x] = \mathbb{E}[Y_{1i} - Y_{0i} | i \in B, \theta_i \le x], \forall x \in [0, \bar{X}]$.

The effect of twinning on fertility for mothers under $X_i = x$ is

$$\begin{aligned}
&\mathbb{E}[D_i|Z_i = 1, X_i = x] - \mathbb{E}[D_i|Z_i = 0, X_i = x] \\
=&\mathbb{E}[D_{1i} - D_{0i}|X_i = x] \\
=&\mathbf{Pr}(D_{0i} = 0|X_i = x) \\
=&\mathbf{Pr}(i \in A|X_i = x) + \mathbf{Pr}(D_{0i} = 0|i \in B, X_i = x) \cdot \mathbf{Pr}(i \in B|X_i = x) \\
=&\mathbf{Pr}(i \in A) + F_B(x) \cdot \mathbf{Pr}(i \in B), \\
=&P_A + F_B(x) \cdot P_B, \\
=&P_A + P_{B_x},
\end{aligned}$$

where the first equality uses A1 and the second-last equality uses A4'.

The effect of twinning on child quality is

$$\begin{aligned}
&\mathbb{E}[Y_i|Z_i = 1, X_i = x] - \mathbb{E}[Y_i|Z_i = 0, X_i = x] \\
=&\mathbb{E}[Y_{1i} - Y_{0i}|i \in A] \cdot \mathbf{Pr}(i \in A) + \mathbb{E}[Y_{1i} - Y_{0i}|i \in B, \theta_i \leq x] \cdot \mathbf{Pr}(i \in B) \cdot F_B(x) \\
=&P_A \cdot \beta_A + F_B(x) \cdot P_B \cdot \beta_{B_x}, \\
=&P_A \cdot \beta_A + P_{B_x} \cdot \beta_{B_x},
\end{aligned}$$

where $\beta_{B_x} = \mathbb{E}[Y_{1i} - Y_{0i}|i \in B, \theta_i \leq x]$.

Under continuous policy, we not only identify P_A and β_A:

$$\begin{aligned}
P_A &= \mathbb{E}[D_i|Z_i = 1, X_i = 0] - \mathbb{E}[D_i|Z_i = 0, X_i = 0], \\
\beta_A &= \frac{\mathbb{E}[Y_i|Z_i = 1, X_i = 0] - \mathbb{E}[Y_i|Z_i = 0, X_i = 0]}{\mathbb{E}[D_i|Z_i = 1, X_i = 0] - \mathbb{E}[D_i|Z_i = 0, X_i = 0]},
\end{aligned}$$

but also P_{B_x} and β_{B_x} for $x \in [0, \bar{X}]$:

$$P_{B_x} = (\mathbb{E}[D_i|Z_i = 1, X_i = x] - \mathbb{E}[D_i|Z_i = 0, X_i = x]) - P_A,$$
$$\beta_{B_x} = \frac{1}{P_{B_x}}(\mathbb{E}[Y_i|Z_i = 1, X_i = x] - \mathbb{E}[Y_i|Z_i = 0, X_i = x] - P_A \cdot \beta_A).$$

We parametrically implement the estimations for P_A, β_A, P_{B_x}, and β_{B_x}, using polynomials in X_i. We have discussed estimation results in Section A3.5 in the main text.

A3 Further Details of the Empirical Analysis

A3.1 Covariates

We now discuss the vector of covariates in Eqs. (7) and (8), the inclusion of which is to ensure that Assumptions 1, 2, and 5 in Section 3.2 are plausible. First, we discuss the covariates related to Assumptions 1 and 2 on the independence and the exclusion restriction of twinning. Following the literature on the QQ effects using twinning as a natural experiment, this category of covariates includes maternal age at the second birth, maternal education, the first-born child's gender, age, and age squared.

Second, Assumption 5 states that the policy can be excluded from the average treatment effect of fertility on child quality for each type of mothers. We use fines as a measure of the coercive policy, which reflects the strength of fertility rationing. However, cross-province and over-cohort variations in the fines may capture factors beyond fertility rationing, especially factors that may correlate with fertility and child quality simultaneously. Before the policy, the socioeconomic conditions in provinces that set the fines earlier or higher may be different from those in other provinces. If this is the case, coefficients on the policy intensity may capture the influence of predetermined provincial characteristics.

To address this concern, we add three sets of control variables step by step in our main regressions. First, in our baseline estimate (column (1) of Table 3), we include province and maternal cohort fixed effects, as well as province-specific linear trends, as in the standard province-by-cohort data sets. By doing so, we remove the province-specific trends in policy strength, and exploit within-province and over-time variations in the residuals, which reflect the changing strength of rationing within each province. In the second step, we further include province-specific quadratic trends to capture unobserved confounders which may potentially correlate with the policy intensity in a nonlinear manner (column (1) of Table A5). In the third step, we include interactions between maternal birth cohort dummies and provincial economic/population growth rates during 1970–1975, prior to the policy (columns (2) and (3) of Table A5).[11] The specification allows predetermined economic/population growth rate to have different effects on fertility and child quality over cohorts. The signs and magnitude of our estimates remain largely unchanged.

As Eqs. (7) and (8) include Z_i, X_i, and $Z_i \cdot X_i$, we should control for the vector of control variables ($\mathbf{C_i}$) and its interactions with Z_i, X_i, and $Z_i \cdot X_i$. By controlling for these interactions, we allow that all control variables, including fixed effects and trends, to have differential effects on fertility and child quality between twin and non-twin families across different policy intensities. However, it is inappropriate to estimate the model with full interactions, because of the large number of controls.[12]

We use a parsimonious specification to deal with the control variables in two steps. Take the fertility equation (Eq. (7)) for example. In the first step, we regress D_i on X_i and $\mathbf{C_i}$ using the sample of non-twin households:

$$D_i = \delta_0 + \delta_1 X_i + \mathbf{C_i}\delta_2 + \epsilon_i.$$

Using the estimated coefficients, we obtain the predicted D_i when $X_i = 0$: $\hat{P}_i^D = \hat{\delta}_0 + \mathbf{C_i}\hat{\delta}_2$. The predicted variable $\hat{P}_i^D$ measures the propensity of mother i to give the third birth if she is not exposed to the birth-control policy and does not have twin children. We subtract $\hat{P}_i^D$ by its sample median

[11]The level terms of economic and population growth rates are absorbed by province fixed effects. In addition, because we use two censuses, we add a census wave dummy.

[12]In our sample, the total number of twin families is 1057 (Table 2). The power of our statistical tests becomes weak if we include the vector of controls and its interactions with Z_i, X_i, and $Z_i \cdot X_i$.

to obtain $\hat{P}_i^{Ddm}$.

In the second step, we estimate the fertility equation, including $\hat{P}_i^{Ddm}$ and its interactions with Z_i, X_i, and $Z_i \cdot X_i$:

$$\begin{aligned} D_i = & \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_i \cdot X_i + \alpha_3 X_i \\ & + \alpha_4 \hat{P}_i^{Ddm} \cdot Z_i + \alpha_5 \hat{P}_i^{Ddm} \cdot X_i + \alpha_6 \hat{P}_i^{Ddm} \cdot Z_i \cdot X_i + \alpha_7 \hat{P}_i^{Ddm} + \epsilon_i. \end{aligned}$$

By including $\hat{P}_i^{Ddm}$ and its interactions with Z_i, X_i, and $Z_i \cdot X_i$, we allow $\hat{P}_i^{Ddm}$ to have differential effects on fertility and child quality between twin and non-twin families across different policy intensities. Given that $\hat{P}_i^{Ddm}$ is de-medianed, coefficients α_1, α_2, and α_3 represent the effects of Z_i, $Z_i \cdot X_i$, and X_i on D_i for people with the median propensity to give the third birth.

We follow the same procedure to generate the four terms of controls for the child-quality equation (Eq. (8)). As our regressions include generated regressors, we use block bootstrap to compute clustered standard errors.

A3.2 Concerns on Twinning as the Instrumental Variable

The Determinants of Twinning

We have made Assumptions 1–3 on twinning. Assumption 3 (monotonicity) naturally holds. We discuss Assumptions 1 and 2 on conditional independence and exclusion.

Table A1 examines the determinants of twinning at the second birth. We regress the twinning indicator on maternal age at the second birth, maternal years of schooling, paternal years of schooling, gender of the first-born child, fines, province-specific linear trends, and fixed effects for province, maternal birth year, and census wave. Across all columns, the R^2 is as low as 0.001; all the coefficients, except one, are statistically insignificant. Consistent with the twin literature, older mothers are more likely to have twins (Rosenzweig and Wolpin, 2000). To safeguard the assumption of conditional independence, we control for maternal age at the second birth in all regressions.

The Concern of Misreported Twins

One concern, as raised by Huang, Lei, and Zhao (2016) (HLZ thereafter), is that the birth-control policy incentivizes parents to misreport regularly spaced siblings as twins to evade the penalty. If twins are misreported, both Assumptions 1 and 2 can fail.

Misreported twins are less of a concern in our study. First, as shown in Table A1, the policy intensity is not correlated with the twinning indicator. Second, parents with first-born daughters have more incentive to misreport twins, but we do not detect a correlation between the gender of the first-born child and the twinning indicator. The differences between our results and those of HLZ may lie in the different data sets: They use the 1982, 1990, 2000, and 2005 censuses, while we only use the 1982 and 1990 censuses. They do not separately conduct the estimation using the sample of censuses 1982 and 1990 only, based on which we derive the estimation sample for our study. To examine the issue of misreported twins, we repeat key estimations of HLZ using censuses 1982 and 1990 only. We find that, using the 1982 and 1990 censuses, key estimates of HLZ become small in magnitude and statistically insignificant.

Following HLZ, we build a data set at the birth level. We take twins as a single observation because twins are in the same birth. We restrict to mothers whose oldest child is not older than 17, and whose children all reside with her. As in our main sample, we restrict to mothers born in 1940–1960 in the 1982 and 1990 censuses. Table A2 summarizes our birth-level data. In comparison with HLZ, our sample includes older cohorts because we use the two early censuses. As the twining rate is rising over time (Figure 1 of HLZ), our older cohorts present a lower twinning rate (0.47%) than those in HLZ (0.58%). Our sample also includes a higher proportion of rural residents (83%) compared with theirs (73%), as China had undergone rapid urbanization. In both our sample and theirs, 93% of births have Han parents, and children are on average eight years old. Our sample size is approximately half of theirs.

To examine the effects of the "One-child" policy on the twinning rate, HLZ exploit within-province and over-time variations in fines for unauthorized births. The main estimation equation is

$$Twin_{ijky} = \beta_0 + \beta_1 Fines_{jy} + \delta_k + \delta_y + \delta_{ky} + \delta_j + X_{ij} + \varepsilon_{ij}, \tag{A3.1}$$

where $Twin_{ijky}$ denotes whether birth i in year y and province j is a twin birth in survey year k; $Fines_{jy}$ is the fine rate in province j one year before the child birth year y; δ_k, δ_y, and δ_{ky} are indicators for year of birth y, survey year k, and their combinations, respectively; δ_j represents province dummies; and X_{ij} is a set of covariates, including dummies for residence type (urban/rural), parents' ethnicity (both Han or either a minority), birth order, mother's education level, and mother's age at childbirth, as well as province-specific linear trends in birth cohorts. Coefficient β_1 captures the effect of the policy intensity, as measured by fines, on the reported twinning rate. They detect $\beta_1 > 0$, suggesting that stricter enforcement of the "One-child" policy causes a rise of man-made twins.

As an alternative strategy, HLZ also exploit differential policy enforcement on Han versus minorities. HLZ estimate

$$\begin{aligned} Twin_{ijky} =& \beta_0 + \beta_2 Post_{y \geq 1980} \times Han_i + \beta_2' Han_i \\ &+ \delta_k + \delta_y + \delta_{ky} + \delta_j + X_{ij} + \varepsilon_{ij}, \end{aligned} \tag{A3.2}$$

where $Post_{y \geq 1980}$ denotes an indicator of whether birth i was in 1980 or later; and Han_i is an indicator for Han ethnicity of both parents. The interpretation of β_2 in Eq. (A3.2) is similar to that of β_1 in Eq. A3.1, because the birth-control policy is stricter for Han than for minorities. HLZ also detect $\beta_2 > 0$.

HLZ then explore the mechanism behind man-made twins. Parents may either "report single children as twins ex post or take fertility drugs ex ante to raise the probability of multiple children in a single birth (HLZ, p. 470)." They argue that, if parents misreport the second- and third-born single children as second-born twins, the registered birthdate of the misreported twins tend to follow that of the third-born singleton. Therefore, if misreporting single children as twins is the main mechanism, the birth spacing between the first birth and the second twin birth should increase after the “One-child” policy. After restricting the sample to the second births, they estimate a third equation,

$$\begin{aligned} BirthGap_i = & \beta_0 + \beta_3 Post_{y\geq 1980} \times Twins_i + \beta_3' Twins_i \\ & + \delta_k + \delta_y + \delta_{ky} + \delta_j + X_{ij} + \varepsilon_{ij}, \end{aligned} \tag{A3.3}$$

where $BirthGap_i$ denotes the observed birth spacing between the current (second) and previous (first) delivery for birth i; and $Twins_i$ denotes whether the second birth is a reported twin birth. Coefficient β_3 on the interaction term reflects how much additional time is needed to give birth to twins than to a single child after the “One-child” policy was implemented. HLZ find $\beta_3 > 0$, and suggest that man-made twins reflect the misreporting of singletons as twins.

Column (1) of Table A3 shows our estimates of Eq. (A3.1), using the 1982 and 1990 censuses. Different from HLZ’s findings, the policy intensity, as measured by fines for unauthorized births, does not appear to affect the incidence of twinning. We find $\beta_1 = 0.021$, which is only one third of the estimated 0.066 in HLZ. Our estimated standard error (0.057) is also larger than the standard error (0.038) in HLZ. Columns (2)–(5) further show the estimates of Eq. (A3.1) using different subsamples. Column (2) presents results for the sample of Han parents, column (3) for the sample in which either parent is a minority, column (4) for Han parents in urban areas, and column (5) for Han parents in rural areas. In columns (2)–(5), the estimates of β_1 are all small and statistically insignificant. We then estimate Eq. (A3.2), and detect a small and statistically insignificant estimate of β_2, as shown in column (6). Table A4 shows estimates of Eq. (A3.3) using the full sample, the Han subsample, and the minority subsample. The estimates of β_3 are statistically insignificant in all columns.

In sum, using 1982 and 1990 censuses only, we are not able to replicate the key results of HLZ. We do not detect a significant rise of misreported twins following a stricter enforcement of the birth-control policy in our sample.

Why does our sample yield different results? The restricted Hukou (household registration) system explains why misreporting regularly spaced siblings as twins was difficult for parents before the 1990s. In 1955, the State Council issued "A Directive to Build a Regular Household Registration System" to regulate population flow in the planned economy. In 1958, the Hukou system was enacted by a national legislation. Under the Hukou system, each household has a Hukou certificate, which records the birth day, gender, and the registration place of every household member. Hukou has two types: agriculture and non-agriculture. Based on Hukou, the local government administers land distribution, school enrollment, medical care, and old-age pensions. Changing the type and location of Hukou was difficult, if not impossible. Moves across localities were greatly restricted by the Hukou system.

The Hukou system gradually relaxed its migration restrictions from the 1990s onwards. In the early 1990s, large coastal cities, such as Shanghai, Shenzhen, and Guangzhou, piloted local registration permits, which allowed children of migrant parents to enroll in local schools. After 1998, a series of Hukou reforms, both at the national and city levels, made it easier to change Hukou type and location, and greatly increased the extent of internal migration in China (Tombe and Zhu, 2019; Wang, Milner, and Scheffel, 2021). The lifting of migration restrictions helps explain why parents are more likely to misreport twins in later periods. Before the 1990s, as the Hukou system strictly restricted moves across localities, local family-planning officers could verify on site if the reported twins were fake. After the lifting of migration restrictions, parents who misreport twins could move to and hide in other localities, and evade the on-site inspection of family-planning officers. This explanation is consistent with Figure 1 of HLZ, which shows that twinning rates were considerably higher in 1990–2005 than those in earlier periods.

The Use of Low-parity Children and High-parity Twins

When estimating Eq. (8), we include first-born children who were born prior to the births of second-born twins. The inclusion of children born prior to the parity of twinning has been widely used in the literature (Black, Devereux, and Salvanes, 2005; Angrist, Lavy, and Schlosser, 2010; Mogstad and Wiswall, 2016; Brinch, Mogstad, and Wiswall, 2017). This practice prevents direct comparison of twin children and singleton children, and mitigates the potential confounder of the birth order effect.

This practice is not free of problems. For example, Rosenzweig and Zhang (2009) show that the inferior endowments of twins compared with singletons induce parents to reallocate family resources toward the first-born child. This resource reallocation reduces the negative effect of twinning-induced fertility increase on the first-born child's education. In addition, the exclusion assumption of twinning may not hold. Bhalotra and Clarke (2019, 2020) show that healthier mothers are more likely to have live twin births, and maternal health status correlates with child quality. These issues may bias estimates of the QQ effects. However, as the rationing income effect mainly predict the difference in QQ effects between the two types of fertility increases, potential biases largely cancel out in the difference.

A3.3 Concerns on Policy Intensity as the Stratifying Variable

In our multivalued-treatment-effect framework, the policy intensity is a stratifying variable as in Hull (2018), not an IV. Thus, our identifying assumption about the policy (Assumption 5) is weaker than the standard assumptions for IVs. We only need to assume that the policy is conditional independent from differences in potential outcomes ($Y_{1i} - Y_{0i}$) for each type of mothers. In our main regressions, as discussed in Section A3.1, we have comprehensively controlled three sets of variables to ensure that Assumption 5 is plausible.

By Assumption 5, the stratifying variable, namely, the policy intensity, enables the identification of QQ effects for type-A and type-B mothers. The policy rations fertility for type-B mothers, but not for type-A mothers; twinning-induced fertility increase is desired for type-B mothers, but undesired for type-A mothers. Referring to Eq. (2) in our theoretical analysis, if price and substitution effects do not vary considerably *between* the two types of mothers, the estimated differential QQ effects reflect the between-type difference in the rationing income effects.[13]

[13] We examine between-type stability of price and substitution effects in Sections 4.4 and 5.

Assumption 5 would fail if the policy correlates with factors that affect price and substitution effects *within* mother type. For example, if a higher policy intensity, net of the controls, correlates with a lower educational cost for type-B mothers only, then the magnitude of the negative price effect declines for type-B mothers when the policy intensity grows. Even if price and substitution effects are similar *between* type-A and type-B mothers in the absence of the policy, a growing policy intensity may induce unequal *within*-type variations in price or substitution effects. Consequently, the estimated $\beta_B - \beta_A$ reflect not only the rationing income effect, but also policy-induced within-type changes in price and substitution effects.[14]

We conduct a robustness analysis by controlling for factors that may simultaneously correlate with the policy intensity and the price and substitution effects. Specifically, we control for variables that determine educational costs and parental preferences.

Educational Costs

We control for three variables on educational costs. First, *Compulsory Education Law* of China, passed in 1986, required each province to implement a system of 9 years of compulsory education. We follow Ma (2019) and generate an indicator variable CEL_i on whether the child was affected by the law, and find that 57% of the children in our sample were affected. We include CEL_i in the vector of control variables and report the result in column (2) of Table A6.

Second, the number of secondary school teachers increases by almost five times from 1970 to 2005 (Appendix Figure A2). We control for the number of secondary school teachers in each province when the child was aged 13, and report the result in column (3).

Third, 18 million urban "educated youths" had been sent down to rural areas in 1962-1979, and their arrival may have improved education in rural areas. We obtain the total number of sent-down youths received by each province from Gu (2009), and divide it by the rural population born in 1962-1979 as a measure of the intensity of sent-down youths. We include interactions of the measure with maternal birth year dummies in the vector of control variables and report the results

[14] If the policy induces within-type changes in price and substitution effects that are the same for the two types of mothers, the estimated $\beta_B - \beta_A$ would remain the same.

in column (4).[15]

Columns (2)–(4) show that our estimation results are robust to factors that affect educational costs.

Parental Preference

We consider two other historical events—the great famine and the cultural revolution—which are believed to affect parental preferences. First, the great famine, which caused approximately 30 million unnatural deaths in 1959-1961, has negative effects on mental health (St Clair et al., 2005). Following Meng, Qian, and Yared (2015), we construct a measure of the severity of the famine by province. We then interact the measure with maternal birth year dummies, and include the interaction terms in the vector of control variables. The level term of the measure is absorbed by province fixed effect.

Second, the conflicts in the cultural revolution lower people's trust on others (Bai and Wu, 2020). We follow Walder (2015) and construct a measure of exposure to the cultural revolution by province. We then interact the measure with maternal birth year dummies, and include the interaction terms in the vector of control variables. Columns (5) and (6) show that our results are robust to the influences of the great famine and the cultural revolution, respectively.

A3.4 Additional Robustness Analyses

Gender Difference

Son preference is still prevalent in China, and quality investment in children is likely to be gender dependent (Qian, 2009; García, 2022). We now discuss the implication of son preference and child gender for the estimated differential QQ effects.

In China, the gender of the first-born child is plausibly random. Data from population censuses (1982, 1990, and 2000) reveal that high sex ratios in China are driven by imbalances in second- and higher-order births, while the sex ratio for first births is stable and falls in the biologically normal

[15] The level term of the measure was absorbed by the province fixed effect.

range (Ebenstein, 2010). On the other hand, parents with first-born sons have lower desired fertility level (Wei and Zhang, 2011).

In our econometric framework, the desired fertility level of type-A mothers is two, which is lower than the desired fertility level of type-B mothers. As mother type is defined by the desired fertility level, type-A mothers may consist of a higher proportion of mothers with first-born sons. Therefore, our estimated differential QQ effects between type-A and type-B mothers are potentially confounded by the difference in child gender.

We conduct a robustness analysis to examine this possibility. We estimate the fertility and child-quality equations by adding the gender of the first-born child interacting with all other independent variables. This method is equivalent to separate estimations using two subsamples classified by child gender. The estimation results show that $\beta_A < 0$, $\beta_B > 0$, and $\beta_B - \beta_A > 0$ for both sons and daughters (Table A7 in the Appendix). Gender differences in the estimates of β_A, β_B, and $\beta_B - \beta_A$ are statistically insignificant. The robustness analysis shows that our estimated differential QQ effects between type-A and type-B mothers are unlikely driven by the difference in child gender.

The "Later, Longer, and Fewer" Campaign

The "One-child" policy is not the only birth-control policy in China. During the early1970s, China implemented the "Later, Longer, and Fewer" (LLF) campaign, which encouraged couples to marry and give birth at older ages, to have longer spacing between births, and to have fewer children. The LLF campaign explains the fertility decline in the 1970s (Wang, 2016; Chen and Huang, 2020; Chen and Fang, 2021).

We use the "One-child" policy rather than the LLF campaign for two reasons. First, the "One-child" policy features "mandatory" reductions in fertility, which is in line with our theory of rationed fertility. Although the LLF campaign also contains some coercive elements, it is believed to be "softer" compared with the "One-child" policy (Whyte, Feng, and Cai, 2015; Wang, 2016; Zhang, 2017). Second, the LLF campaign reduced fertility at high parities. We are not sure which parity of twinning should be used if we exploit the LLF campaign. Moreover, the sample of mothers with three or more children is selective if we use twinning at high birth parities.

To address the concern that the LLF campaign may confound the "One-child" policy, we conduct the following robustness analysis. First, we follow Chen and Fang (2021) to construct a measure of exposure to the LLF campaign:

$$LLF_i = \sum_{a=16}^{45} ProbBirth(a) \cdot I[t + a \geq PolicyYear_p],$$

where $PolicyYear_p$ is the year the LLF campaign was launched in province p; $I[t+a \geq PolicyYear_p]$ is equal to one if a mother born in year t and province p was subject to the LLF campaign at age a, and otherwise zero; $ProbBirth(a)$ is the probability of giving birth at the third parity at age a. We calculate $ProbBirth(a)$ using a sample of mothers born in 1930–1939 in the 1% sample of the 1982 wave of the census. We then include LLF_i in the vector of control variables. We find robust results when controlling for the LLF campaign (column (1) of Table A6).

Fertility Changes at Higher Parities

Our baseline specification considers the fertility increase from two to three. We exploit second-born twins, which breaks the policy and increases fertility to at least three children.

In this part, we consider fertility increases at higher birth parities. We note that the incidence of second-born twins also increases fertility beyond the parity of twinning. For example, the incidence of second-born twins also increases the likelihood that mothers have at least four children (Angrist, Lavy, and Schlosser, 2010). But the three-to-four fertility increase induced by second-born twins does not represent an "exogenous break of rationing," as this three-to-four fertility increase represents continued birth giving at a parity higher than the parity of twinning. The fourth child is almost surely a singleton birth for mothers with second-born twins. To be consistent with our theory, if we explore fertility increases at higher birth parities, we should use twinning at higher birth parities.

Appendix Table A8 reports estimates of Eq. (7) (Panel A) and Eq. (8) (Panel B) using higher-parity twins. Column (1) presents results for the three-to-four fertility increase induced by third-born twins. The sample includes mothers with at least three children. With a 0.48% twinning rate, we obtain 860 pairs of twins at the third birth parity. Unsurprisingly, we have fewer third-born twins (860 pairs) than second-born twins (1057 pairs), because only a selective set of mothers proceed to the third birth. The estimated QQ effects using third-born twins are qualitatively similar to our baseline results using second-born twins. We continue to observe $\beta_A < 0, \beta_B > 0$, and $\beta_B - \beta_A > 0$, with statistical significance at least at the 5% level.

Column (2) presents results for the four-to-five fertility increase induced by fourth-born twins. The sample includes mothers with at least four children. We obtain 377 pairs of fourth-born twins. The sample size decreases sharply, because only a very selective set of mothers managed to proceed to the fourth birth. The estimated $\beta_B - \beta_A$ is close to zero with a very large standard error.

In summary, results on fertility increases at higher birth parities further substantiate our understanding of the data at hand. As most mothers obtain at least two children in rural China, our primary working sample using second-born twins are largely representative of rural China in our study period. Our primary working sample consists of 264,013 mothers with at least two children, among whom 1057 mothers have second-born twins. When we consider the three-to-four fertility increase, the sample size declines modestly. We have 860 mothers with third-born twins out of 179,453 mothers with at least three children. The results remains robust for the three-to-four fertility increase. As we further consider the four-to-five fertility increase, the sample size declines sharply: 377 mothers with fourth-born twins out of 80,413 mothers with at least four children. The results become statistically insignificant in this small selective sample.

A3.5 Heterogeneous Treatment Effects by the Continuous Policy Intensity

As discussed in Section 3.2, the policy intensity is a stratifying variable instead of an IV. Our identification does not require strong assumptions on the variable of the policy intensity, such as the independence and exclusion. Thus, we do not use the MTE framework (Heckman and Vytlacil, 1999, 2005, 2007; Heckman, 2010). However, given that the policy intensity is a continuous variable, this section conducts a heterogeneity-treatment-effect analysis, which is in the style of the MTE analysis. Appendix Section A2.6 derives our econometric model with continuous policy intensity.

Consider a continuous policy that ranges from zero to $\bar{X}$, $X_i \in [0, \bar{X}]$. The definitions of type-A and type-C mothers remain the same. The definition for type-B mothers becomes subtler. A type-B mother, who desired three children, will have two children only when the policy is strong enough. We use θ_i to denote the minimal policy intensity that reduces a type-B mother's fertility from three to two. Let $F_B(\cdot)$ denote the cumulative distribution function of θ_i among type-B mothers. That is, $F_B(x) \equiv \mathbf{Pr}(\theta_i \leq x | i \in B, X_i = x) = \mathbf{Pr}(D_{0i} = 0 | i \in B, X_i = x)$. We use type-$B_x$ mothers to denote type-B mothers with $\theta_i \leq x$.

When the policy intensity equals x, the effect of twinning on fertility is

$$\mathbb{E}[D_i | Z_i = 1, X_i = x] - \mathbb{E}[D_i | Z_i = 0, X_i = x] = P_A + P_{B_x},$$

and the effect of twinning on child quality is

$$\mathbb{E}[Y_i | Z_i = 1, X_i = x] - \mathbb{E}[Y_i | Z_i = 0, X_i = x] = P_A \cdot \beta_A + P_{B_x} \cdot \beta_{B_x},$$

where $\beta_A = \mathbb{E}[Y_{1i} - Y_{0i} | i \in A]$ and $\beta_{B_x} = \mathbb{E}[Y_{1i} - Y_{0i} | i \in B, \theta_i \leq x]$.

In this case, we not only can identify P_A and β_A:

$$\begin{aligned}
P_A &= \mathbb{E}[D_i | Z_i = 1, X_i = 0] - \mathbb{E}[D_i | Z_i = 0, X_i = 0], \\
\beta_A &= \frac{\mathbb{E}[Y_i | Z_i = 1, X_i = 0] - \mathbb{E}[Y_i | Z_i = 0, X_i = 0]}{\mathbb{E}[D_i | Z_i = 1, X_i = 0] - \mathbb{E}[D_i | Z_i = 0, X_i = 0]},
\end{aligned}$$

but also P_{B_x} and β_{B_x} for $x \in [0, \bar{X}]$:

$$\begin{aligned}
P_{B_x} &= (\mathbb{E}[D_i | Z_i = 1, X_i = x] - \mathbb{E}[D_i | Z_i = 0, X_i = x]) - P_A \\
\beta_{B_x} &= \frac{1}{P_{B_x}} (\mathbb{E}[Y_i | Z_i = 1, X_i = x] - \mathbb{E}[Y_i | Z_i = 0, X_i = x] - P_A \cdot \beta_A).
\end{aligned}$$

Thus, we nonparametrically identify P_A, β_A, P_{B_x}, and β_{B_x} for $x \in (0, \bar{X}]$.

We parametrically implement the estimations for P_A, β_A, P_{B_x}, and β_{B_x}, using polynomials in X_i. The fertility equation now becomes

$$D_i = \alpha_0 + \alpha_1 Z_i + \sum_{m=1}^{M} \alpha_{2m} Z_i \cdot X_i^m + \sum_{m=1}^{M} \alpha_{3m} X_i^m + v_i,$$

and the child-quality equation becomes

$$Y_i = \rho_0 + \rho_1 Z_i + \sum_{m=1}^{M} \rho_{2m} Z_i \cdot X_i^m + \sum_{m=1}^{M} \rho_{3m} X_i^m + \varepsilon_i,$$

where M denotes the polynomial order. For any $x \in (0, \bar{X}]$, we have $P_A = \alpha_1$, $\beta_A = \frac{\rho_1}{\alpha_1}$, $P_{B_x} = \sum_{m=1}^{M} \alpha_{2m} \cdot x^m$, and $\beta_{B_x} = \sum_{m=1}^{M} \rho_{2m} \cdot x^m$.

In Figure A4, the solid line denotes estimates of $\beta_{B_x} - \beta_A$ using third-order polynomials for the policy intensity, which is measured by the expected fines for unauthorized births in multiples of local household annual income, ranging from 0 to 1.27. The two dashed lines mark the 90% confidence intervals.

The estimates of $\beta_{B_x} - \beta_A$ are all positive, and the size generally increases with the policy intensity. The finding is consistent with our theory. Because β_A remains constant, the shape of $\beta_{B_x} - \beta_A > 0$ in Figure A4 is driven by β_{B_x}. As x increases, the measure of the policy intensity better reflects the coercive nature of the birth control, and the estimated β_{B_x} more likely captures the positive rationing income effect for the desired fertility increase. Our results remain robust when using fourth- or fifth-order polynomials in the policy intensity (Appendix Figure A5).

A3.6 Parental Responses in Consumption and Labor Supply

The results show that the QQ effect of desired fertility increases is positive. Angrist, Lavy, and Schlosser (2010) and Galor (2012) conjecture that parents may adjust their consumption and labor supply to boost investment in child quality. Appendix A1.6 formulates a model with parental labor supply, and derives the comparative statics of rationed fertility on parental consumption and labor supply, which we now examine empirically.

While the literature has comprehensively examined the effect of fertility on parental labor supply (Rosenzweig and Wolpin, 2000), few studies examine the effect on parental consumption. Data on parental consumption rarely include a large sample of twins. We use the Chinese Child Twins Survey (CCTS), which contains rich information on parental consumption and labor supply. The CCTS was carried out by the Urban Survey Unit of the National Bureau of Statistics in late 2002 and early 2003 in Kunming, the capital city of an underdeveloped province in China. The Urban Survey Unit initially identified 2,300 households with twins between the ages of 7 and 18 from the 2000 population census as the target sample. Of this target sample, 1,694 twin households were successfully interviewed. As a comparison group, 1,693 non-twin households with children in the same age group were also interviewed. We use the rural sample of the CCTS. Appendix Table A9 reports the summary statistics. Using the CCTS, Rosenzweig and Zhang (2009) find that twinning has negative effects on children's education and health, but the magnitude of the effects is modest. We can explain the modest effect by twinning inducing a mixture of desired and undesired fertility increases in the CCTS, since all children in the sample were born after the launch of the "One-child" policy.

We estimate the following equation:

$$Y_j = \theta_0 + \theta_1 Z_j + \mathbf{C}\theta_2 + \epsilon_j, \tag{A3.4}$$

where Y_j is a measure of parental consumption or labor supply in family j; Z_j is a dummy variable indicating twinning at the second birth; $\mathbf{C}$ is a vector of controls, including maternal age at the second birth, parents' years of schooling, age, and age squared; and ϵ_j is the error term.[16] Because the CCTS covers one city only, we are unable to explore variations in the policy intensity, as in our previous specifications. The estimate of θ_1 reflects the effects on parental consumption and labor

[16]Similar to Rosenzweig and Zhang (2009), the estimated coefficient of twinning on having at least three children is 0.98. The birth-control policy strictly enforces two children per mother in rural Kunming.

supply of a mixture of desired and undesired fertility increases.

Panel A of Table A10 presents the effect of twinning on parental consumption. We have information on three categories of paternal consumption—cigarettes, alcohol, and clothing; and two categories of maternal consumption—cosmetics and clothing. In comparison with fathers of non-twins, fathers of twins spend 61% less on clothing. In comparison with mothers of non-twins, mothers of twins spend 47% less on cosmetics and 61% less on clothing. Panels B and C reports the effects of twinning on paternal labor supply and maternal labor supply, respectively. We have four measures of parental labor supply: (i) an indicator variable of being employed, (ii) number of days worked in the last month, (iii) an indicator variable of managing a private business, and (iv) an indicator variable of leaving home for more than 30 days in the last 180 days. All coefficient estimates are positive. Specifically, compared with fathers of non-twins, the probability of setting up a private business for fathers of twins increases by 11 percentage points. In comparison with mothers of non-twins, the probability of setting up a private business for mothers of twins increases by 14 percentage points; and the probability of migration increases by 3 percentage points. Our results suggest that twinning under the "One-child" policy induced parents to work harder and consume less, which may have facilitated the increase in child quality.

A3.7 Data for Structural Estimation

In this subsection, we provide further details on the data and statistics we use in the structural model. First, we estimate the mean and variance of logarithm household income form China Health and Nutrition Survey (CHNS, 1991 and 1993). Specifically, we apply the same sample selection criteria as we did for the Census data, and then use the selected sample to calculate the mean and standard deviation of logarithm household income, which is 9.713 and 0.497 respectively.

Second, we calibrate the parameter γ. We use a two-way fixed-effect regression, $\ln(S_{pc}) = \delta + \gamma Y_{pc} + \epsilon_{pc}$, where Y_{pc} is the middle school attendance rate in 13–17 of cohort c in province p, and S_{pc} is the average completed schooling years of cohort c in province p. We calculate Y_{pc} when our sample cohorts are 13–17 years old, using the same data source as in our main analysis (1982 and 1990 China population censuses). We obtain S_{pc} from the 2000 China population census, when our sample cohorts were at least 23 years old. The estimate of γ is 0.424, suggesting that a ten-percentage-point increase in the probability of middle school attendance is associated with a 4.24% increase in completed schooling years. Because the mean of completed schooling years is 8.7 years, the 4.24% increase represents 0.369 (= 0.0424×8.7) year of completed schooling.

Third, our targeted statistics in the structural model including β_A, β_B, α_A, α_B, the correlation between household income and a dummy variable of whether the household has 3 or more children ($corr(y_i, I_{n_i \geq 3})$), the correlation between household income and quality of children ($corr(y_i, q_i)$), share of expenditure on basic children costs over household income $\frac{\pi_n n}{y}$, share of expenditure on private goods of children over household income $\frac{\pi_{nq} nq}{y}$. The statistics of (β_A, β_B, α_A, α_B) are from our reduced form estimates form the paper. We calculate $corr(y_i, I_{n_i \geq 3})$ using CHNS (1991 and 1993), again applying the same selected sample as our Census data. Then we use the Chinese Household Income Project (CHIP, 1988 and 1995) to calculate $corr(y_i, q_i)$. We measure quality of children as whether the first child of the family attended middle school. Finally, we use CHIP (1988 and 1995) household expenditure data to calculate the share of expenditure on basic children costs over household income $\frac{\pi_n n}{y}$, and share of expenditure on private goods of children over household income $\frac{\pi_{nq} nq}{y}$. The expenditure on basic children costs include food, clothing, durable and daily consumption. The expenditure on children's private goods includes medical and educational expenses. The values of the data moments and model simulated moments are presented in A11.

References

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434):444.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics* 28 (4):773–824.

Bai, Liang and Lingwei Wu. 2020. "Political Movement and Trust Formation: Evidence from the Cultural Revolution (1966-76)." *European Economic Review* 122.

Becker, Gary S. 1991. *A Treatise on the Family*. Cambridge, Massachusetts: Harvard University Press, enlarged e ed.

Becker, Gary S. and H. Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81 (2):S279–S288.

Bhalotra, Sonia and Damian Clarke. 2019. "Twin Birth and Maternal Condition." *The Review of Economics and Statistics* 101 (5):853–864.

———. 2020. "The Twin Instrument: Fertility and Human Capital Investment." *Journal of the European Economic Association* 18 (6):3090–3139.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics* 120 (2):669–700.

———. 2010. "Small Family, Smart Family? Family Size and the IQ Scores of Young Men." *Journal of Human Resources* 45 (1):33–58.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. 2017. "Beyond LATE with a Discrete Instrument." *Journal of Political Economy* 125 (4):985–1039.

Chen, Yi and Hanming Fang. 2021. "The long-term consequences of China's "Later, Longer, Fewer" campaign in old age." *Journal of Development Economics* 151 (March):102664.

Chen, Yi and Yingfei Huang. 2020. "The power of the government: China's Family Planning Leading Group and the fertility decline of the 1970s." *Demographic Research* 42:985–1038.

Ebenstein, Avraham. 2010. "The 'Missing Girls' of China and the Unintended Consequences of the One Child Policy." *Journal of Human Resources* 45 (1):87–115.

Galor, Oded. 2012. "The Demographic Transition: Causes and Consequences." *Cliometrica* 6 (1):1–28.

García, Jorge Luis. 2022. "Pricing Children, Curbing Daughters: Fertility and the Sex-Ratio During China's One-Child Policy." *Journal of Human Resources (forthcoming)* .

Gu, Hongzhang. 2009. *Chinese Educated City Youth: The Whole Story (Zhongguo Zhishi Qingnian ShangShan Xiaxiang Shimo, in Chinese).* China's Daily Publishing House.

Heckman, James J. 2007. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences* 104 (33):13250–13255.

———. 2010. "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature* 48 (2):356–98.

Heckman, James J. and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3):669–738.

Heckman, James J. and Edward J. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences of the United States of America* 96 (8):4730–4734.

———. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments." In *Handbook of Econometrics*, vol. 6, chap. 71. Elsevier, 4875–5143.

Huang, Wei, Xiaoyan Lei, and Yaohui Zhao. 2016. "One-Child Policy and the Rise of Man-Made Twins." *Review of Economics and Statistics* 98 (3):467–476.

Hull, Peter. 2018. "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons." *Woring Paper* .

Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467.

Ma, Mingming. 2019. "Does Children's Education Matter for Parents' Health and Cognition? Evidence from China." *Journal of Health Economics* 66:222–240.

Meng, Xin, Nancy Qian, and Pierre Yared. 2015. "The Institutional Causes of China's Great Famine, 1959–1961." *The Review of Economic Studies* 82 (4):1568–1611.

Mogstad, Magne and Matthew Wiswall. 2016. "Testing the Quantity-Quality Model of Fertility: Estimation using Unrestricted Family Size Models." *Quantitative Economics* 7:157–192.

National Bureau of Statistics. 2009. *China Compendium of Statistics (1949-2008).* Beijing: China Statistics Press.

Neary, J. P. and K. W. S. Roberts. 1980. "The Theory of Household Behaviour under Rationing." *European Economic Review* 13 (1):25–42.

Qian, Nancy. 2009. "Quantity-Quality and the One Child Policy:The Only-Child Disadvantage in School Enrollment in Rural China." *NBER Working Paper* .

Rosenzweig, Mark R. and Kenneth I. Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica* 48 (1):227.

Rosenzweig, Mark R and Kenneth I Wolpin. 2000. "Natural "Natural Experiments" in Economics." *Journal of Economic Literature* 38 (4):827–874.

Rosenzweig, Mark R. and Junsen Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-child" Policy." *Review of Economic Studies* 76 (3):1149–1174.

St Clair, David, Mingqing Xu, Peng Wang, Yaqin Yu, Yourong Fang, Feng Zhang, Xiaoying Zheng, Niufan Gu, Guoyin Feng, Pak Sham, and Lin He. 2005. "Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961." *Journal of the American Medical Association* 294 (5):557–62.

Tobin, James and H. S. Houthakker. 1950. "The Effects of Rationing on Demand Elasticities." *Review of Economic Studies* 18 (3):140–153.

Tombe, Trevor and Xiaodong Zhu. 2019. "Trade, migration, and productivity: A quantitative analysis of China." *American Economic Review* 109 (5):1843–1872.

Walder, Andrew G. 2015. "Rebellion and Repression in China, 1966–1971." *Social Science History* 38 (3-4):513–539.

Wang, Fei. 2016. "Using New Measures to Reassess the Fertility Effects of China's Family Planning Policies Fei Wang." *Working Paper* :1–40.

Wang, Feicheng, Chris Milner, and Juliane Scheffel. 2021. "Labour market reform and firm-level employment adjustment: Evidence from the hukou reform in China." *Journal of Development Economics* 149:102584.

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 (3):511–564.

Whyte, Martin King, Wang Feng, and Yong Cai. 2015. "Challenging myths about China's one-child policy." *China Journal* 74 (74):144–159.

Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and Its Effects on Family Outcomes." *Journal of Economic Perspectives* 31 (1):141–160.

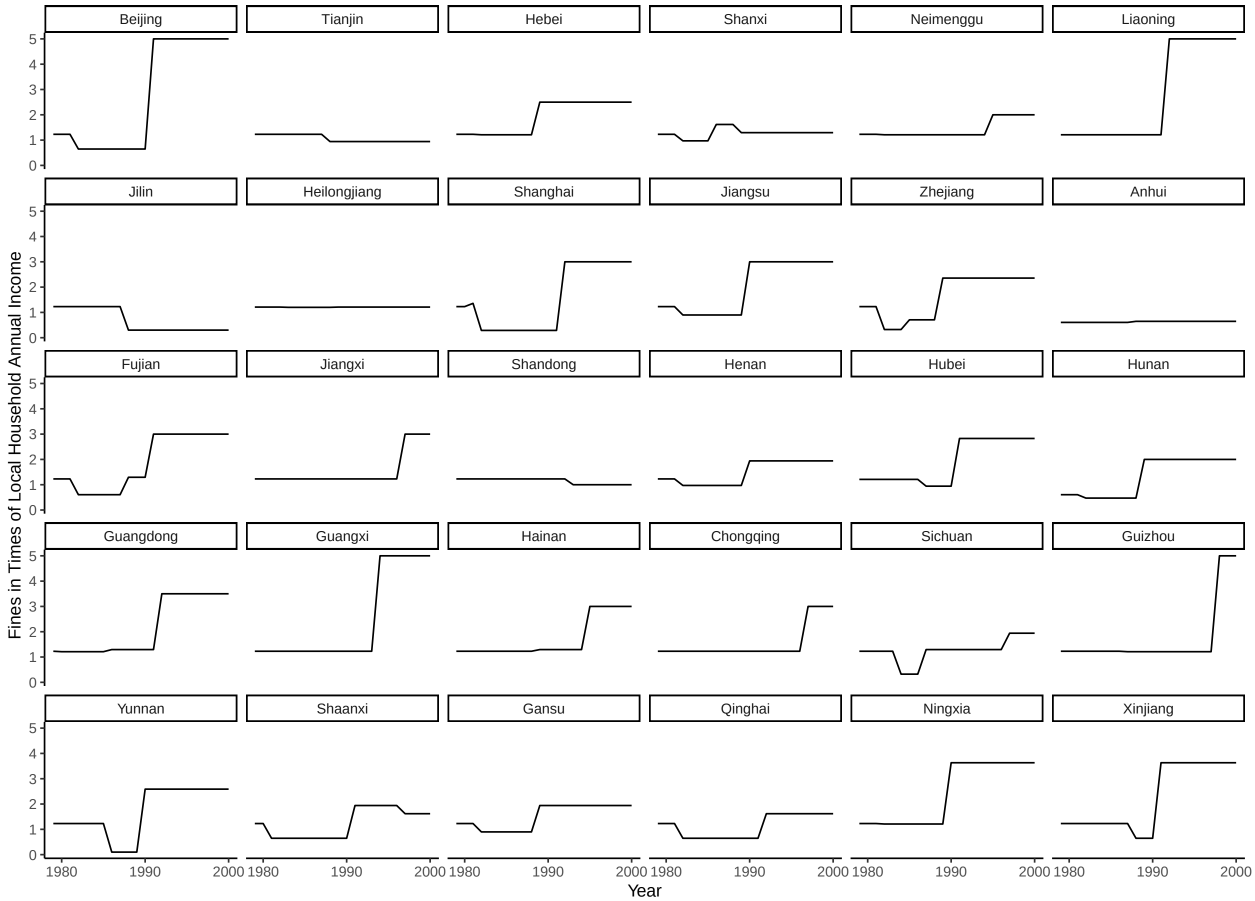

Figure A1: Fines by Province

Notes: This figure shows fines for unauthorized births from 1979 to 2000 in Chinese provinces, excluding Tibet. Fines are measured by multiples of local household annual income. We obtain fines data from Ebenstein (2010).

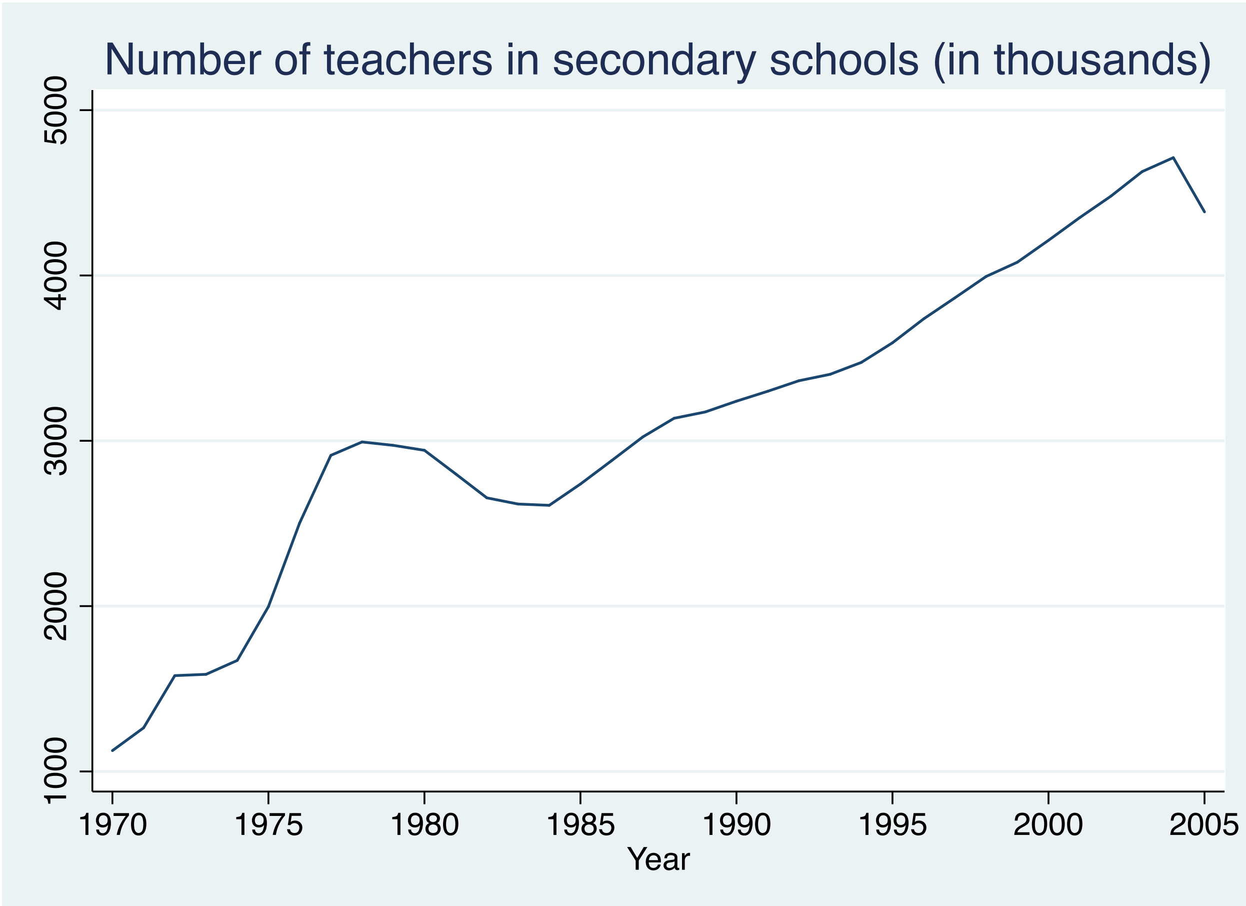

Figure A2: Trend in the number of teachers in secondary schools

Notes: This figure shows the number of teachers in secondary schools from 1970 to 2005 in China. The data source is the National Bureau of Statistics (2009) "China Compendium of Statistics (1949-2008)."

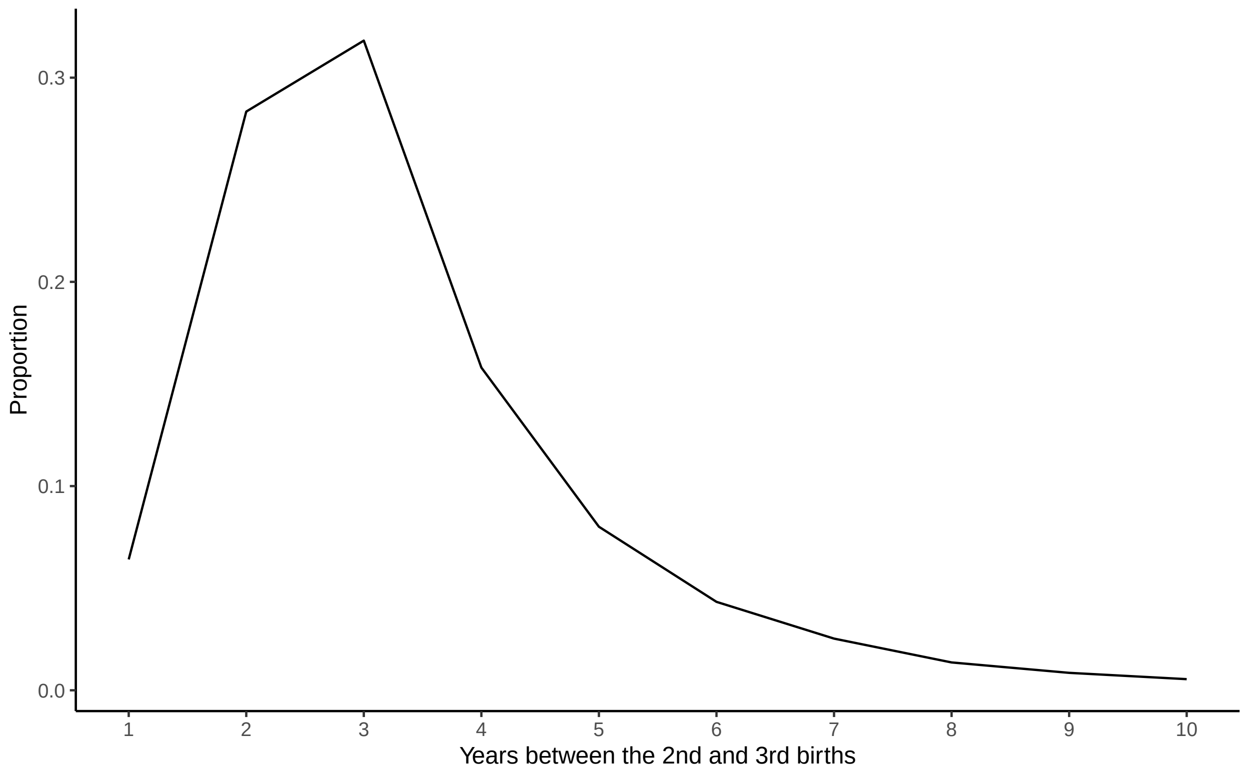

Figure A3: Birth spacing between the second and third births

Notes: This figure shows the distribution of birth spacing in years between the second and third births. We calculate the distribution using a sample of mothers born in 1930–1939 in the 1% sample of the 1982 wave of the China population census.

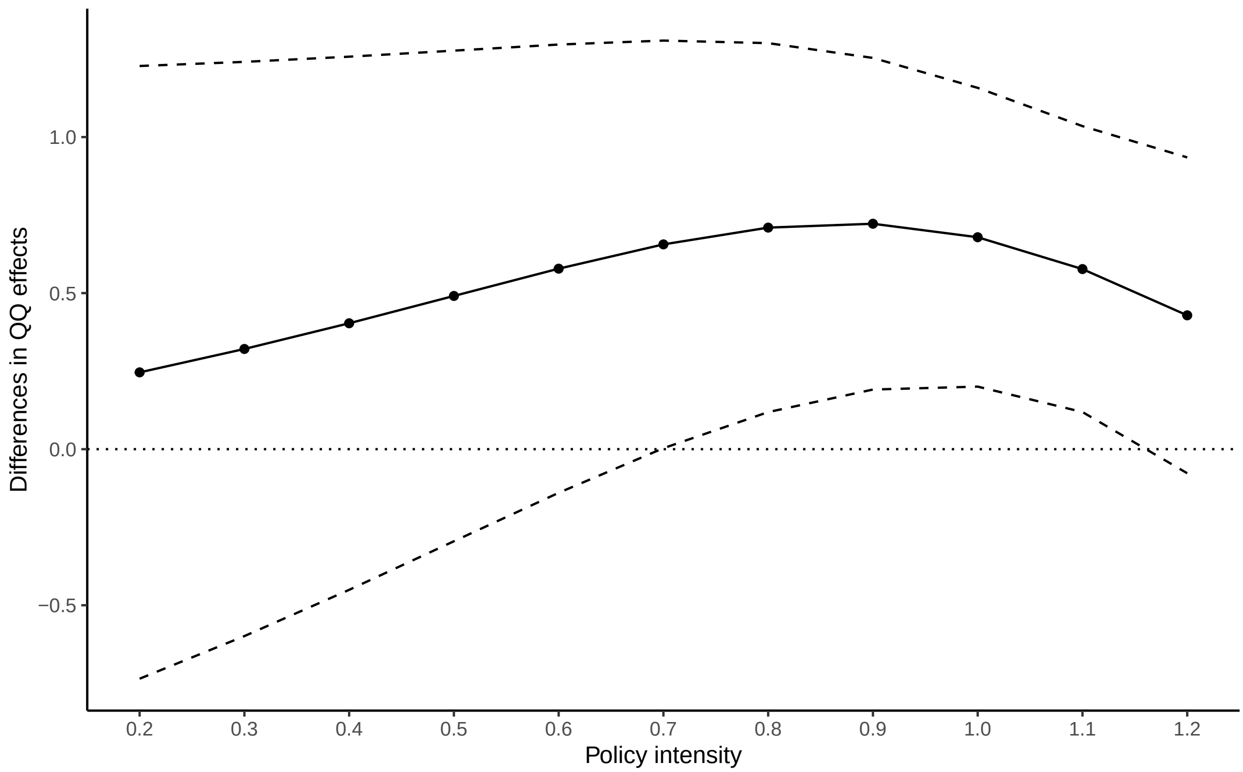

Figure A4: Estimates of differential QQ effects at different levels of policy intensity

Notes: This figure shows estimates of $\beta_{B_x} - \beta_A$ using third-order polynomials in the policy intensity. The continuous policy intensity (X_i) ranges from 0 to 1.27 in multiples of local household annual income. The solid line denotes estimates of $\beta_{B_x} - \beta_A$ for X_i between 0.2 and 1.2 in the interval of 0.1. The two dashed lines mark the 90% confidence intervals.

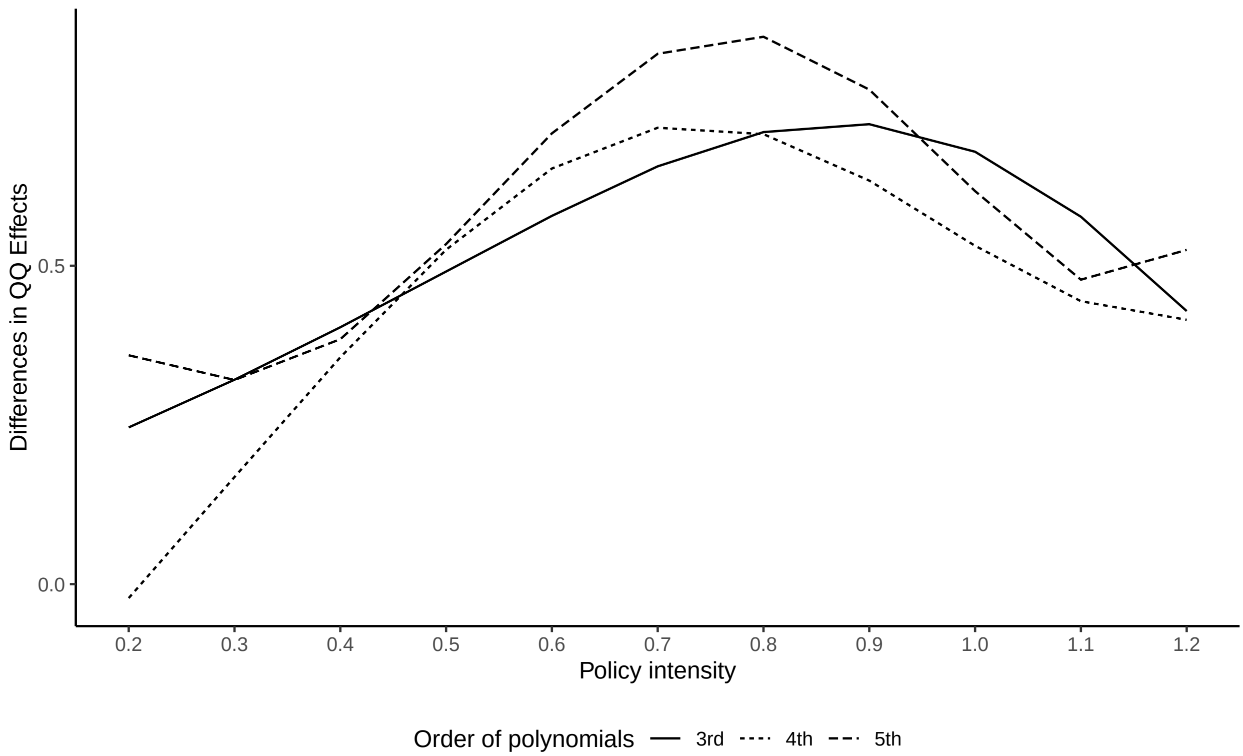

Figure A5: Comparing the use of 3rd, 4th, and 5th orders of polynomials

Notes: This figure shows estimates of $\beta_{B_x} - \beta_A$ using third-, fourth-, and fifth- order polynomials in the policy intensity. The solid line denotes estimates of $\beta_{B_x} - \beta_A$ for X_i between 0.2 and 1.2 in the interval of 0.1. The two dashed lines mark the 90% confidence intervals.

Table A1: Determinants of twinning at the second birth

Dependent variable	Twinning at the second birth					
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal age at the 2nd birth	0.00070***	0.00070***	0.00070***	0.00070***	0.00075***	0.00077***
	(0.00007)	(0.00007)	(0.00008)	(0.00007)	(0.00010)	(0.00011)
Maternal schooling years		0.00006	0.00003	0.00003	0.00003	0.00003
		(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Paternal schooling years			0.00001	0.00001	0.00001	0.00002
			(0.00004)	(0.00004)	(0.00004)	(0.00004)
First-born son				-0.00028	-0.00028	-0.00028
				(0.00023)	(0.00023)	(0.00023)
Policy					-0.00068	-0.00100
					(0.00079)	(0.00086)
Province-specific linear trends						X
Observations	264,013	264,013	235,638	235,638	235,638	235,638
R-squared	0.001	0.001	0.001	0.001	0.001	0.001

Notes: This table examines the determinants of twinning at the second birth. The sample includes mothers born in 1940–1960 with at least two children in the 1982 and 1990 waves of the China population census. "Policy" is the expected fines for third-born children. In all columns, we include fixed effects on province, mother's birth year, and census wave. Column (5) includes province-specific linear trends. Standard errors in parentheses are clustered by province and maternal education level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Summary statistics of the sample for the misreported-twin analysis

Sample	(1) Full	(2) Parents are Han	(3) Either parent is minority
Twinning rate (%)	0.47	0.47	0.43
	(6.85)	(6.87)	(6.52)
Rural (dummy)	0.83	0.82	0.86
	(0.38)	(0.38)	(0.35)
Both parents are Han	0.93		
	(0.26)		
Age in years	7.86	7.89	7.50
	(4.44)	(4.44)	(4.42)
Mother's age at childbirth	25.57	25.53	26.04
	(3.67)	(3.63)	(4.07)
Birth order			
First (dummy)	0.45	0.46	0.39
	(0.50)	(0.50)	(0.49)
Second (dummy)	0.33	0.33	0.32
	(0.47)	(0.47)	(0.46)
Third or above (dummy)	0.22	0.21	0.29
	(0.41)	(0.41)	(0.45)
Observations	3,161,322	2,927,861	233,461

Notes: This table replicates Table A1 in Huang, Lei, and Zhao (2016) using the 1982 and 1990 waves of the China population census. Standard deviations in parentheses.

Table A3: Effect of the policy on the reported birth of twins

Sample	Full	Han	Minority	Urban Han	Rural Han	Full
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Reported twining birth (Yes = 100)					
Fines	0.021	0.023	0.003	-0.026	0.039	
	(0.057)	(0.059)	(0.172)	(0.100)	(0.063)	
Post × Han						0.041
						(0.025)
Observations	3,161,322	2,927,861	233,461	512,547	2,415,314	3,161,322
R-squared	0.001	0.001	0.001	0.001	0.001	0.001

Notes: The data set is derived from the 1982 and 1990 China population censuses. The dependent variable is whether the birth is twins or not (yes = 100). Control variables include fixed effects for province, birth year, survey year, and for the combinations of birth year and survey year. Other covariates include dummies for residence type (urban/rural), parents' ethnicity (both Han or either a minority), birth order, mother's education level, and mother's age at childbirth, as well as province-specific linear trends in birth cohorts. Robust standard errors in parentheses are clustered at the province level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Effect of the policy on the age gap between the first and second births

	(1) Full sample	(2) Parents Han	(3) Either parent Minority
Dependent variable	Age gap between first and second births		
Twinning in the second birth × Post	0.063	0.054	0.140
	(0.039)	(0.043)	(0.162)
Twinning in the second birth (Yes = 1)	0.204***	0.204***	0.219**
	(0.025)	(0.026)	(0.091)
Observations	1,035,323	961,589	73,734
R-squared	0.274	0.276	0.263

Notes: The data set is derived from the 1982 and 1990 China population censuses. The dependent variable is the age gap between the first and second births (in years). The sample is restricted to second births. Control variables include fixed effects for province, birth year, survey year, and for the combinations of birth year and survey year. Other covariates include dummies for residence type (urban/rural), parents' ethnicity (both Han or either a minority), birth order, mother's education level, and mother's age at first childbirth, as well as province-specific linear trends in birth cohorts. Robust standard errors in parentheses are clustered at the province level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Robustness tests: Pre-trends

	(1)	(2)	(3)
Panel A. Dependent variable: three or more children			
Twin	0.212***	0.212***	0.211***
	(0.021)	(0.021)	(0.021)
Policy × Twin	0.120***	0.121***	0.119***
	(0.011)	(0.010)	(0.011)
Policy	-0.120***	-0.121***	-0.119***
	(0.011)	(0.010)	(0.011)
F-statistic	140.83	139.60	133.46
R-squared	0.32	0.32	0.32
Panel B. Dependent variable: middle school attendance			
Twin	-0.033**	-0.033**	-0.033**
	(0.016)	(0.016)	(0.016)
Policy × Twin	0.049**	0.050**	0.049**
	(0.024)	(0.024)	(0.024)
Policy	0.026***	0.026***	0.025***
	(0.006)	(0.006)	(0.006)
β_A	-0.155**	-0.157**	-0.157**
	(0.077)	(0.078)	(0.078)
β_B	0.408*	0.409*	0.412*
	(0.213)	(0.209)	(0.213)
$\beta_B - \beta_A$	0.563**	0.566**	0.569**
	(0.272)	(0.268)	(0.271)
R-squared	0.18	0.18	0.18
Prov-specific quadratic trends	X		
GDP growth		X	
Pop growth			X
Observations	264,013	264,013	264,013

Notes: The sample includes mothers born in 1940–1960 with at least two children in the 1982 and 1990 waves of the China population census. Dependent variables are whether the mother has three or more children (Panel A) and whether the first-born child of the mother has ever attended middle school (Panel B). "Twin" is an indicator variable on whether the mother has second-born twins. "Policy" is the expected fines for third-born children. In all columns, control variables include maternal age at the second birth, maternal education, the child's gender, age, and age squared, fixed effects for province, maternal birth year, and census wave, as well as province-specific linear trends. Columns (1) includes province-specific quadratic trends. Column (2) includes predetermined GDP growth interacted with dummies for maternal birth year and twinning status. Column (3) includes predetermined population growth interacted with dummies of maternal birth year and twinning status. We use block bootstrap with 100 repetitions to obtain standard errors clustered by province and maternal education (in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Robustness tests: LLF, Education Costs, and Parental Preferences

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Dependent variable: three or more children					
Twin	0.212***	0.213***	0.216***	0.211***	0.211***	0.212***
	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)
Policy × Twin	0.121***	0.118***	0.111***	0.121***	0.122***	0.121***
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)
Policy	-0.121***	-0.118***	-0.111***	-0.121***	-0.122***	-0.121***
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)
F-statistic	136.66	132.01	105.02	136.15	137.81	138.70
R-squared	0.32	0.32	0.32	0.32	0.32	0.32
	Panel B. Dependent variable: middle school attendance					
Twin	-0.033**	-0.033**	-0.033**	-0.033**	-0.033**	-0.033**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Policy × Twin	0.050**	0.050**	0.049**	0.049**	0.049**	0.050**
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Policy	0.026***	0.025***	0.023***	0.026***	0.026***	0.026***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
β_A	-0.158**	-0.156**	-0.152**	-0.158**	-0.157**	-0.157**
	(0.078)	(0.077)	(0.076)	(0.078)	(0.078)	(0.078)
β_B	0.411*	0.421*	0.443*	0.409*	0.405*	0.411*
	(0.210)	(0.217)	(0.226)	(0.210)	(0.210)	(0.210)
$\beta_B - \beta_A$	0.568**	0.578**	0.595**	0.567**	0.562**	0.568**
	(0.269)	(0.275)	(0.283)	(0.269)	(0.270)	(0.269)
R-squared	0.18	0.18	0.18	0.18	0.18	0.18
Later, Longer, Fewer	X					
Compulsory education		X				
Secondary school teachers			X			
Send-down ratio				X		
Famine loss					X	
Cultural revolution death						X
Observations	264,013	264,013	264,013	264,013	264,013	264,013

Notes: The sample includes mothers born in 1940–1960 with at least two children in the 1982 and 1990 waves of the China population census. Dependent variables are whether the mother has three or more children (Panel A) and whether the first-born child of the mother has ever attended middle school (Panel B). "Twin" is an indicator variable on whether the mother has second-born twins. "Policy" is the expected fines for third-born children. In all columns, control variables include maternal age at the second birth, maternal education, the child's gender, age, and age squared, fixed effects for province, maternal birth year, and census wave, as well as province-specific linear trends. Column (1) controls for the "Later, Longer, and Fewer" campaign. Column (2) controls for China's Compulsory Education Law. Column (3) controls for the number of teachers in secondary school when the child was aged 13. Column (4) controls for the send-down event. Column (5) controls for the population loss in the great famine. Column (6) controls for deaths during the cultural revolution. We use block bootstrap with 100 repetitions to obtain standard errors clustered by province and maternal education (in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Estimates of the gender-interactive model

Panel A. Estimates of the fertility and child-quality equations		
Dependent variable	Having three or more children	Middle school attendance
	(1)	(2)
Twin	0.251***	-0.045*
	(0.025)	(0.027)
Policy × Twin	0.153***	0.074**
	(0.013)	(0.035)
Policy	-0.153***	0.003
	(0.013)	(0.007)
Female × Twin	-0.077***	0.022
	(0.009)	(0.035)
Female × Policy × Twin	-0.063***	-0.049
	(0.008)	(0.046)
Female × Policy	0.063***	0.053***
	(0.008)	(0.007)
Female	0.077***	-0.146***
	(0.009)	(0.010)
R-squared	0.32	0.18
Observations	264,013	

Panel B. Estimates of the QQ effects		
Child gender	Son	Daughter
	(1)	(2)
β_A	-0.179	-0.130
	(0.114)	(0.120)
β_B	0.487**	0.281
	(0.242)	(0.388)
$\beta_B - \beta_A$	0.665**	0.411
	(0.321)	(0.470)
Gender difference in β_A	-0.049	
	(0.169)	
Gender difference in β_B	0.205	
	(0.436)	
Gender difference in $\beta_B - \beta_A$	0.254	
	(0.531)	

Notes: The sample includes mothers born in 1940–1960 with at least two children in the 1982 and 1990 waves of the China population census. Panel A shows the estimates of the gender-interactive model, where the dependent variables are whether the mother has three or more children (column (1)), and whether the first-born child of the mother has ever attended middle school (column (2)). "Twin" is an indicator variable on whether the mother has second-born twins. "Policy" is the expected fines for third-born children. Control variables include maternal age at the second birth, maternal education, the child's age, and age squared, fixed effects for province, maternal birth year, and census wave, as well as province-specific linear trends. Panel B shows estimates of the QQ effects for sons (column (1)) and daughters (column (2)). We use block bootstrap with 100 repetitions to obtain standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Higher-parity fertility increases

Sample	At least three children	At least four children
Birth parity of twins	Third-born twins	Fourth-born twins
	(1)	(2)
Panel A. Estimates of the fertility equation		
Dependent variable	Four or more children	Five or more children
Twin	0.453***	0.634***
	(0.018)	(0.013)
Policy × Twin	0.185***	0.111***
	(0.014)	(0.014)
Policy	-0.185***	-0.111***
	(0.014)	(0.014)
R-squared	0.24	0.15
Panel B. Estimates of the child-quality equation		
Dependent variable	Middle school attendance	
Twin	-0.040**	-0.038
	(0.019)	(0.033)
Policy × Twin	0.085***	-0.004
	(0.021)	(0.035)
Policy	-0.015**	-0.016**
	(0.007)	(0.007)
β_A	-0.088**	-0.060
	(0.042)	(0.053)
β_B	0.457***	-0.038
	(0.127)	(0.309)
$\beta_B - \beta_A$	0.545***	0.022
	(0.154)	(0.354)
R-squared	0.17	0.17
Number of twin pairs	860	377
Twinning rate	0.48%	0.47%
Observations	179,453	80,413

Notes: Column (1) includes mothers with at least three children, in which "Twin" is an indicator variable on whether the mother has third-born twins; "Policy" is the expected fines for fourth-born children. Column (2) includes mothers with at least four children, in which "Twin" is an indicator variable on whether the mother has fourth-born twins; "Policy" is the expected fines for fifth-born children. Both samples include mothers born in 1940–1960 in the 1982 and 1990 waves of the China population census. Control variables include maternal age at the order of twin birth, maternal education, the child's age, and age squared, fixed effects for province, maternal birth year, and census wave, as well as province-specific linear trends. Panel A shows estimates of the fertility equation. Panel B shows estimates of the child-quality equation, and the QQ effects. We use block bootstrap with 100 repetitions to obtain standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Summary statistics of the Chinese Child Twins Survey

	Non-twin		Twin	
	Mean	S.D.	Mean	S.D.
	(1)	(2)	(3)	(4)
Basic characteristics				
Paternal age	37.20	4.72	40.20	4.74
Maternal age	35.28	4.41	37.99	4.35
Paternal schooling years	8.42	2.65	8.02	2.43
Maternal schooling years	7.36	2.50	6.83	2.39
Family income (¥/year)	10,704.67	8,845.79	11,032.37	9,888.22
Paternal consumption				
Cigarette expense (¥/month)	58.82	65.51	47.49	54.93
Alcohol expense (¥/month)	11.69	15.23	13.15	23.42
Clothing expense (¥/six months)	124.20	177.34	87.77	168.39
Maternal consumption				
Cosmetic expense (¥/six months)	18.12	39.14	10.10	32.93
Clothing expense (¥/six months)	118.55	146.37	82.32	109.33
Paternal labor supply				
Employment (dummy)	0.83	0.37	0.82	0.38
Days worked last month	25.88	4.32	26.38	4.64
Private business (dummy)	0.21	0.41	0.33	0.47
Migration (dummy)	0.10	0.30	0.12	0.32
Maternal labor supply				
Employment (dummy)	0.78	0.42	0.78	0.41
Days worked last month	25.76	4.96	26.53	4.57
Private business (dummy)	0.20	0.40	0.35	0.48
Migration (dummy)	0.02	0.14	0.04	0.20
Observations	364		278	

Notes: ¥ stands for Chinese yuan. "Private business" is an indicator variable that equals one if the father or mother has a private business. "Migration" is an indicator variable that equals one if the father or mother has left home for more than 30 days in the last 180 days.

Table A10: The effect of twinning on parental consumption

	(1)	(2)	(3)	(4)	(5)
	Panel A. Consumption				
Dependent	Paternal consumption			Maternal consumption	
variable	Cigarette	Alcohol	Cloth	Cosmetics	Cloth
Twin	-0.091	0.027	-0.611***	-0.470***	-0.614***
	(0.141)	(0.119)	(0.167)	(0.129)	(0.165)
R-squared	0.02	0.03	0.10	0.10	0.12
Observations	642	642	642	642	642

	Panel B. Paternal labor supply			
Dependent variable	Employment	Days worked last month	Private business	Out home one month
Twin	0.011	0.005	0.114***	0.041
	(0.034)	(0.018)	(0.041)	(0.026)
R-squared	0.02	0.02	0.05	0.02
Observations	642	530	532	642

	Panel C. Maternal labor supply			
Dependent variable	Employment	Days worked last month	Private business	Out home one month
Twin	0.002	0.016	0.142***	0.031**
	(0.036)	(0.020)	(0.041)	(0.016)
R-squared	0.02	0.01	0.07	0.02
Observations	640	498	500	642

Notes: This table presents OLS estimates of the effect of twinning on parental consumption and labor supply using the Chinese Child Twins Survey. Panel A shows results for parental consumption. Consumption variables are in the log scale. We add one to these outcome variables before taking logs to exploit information on zero expenditure. The results are robust if we do not add one—that is, if we use the truncated samples. Panels B and C present results for paternal labor supply and maternal labor supply, respectively. Control variables include maternal age at the second birth, parental years of schooling, age, and age squared. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: Targeted moments

Moments	Data	Model
β_A	-0.157	-0.156
β_B	0.412	0.411
α_A	0.212	0.212
α_B	0.121	0.120
$\frac{\pi_n n}{y}$	0.214	0.212
$\frac{\pi_{nq} nq}{y}$	0.111	0.111
$corr(y_i, I_{n_i \geq 3})$	0.128	0.128
$corr(y_i, q_i)$	0.296	0.296

Notes: This table shows the value of targeted moments from the data and from the model simulation.

Table A12: Sensitivity analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Sensitivity to τ					
Pre-set value of τ	0.05	0.06	0.07	0.08	0.09	0.10
QQ effect heterogeneity explained by QQ^I	78.5%	76.6%	78.1%	79.9%	81.2%	84.4%
	Panel B. Sensitivity to γ					
Pre-set value of γ	0.30	0.35	0.40	0.45	0.50	0.55
QQ effect heterogeneity explained by QQ^I	77.8%	85.0%	88.5%	85.9%	88.8%	91.7%

Notes: This table presents the robustness of the decomposition results derived from structural estimations, by varying the pre-set values of τ and γ. Panel A explores values of τ ranging from 0.05 to 0.10, centered around the baseline calibration of $\tau = 0.075$. Similarly, Panel B examines values of γ ranging from 0.30 to 0.55, centered around the baseline calibration of $\gamma = 0.424$. The last row of each panel reports the proportion of the difference in QQ effects between type-A and type-B mothers (QQ effect heterogeneity) that is explained by the rationing income effects (QQ^I).