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

Rufei Guo* Junjian Yi† Junsen Zhang‡

June 15, 2020

Abstract

We develop a generalized theory of rationed fertility to analyze treatment effect heterogeneity in the child quantity-quality (QQ) tradeoff. 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. Our theory derives a new positive rationing income effect on child quality for desired fertility increases, but a negative rationing income effect for undesired fertility increases. We explore the natural experiment of twin births and China's "One-child" policy to test the differential treatment effects between desired and undesired fertility increases. Consistent with our theoretical predictions, the estimated QQ effect for undesired fertility increases is negative, and the estimated QQ effect for desired fertility increases is positive. Our study highlights the importance of distinguishing between desired and undesired changes when evaluating social programs.

JEL Classification: C31, J13

Key words: Treatment effect heterogeneity; rationing theory; child quantity-quality tradeoff; birth-control policy; twinning

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

†Department of Economics, National University of Singapore. Email: junjian@nus.edu.sg.

‡School of Economics, Zhejiang University; Department of Economics, Chinese University of Hong Kong. Email: jszhang@cuhk.edu.hk.

1 Introduction

Since fertility entered the scope of economic analysis, economists have considered fertility to be a human choice (Becker, 1960). Not infrequently, real-world constraints prevent people from achieving their preferred fertility level. Contraceptives can fail, leading to unwanted births. Infertility may not be curable. Forced sterilization once prevailed in India and China (Panandiker and Umashankar, 1994; Zhang, 2017). In these cases, fertility is "rationed"—that is, fertility is determined by external forces outside of an individual's own control. Rationed fertility is a real-world constraint and has profound economic implications.

We derive a generalized theory of rationed fertility to analyze treatment effect heterogeneity in the child quantity-quality (QQ) tradeoff. Becker and Lewis's (1973) theory of the child QQ trade-off predicts that the average quality of children is lower in larger families than in smaller ones.[1] Estimates of the treatment effect of fertility on child quality, i.e., the QQ effect, fall into a wide range.[2] For example, Rosenzweig and Wolpin (1980), Hanushek (1992), and Rosenzweig and Zhang (2009) find negative QQ effects; in 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.

We use rationing theory to extend the QQ model, and find a new rationing income effect that offers clear predictions on the differential QQ effects between desired and undesired fertility increases. Despite a large body of empirical studies, theoretical analysis of the QQ effect is in its nascency. In the original Becker-Lewis setup, child quantity and quality enter the household budget constraint in a multiplicative manner, so that an increase in child quantity raises the cost of child quality (Becker and Lewis, 1973). If the

[1] We use child quantity, family size, and fertility interchangeably in this paper.
[2] Appendix Table A1 reviews estimates of the QQ effect in the literature.

income elasticity of child quality is larger than the income elasticity of child quantity, an increase in income lowers parental demand for child quantity and raises demand for child quality, which generates a negative correlation between child quantity and quality. One is unable to 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, child quantity is treated as a parameter, as in the rationing theory of Tobin and Houthakker (1950). When 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. It is negative if child quantity and quality are net substitutes, and positive if they are net complements.

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. Although the sign of the QQ effect is not unambiguously determined, our theory predicts that the QQ effect for desired fertility increases is less negative than that for undesired fertility increases.

We test our theoretical prediction on differential QQ effects between desired and undesired fertility increases. For brevity, we assume a pool of three types of mothers with either two or three children. With neither a birth-control policy nor twinning at the second birth, all mothers achieve their optimal fertility levels: Type-A mothers have two children, and type-B and type-C mothers have three children. With the birth-control policy, which targets two children per family, type-A mothers still have two children, type-B mothers comply with the policy and have two children, and type-C mothers do not comply with the policy and have three children.

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). Although our observed treatment is binary—mothers have either two or three children—it represents two types of fertility increases: desired or undesired.[3] We show that desired and undesired fertility increases can have different effects on child quality.

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. To estimate QQ effects for the two types of fertility increases, we use multiple waves of the Chinese population census. We use fines for unauthorized births compiled by Ebenstein (2010) to measure the intensity of the "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. When fines are zero, twinning at the second birth increases the proportion of mothers with three children by 23 percentage points. Compliers without the policy are mothers who experience undesired fertility increases. When fines equal the local household annual income, twinning at the second birth increases the proportion of mothers with three children by

[3] Lin et al. (2019) use the framework of Heckman and Pinto (2018) with unordered monotonicity to study heterogeneous QQ effects. Using multivalued instruments, they identify the effects of planned and unplanned expansions in family size on outcomes for existing children for the same unobserved type of households. 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 econometrics on estimation of the QQ effect. Second, we use the twin instrument and China's "One-child" policy, while they use the same-sex instrument.

an additional 12 percentage points. The incremental 12 percentage points of compliers are mothers who experience desired fertility increases.

The estimated QQ effect for undesired fertility increases is negative and statistically significant. The probability of middle school attendance falls by 15 percentage points when fertility is forced to increase from two to three. In contrast, the estimated QQ effect for desired fertility increases is positive and statistically significant. The probability of middle school attendance rises by 42 percentage points when twinning nullifies the policy restriction and shifts fertility from two to three. The positive QQ effect for desired fertility increases is consistent with Pantano (2016), who exploits a desired fertility reduction induced by access to birth control pills in the US, and finds that children born to families with less unwanted fertility commit less crime. The difference in QQ effects between desired and undesired fertility increases is positive and statistically significant. These results are consistent with our theoretical predictions that the rationing income effect is negative for undesired fertility increases, but positive for desired fertility increases.

Since we focus on difference in QQ effects between the two types of fertility increases, potential bias associated with twinning in the levels of QQ effects is largely cancelled out by the difference in QQ effects (Rosenzweig and Zhang, 2009; Huang, Lei, and Zhao, 2016; Bhalotra and Clarke, 2019). We note that our identification relies on the assumption that the policy does not affect average QQ effects for different types of fertility changes, which is weaker than the conventional assumptions of independence, exclusion, and monotonicity for instruments. We also conduct a series of sensitivity analyses to check whether the timing and amount of fines may capture other socioeconomic factors, which may correlate with fertility or child quality. Our results remain robust in these checks. We also find that the rationing income effect is stronger for sons and for children from disadvantaged families.

Although our study focuses on the child QQ tradeoff, rationing theory, and the distinction between desired and undesired changes, has broader implications 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 instrumental variables (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. For illustration, 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.

Second, when we evaluate the treatment effects of a population policy, it is important to know whether the policy is compulsory or voluntary. A voluntary population policy, such as distribution of knowledge about contraception, helps parents achieve their desired fertility level. A compulsory population policy may induce desired fertility changes for some people, but also undesired changes for others. The treatment effects can differ substantially between the two types of changes.

The differential treatment effects between desired and undesired changes also have major implications for program evaluation based on 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 involves compulsory or mandatory policies, which induce both desired and undesired changes.

Finally, the composition of compliers with the same instrument differs with or without a binding birth-control policy. In our setup, without the birth-control policy, only type-A mothers have two children in the absence of twinning. With the policy, both type-A and type-B mothers have two children in the absence of twinning. Twinning raises fertility for type-A mothers only without the policy, but raises fertility for both type-A and type-B mothers with the policy. The composition of compliers who experience desired or undesired changes varies across institutions. To fully understand treatment effect heterogeneity across societies, we need to understand who complies and why, which depends on the preferences, technology, budget, information, and institution (Heckman, 2001; Deaton, 2010a,b).

The rest of the paper is organized as follows. Section 2 formulates the theory of rationed fertility; Section 3 derives the empirical strategy; Section 4 describes the institutional background, data, and regression specification; Section 5 presents the estimation results; and Section 6 concludes.

2 Theory

We extend Becker and Lewis (1973) and Rosenzweig and Wolpin (1980) to derive 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 new rationing income effect, which predicts differential QQ effects for different types of fertility increases. We also conduct a simulation to show that the rationing income effect can drive the QQ effect in a surprising way.

2.1 The 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),

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

where y is the monetary income of the family. The price of child quality, i.e., the cost of increasing q by one unit, is

$$p_q = \pi_q + \pi_{nq} n,$$

where π_q represents the costs of "public goods" or "family goods," such as "some aspects of training in the home and the 'handing down' of some clothing" (Becker and Lewis, 1973, p. S283); $\pi_{nq} n$ 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 $p_n = \pi_n + \pi_{nq} q$, 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 cost of private goods, which increases in child quality; and 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).

Because n is part of the price of q, a decline in n caused by a rise in π_n would raise 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 intuitively seem appealing, because they may inform population control policies that target fertility. The estimates, however, 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) appear to be 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):

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

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 costs of all choice variables: $v_n = p_n$, $v_q = p_q$, and $v_s = p_s$.

2.2 The 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

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

P2 differs from P1, since 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[4] and p_q^* is the price of child quality when child quantity is rationed such that $p_q^* = \pi_{nq}\bar{n} + \pi_q$.[5]

2.3 Generalized Comparative Statics of the Child Quantity-quality Trade-off

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 in non-optimal fertility levels. We find a rationing income effect, which has not been identified in the literature. The rationing income effect can explain heterogeneous effects of fertility on child quality for different types of fertility increases.

[4] 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 $p_q^* = \pi_{nq}\bar{n} + \pi_q$, $p_n^* = \pi_n$, $p_s^* = \pi_s$.

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

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

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

where q^* is the Walrasian demand for child quality;[6] 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. Eq. (2) is a generalized version of Eq. (1).[7]

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

The second term arises because $\bar{n}$, as part of the price of child quality, directly enters the utility function. The second term 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 the second term 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

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

[7] 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.

A1). The second component can be further decomposed as

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

where $q^{*c}(p^*, u)$ and $n^{*c}(p^*, u)$ are the Hicksian demand functions for child quantity and quality.[8] 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, 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. 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.

Consider a one-unit increase in $\bar{n}$ that is financed by a reduction in s. The one-unit

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

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.

The third term, 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 the third term 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:

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

Figure 1 illustrates that a fertility increase when $\bar{n} < n^o$ and a fertility reduction when $\bar{n} > n^o$ are desired; a fertility increase when $\bar{n} \geq n^o$ and a fertility reduction when $\bar{n} \leq n^o$ are undesired. 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 sign of the QQ effect in Eq. (2) is not unambiguously determined. The price effect is negative. The substitute effect depends on parental preference. The rationing income effect is determined by the type of fertility change. Mogstad and Wiswall (2016) suggest that most empirical studies neglect the substitution effect and interpret the QQ effect using the negative price effect. They emphasize the role of the substitution effect in interpreting heterogeneous QQ effects. However, they note that the substitution effect hardly predicts the sign of the QQ effect a priori, because the substitution effect hinges on parental preference, about which we have little information. By contrast, 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. The QQ effect for desired fertility changes is less negative than the QQ effect for undesired fertility changes, and could even reverse the usual negative QQ effect.

2.4 A Parametric Simulation

We simulate a parametric model of P2 to better understand heterogeneous QQ effects. As in Mogstad and Wiswall (2016), we adopt a nested constant elasticity of substitution (CES) form for parental utility,

$$U(\bar{n}, q, s) = U_1^{\theta} s^{1-\theta},$$

where

$$U_1 = (\alpha \bar{n}^{\rho} + (1-\alpha) q^{\rho})^{\frac{1}{\rho}},$$

where $\theta \in (0,1)$, $\alpha \in (0,1)$, and $\rho \in (-\infty, 1)$. U_1 is a CES aggregate of rationed child quantity $\bar{n}$ and child quality q. The elasticity of substitution between $\bar{n}$ and q is given by $\sigma = \frac{1}{1-\rho}$. The budget constraint is $\pi_{nq}\bar{n}q + \pi_n \bar{n} + \pi_q q + p_s s \leq y$. We set $\theta = 0.5$, $\alpha = 0.5$, $\pi_{nq} = 1$, $\pi_q = 0$, $\pi_n = 0$, $p_s = 1$, and $y = 10$. We further set $\rho = -3$ such that $\bar{n}$ and q are net complements; the elasticity of substitution between $\bar{n}$ and q equals 0.25. The substitution effect is positive. Under this parametric setting, the optimal fertility level n^o is 1.83 children in P1 when fertility is not rationed.

Figure 2a plots optimal child quality in P2 when fertility is rationed from one to five. The figure shows an inverted U-shaped relationship between q and $\bar{n}$. When $\bar{n} < n^o$, q rises with $\bar{n}$; by contrast, when $\bar{n} > n^o$, q declines with $\bar{n}$; and q reaches the peak at $\bar{n} = n^o$.

The inverted U-shaped relationship is consistent with the simulation result of Mogstad and Wiswall (2016, p. 161). They emphasize that the substitution effect, which was first observed by Rosenzweig and Wolpin (1980) but received little attention in the empirical literature, can explain the inverted U-shaped relationship. However, Eq. (1), proposed by Rosenzweig and Wolpin (1980), is inappropriate for interpreting the inverted U-shaped relationship in Figure 2a. Eq. (1) applies to the special case of $\bar{n} = n^o$, but Figure 2a shows the changes in $\bar{n}$ in the range of one to five.

Eq. (2) provides the comparative statics of child quantity on child quality at any possible value of $\bar{n}$, and offers the appropriate framework for interpreting the inverted U-shaped relationship between q and $\bar{n}$ in Figure 2a. Based on Eq. (2), we simulate the QQ effect and decompose it into three components. The solid line in Figure 2b shows that the QQ effect is positive when $\bar{n} < n^o$ and negative when $\bar{n} > n^o$. The price effect (dashed line) is negative; its magnitude declines with $\bar{n}$ but remains sizable for large $\bar{n}$. The substitution effect (line with triangles) is positive; its magnitude declines with $\bar{n}$ and approaches zero for large $\bar{n}$. The rationing income effect (line with squares) is positive for desired fertility increases ($\bar{n} < n^o$) and negative for undesired fertility increases ($\bar{n} > n^o$).

Figure 2b shows that the rationing income effect appears to drive the QQ effect. The negative price effect and the positive substitution effect largely cancel out. The parameterization merely illustrates the possibility that the rationing income effect can easily become the dominating force in the QQ effect. The importance of the rationing income effect in reality requires a rigorous empirical investigation.

3 Identification

Our empirical analysis focuses on testing differential treatment effects between desired and undesired fertility increases, as implied by the generalized theory of rationed fertility. Because of the rationing income effect, the effect on child quality of a desired increase in fertility is less negative than that of an undesired increase. To achieve this end, we need two different types of fertility increases, which enable us to estimate the QQ effects for desired and undesired fertility increases. The difference between QQ effects informs the rationing income effect.

3.1 Distinguishing between 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). In particular, 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 (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-

[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.

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 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, 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.[10] 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).[11]

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

$$D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_i \cdot X_i + \alpha_3 X_i + v_i, \quad (3)$$

where D_i is an indicator equal to one if mother i has three children and zero otherwise, and v_i represents idiosyncratic fertility preference shocks.[12] 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 $X_i = 1$,

[10] 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.

[11] 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 fertility from three to two, which represents an undesired decrease.

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

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.[13]

3.2 Identify QQ Effects for Desired and Undesired Fertility Increases

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.

To interpret the coefficients in Eq. (4), we adopt the framework from the literature on the treatment effect. 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 otherwise $D_i = 0$). 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 otherwise $Z_i = 0$). 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.* $Y_i(d, 0) = Y_i(d, 1) = Y_i(d) = Y_{di}\Big|_{X_i}, \forall d \in \{0, 1\}$

[13] 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, $\alpha_3 = -P_B$.

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 4.5. 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 mother.

Under Assumptions 1–5, 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, and $\beta_B = \mathbb{E}[Y_{1i} - Y_{0i}|i \in B]$ 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 mother, 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.

Our theory derives a negative rationing income effect for type-A mothers and a positive rationing income effect for type-B mothers. If the difference in the rationing income effect dominates the differences in price and substitution effects, we predict $\beta_A < \beta_B$. That is, the effect of twinning-induced fertility increase on child quality becomes less negative under the policy. Our empirical analysis tests the prediction $\beta_B - \beta_A > 0$.

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:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 D_i \cdot X_i + \gamma_3 X_i + \epsilon_i, \tag{7}$$

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, $\gamma_1 = \beta_A$. 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).

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).[14] However, we do not use the framework for the marginal treatment effect. Instead, we develop the current identification framework for three reasons. First, our identification does not impose the restrictive assumptions on the birth-control policy. 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 Background, Data, and Specification

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.

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 most 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,

[14] 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.

are exempted from the policy.

Figures 3a 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 3b plots completed fertility for urban Han mothers. Different from Figure 3a, 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. 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 families with or without second-born twins. We discuss potential identification concerns regarding the specification in Section 4.5.

4.2 Data

Our primary data source is the 1% samples of the 1982 and 1990 waves of China's population censuses.[15] 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

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

sample covers 11,835,947 individuals from 3,152,818 households. The census includes information at both household and individual level. 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.

1. 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.

2. 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).

3. Drop mothers in Tibet, because the population-control policy in Tibet differs from that in the rest of China.

4. Drop mothers with first-born twins.

5. 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 restrictions, we obtain a sample of 264,013 mothers that will be used in the analysis below.

4.3 Regression Specifications

Empirical implementation of the fertility equation (Eq. (3)) is specified as

$$D_i = \alpha_0 + \alpha_1 Twin_i + \alpha_2 Twin_i \cdot Fines_i + \alpha_3 Fines_i + \mathbf{C_i}\alpha_4 + \epsilon_i, \tag{8}$$

and the empirical implementation of the child-quality equation (Eq. (4)) is specified as

$$Y_i = \rho_0 + \rho_1 Twin_i + \rho_2 Twin_i \cdot Fines_i + \rho_3 Fines_i + \mathbf{C_i}\rho_4 + \varepsilon_i, \tag{9}$$

where D_i is an indicator variable on whether mother i has three or more children; Y_i is an indicator variable on whether the first-born child has ever attended middle school; $Twin_i$ is an indicator on whether mother i has second-born twins; and $Fines_i$ are fines for unauthorized births, which measures the intensity of the "One-child" policy, as discussed below. The variable $Twin_i$ is the empirical counterpart of Z_i, and $Fines_i$ are the empirical counterpart of X_i. Although X_i in Section 3 is a dichotomous variable, Appendix A2.6 shows that interpretation of the estimation coefficients under a continuous policy is similar to that under a discrete policy. The vector $\mathbf{C_i}$ includes an estimated propensity score and its interactions with $Twin_i$ and $Fines_i$, which we discuss below. Bootstrapped standard errors are clustered by province and maternal education.

As discussed in Section 3, α_1 captures the proportion of type-A mothers who experienced undesired fertility increases; α_2 captures the proportion of type-B mothers who experienced desired fertility increases; $\beta_A = \frac{\rho_1}{\alpha_1}$ is the QQ effect for undesired fertility increases; and $\beta_B = \frac{\rho_2}{\alpha_2}$ is the QQ effect for desired fertility increases. We expect that $\beta_B - \beta_A > 0$.

4.4 Variable Construction and Summary Statistics

Child Quality (Y_i)

Child quality is measured by the first-born child's middle school attendance. Figure 4a 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 4b 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).

Fines (X_i)

The "One-child" policy is an umbrella term for a package of birth-control programs, including sterilization, fines on unauthorized births, bonuses for policy compliance, and several exemptions. Policy enforcement was carried out by provincial governments, and the strength of the policy varied over time (Zhang, 2017). We use fines for unauthorized births to measure across-province and over-time enforcement of the policy. Ebenstein (2010) was the first to compile fines on unauthorized births as multiples of local household annual income over time. We plot the over-time variation of fines by province in

Appendix Figure A2. Huang (2017) discusses the fines data in detail.

We construct a measure of a mother's fines for her third birth as the empirical counterpart of X_i. This 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 5 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 variable of fines is defined as

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

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

Control Variables (C_i)

The vector of control variables includes maternal age at the second birth, maternal education, the first-born child's gender, age, and age squared, fixed effects for province, maternal birth year, and census wave, as well as province-specific linear trends. As Eqs. (8) and (9) include the interactions between $Twin_i$ and $Fines_i$, we should include the interactions of $Twin_i$ with the vector of control variables, and the interactions of $Fines_i$ with the vector of control variables.

Because of the large number of fixed effects, we use a parsimonious specification to deal with the control variables. Take Eq. (8) as an example. In the first step, we generate a summary index of the control variables. We regress D_i on $Fines_i$ and the control variables using the sample of non-twin households. Using the estimated coefficients, we obtain the predicted D_i when $Fines_i = 0$. We de-median the predicted D_i to generate a summary index of the control variables. In the second step, we separately interact the index with $Twin_i$ and $Fines_i$. Our final control variables include the index and the two interaction terms. Because the index is de-medianed, the coefficient α_1 (α_3) represents the effect of $Twin_i$ ($Fines_i$) on D_i when the index equals its median value. We follow the same procedures to generate three control variables for Eq. (9). As our regressions include generated regressors, we use block bootstrap to compute clustered standard errors.

4.5 Potential Concerns

We now discuss three potential concerns with our empirical analysis: the first-born children sample, quasi-experimental variations by twinning, and the measure of the policy.

First-born Children

We only use first-born children to estimate Eq. (9). 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. However, 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. This problem is less of a concern in our analysis, as the effect of resource reallocation in response to child endowment largely cancels out in the difference of QQ effects between two types of fertility increases.

Quasi-experimental Variations by Twinning

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

Table 3 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 coefficients except one are statistically insignificant. Consistent with the twin literature, older mothers are more likely to have twins (Rosenzweig and Wolpin, 2000). We control for maternal age at the second birth in all regressions.

One concern, as raised by Huang, Lei, and Zhao (2016), is that the birth-control policy incentivizes parents to report regularly spaced siblings as twins to evade the penalty. Fake twins are less of a concern in our study. First, as shown in Table 3, fines are not correlated with the twinning indicator. Second, parents with first-born daughters have more incentive to report fake twins, but we do not detect a correlation between the gender of the first-born child and the twinning indicator. The difference between our result and those of Huang, Lei, and Zhao (2016) may lie in the different data sets: They use the 2000 and 2005 waves of the census, while we use the 1982 and 1990 waves.

We then discuss the exclusion assumption of twinning. The literature finds two channels through which twinning directly affects child quality. First, as discussed above, Rosenzweig and Zhang (2009) consider intrahousehold reallocation based on child endowment. Second, Bhalotra and Clarke (2019) 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 we focus on the difference in QQ effects between the two types of fertility increases, potential biases largely cancel out in the difference.

The Birth-control Policy

We consider the birth-control policy as a dichotomous variable in Section 3. In the empirical analysis, we use a continuous variable of $Fines_i$ to measure the policy. Appendix

A2.6 shows that interpretation of the coefficients in Eqs. (8) and (9) is similar to that of the coefficients in Eqs. (3) and (4).

Assumption 5 requires that the fines do not affect QQ effects for either type of fertility increases. We will conduct a series of robustness analyses to check whether our results are sensitive to this assumption.

5 Results

We present the estimates of Eqs. (8) and (9) and the estimated QQ effects for desired and undesired fertility increases. The results are robust to potential confounding effects of differential pre-policy trends, an alternative birth-control program, public policies on education, and several historical events. We conduct a subsample analysis by child gender and maternal education. We also find that twinning induces parents to work harder and consume less.

5.1 Baseline Results

Panel A of Table 4 reports estimates of the fertility equation (Eq. (8)). Column (1) shows that the estimated coefficient on twinning (α_1) is positive and statistically significant. Without the policy ($Fines_i = 0$), twinning increases the proportion of mothers with at least three children by 23 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 ($Twin_i = 0$), the proportion of mothers with at least three children decreases by 12 percentage points when $Fines_i$ increase by one. $Fines_i$ 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 $Fines_i = 0$, when $Fines_i = 1$ 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.

Panel B of Table 4 reports estimates of the child-quality equation (Eq. (9)). Column (1) shows that the estimated coefficient before twinning (ρ_1) is negative and statistically significant, which implies a negative QQ effect for undesired fertility increases. Based on Eq. (5), the estimated QQ effect for undesired fertility increases (β_A) is -0.15 (= -0.033/0.229). The probability of attending middle school decreases by 15 percentage points when fertility is forced to increase from two to three by twinning.

The estimated coefficient on the interaction term (ρ_2) is positive and statistically significant, which implies a positive QQ effect for desired fertility increases. Based on Eq. (6), the estimated QQ effect for desired fertility increases (β_B) is 0.42 (= 0.051/0.122). The probability of attending middle school increases by 42 percentage points when twinning circumvents the policy restriction and shifts fertility from two to the desired level of three.

The QQ effect is positive for desired fertility increases and negative for undesired fertility increases. The difference between the QQ effects is statistically significant. Eq. (2) shows that the QQ effect is a combination of the price effect, substitution effect, and rationing income effect. The price effect is always negative. The substitution effect depends on parental preference, on which we have no information a priori. Consistent with the estimated QQ effects, the rationing income effect is positive for desired fertility increases and negative for undesired fertility increases. The rationing income effect can explain the heterogeneous treatment effects.

5.2 Robustness

We now discuss the robustness of our baseline results. A major concern is that the timing and amount of the fines may capture other socioeconomic factors, which may correlate with fertility or child quality.

Pre-policy Trends

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, the coefficient on the fines may capture the influence of predetermined provincial characteristics. We adopt three specifications to address this concern. First, we include province-specific quadratic trends as additional control variables. The results are reported in column (2) of Table 4. Second, we include the interaction terms between predetermined provincial economic growth rate and maternal birth cohort dummies.[16] The specification allows predetermined economic growth rate to have different effects on fertility and child quality for different cohort of mothers. The results are reported in column (3). Third, we include the interaction terms between the predetermined provincial population growth rate and maternal birth cohort dummies in column (4). The signs and magnitude of our estimates remain largely unchanged in all three alternative specifications.

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 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, 2018; Chen and Fang, 2018).

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

[16] The interaction terms are, $\sum_{t=1940}^{1960} GDPgrowth_p \cdot \mathbf{I}[MoBirthYear_i = t]$, where $GDPgrowth_p$ is the average provincial GDP growth rate in years 1970–1975, and $MoBirthYear_i$ is mother's birth year.

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 highly 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 (2018) 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. Column (1) of Table 5 shows that our estimation results remain robust.

Middle School Attendance

We model the child's middle school attendance as a family choice from the demand side. The child's middle school attendance also depends on educational policies and factors from the supply side, which may also correlate with fines. The first is the *Compulsory Education Law* of China. The law, 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 5.

Second, the number of secondary school teachers increases by almost five times from 1970 to 2005 (Appendix Figure A3). We include the number of secondary school teachers in each province when the child was aged 13 in the vector of control variables, 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 in column (4). Columns (2)–(4) show that our estimation results are robust to educational policies and factors from the supply side.

Historical Events

Finally, we consider two other historical events—the great famine and the cultural revolution—which may correlate with fertility, child education, and fines. First, the great famine, which caused approximately 30 million unnatural deaths in 1959-1961, has had long-run consequences in China (Almond et al., 2010). 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. Second, the cultural revolution has had long-term impacts on economic development (Bai and Wu, 2019, 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.

5.3 A Subsample Analysis

We now investigate QQ effects by child gender and maternal education. We first split the sample by child gender and estimate Eqs. (8) and (9). Columns (1) and (2) of Table 6 report results for the samples of mothers with first-born sons and daughters, respectively. The estimate of α_1 is bigger in column (1) than in column (2), indicating that the effect of twinning on fertility is smaller for mothers with first-born daughters without the policy. This is because the probability of having three children is higher for mothers with first-born daughters. The absolute value of the estimate of α_3 in column (1) is bigger than in column (2). Fines reduce the probability of having three children more for mothers with first-born sons. Correspondingly, the estimate of α_2 in column (2) is bigger than in column (2). The estimates of β_A are negative for both samples, but they are statistically insignificant. Splitting the sample reduces the statistical power. The estimate of β_B is bigger in column (1) than in column (2), implying that the QQ effect of desired fertility increases is bigger for first-born sons than for first-born daughters. The results suggest that the rationing income effect is larger for sons than daughters.

Columns (3) and (4) report the results for mothers who have not completed primary school education and mothers with at least primary school education, respectively. The estimate of α_1 is smaller in column (3) than in column (4), indicating that the effect of twinning on fertility is smaller for less educated mothers, for whom the probability of having three children is higher. The estimate of β_B is bigger in column (3) than in column (4), implying that the QQ effect of desired fertility increases is bigger for children of less educated mothers. The results suggest that the rationing income effect is larger for children born to less educated mothers.

5.4 Parental Responses

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 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 A2 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:

$$Y_j = \theta_0 + \theta_1 Z_j + \mathbf{C}\theta_2 + \epsilon_j, \tag{11}$$

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; **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.[17] Because the CCTS covers one city only, we are unable to explore variations

[17] Similar to Rosenzweig and Zhang (2009), the estimated coefficient of twinning on having at least three

in the "One-child" policy, as in our previous specifications. The estimate of θ_1 reflects the effects on parental consumption and labor supply of a mixture of desired and undesired fertility increases.

Table 7 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. Compared with fathers of non-twins, fathers of twins spend 61% less on clothing. Compared with mothers of non-twins, mothers of twins spend 47% less on cosmetics and 61% less on clothing.

Table 8 reports the effects of twinning on parental labor supply. 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; compared 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 the child quality.

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

children is 0.98. The birth-control policy strictly enforces two children per mother in rural Kunming.

quality, an undesired fertility increase has a negative rationing income effect. The theory predicts that QQ effects differ between the two types of fertility increases.

Our empirical analysis tests differential treatment effects between desired and undesired fertility increases. We explore the natural experiment of twin births and China's "One-child" policy. We find that the estimated QQ effect for undesired fertility increases is negative; in contrast, the estimated QQ effect for desired fertility increases is positive. The difference in QQ effects between desired and undesired fertility increases is positive. The results are consistent with our theoretical predictions.

Differential treatment effects between desired and undesired changes have major implications for program evaluation. A voluntary policy mainly induces desired changes; a compulsory policy can induce desired changes for some people and undesired changes for others. It is important to know whether the policy is compulsory or voluntary. As participation in most RCTs is voluntary, one should exercise caution when extrapolating IV estimates based on RCTs to contexts that involve compulsory policies.

The same instrument induces different groups of compliers across different societies. Estimating a particular treatment effect is only a starting point. To understand which policy works in what context, we need to investigate who complies and why. Here comes economic theory.

References

Almond, Douglas, Lena Edlund, Hongbin Li, and Junsen Zhang. 2010. "Long-term Effects of the 1959-1961 China Famine: Manland China and Hong Kong." In *The Economic Consequences of Demographic Change in East Asia*, edited by Takatoshi Ito and Andrew K. Rose. Chicago: University of Chicago Press.

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. 2019. "Political Conflict and Development Dynamics: Economic Legacies of the Cultural Revolution." *Working Paper* .

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

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.

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

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. 2018. "The Long-Term Consequences of Having Fewer Children in Old Age: Evidence from China's Later, Longer, Fewer Campaign." *NBER Working Paper Series* .

Chen, Yi and Yingfei Huang. 2018. "The Power of the Government: China's Family Planning Leading. Group and the Fertility Decline since 1970." *GLO Discussion Paper* .

Deaton, Angus. 2010a. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2):424–455.

———. 2010b. "Understanding the Mechanisms of Economic Development." *Journal of Economic Perspectives* 24 (3):3–16.

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.

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.

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

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

Heckman, James J. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109 (4):673–748.

———. 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. 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. 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.

Heckman, JamesJ. and JamesR. Walker. 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 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.

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* .

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

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

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.

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.

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* .

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.

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

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.

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.

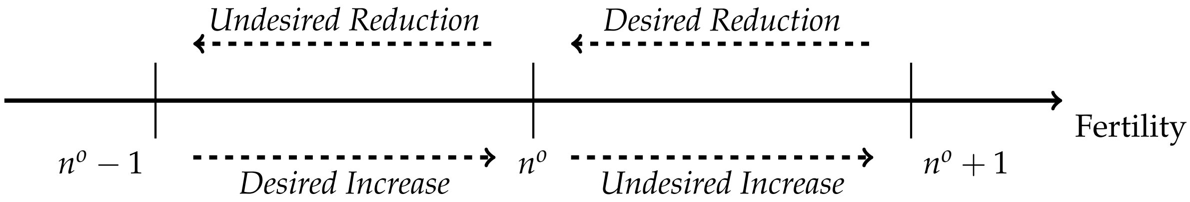

Figure 1: Undesired versus desired fertility changes

Note: n^o is the optimal fertility level in the solution of the utility maximization problem in Problem (P1) when fertility is not rationed.

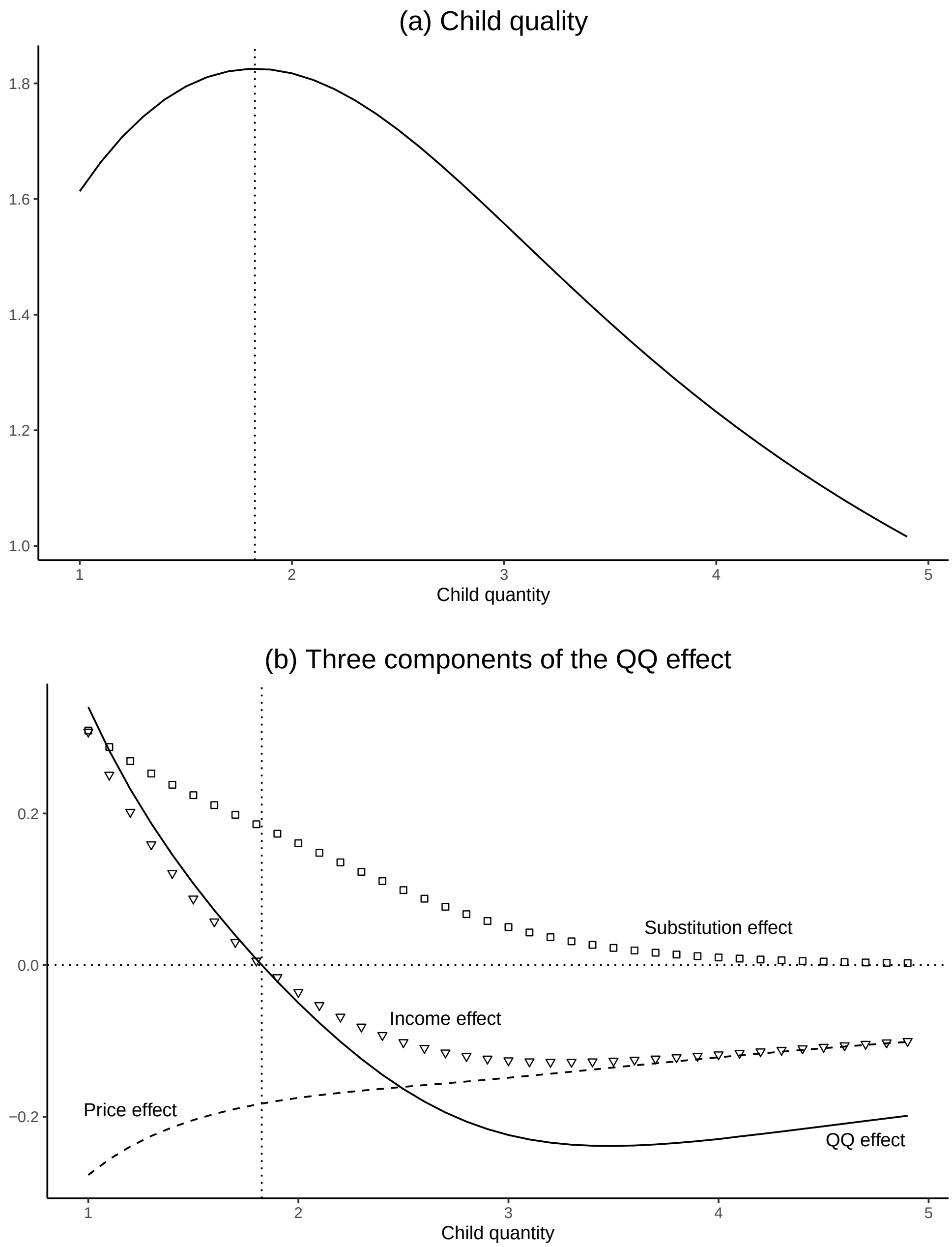

Figure 2: Simulation results for child quality

Notes: The x-axis is the rationed fertility level. In subfigure (a), the y-axis shows optimal child quality q in P2. In subfigure (b), the y-axis shows the semi-elasticity of the effect of rationed fertility on child quality ($\frac{\partial q/q}{\partial \bar{n}}$, solid line) and its three components: price effect (dashed line), substitution effect (line in squares), and income effect (line in triangles). Using a semi-elasticity form, the results are insensitive to the unit of measurement of child quality. Parental utility is in the form of nested CES, $U(n,q,s) = U_1^{\theta} s^{1-\theta}$, where $U_1 = (\alpha n^{\rho} + (1-\alpha) q^{\rho})^{\frac{1}{\rho}}$. The budget constraint is $\pi_{nq} nq + \pi_n n + \pi_q q + \pi_s s \leq y$. We set $\theta = 0.5$, $\alpha = 0.5$, $\rho = -3$, $\pi_{nq} = 1$, $\pi_q = 0$, $\pi_n = 0$, $\pi_s = 1$, and $y = 10$. The optimal fertility in P1 when fertility is not rationed, n^o, is 1.83, which is denoted by the vertical dotted line.

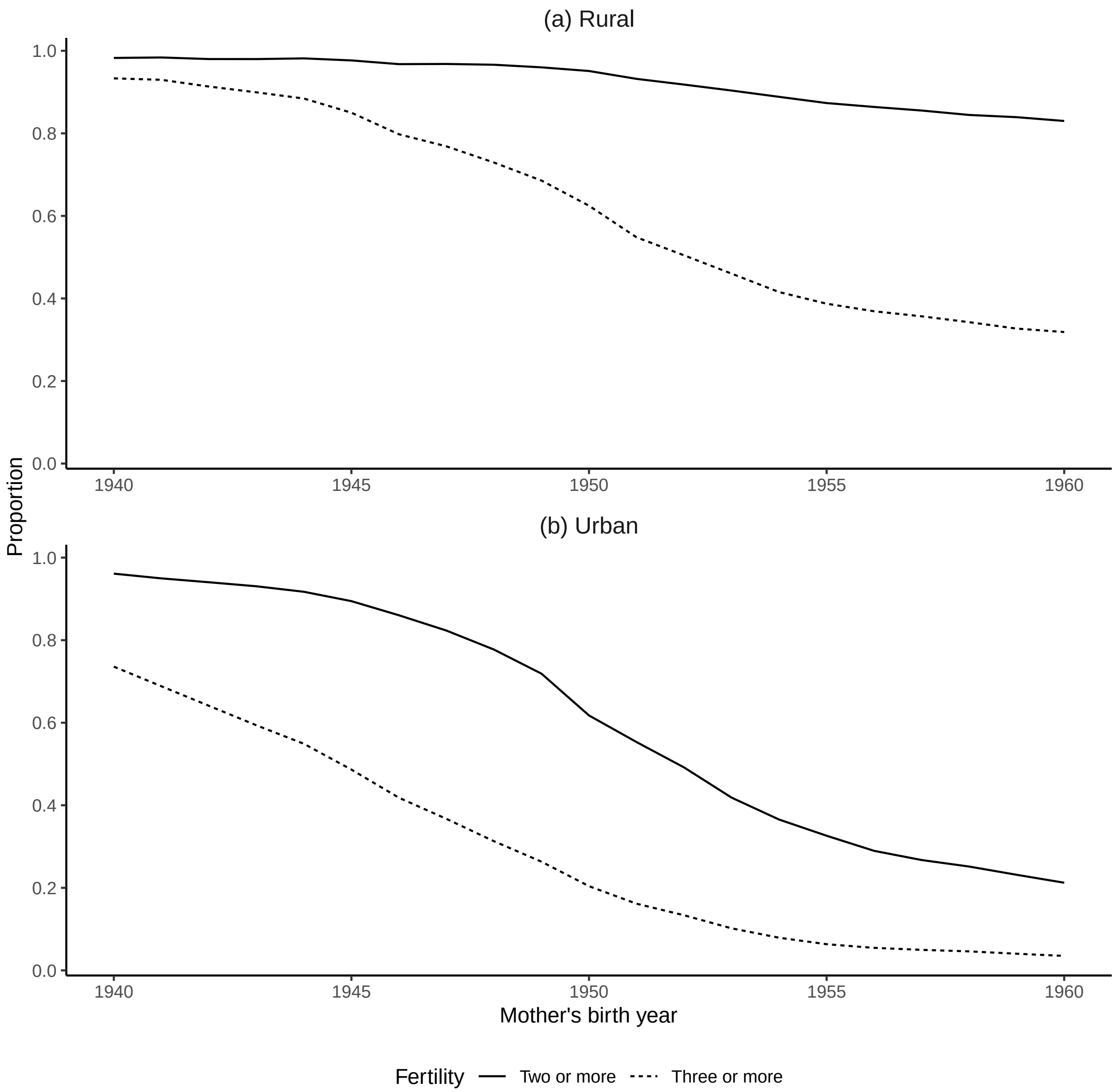

Figure 3: Trends in fertility

Notes: This figure shows the probability of a mother having at least two, three, or four children by the mother's birth year. Subfigure (a) includes rural mothers, and subfigure (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.

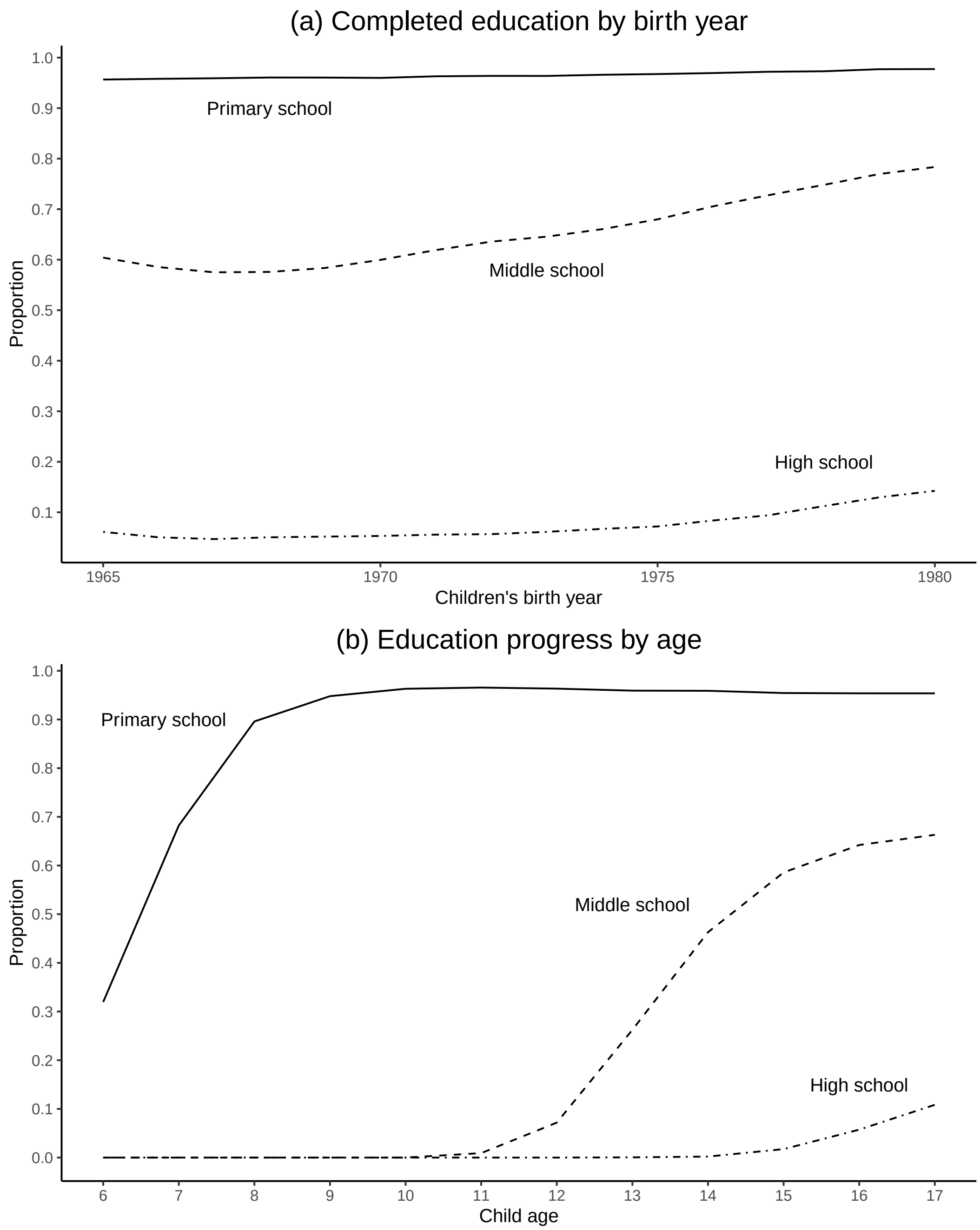

Figure 4: Children's education attainments

Notes: Subfigure (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. Subfigure (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.

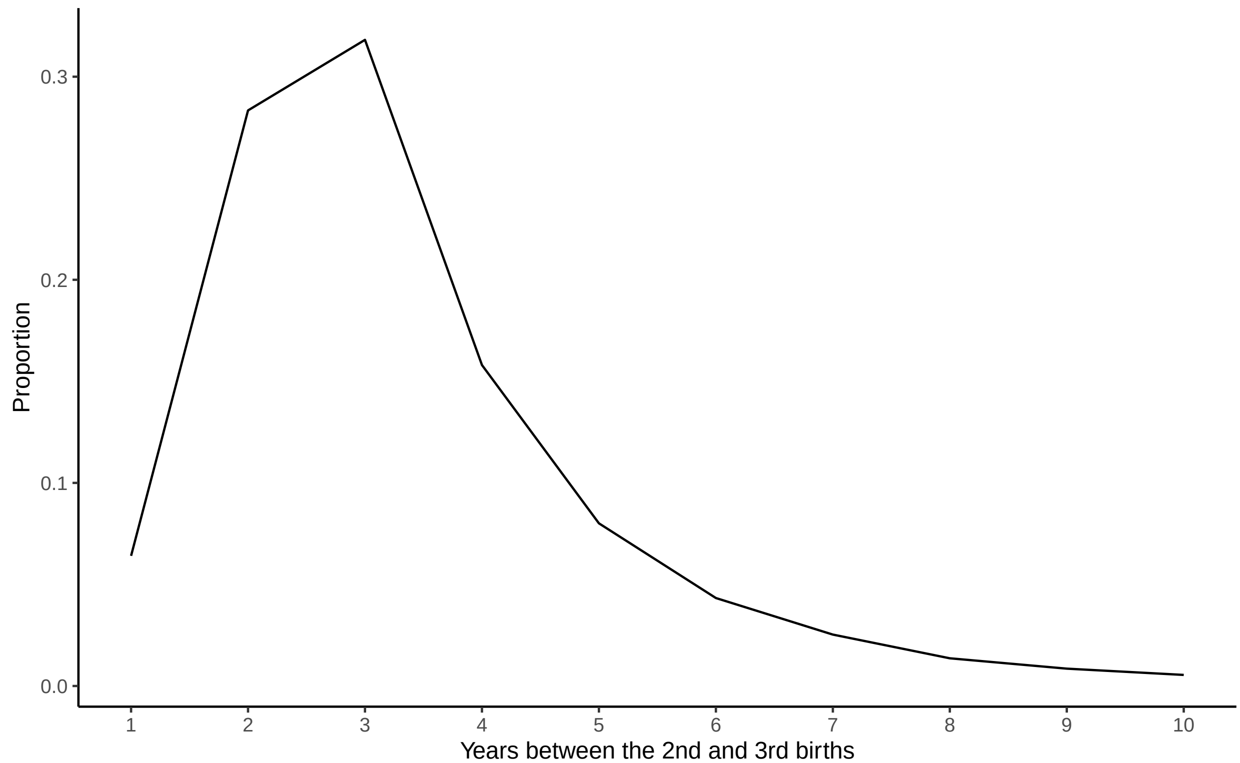

Figure 5: Birth spacing between the second and third births

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.

Table 1: Realized fertility for three types of mothers

	Fertility						Proportion
Policy	$X_i = 0$			$X_i = 1$			
Twinining	$Z_i = 0$		$Z_i = 1$	$Z_i = 0$		$Z_i = 1$	
	(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						
Fines for the third birth	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: 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 years of schooling		0.00006	0.00003	0.00003	0.00003	0.00003
		(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Paternal years of schooling			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)
Fines					-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. "Fines" are the 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 4: Baseline estimates

	(1)	(2)	(3)	(4)
Panel A. Dependent variable: three or more children				
Twin	0.229***	0.230***	0.228***	0.228***
	(0.021)	(0.021)	(0.021)	(0.021)
Fines × Twin	0.122***	0.121***	0.123***	0.121***
	(0.011)	(0.011)	(0.011)	(0.011)
Fines	-0.121***	-0.120***	-0.121***	-0.119***
	(0.011)	(0.011)	(0.010)	(0.011)
F-statistic	183.97	188.25	189.31	180.83
R-squared	0.32	0.32	0.32	0.32
Panel B. Dependent variable: middle school attendance				
Twin	-0.033**	-0.033**	-0.033**	-0.033**
	(0.016)	(0.016)	(0.016)	(0.016)
Fines × Twin	0.051**	0.050**	0.051**	0.051**
	(0.024)	(0.024)	(0.024)	(0.024)
Fines	0.026***	0.025***	0.026***	0.025***
	(0.006)	(0.006)	(0.006)	(0.006)
β_A	-0.146**	-0.144**	-0.147**	-0.147**
	(0.072)	(0.072)	(0.072)	(0.072)
β_B	0.419**	0.416**	0.418**	0.421**
	(0.206)	(0.207)	(0.203)	(0.208)
$\beta_B - \beta_A$	0.566**	0.560**	0.564**	0.568**
	(0.260)	(0.261)	(0.258)	(0.261)
R-squared	0.18	0.18	0.18	0.18
Prov-specific quadratic trends		X		
GDP growth			X	
Pop growth				X
Observations	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. "Fines" are 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 (2) includes province-specific quadratic trends. Column (3) includes predetermined GDP growth interacted with dummies for maternal birth year and twinning status. Column (4) 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 5: Robustness tests

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Dependent variable: three or more children					
Twin	0.229***	0.230***	0.233***	0.228***	0.228***	0.228***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Fines × Twin	0.122***	0.119***	0.112***	0.122***	0.123***	0.122***
	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.011)
Fines	-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	185.16	177.93	139.57	184.56	184.82	185.64
R-squared	0.32	0.32	0.32	0.32	0.32	0.32
	Panel B. Dependent variable: middle school attendance					
Twin	-0.034**	-0.033**	-0.033**	-0.033**	-0.033**	-0.033**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Fines × Twin	0.051**	0.051**	0.050**	0.051**	0.051**	0.051**
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Fines	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	-0.147**	-0.146**	-0.142**	-0.147**	-0.146**	-0.147**
	(0.072)	(0.071)	(0.070)	(0.072)	(0.072)	(0.072)
β_B	0.418**	0.429**	0.451**	0.416**	0.413**	0.421**
	(0.205)	(0.211)	(0.221)	(0.204)	(0.206)	(0.206)
$\beta_B - \beta_A$	0.565**	0.575**	0.593**	0.563**	0.559**	0.568**
	(0.259)	(0.264)	(0.273)	(0.259)	(0.261)	(0.260)
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. "Fines" are 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 6: Heterogeneous effects by child gender and maternal education

Subsample	By child gender		By maternal education	
	Son	Daugther	Low	High
	(1)	(2)	(3)	(4)
Panel A. Dependent variable: three or more children				
Twin	0.266***	0.184***	0.159***	0.291***
	(0.024)	(0.019)	(0.012)	(0.015)
Fines × Twin	0.141***	0.099***	0.124***	0.116***
	(0.012)	(0.012)	(0.013)	(0.010)
Fines	-0.141***	-0.098***	-0.109***	-0.123***
	(0.012)	(0.011)	(0.010)	(0.010)
F-statistic	201.21	133.27	364.87	302.76
R-squared	0.34	0.28	0.28	0.32
Panel B. Dependent variable: middle school attendance				
Twin	-0.039	-0.026	-0.039	-0.029
	(0.026)	(0.021)	(0.030)	(0.032)
Fines × Twin	0.082**	0.030	0.080**	0.041
	(0.036)	(0.032)	(0.041)	(0.035)
Fines	0.022***	0.031***	0.046***	0.020***
	(0.007)	(0.007)	(0.010)	(0.006)
β_A	-0.148	-0.142	-0.247	-0.101
	(0.103)	(0.113)	(0.195)	(0.110)
β_B	0.579**	0.308	0.644*	0.356
	(0.275)	(0.342)	(0.350)	(0.311)
$\beta_B - \beta_A$	0.727**	0.450	0.891*	0.457
	(0.345)	(0.423)	(0.491)	(0.407)
R-squared	0.18	0.16	0.14	0.17
Observations	133,126	130,887	109,048	154,965

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. The 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. "Fines" are the 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) and (2) show the results for sons and daughters, respectively. Column (3) shows the results for mothers with below-primary school education, and column (4) for mothers with primary school education or above. We use block bootstrap with 100 repetitions to obtain standard errors in parentheses. In columns (1) and (2), standard errors are clustered by province and maternal education. In columns (3) and (4), standard errors are clustered by province and maternal cohort. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: The effect of twinning on parental consumption

Dependent variable	Paternal consumption			Maternal consumption	
	Cigarette	Alcohol	Cloth	Cosmetics	Cloth
	(1)	(2)	(3)	(4)	(5)
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

Notes: This table presents OLS estimates of the effect of twinning on parental consumption using the Chinese Child Twins Survey. Outcome variables are in 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. 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 8: The effect of twinning on parental labor supply

Dependent variable	Employment	Days worked last month	Private business	Migration
	(1)	(2)	(3)	(4)
	Panel A: Paternal labor supply			
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 B: Maternal labor supply			
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 twinning on parental labor supply using the Chinese Child Twins Survey. Panels A and B 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.