Secular Fertility Declines Hinder Long-Run Economic Growth

Huang, Kaixing

School of Economics, Nankai University

3 November 2018

Secular Fertility Declines Hinder Long-Run Economic Growth

HUANG KAIXING

Current version: April, 2021
Last version: November, 2018

Declining fertility is among the most salient features of global demography. By examining the lagged effects of fertility on the economic growth of 164 countries over the last half-century, this study found that the effect of a fertility decline lasts for more than three decades and that the long-run average effect is strongly negative for most countries. This finding was confirmed by using the plausibly exogenous fertility declines from the global family planning campaign since the mid-1960s. Within-country evidence from China's one-child policy also confirmed this finding. Therefore, secular fertility declines represent a strong force driving down global economic growth.

Keywords: Secular fertility declines, economic growth, birth control
JEL: J13, O47, N30

Corresponding information: School of Economics, Nankai University (Kaixing.huang@nankai.edu.cn)

1. Introduction

Declining fertility is among the most salient features of global demography. As presented in Figure 1, the world total fertility rate (TFR) dramatically declined from 4.98 children per woman in 1960 to 2.41 children per woman in 2018. Although global fertility decline was mainly driven by middle- and low-income countries, fertility in high-income countries was also nearly halved during this period.

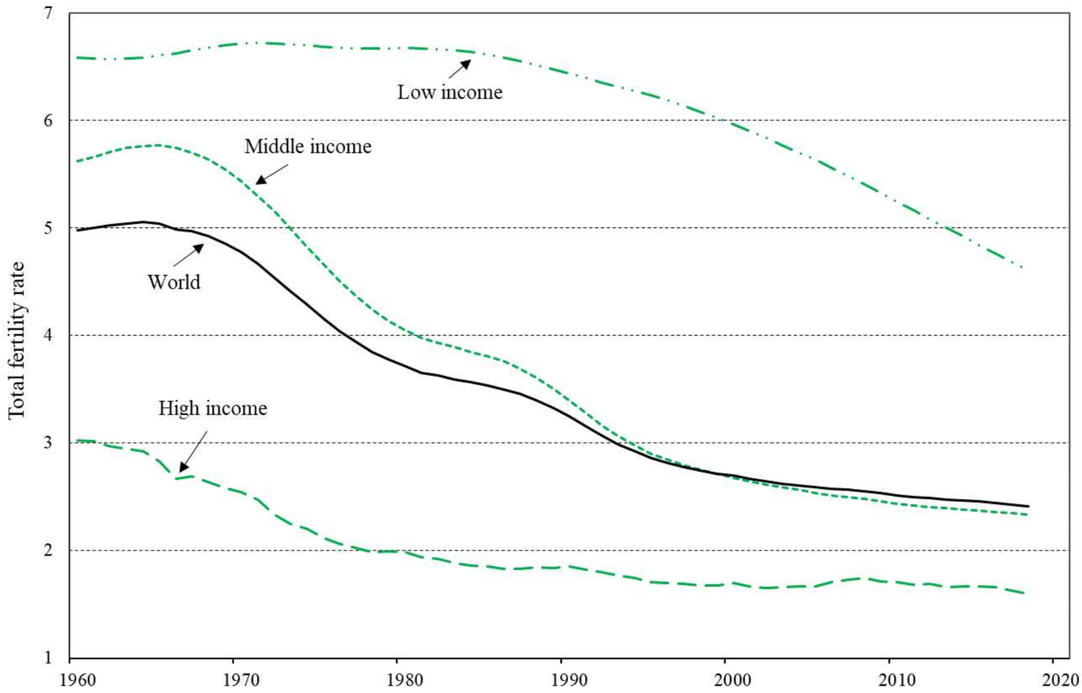

Figure 1. Fertility Trends for the Past 60 Years

Notes: The data are derived from the World Development Indicators of the World Bank. The dataset classifies a country as high income, middle income, or low income based on its 2016 gross national income per capita.

What is the impact of secular fertility declines on economic growth? This question has been closely related to the population-growth debate since the time of Thomas Malthus (1798). This debate has generated a significant amount of empirical literature on the effect of fertility rates and other measures of population growth on economic growth. However, almost all surveys of the empirical literature have yielded mixed results.[1] For example, Julian Simon (1992, p.ix) summarizes that "the most important

[1] Important surveys including Kuznets (1967), Kelley (1988), Simon (1992), Kelley and Schmidt (1994), Dasgupta (1995), Paul Schultz (2008), and Headey and Hodge (2009).

fact in today's population economics is the lack of observed correlation between the rate of a country's population growth and the rate of its economic development." Similarly, Paul Schultz (2008, p.01) concludes that "the evidence linking rapid population growth to slower economic development was not confirmed to be a major impediment to modern economic growth."

This paper argues that the lack of predominant evidence of the effect of fertility on economic growth is due to two major empirical challenges. The first is to address endogeneity bias arising from reverse causation and incidental association. Strong evidence suggests that higher incomes lead to lower fertility (e.g., Chatterjee and Vogl 2018, Herzer et al. 2012). Since Becker and Lewis's (1973) study, it has been well recognized that fertility decisions are jointly made with other choices that can affect economic outcomes, such as human capital investments and savings. Therefore, to the extent that endogeneity bias is addressed, studies could find different effects of fertility on economic growth.

The second challenge is to estimate the long-run effect of a change in fertility. An individual's interactions with the economy over a lifetime suggest that the effect of a change in fertility can last for decades and differ substantially in the short- and long-run.[2] For example, Simon (1989) asserted that the short-run economic effects of population growth operate mainly through capital dilution and the cost of raising children, but population growth has more pronounced effects in the long-run through channels like the contribution of new ideas. As such, to the extent that the long-run effects of fertility change can be determined, studies could potentially reach significantly different conclusions.

[2] In this paper, the long-run effect on growth refers to the effect over several decades (i.e., over the human lifecycle), but not the effect during the steady state. Because fertility declines continuously over time, steady-state effects cannot be estimated without making very strong assumptions.

Thus far, it has been difficult to address these two challenges simultaneously. Cross-sectional studies have the potential to capture the long-run effects of fertility by comparing long-run cross-sectional differences in fertility and economic growth, but the results are vulnerable to omitted variable bias. Similarly, time-series analyses can capture long-run effects by employing long-run fertility trends, but they would only be able to identify Granger causality, which is subject to the post hoc fallacy. A standard method for addressing endogeneity bias is to employ a panel model that uses fixed effects to eliminate time-invariant confounding factors, and the remaining endogeneity in the model is addressed by using plausibly exogenous fertility changes for the instrumental variables estimation.[3] This paper explains in great detail, however, that because fixed effects eliminate long-run fertility differences, it is mostly the short-run effects of fertility changes that are captured in the fixed-effects panel model.

This paper attempts to concurrently address these two challenges and thus reveal the long-run causal effect of fertility on economic growth. Specifically, I estimate a series of panel models, each of which regresses the income growth rate on the fertility rate lagged by different years, ranging from zero to decades.[4] This model specification is able to uncover the long-run dynamic effects of fertility change on income growth. Endogeneity bias is then addressed by the fixed effects in these panel models and by the instrumental variables (IVs) constructed from plausibly exogenous intensity measures of birth control policies.

Depending on national data from 164 countries during1960–2016, the estimation leads to four important conclusions: (i) the effect of a higher fertility rate on income

[3] Note that for cross-sectional models, which cannot include fixed effects at the observational level, it is nearly impossible to find a valid IV for fertility, because potential IVs are most likely correlated with unobservable, time-invariant determinants of economic growth.

[4] This model is a modification of the distributed-lag model, which includes all lags of the key explanatory variable in a single regression, in order to avoid collinearity. As detailed later, this modification does not change the estimated effect pattern and long-run average effect.

growth lasts for more than three decades; (ii) although the short-run effects of higher fertility are negative, the lagged long-run effects are positive; (iii) the lagged positive effect increases with the economic development level; and (iv) the long-run *average* effect of a higher fertility rate is significantly positive for most countries, but significantly negative for stagnant low-income countries. A similar dynamic effect pattern is found when using Chinese provincial data to estimate the model, and the IV estimates based on Chinese data are comparable to those based on the global data. A brief review of the theoretical literature finds predictions consistent with these findings: a larger new cohort immediately (and mechanically) dilutes per capita income (Malthus 1798), but latter on promotes physical capital accumulation (Solow 1956) and accelerates innovation (Jones 1995); to the extent that the efficiency of capital accumulation and innovation increases with the economic development level, the lagged positive effects should be larger in more developed countries.

Using relevant long-run average marginal effect estimates, this paper roughly calculated that the fertility declines observed during 1960–2016 might have *reduced* the average annual GDP per capita growth rate by 0.70, 0.22, and 0.35 percentage points, respectively, in countries with 1960 GDP per capita ranking in the top one-third, middle one-third, and bottom one-third but which experienced economic takeoffs. In contrast, the calculation also finds that the observed fertility declines *increased* the average annual growth rate by 0.24 percentage points in the stagnant bottom one-third of countries. These findings lead to two important implications: (i) the secular fertility declines observed worldwide represent a strong force driving down global economic growth; and (ii) pro-natalist policies can increase long-run economic growth for most countries, except for those stagnant low-income countries where anti-natalist policies are likely to be growth-enhancing.

This paper is among only a few to estimate the *long-run causal* effect of population growth on economic growth. Literature reviews (e.g., Kelley and Schmidt 1994, Headey and Hodge 2009) have found that many population-growth studies are cross-country studies that fail to adequately address endogeneity bias. Several studies have attempted to address endogeneity bias using IV estimations of panel models with fixed effects and have mainly concluded that the (short-run) effect of higher fertility on economic growth is negative or insignificant (e.g., Hongbin Li and Junsen Zhang 2007). The current paper complemented these studies by examining the *long-run* lagged effects and found that the long-run positive effects generally overcome the short-run negative effects. The existing long-run evidence was primarily derived from time-series analyses that attempted to uncover Granger causality between population growth and economic growth (e.g., Thornton 2001). This paper complements these studies by examining causality with a panel framework that could more adequately address a potential post hoc fallacy. By addressing endogeneity bias and identifying long-run effects, this paper finds strong evidence supporting the seminal work by Kremer (1993) and Jones (1995) that established a positive link between population, technological change, and economic growth in the long run.

Moreover, this study contributes to evaluating the economic impact of family planning programs, which are prevalent in developing countries.[5] Substantial effort has been made to evaluate the impact of family planning interventions on development and well-being in recent years (e.g., Ashraf et al. 2014, Cavalcanti et al. 2020). However, the current understanding of the effect of family planning programs on income largely comes from family-level studies, which generally find that families with smaller

[5] According to the World Population Policies Database (United Nations 2015), the number of countries that adopted national family planning programs reached 95 by 1976 and increased to 160 by 2013.

numbers of children have higher per capita income. Micro-level studies are incapable of capturing the positive spillovers associated with population growth, however, because these effects are economy-wide (Dasgupta 1995). Unfortunately, existing macro-level studies are either biased due to endogenous fertility or fail to account for the long-run effects. Using macro-level data, this paper finds that, although family planning programs are growth-enhancing for stagnant low-income countries, they are growth-hindering for countries in which economic takeoffs have been achieved.

The remainder of this paper is organized as follows: Section 2 briefly reviews the theoretical literature that guides the empirical strategy of this paper, Section 3 details the identification strategy, Section 4 presents global evidence, Section 5 presents evidence from China, Section 6 calculates the total impact of secular fertility declines, and the final section concludes this paper.

2. A Brief Review of the Theoretical Literature

Although the contribution of this paper is not theoretical, a brief review of the standard growth models' predictions on the effects of population growth on economic growth could guide the empirical analysis and facilitate understanding of the empirical findings.

Existing growth theories have suggested at least three major channels through which fertility can affect income growth. First, the classical growth theory of Thomas Malthus (1798) highlighted that per capita income is simply the ratio of output to population, meaning that a larger newborn cohort corresponds to a lower per capita income. Second, while recognizing the income-dilution effect of increased population growth, neoclassical growth theories (e.g., Solow 1956, Cass 1965) also suggested that the higher demand arising from population growth might promote income growth by inducing physical capital accumulation. Third, most R&D-based growth models have predicted that a larger population, which means more potential innovators, can improve income growth by promoting technological progress (Jones 1995).[6]

Note that the timing of the effects implied by these channels differs. A larger newborn cohort should immediately (mechanically) reduce per capita income through the first channel. The effect through the second channel—the inducement of capital accumulation—is likely to increase over time because a newborn's demands and thus pressure on the economy increase with age (Boserup 1981, Simon 1992). The effect through the last channel—the promotion of technological progress—becomes possible only after the cohort enters the labor force.

Note also that the magnitude (and even the likelihood) of the effects implied by the latter two channels depends on the economic development level of the economy considered. Growth models with endogenous savings (e.g., the Ramsey–Cass–Koopmans models) generally predict that the rate of physical capital accumulation increases when an economy emerges from a relatively low-income level. Similarly, growth models with endogenous human capital (e.g., Becker et al. 1990) predict that the accumulation of human capital (which fosters idea creation) accelerates only after the economy develops to a level with sufficiently high returns on human capital. Therefore, the positive effects through the latter two channels are likely to be larger in more developed countries and may have no effect in low-income countries with stagnant physical and human capital accumulation.

Besides the above three channels, there is an even more natural fourth channel that

[6] Another well-known channel connecting fertility and economic growth is the trade-off between the quality and quantity of children (Becker et al. 1990). This channel is less relevant when analysing the causal effect of fertility on economic growth because it emphasizes reverse causality: economic development increases human capita returns, which in turn motivates parents to invest more in each child's human capital and to reduce their numbers of children.

could generate time-varying effects: the effect of fertility on labor supply (Galor and Weil 1996, Bloom et al. 2009). Newborns initially reduce parental labor-market participation because child rearing is time consuming, but parents spend less time on each child as the child's age increases. Eventually, when children reach working age, they directly contribute to the labor force. Therefore, by affecting labor supply, a larger newborn cohort initially has a negative effect and then a positive effect on per capita income.

The dynamic effects of fertility change through these four channels can be formalized by a simple Cobb–Douglas production function in an overlapping-generation economy consisting of parents and children. The time horizon examined is $t \in [0, T]$, where $t = 0$ is the birth year of a new cohort and $t = T$ is the life expectancy thereof. The total output in year t is assumed to be:

$$Y_t = A_t L_t^{\alpha} K_t^{1-a},$$

where A_t is the technological level, L_t is the total labor supply, K_t is the total physical capital of production, and $\alpha \in (0,1)$. Per capita aggregate income is

$$y_t = \frac{A_t L_t^{\alpha} K_t^{1-a}}{P(1+n)} \quad , \tag{1}$$

where P is the parental population and n is the number of children per parent. Naturally, equation (1) shows that a higher birth rate n leads to a lower per capita income, which reflects the effect of the first channel.

The effect of the second channel, the inducement of physical capital accumulation, can be characterized by:

$$K_t = K_0(1 + x_t n) \quad , \tag{2}$$

where K_0 is the initial capital stock, and $x_t > 0$ measures the demand (or the incentive for capital accumulation) of the new cohort. The demand is assumed to increase with the age of the newborns (i.e., $dx_t/dt > 0$).

The third channel, which asserts that higher fertility rates lead to the promotion of technological progress when the new cohort reaches working age, can be characterized by:

$$A_t = P(1+s_t n) \quad , \tag{3}$$

where s_t measures each newborn's labor supply (or contribution to innovation). For simplicity, it can be assumed that $s_t = 0$ before the newborn reaches working age t^*, and $s_t > 0$ thereafter. The labor supply s_t is likely to eventually decline due to the aging of individuals.

The fourth channel can be characterized by:

$$L_t = P\left(1-r_t n\right)(1+s_t n) \quad , \tag{4}$$

where r_t is the share of parental time spent on raising each child, which generally declines with the child's age (i.e., $dr_t/dt < 0$). In equation (4), $1-r_t n$ captures that a newborn reduces the parental labor supply, and $1+s_t n$ captures that a newborn directly contributes to the labor force (when $t > t^*$).

Combining equations (2)–(4) with equation (1) results in:

$$y_t = \begin{cases} \dfrac{\left(P\left(1-r_t n\right)\right)^{\alpha}\left(K_0(1+x_t n)\right)^{1-\alpha}}{1+n}, & t < t^* \\ \dfrac{(1+s_t n)\left(P(1-r_t n)(1+s_t n)\right)^{\alpha}\left(K_0(1+x_t n)\right)^{1-\alpha}}{1+n}, & t \geq t^* \end{cases} \quad . \tag{5}$$

Equation (5) illustrates the complex dynamic effects of fertility on income growth. The immediate effect ($t = 0$) of a larger newborn cohort depends on the effects of three channels: the negative effects of $1+n$ (income dilution) and $(1-r_0 n)^{\alpha}$ (labor reduction), and the positive effect of $(1+x_0 n)^{1-\alpha}$ (induction of capital accumulation). The initial net effect tends to be negative if the induced physical capital accumulation is small. As the growth of the newborn cohort, r_t declines and x_t increases, the net effect increases in a positive direction over time. After the cohort reaches working age ($t \geq t^*$), two additional positive effects occur: the effects from additional innovators, $1+s_t n$, and additional workers, $\left(1+s_t n\right)^{\alpha}$. As a result, the positive effect after t^* further increases. Since an individual's labor supply, s_t, is likely to eventually decline (and may decline to negative in old age, reflecting a burden on the economy), the net effect should peak and then decline.

In summary, the existing theoretical studies suggested three empirically testable predictions: (i) the effect of a change in fertility rate on income growth could last for decades, throughout the life cycle of an individual; (ii) a higher fertility rate is likely to first reduce and then promote per capita income growth; and (iii) the effect of a higher fertility rate could be more positive (or less negative) in high-income countries than in low-income countries if the latter are less capable of accumulating physical and human capital.

3. Identification Strategy

This section details how this paper addressed endogeneity bias and captured the long-run dynamic effects. A starting point for addressing endogeneity bias is to estimate the following panel data model with fixed effects:

$$y_{ct} = \nu_c + \tau_t + \beta_0 TFR_{ct} + Z_{ct}\lambda + \varepsilon_{ct} \quad , \tag{6}$$

where y_{ct} is the growth rate of GDP per capita in country c and year t, TFR_{ct} is the total fertility rate, Z_{ct} is a vector of time-varying control variables, ν_c and τ_t

denote country fixed effects and year fixed effects, respectively, β_0 and λ are both coefficients, and ε_{ct} is an error term.

3.1 Omitted Variables

Model (6) should be able to substantially reduce endogeneity bias arising from omitted variables. The country fixed effects ν_c eliminate the confounding effect of all country-specific, time-invariant factors. The year fixed effects τ_t eliminate the confounding effect of annual shocks that are common to all countries. Since most factors that could affect both fertility and income growth are likely time persistent (such as culture or geographic features) or change similarly over time across countries (such as life expectancy or food prices), the two-way fixed effects should be able to account for most of the confounding factors.

The remaining omitted variables pertain only to *country-specific, time-varying* factors that happened to affect both fertility and income growth. This paper will show that the fertility estimates from the fixed-effect panel model are not sensitive to excluding even the most important time-varying confounding factors or to including various other time-varying factors, suggesting that the remaining country-specific, time-varying omitted factors are unlikely to substantially bias the fertility estimates. Note that this model specification does not address the potential bias arising from reverse causality (i.e., simultaneity bias, a point I will address shortly).

3.2 Long-Run Dynamic Effects

A major argument of this paper is that, although the fixed-effect panel model helps to address omitted variable bias, it mainly captures the short-run effects of fertility changes. This is because the fixed effects eliminate most long-run inter-country fertility differences and common fertility trends, so that the identification depends primarily on country-specific fertility changes. A substantial portion of the long-run effect could be captured by the fixed-effect model only in an extreme case: if the time-series of the panel data is sufficiently long and if the country-specific fertility changes primarily occur in early periods of the time-series.[7] However, the fact is that fertility has declined continuously over time (Figure 1). As illustrated in Appendix B1, even for panel data that cover 50 years, the identification of the fixed-effect model still heavily depends on short-run fertility changes and thus mainly captures the short-run effects.

A natural way to capture the long-run effect with the fixed-effect panel model is by extending it to a distributed-lag model:

$$y_{ct} = v_c + \tau_t + \sum_{s=0}^{T} \beta_s TFR_{c(t-s)} + Z_{ct}\lambda + \varepsilon_{ct} \quad , \tag{7}$$

where T denotes the maximum lag length of the TFR. The main difference between model (6) and model (7) is that the latter includes both the current TFR ($s=0$) and the TFR lagged by up to T years ($1 \le s \le T$). If the TFRs lagged by different years are independent of each other, β_s captures the s-year lagged effect of a TFR change on income growth.

However, model (7) suffers from the serious problem of collinearity because fertility rates in successive years are highly correlated with one another.[8] Unless collinearity is adequately addressed, the estimate of β_s could be imprecise and have incorrect signs. A standard method for addressing collinearity is to use a restricted least

---

[7] This extreme case is similar to capturing the long-run average effect of a one-off treatment in a standard difference-in-differences (DID) model. To see this, assume that there are panel data for 100 countries over a 60-year period and that only a one-off exogenous fertility decline in the tenth year occurs in half of the sample countries. A standard DID model that compares income growth before and after the fertility decline across the 100 countries could capture the long-run average effect of the fertility decline over the 50 years that followed the fertility decline.

[8] If $TFR_{c(t-i)}$ follows a pattern over time, then $TFR_{c(t-i+1)}$ will follow a similar pattern, thus causing $TFR_{c(t-i)}$ and $TFR_{c(t-i+1)}$ to be strongly correlated. For example, the data from 164 countries examined in this paper show that the bivariate correlation between the current TFR and the TFR lagged by 3, 5, and 10 years are 0.99, 0.98, and 0.96, respectively.

squares estimator depending on a polynomial distributed lag, which was first explored by Almon (1965). To do this, however, one must first know the pattern of the time effects, which can then be translated into parameter restrictions. Unfortunately, the time evolution for the effects of fertility change is too complicated to be characterized by a tractable functional form.[9] Since imposing incorrect restrictions on parameters can lead to additional biases, this paper does not seek to solve the collinearity problem by using restricted least squares estimators.

Instead, estimations of a series of equations were conducted, each of which only included one of the lag terms for the TFR, with a lag length that ranged from 0 to T years:

$$y_{ct} = \nu_c + \tau_t + \beta_s TFR_{c(t-s)} + Z_{c(t-s)}\lambda + \varepsilon_{ct}^s \ , \quad s = 0,1,2,\cdots,T \ . \tag{8}$$

To the extent that the nearby lags are correlated with one another, the coefficient of the included lag, β_s, captures the effect of the "omitted" nearby lags. Specifically, the estimate of β_s in model (8) can be seen as the weighted average of the effects of the included lag and the omitted nearby lags, and the weighting is the strength of the correlation. The bias from this kind of omitted variable therefore makes the estimated dynamic effects smoother but does not change the estimated effect pattern. Moreover, the simple average of β_s (over $s = 0,1,2,\cdots,T$) approximately equals the long-run average effect of a fertility increase on income growth. It should be caution that, as detailed in Appendix B1, the estimate of β_s in model (8) *cannot* be interpreted as the effect of a fertility increase on the income growth rate s years later. Instead, it should be understood as the weighted average of the effects of an increase in fertility on the income growth rates a minimum of s years later, and the weighting declines with the

[9] Note that the lagged effects of fertility change could be more complicated than the lagged effects of other factors, such as a fiscal policy, because an individual's interaction with the economy lasts for many decades and changes non-linearly over time.

distance to s.

3.3 Reverse Causality

By including fixed effects and lagged fertility, panel model (8) has the advantage of addressing omitted variable bias and capturing the long-run dynamic effects, but the issue of reverse causation is not adequately addressed. The lagged model specification is insufficient for addressing reverse causality because fertility could be affected by expectations of future income growth. Since income growth generally has a negative effect on fertility in modern societies (and, as found in this paper, the long-run average effect of fertility on income growth is generally positive), the ordinary least squares (OLS) estimate of β_s from model (8) tends to be downwardly biased. The only way to fully address reverse causality bias is to look for exogenous IVs for the TFR. This paper adopts plausibly exogenous intensity measures of birth control policies as IVs for the TFR in model (8). Details of the IV estimation will be presented later.

3.4 The Question of Interest and the Relevant Estimates

One of the main purposes of this paper is to evaluate the impact of secular fertility declines (as presented in Figure 1) on income growth. For this purpose, the IV estimate of β_s from model (8) is obviously relevant. If the average IV estimate of β_s over $s = 0,1,2,\cdots,T$ is significantly positive (or negative), it could be concluded that the secular fertility declines reduced (or increased) the long-run average income growth, respectively. However, the IV estimate, which removed the effect from reverse causality, could not be used to evaluate the *magnitude* of the effect of the observed fertility declines, because a substantial portion of the observed fertility declines were (directly or indirectly) caused by income growth (Herzer et al. 2012, Chatterjee and Vogl 2018).

To evaluate the magnitude of the effect of the observed endogenous fertility declines, the OLS estimate of β_s (which addressed the omitted variable bias but did not remove the effect from reverse causation) is more relevant because it captures the net outcome of the endogenous interaction between fertility and income growth. The interactions between fertility and income are most convincingly demonstrated by Becker et al. (1990) and Galor and Weil (1996): economic growth raises returns to human capital and the opportunity cost of a mother's time, thus leading parents to have a smaller number of children, invest more in the human capital of each child, and allocate more time to the labor market. In other words, income growth could affect fertility by affecting human capital accumulation and parental labor supply. Therefore, the impact of the endogenous fertility *declines* on income growth (which is negative according to the IV estimate of this paper) can be seen as the "cost" of the additional human capital accumulation and labor supply, which in turn "benefit" income growth. While the IV estimate captures only the "cost" of the endogenous fertility decline, the OLS estimate captures the net outcome of the "cost" and "benefit." The OLS estimate is positive (i.e., a lower fertility rate corresponds to lower income growth) if the "cost" outcomes the "benefit," and vice versa.

4. Global Evidence

4.1 Lagged Dynamic Effects

Figure 2 presents the OLS estimate of β_s (dot on the solid line) from model (8) and the corresponding 95% confidence intervals (broken lines). I estimated each equation in model (8) separately based on annual data from 164 countries (listed in Table A2) during the 1960–2016 period. The data for GDP per capita (in 2011 US$) were retrieved from the Maddison Project Database 2018, and the data for the TFR were derived from the World Bank's World Development Indicators. The longest TFR lag length in these estimations is 50 years, but this figure only presents the first 41 estimates, because the remaining estimates are mainly statistically insignificant and with wide confidence intervals; the full results are presented in Table B2. The same effect pattern is found in Figure B1 when using the level (instead of the growth rate) of GDP per capita as the dependent variable for model (8).

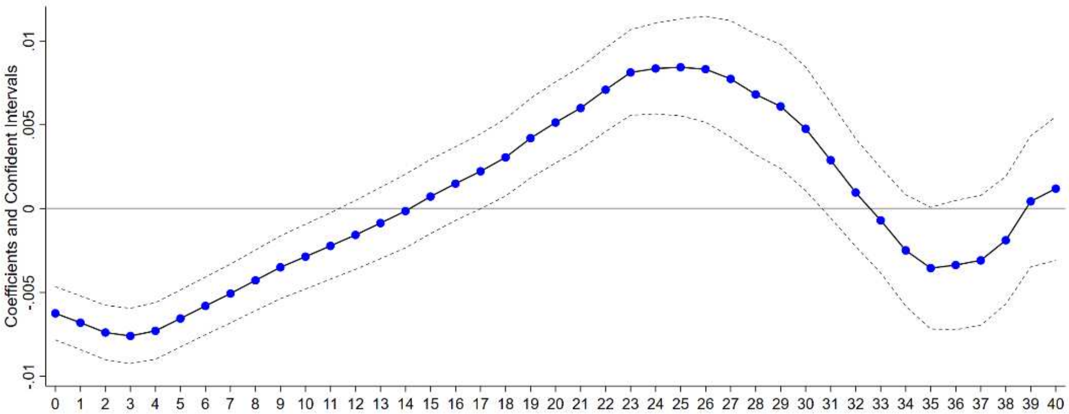

Figure 2. Current and Lagged Effects of a Higher Fertility Rate on the Growth Rate of GDP per capita

Notes: I regress the annual growth rate of real GDP per capita on the TFR in a panel model (i.e., model (8)) with country and year fixed effects based on data from 164 countries during 1960–2016. This figure presents estimates from 41 separate regressions. In each regression, the TFR was lagged by 0–40 years. Each dot on the solid line is the point estimate of the coefficient of the TFR lagged by the year indicated by the x-axis, and the broken lines indicate the corresponding 95% confidence intervals.

All regressions include the four time-varying control variables that are most frequently used in the population-growth literature: the five-year lagged log GDP per capita, years of schooling for individuals aged 25 or over, infant mortality rate, and the share of urban population. Other control variables are set aside for robustness checks. All control variables are lagged by the same years as the TFR in each regression in order to separate their effects from the effect of the TFR; this is necessary because, for example, it is the infant mortality rate in the same period as the TFR (rather than the future infant mortality rate) that could affect the TFR.[10] The results are comparable,

[10] If the control variables are not lagged, which means that they are in the same time period as the dependent variable, the model actually controls for the future values of the control variables relative to the lagged TFR. If there are any correlations between the future values of control variables and the lagged

however, when all control variables are excluded from the estimation (Figure B2) or not lagged (Figure B7). Details of the control variables are presented in Table A1.

Figure 2 verifies the theoretical predictions that the effect of a higher fertility rate lasts for decades and that a higher fertility rate first reduces and then promotes income growth. Specifically, the figure shows that a higher TFR has a statistically significant effect on the growth rate of GDP per capita over a period of more than three decades. While the initial effect is negative, the effect becomes positive when the lag length is longer than 14 years. The effect peaks when the lag length is 25 years, after which it declines and eventually becomes statistically insignificant (see Table B2 for estimates up to 50 lagged years). The decline of the effect after 25 lagged years is understandable in terms of an individual's life cycle: 25 years old is the usual reproductive age of individuals, and the initial effect of a newborn cohort on income growth is negative.

Here, I only briefly discuss the estimated marginal effects, and a more detailed discussion is postponed to Section 6. As mentioned above, the average of β_s is an approximation of the long-run average effect. It can be calculated that the average marginal effect for the lag length ranging from zero to 31 years (after which the effects are statistically insignificant) is only 0.07 percentage points, with a 95% confidence interval (0.03, 0.11). The small average effect covers up the substantial heterogeneity: the most negative (or most positive) estimate observed in the 3 lagged years (or 25 lagged years) suggests that the effect of a one-unit higher TFR on the growth rate of GDP per capita is -0.76 percentage points (or 0.84 percentage points), respectively. This finding highlights the importance of identifying the long-run dynamic effects and

---

TFR, it is most likely that the lagged TFR is the cause (i.e., the lagged TFR affects income growth through these control variables). Therefore, controlling for future values of these correlated variables could partly account for the true effect of the TFR and thus bias the TFR estimate toward zero. The robustness check presented in Figure B7 indeed shows that, although the effect pattern is the same, replacing the lagged controls by their current values leads to smaller estimated fertility coefficients.

explains why previous studies focusing on different time spans obtained substantially different results. Later, I will show that the average effect is much more positive when excluding the stagnant low-income countries from the sample.

Appendix B4 provides various robustness checks supporting that the OLS estimates are not sensitive to omitted variables. Figure B2 removes the four time-varying control variables, Figure B3 controls for five additional time-varying factors, Figure B4 controls for the interactions between a full range of year dummies and three time-invariant factors, Figure B5 controls for country-specific linear year trends, and Figure B6 clusters the error at the region-year level. All these robustness checks show the same effect pattern and comparable marginal effects as Figure 2. This finding is not surprising because the country and year fixed effects included accounted for the confounding effects of most factors that could jointly affect fertility and income growth.

4.2 Differential Effects across Countries

I classify the sample countries into three equal-sized groups according to their 1960 GDP per capita and estimate model (8) for each country group. As presented in each of the three panels in Figure 3, the subsample regressions reveal the same effect pattern: the effect of higher fertility is first negative and then positive, and the effect lasts for more than three decades. In addition, the turning points of the dynamic effects are similar across country groups: the effect turns from negative to positive at around 15 lagged years, peaked at around 25 lagged years, and declines to zero at around 35 lagged years. Thus, the dynamic effect pattern is common across countries with different economic development levels.[11]

[11] Another interesting observation in Figure 3 is that, after the TFR is lagged by about 35 years, the effect in the middle one-third of countries (Panel B) declines further, while the effects in the bottom one-third and top one-third countries (Panels A and C) begin to increase. A potential explanation is that the cost of human capital investment is relatively high in middle-income countries: parents in low-income

More importantly, Figure 3 provides evidence for the theoretical prediction that the effect of higher fertility tends to be more positive in more developed countries. Specifically, it shows that the initial negative effect is much smaller in the high-income group (Panel C) than in the low- and middle-income groups (Panels A and B), while the peaked positive effect is about twice larger in the high-income group than in the low- and middle-income groups. The average marginal effects calculated from zero to 31 lagged years are 0.02, 0.17, and 0.95 percentage points, respectively, for the low-, middle-, and high-income groups. These different effects are consistent with the hypothesis that more developed countries are more efficient in accumulating physical and human capital in response to population growth. Similar results are obtained when classifying the sample countries according to their GDP per capita from other years.

To verify this finding, I further disaggregate the one-third of countries in the bottom income group (Panel A of Figure 3) into three subgroups based on whether the country's average growth rate of GDP per capita (denoted by g) during 1960–1979 was below 0%, between 0% and 2%, or above 2%. If the low-income countries that experienced faster income growth were those more efficient in accumulating physical and human capital, then one should observe that the lagged positive effect of higher fertility increases with the average growth rate during 1960–1979. In addition, if stagnant low-income countries ($g < 0$) were also stagnant in accumulating physical and human capital, a very limited lagged positive effect of higher fertility in those countries should be observed.

countries may invest little on the formal education of children, while the cost of education in the high-income countries is low relative to parents' incomes.

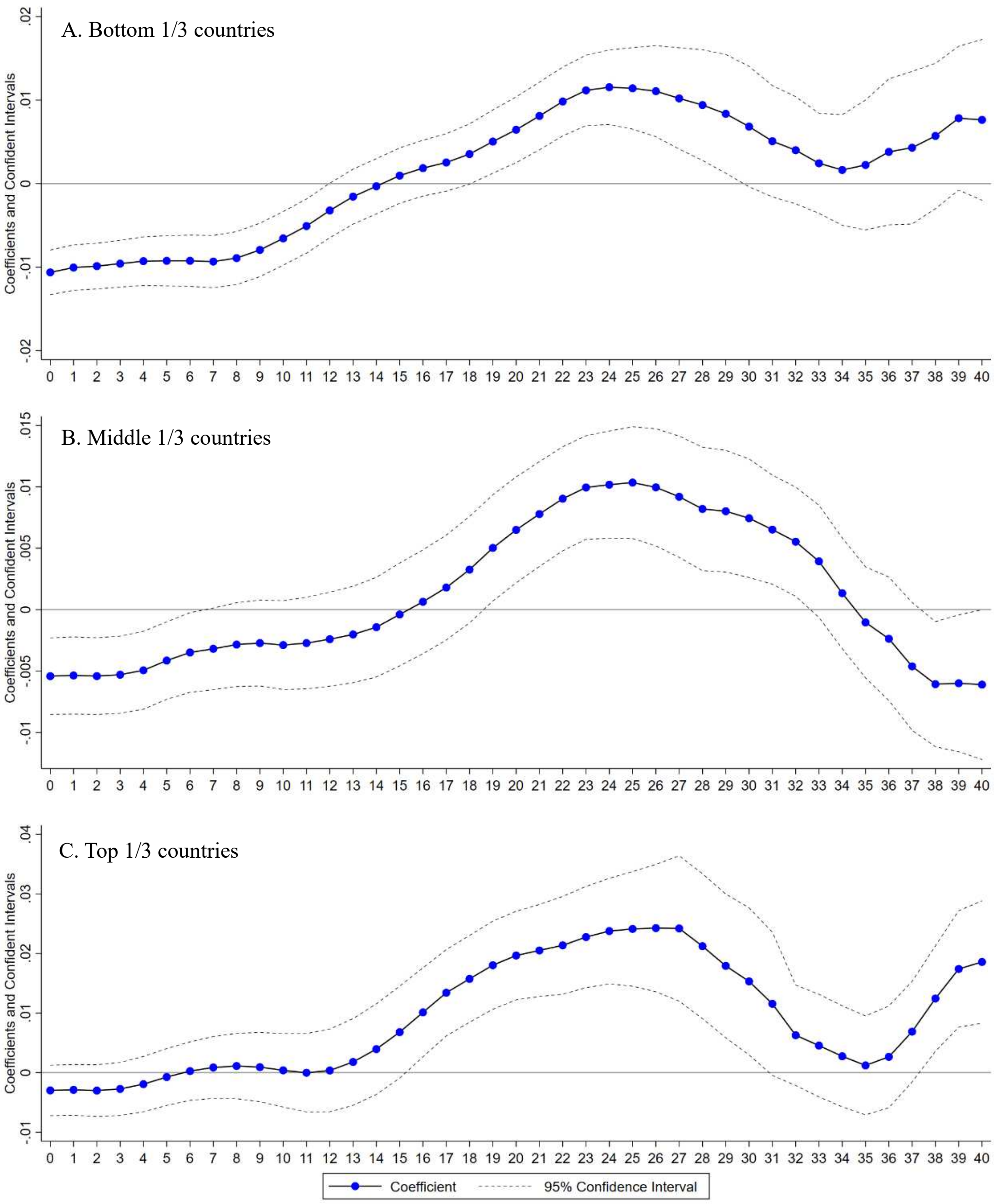

Figure 3. Current and Lagged Effects of a Higher Fertility Rate on the Growth Rate of GDP per capita (by development level)

Notes: This figure replicates Figure 2. The difference is that the sample countries are classified into three equal-sized groups according to their 1960 GDP per capita and then model (8) is estimated separately for each country group.

Figure 4 presents the estimates of model (8) based on data from each of the three low-income country groups. Consistent with the prediction, almost no significantly positive lagged effects are found for the country group with stagnant income growth ($g<0$, Panel A). For low-income countries with moderate income growth ($0 \le g \le 2$, Panel B) or economic takeoff ($g>2$, Panel C), the lagged effects are much more positive. To facilitate comparison, I calculated the average marginal effects from zero to 31 lagged years and found that the effects are -0.28, 0.01, and 0.30 percentage points, respectively, for the stagnant, moderate-growth, and takeoff country groups. The confidence intervals are wide in Figure 4, because the sample countries used in each panel are small (less than 20). The results are similar when low-income countries are classified based on their average income growth rates during 1960–1969 or 1960–1989.[12]

In summary, the subsample regressions show that the lagged positive effect of higher fertility increases with the income level, and no obvious positive lagged effects are found for stagnant low-income countries. The long-run average effects of higher fertility are strongly positive in the middle- and high-income countries as well as in the low-income takeoff countries (according to Panels B and C of Figure 3 and Panel C of Figure 4), but are strongly negative or statistically insignificant in the low-income countries with limited income growth (according to Panels A and B of Figure 4). These findings have an important policy implication: while pro-natalist policies are growth enhancing in most countries, anti-natalist policies may improve income growth in stagnant low-income countries.

[12] I also conducted subsample regressions for countries within the middle- or top-income groups (also classified based on the average income growth rate) and found estimates similar to those presented in Panels C and D of Figure 3.

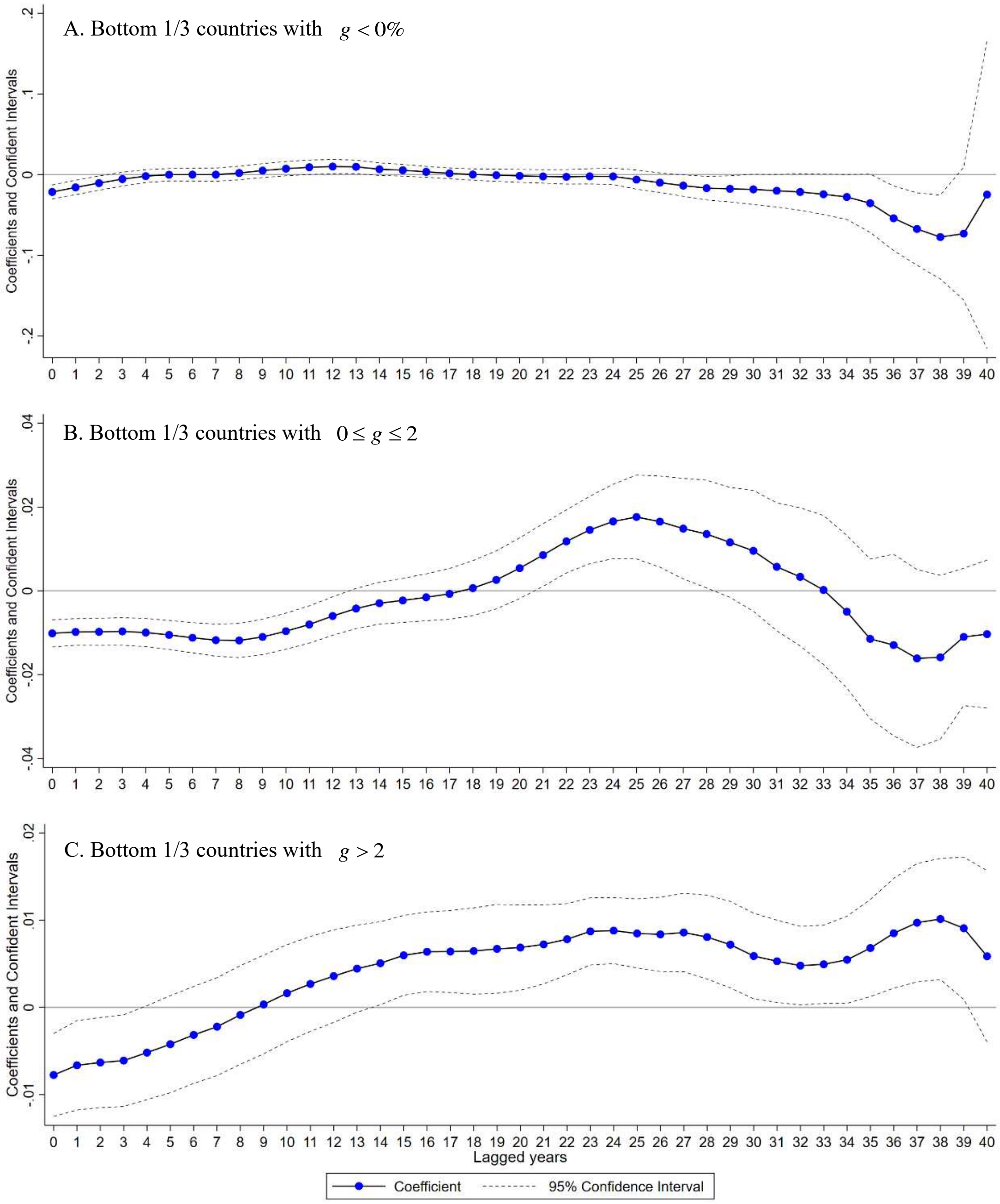

Figure 4. Current and Lagged Effects of a Higher Fertility Rate on the Growth Rate of GDP per capita (subgroup regressions for the bottom one-third of countries)

Notes: This figure replicates Panel A of Figure 3. The difference is that Panels A, B, and C of this figure only use data from the one-third of lowest-income countries (i.e., where the 1960 GDP per capita ranked in the bottom one-third of all sample countries) with average income growth rates during 1960–1979 below 0, between 0 and 2%, and above 2%, respectively.

4.3 Evidence from the Global Family Planning Campaign

This section addresses the reverse causality bias of the OLS estimates. Although various pieces of evidence (Appendix B4) show that the OLS estimates of the fixed-effect panel model are not substantially biased by omitted variables, the issue of reverse causality has not been adequately addressed. This section adopts three plausibly exogenous intensity measures of national family planning programs to address the issue of reverse causality. However, due to two major limitations discussed at the end of this section, the findings can only be taken as suggestive evidence. More credible causal evidence will be provided later based on within-country intensity measures of the one-child policy in China.

Since the mid-1960s, increasing concerns over the unprecedented levels of population growth in developing countries has led to a global family planning campaign (Robinson and Ross 2007). The number of countries that adopted national family planning programs reached 95 by 1976 and increased to 160 by 2013 (United Nations 2015). De Silva and Tenreyro (2017) found strong evidence that national family planning programs significantly reduced fertility rates in developing countries. Following De Silva and Tenreyro (2017), this section uses three intensity measures for national family planning programs: the family planning program effort score (effort score), the percentage of women exposed to family planning messages (message exposure), and funds for family planning per capita (funds per capita).

The data for these intensity measures were compiled by De Silva and Tenreyro (2017) from various sources. The definition and summary statistics of these intensity measures are presented in Table A1. Data on the effort score are available for 95 countries for 1972, 1982, 1989, 1994, and 1999; data on the message exposure are available for 57 countries over various years from 1993 to 2013; and data on the funds per capita are available for 58 countries over various years starting in 1972 and going up to 1992. Because the intensity measures are only sparsely available over narrow periods, it is not feasible to examine the long-run dynamic effects when using these intensity measures as the IVs.

As such, this section only uses these intensity measures as IVs for the TFR in model (6), which is a version of mode (8) when $s=0$. I conduct a two-stage least squares (2SLS) estimation with model (6) as the second-stage regression and the following equation as the first-stage regression:

$$TFR_{ct} = \nu_c + \tau_t + \gamma IV_{ct} + Z_{ct}\lambda + \omega_{ct} \quad , \tag{9}$$

where IV_{ct} is *one* of the three intensity measures in country c and year t, and ω_{ct} is the error term. All other variables are defined as before. The 2SLS estimation only use one of the three intensity measures each time, because these measures are mostly available for different years and/or different countries. Each regression only includes countries for which the data on the corresponding IV for at least two years are available. Missing values between the beginning and ending sample years of the IVs are filled in by a linear interpolation.

The identification assumption is that, conditional on the fixed effects and control variables, the IVs are not correlated with income growth except through fertility. This assumption could be violated in two cases: first, if the IVs are correlated with omitted determinants of income growth, and second, if the IVs are directly correlated with income growth. The model substantially reduced the concern about the first case by including the country and year fixed effects; the remaining omitted variables pertain only to country-specific, time-varying factors, which are not very likely to have systematic effects on both birth control intensity and income growth. The concern about the second case is relieved by the fact that past income growth rates have no predictive power on future birth control intensities (detailed in Appendix B5).

The 2SLS estimation results are reported in Table 1. The first-stage estimates presented in Panel A show that, as expected, all three intensity measures have significantly negative effects on the TFR. The second-stage estimates presented in Panel B show that the causal effect of the TFR on income growth is positive and large, although imprecisely estimated when the IV used is funds per capita. These estimates are not sensitive to the same robustness checks as presented in Appendix B4. The 2SLS estimate varies when different IVs are used, potentially because the IVs are mostly available for different countries and years. Recall that the OLS estimates are significantly negative when the TFR lag length is small, the positive 2SLS estimates confirm that the OLS estimates are downwardly biased. I postpone the marginal effect comparison to Section 6.

Table 1. Causal Effect of the TFR on the Growth Rate of GDP per capita

	(1) IV = Effort score	(2) IV = Message exposure	(3) IV = Funds per capita
Panel A. ***First stage*** (Dependent variable: TFR)			
IV	-0.010***	-0.008***	-0.003***
	(0.001)	(0.002)	(0.001)
Panel B. ***Second stage*** (Dependent variable: Growth rate of GDP per capita)			
TFR (instrumented)	0.030***	0.043***	0.018
	(0.009)	(0.019)	(0.017)
Four control variables	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
First-stage *F*	84.7	22.0	26.3
Observations	2,072	561	468
Number of Countries	81	44	45

Notes: All regressions include country and year fixed effects as well as the four time-varying control variables. The standard errors (in parentheses) are clustered at the country level, based on the bootstraps procedure suggested by Cameron et al. (2008). Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

There are two major limitations to using these country-level IVs. First, because countries differ widely, these IVs may still be endogenous if there were omitted country-specific, time-varying factors that could jointly affect the IVs and income growth. Additionally, due to limited data, the dynamic effects (and thus the long-run average effect) of fertility cannot be estimated using these IVs. The remainder of this paper employs provincial data from China's one-child policy (OCP) to overcome these limitations. Provinces within a country are more comparable to each other than the country would be to other countries. Additionally, detailed yearly intensity measures of the OCP are available at the province level, which enables the estimation of the dynamic causal effects.

5. Evidence from China

China implemented the OCP in 1979 to curb its population explosion (Coale 1981). The OCP lasted for three decades and was significantly modified in 2011. During 1979–2010, the OCP generally allowed each couple to only have one child, but set several exemption rules.[13] Residents who violated the OCP not only faced a stiff fine but also risked losing their employment and not being able to register their children for health care and education (Feng et al. 2013). Rigorous empirical studies found that the OCP significantly reduced birth rates in China.[14] This section depends on the 1980–2010 data for 27 of the 31 mainland Chinese provincial districts that enforced the OCP.[15]

[13] The three most important exemptions were (1) couples with an agricultural hukou (a system of household registration) were allowed to have a second child if their first child was a girl; (2) residents who belonged to an ethnic minority group were allowed to have more than one child; and (3) residents in Xinjiang and Tibet were not subject to the OCP until the early 1990s (Baochang et al. 2007).

[14] For example, Goodkind (2017) found that China's low fertility was achieved two or three decades earlier than would be expected given its level of development. McElroy and Yang (2000) estimated that a complete removal of monetary penalties for violating the OCP would have increased fertility in rural China by 0.33 cumulative births per woman by 1992. Bingjing Li and Hongliang Zhang (2017) estimated that a one-percentage-point increase in the enforcement intensity of the OCP reduced family size from 1981–1999 by approximately 0.05.

[15] The provinces of Xinjiang and Tibet are excluded because they were not subjected to the OCP until the early 1990s, and Hainan and Chongqing are excluded because they were separated from Guangdong

Data sources and summary statistics of all variables used in this section are presented in Table A3.

Three intensity measures of the OCP that have been used as the IVs for fertility in the literature. The first is the policy-violation fine rate (measured in times of local yearly household income; see Appendix C1), which has been used to examine the effects of birth control on various outcomes, such as the sex ratio (Ebenstein 2010) and man-made twins (Huang et al. 2016). The second is the excess fertility rate (measuring local violations of the OCP; see Appendix D), which has been used to examine the effect of child quantity on child quality (Bingjing Li and Hongliang Zhang 2017). The third is the ethnic minority population share (ethnic minorities were subject to less-strict birth control measures; see Footnote 13), which has been used to examine the effect of fertility on income growth (Hongbin Li and Junsen Zhang 2007). This paper depends mainly on the first intensity measure to identify dynamic causal effects; the second measure is only available for two census years, and the third measure is proved to be endogenous. However, similar results are obtained when using the latter two intensity measures in complementary analyses. Details of these intensity measures will be presented when introduced in the analysis.

5.1 The OLS Estimates

Before moving on to the IV estimation, I first provide the OLS estimates in order to check whether the same dynamic effect pattern in the global data is observed in China. I estimate the following version of model (8) based on province-year data from the 27 Chinese provinces during 1980–2010:

---

and Sichuan in 1988 and 1997, respectively. The year of 1979 is excluded because the OCP was implemented at the end of 1979 and thus have no effect on the fertility in 1979 due to the nine-and-a-half-month length of a pregnancy. Data after 2010 are excluded because the OCP was significantly modified in 2011.

$$y_{pt} = \nu_p + \tau_t + \beta_s CBR_{p(t-s)} + Z_{p(t-s)}\lambda + \varepsilon^s_{pt} \text{ , } s = 0,1,2,\cdots,T \quad . \tag{10}$$

The key explanatory variable in equation s, $CBR_{p(t-s)}$, is the crude birth rate (CBR) in province p lagged by s years. $Z_{p(t-s)}$ is a vector of seven time-varying control variables: five-year lagged GDP per capita, share of labor with secondary education, share of urban population, crude death rate, out-migration rate, trade share in GDP, and government spending share. These control variables are included because previous studies showed them to be important determinants of fertility and/or growth, and I will present evidence to show that the estimates are robust when excluding these controls or including additional controls. All other variables are the same as has been previously defined, although at the province level. Note that while the global analysis measures fertility by the TFR, the China analysis measures fertility by the CBR because province-level TFR data are not available for China.[16]

As presented in Figure 5, the estimated effect pattern is consistent with that based on the global data: with an increase of the lag length, the effect of a higher fertility rate changes from negative to positive, and the effect eventually peaks and declines to statistical insignificance. The average effect calculated from zero to 21 lagged years (i.e., the last significant estimate) suggests that a one-unit higher CBR increases the growth rate of GDP per capita by 0.10 percentage points, with a 95% confidence interval (0.08, 0.12). Interestingly, the marginal effect becomes positive after only 4 lagged years in China, while the corresponding turning point is 14 lagged years in the global data. This could reflect the fact that China implemented a more coercive birth control policy than other countries; estimates based more on exogenous fertility declines are less affected by reverse causality and thus more positive.

[16] The CBR is defined as the annual number of births per thousand *population*, while the TFR is defined as the average number of children that a woman would have over her childbearing years. Using the CBR as the fertility measure in the global analysis yielded a comparable result.

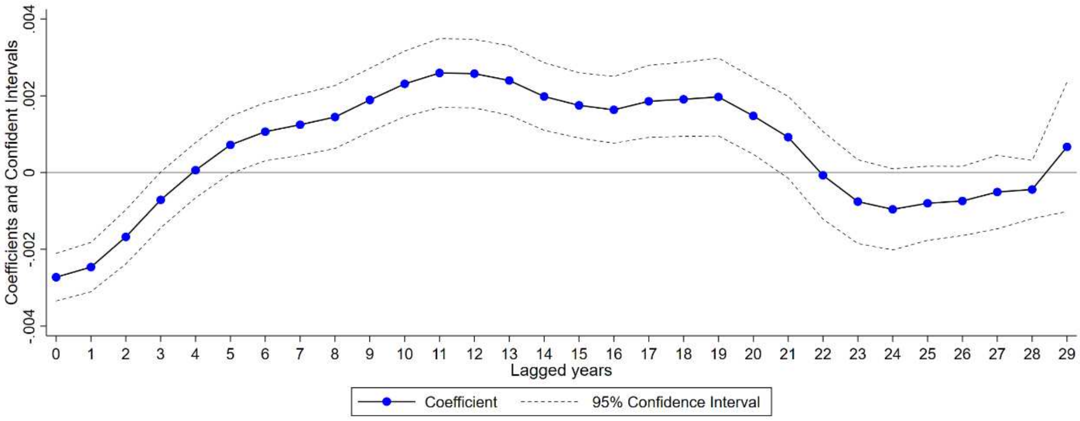

Figure 5. Current and Lagged Effects of Fertility on Income Growth in China

Notes: This figure presents the OLS estimates of model (10) based on China provincial data. The x-axis indicates the lagged years of the CBR. Each dot on the solid line is the point estimate of the coefficient, and the broken lines indicate the corresponding 95% confidence intervals.

Various robustness checks show that the estimates presented in Figure 5 are not sensitive to omitted variables. As presented in Figures C2, C4, and C5, similar results are obtained when excluding the seven time-varying control variables, including province-specific linear time trends, and clustering the error term at the province level, respectively. Additionally, Figure C3 shows that the estimates are robust when controlling for the three most important concurrent events: the reform and opening-up in 1978, the tax system reform in 1994, and joining the World Trade Organization (WTO) in 2001. These events are controlled for by the interactions between the timing and province-level intensity measures of each event.[17] All these findings suggest that the dynamic effects of fertility presented in Figure 5 are unlikely to be primarily driven by omitted variables.

[17] Specifically, China started to open its doors to foreign businesses in 1978, which increased its trade-to-GDP ratio from under 10% to 64% in only two decades (Brandt and Rawski 2008, pp. 1–3). I control for the confounding effect of the opening-up by the interactions between the 1978 dummy and the trade-to-GDP ratio and distance to the nearest port, respectively. The economic reform substantially reduced the central government's budget, which led to the 1994 tax system reform that tripled the central government's share of revenues in GDP from 3% to 9% (Brandt and Rawski 2008, pp. 431–440). I control for the confounding effect of the tax reform by the interaction between the 1994 dummy and the government spending share of the GDP. Joining the WTO in 2001 dramatically increased China's international trade and liberalized the country's service sector (Brandt and Rawski 2008, pp. 657–659). I control for the confounding effect of joining the WTO by the interactions between the 2001 dummy and the trade-to-GDP ratio and contribution of services to the GDP, respectively.

5.2 The IV Estimates

Ebenstein (2010) collected province-level OCP violation fine rates in China from 1979–2000 and found substantial cross-province and temporal variations in the fine rate (Figure C1). Previous studies have used the fine rate as an IV or a proxy variable for fertility when examining the effect of the OCP on the sex ratio (Ebenstein 2010), saving rates (Wei and Zhang 2011), man-made twins (Huang et al. 2016), or various micro-level individual outcomes (Huang et al. 2020). Following the literature, this section uses the policy fine rate data from 1980–2000 collected by Ebenstein (2010) to examine the dynamic causal effect of fertility change on income growth.

Specifically, I use the policy fine rate lagged by s years (denoted by $Fine_{p(t-s)}$) as the IV for $CBR_{p(t-s)}$ in the 2SLS estimation of model (10).[18] As reported in column (1) of Table C1, the first-stage regression results indicate that a one-unit higher policy fine rate *reduces* the CBR by 0.31 (births per thousand population), and this effect is statistically significant at the 1% level. Columns (2) to (5) of the same table present various robustness checks to confirm this strong negative association.

The second-stage IV estimates are presented in Figure 6. The longest examined lag length is 18 years, because policy fine data are for 20 years. Consistent with the OLS estimates, the IV estimates also suggest that the effect of higher fertility increases with the lag length. However, all IV estimates are positive, although statistically insignificant when the lag length is small.[19] In addition, all IV estimates are much larger than the corresponding OLS estimates, confirming that the OLS estimates are downwardly

[18] In the robustness check presented in Figure C10, $Fine_{p(t-s-1)}$ was used as the IV for $CBR_{p(t-s)}$ to allow a lag for the translation of the policy fine rate change to fertility change, and this yielded results similar to those presented in Figure 6.

[19] No obvious declines in the lagged effect are observed with an increase of lag length, potentially due to the relatively short time period examined, although the effect is statistically insignificant when the lag length is 18 years.

biased. The average effect calculated over 0–18 lagged years suggests that a one-unit increase in the CBR promotes the average growth rate of GDP per capita by 0.72 percentage points. As presented in Figures C6–C9, the IV estimates are robust when excluding the seven time-varying control variables, controlling for concurrent events, including province-specific linear time trends, and clustering the error term at the province level, respectively.

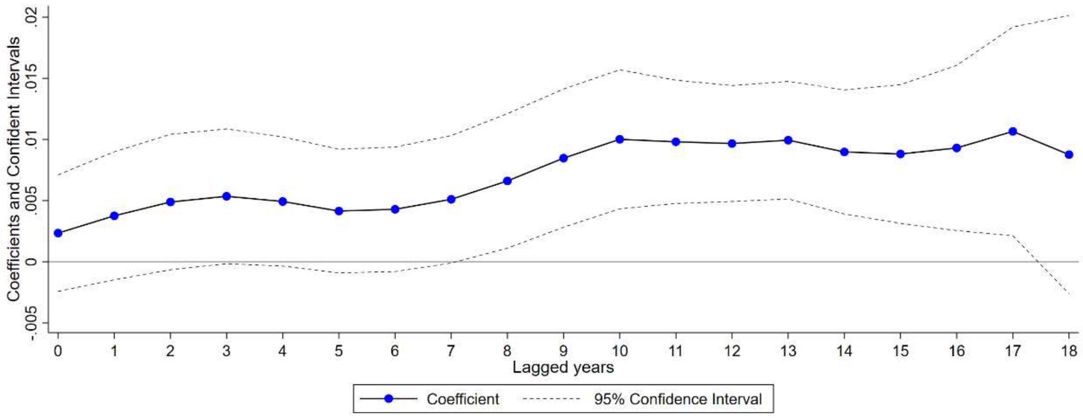

Figure 6. Current and Lagged Effects of Fertility on Income Growth in China (IV Estimates)

Notes: The figure presents the 2SLS estimates based on models (10). The x-axis indicates the lagged years of the CBR. Each dot on the solid line is the point estimate of the coefficient, and the broken lines indicate the corresponding 95% confidence intervals.

The above 2SLS estimation is based on the assumption that the policy fine rate is exogenous. Existing studies using the policy fine rate as the IV generally assume that changes in the province-level enforcement intensity of the national OCP policy were determined by local-specific factors, such as local governor’s preferences, that were not systematically correlated with the outcomes of interest. As shown in Figure C1, changes in provincial policy fine rates indeed seem random: no obvious common patterns are found for the timing, magnitude, or even direction of the changes in the policy fine rate. The following text presents three pieces of evidence supporting the exogeneity assumption: the policy fine rate is neither directly correlated with income growth nor indirectly correlated with it through factors other than fertility.

First, I find that preexisting income levels or growth rates have no predictive power on the policy fine rate. Specifically, I regress the policy fine rates in the next one, three, and five years, respectively, on the current growth rate (or level) of GDP per capita in a panel model with province and year fixed effects. As presented in Table C2, all estimates are small and with p-values greater than 0.1, indicating no predictive power of preexisting income growth rates (or levels) on the policy fine rate. Although this finding does not exclude the possibility that the policy fine rate could be determined by expectations of future income growth, it should substantially reduce this concern: if the policy fine rate is not based on the readily available information of past incomes, it is less likely to be based on the uncertain predictions of future income growth. Therefore, this finding can be taken as suggestive evidence that the policy fine rate is not directly affected by income growth.

Second, I show that the policy fine rate is uncorrelated with important determinants of income growth. The policy fine rate is endogenous if it is correlated with omitted province-specific, time-varying determinants of income growth (note that the time-invariant factors and common changes have been controlled for). This concern can be substantially reduced if the policy fine rate is not correlated with even the most important time-varying determinants of income growth. As presented in Table C3, when examining a set of nine most frequently used determinants of income growth, I find that none of them have predictive power on the policy fine rate. In addition, the same table also shows that changes in the policy fine rate are not correlated with changes in these nine variables.[20]

Finally, I show that the lead of policy fine rate is not correlated with current fertility

[20] Huang et al. (2020) examined a set of 28 macroeconomic indices in a similar way and also found no correlation between them and the policy fine rate.

and income growth. To the extent that the policy fine rate captures the impact of the OCP's strictness on fertility, rather than differential trends across provinces generated by omitted variables, the future policy fine rate should not predict current fertility and income growth. Table C4 examines the effect of the five-year lead policy fine rate on income growth and fertility by including it as a control variable in the first- and second-stage regressions of the 2SLS estimation. The estimated coefficients for the five-year lead policy fine rate are extremely small and statistically insignificant in both the first- and second-stage regressions. Similar results are obtained when using the three- or seven-year lead policy fine rate. This finding further relieves the concern that the policy fine rate is correlated with omitted variables.

5.3 Evidence from Other Intensity Measures

This section presents supplementary causal evidence from two additional intensity measures of the OCP. The first is the excess fertility rate (EFR), which is constructed as the percentage of Han Chinese mothers aged 15–49 years who gave a higher order birth in 1981 (using microdata from the 1982 Chinese Population Census; see Appendix D1 for details). By assuming local violations of the OCP as an exogenous source of fertility variation, the EFR has been used by Bingjing Li and Hongliang Zhang (2017) to examine the causal effects of child quantity on child quality, and by Junsen Zhang (2017) to examine the effect of the OCP on marital status, labor supply, and migration. The main analysis of this paper does not depend on the EFR for two reasons. First, microdata that can be used to construct the EFR is only available for two census years (1982 and 1990), and thus the EFR cannot be used to identify the dynamic causal effects. Second, analyses using the EFR depend mainly on cross-province fertility variation and thus are more vulnerable to endogeneity bias.

For completeness, Appendix D presents the DID estimates from the model that uses the EFR as the intensity measure of the OCP. Figure D1 shows that there are substantial EFR differences across provinces and that the EFR is positively and significantly correlated with the CBR. Table D1 classifies the sample provinces into two groups according to their 1981 EFR and find that, although the income growth rates were statistically identical before the OCP across these two groups, the group with a higher EFR experienced faster income growth after the OCP. Table D2 presents the DID estimate from the model that uses the 1981 EFR as the intensity measure and controls for province and year fixed effects. The DID estimate suggests that a one-percentage-point increase in the EFR significantly increased the growth rate of GDP per capita by 0.62 percentage points. The various robustness checks provided in Tables D1 and D3 and Figure D2 suggest that the DID estimates are robust and that the EFR is likely exogenous. Therefore, the estimation results based on the EFR support the main finding of this paper that the causal effect of higher fertility on income growth is positive. As presented in Table 2 rows 10 and 11, the marginal effect estimated based on the EFR is quite comparable to that based on the policy fine rate.

Another alternative intensity measure is the provincial ethnic minority population share (MPS), which may capture exogenous fertility changes because the ethnic minorities were subject to less-strict birth control measures. The MPS was used as an IV for fertility by Hongbin Li and Junsen Zhang (2007) when they examined the effect of fertility on income growth. However, evidence presented in Figure E1 and Table E1 suggests that the MPS is likely endogenous. In particular, the MPS is strongly correlated with preexisting growth rates even after controlling for province and year fixed effects as well as various time-varying factors. Nevertheless, I still tried to estimate the dynamic effects by using the MPS as the IV for the CBR in model (10). As presented in Figure E2, the estimated effect pattern is generally consistent with that based on the policy fine rate: the effect of higher fertility is statistically insignificant when the lag length is smaller than three years, but the effect becomes significantly positive thereafter. However, the estimated effects are unreasonably large and with wide confidence intervals after 14 lagged years, potentially due to the endogeneity bias of the MPS.

6. Marginal Effects

Table 2 compares the marginal effects estimated from different models (column 1) and datasets (column 2). To facilitate a comparison, all marginal effects are transformed into the format of the effect of a 1% increase in fertility from its mean on the growth rate of GDP per capita, measured in percentage points. Column 3 presents the calculated marginal effects and the corresponding 95% confidence intervals (in parentheses). Column 5 details the original estimates used to calculate the marginal effects.

Specifically, for each model examining the dynamic effects (rows 1 and 3–10), the marginal effect is calculated as the *average* of the effects of the current fertility rate and the fertility rates lagged by up to s_m years (i.e., the average effect for $0 \leq s \leq s_m$), where s_m is the maximum lag length for which the effect of fertility is still statistically significant (to facilitate comparison, the maximum lag length is set to 31 years for all estimates based on the national data). For the 2SLS estimates based on the national data (row 2), the marginal effect is calculated from the estimates in column 1 of Table 1—the column with the largest sample size and thus likely contains the most credible estimates. For the DID estimate (row 11), the marginal effect is calculated based on the estimate from column 5 of Table D2.

Table 2. Marginal and Total Effects Estimated from Different Models and Datasets

	(1) Model	(2) Dataset	(3) Marginal effect (pp)	(4) Total effect (pp)	(5) Calculation based on
			Global evidence		
1	OLS	164 countries	**0.003** (0.001, 0.004)	-0.069	Figure 2, $0 \le s \le 31$ average
2	2SLS	81 developing countries	**0.152** (0.065, 0.238)		Table 1, column 1
Subsample estimation of the 164 countries (classified by 1960 GDP per capita)					
3	OLS	Top 1/3 countries	**0.026** (0.022, 0.029)	-0.701	Panel C of Figure 3, $0 \le s \le 31$ average
4	OLS	Middle 1/3 countries	**0.007** (0.004, 0.010)	-0.216	Panel B of Figure 3, $0 \le s \le 31$ average
5	OLS	Bottom 1/3 countries	**0.001** (-0.002, 0.005)	-0.026	Panel A of Figure 3, $0 \le s \le 31$ average
Subsample estimation of the bottom 1/3 countries (classified by the 1960–1979 average growth rate)					
6	OLS	$g > 2$	**0.015** (0.010, 0.020)	-0.345	Panel C of Figure 4, $0 \le s \le 31$ average
7	OLS	$0 \le g \le 2$	**0.001** (-0.005, 0.006)	-0.010	Panel B of Figure 4, $0 \le s \le 31$ average
8	OLS	$g < 0$	**-0.017** (-0.027, -0.007)	0.241	Panel A of Figure 4, $0 \le s \le 31$ average
			China evidence		
9	OLS	China	**0.020** (0.017, 0.023)	-0.527	Figure 5, $0 \le s \le 21$ average
10	2SLS	China	**0.106** (0.086, 0.126)	-0.371	Figure 6, $0 \le s \le 17$ average
11	DID	China	**0.150** (0.052, 0.247)		Table D2, column 5

Notes: Column 3 presents the marginal effects of fertility calculated based on estimates from different models (column 1) and datasets (column 2). The marginal effects are expressed as the effect of a 1% increase in the fertility rate from its mean on the average growth rate of GDP per capita, in percentage points (pp). The 95% confidence intervals presented in parentheses in column 3 are calculated based on the delta method. The total effects on growth presented in column 4 are calculated by combining the estimated marginal effects with the observed percentage changes in fertility.

Table 2 presents four major findings. First, the marginal effect of higher fertility on income growth is positive for most country groups but is negative for stagnant low-income countries (row 8). As indicated by the 95% confidence intervals in parentheses, these marginal effects are mainly statistically significant, except for the two associated with low-income countries (rows 5 and 7).

Second, the marginal effect of higher fertility on income growth increases with the economic development level. A 1% increase in fertility promotes the income growth rates by 0.001, 0.007, and 0.026 percentage points, respectively, for countries with 1960

GDP per capita ranked in the bottom, middle, and top one-third of the sample countries (rows 5, 4, and 3). Consistent with this, the marginal effects are -0.017, 0.001, and 0.015 percentage points, respectively, for the bottom one-third of countries in which the average growth rate during 1960–1979 was below 0, between 0 and 2%, or above 2% (rows 8, 7, and 6).

Third, the IV estimates are substantially larger than the OLS estimates, confirming that the latter is downwardly biased. The marginal effects calculated based on the 2SLS estimates are 0.152 and 0.106, respectively, for the 81 developing countries and for China (rows 2 and 10). The corresponding marginal effects calculated from the OLS estimates are only 0.003 and 0.020 (Rows 1 and 9).

Fourth, the estimated causal link between fertility and income growth does not change substantially when different sample countries are examined or different estimation methods are used. The marginal effects calculated from the IV estimate and the DID estimate based on Chinese provincial data (0.106 and 0.150, rows 10 and 11) are comparable to that calculated from the IV estimate based on the country-level data (0.152, row 2).

The rest of this section evaluates the total effect of secular fertility declines on income growth rates. As detailed in section 3.4, the OLS estimates are more relevant for evaluating the magnitude of the effect of endogenous fertility declines. By assuming that the observed fertility declines were mainly (directly or indirectly) caused by income growth, I use the OLS estimates to calculate the total effect of the fertility declines on income growth rates. I do this in three steps: first, calculate the average annual fertility declines for each study sample; second, sum up the average annual fertility declines over the number of years during which a fertility change has significant effects (e.g., 31 years in the global data); and third, transform the accumulated fertility decline into a percentage decline and then multiply it by the corresponding marginal effect to obtain the total effect.[21] For example, the accumulated fertility declines over 31 years were 30.9% for the middle one-third of countries, and thus the total effect was calculated as -0.216 percentage points (i.e., -30.9 * 0.007).

The total effects are presented in column (4) of Table 2. The observed fertility declines reduced the average annual growth rates of GDP per capita by 0.701, 0.216, and 0.345 percentage points, respectively, in the top one-third of countries (row 3), middle one-third of countries (row 4), and bottom one-third of takeoff countries (row 6). These negative effects are economically significant because the corresponding average growth rates during the sample periods were 2.18, 1.78, and 2.75 percentage points, respectively. In contrast, in the stagnant low-income countries (row 8), the observed fertility declines *increased* the average annual growth rate by 0.241 percentage points. It is important to highlight that these total effects are calculated based on two strong assumptions: all fertility declines are endogenous, and the marginal effect of fertility does not change with fertility level. Additionally, the total effect reflects the net outcome of the endogenous interaction between fertility and income growth, but not the pure causal effect of fertility on income growth. Therefore, although these calculations are helpful for understanding the economic implications of secular fertility declines, the total effects calculated should be interpreted with caution.

When it comes to evaluating the total effect of exogenous fertility declines, the IV estimates are more relevant. However, the endogenous nature of fertility implies that it is difficult to determine the share of fertility declines caused by exogenous factors, such as birth control policies. As an illustration, I approximately calculate the total effect of

[21] The effect of a fertility change on income growth is not permanent, so the total effect of the secular fertility declines should be calculated as the accumulated effect over the periods during which a fertility change has effects.

the OCP working through the policy-violation fine. Without the OCP violation fine, the average CBR in China during 1980–2000 could have been 3.5% higher.[22] When this metric is combined with the IV estimate of marginal effect in row 10, I find that the policy fine reduced the average annual growth rate of GDP per capita by 0.37 percentage points (i.e., 3.5 * 0.106), which is approximately 4.7% of China's average growth rate of 7.8 during the same period. Considering that the policy violation fine was only one means of enforcing the OCP, the total impact of the OCP on growth might be greater.

7. Concluding Remarks

It is important to understand the effect of fertility on economic growth not only because dramatical fertility declines have been observed in virtually all countries over the past decades, but also because many developing countries still adopt national family planning programs to curb fertility. However, the fertility-growth nexus has been highly debated and inconclusive in the empirical literature.

The existing theoretical models predict different effects of fertility on growth over the human life cycle: the initial effect of higher fertility tends to be negative due to capital dilution and the time costs of child rearing, but the lagged effect tends to be positive due to induced physical capital accumulation and innovation and to additional labor supply. This theoretical prediction has three implications for empirical studies: (i) what really matters is the long-run average effect of fertility on income growth, but not the time-varying short-run effect; (ii) it would be interesting to empirically verify the

[22] According to column 1 of Table C1, a one-unit increase in the policy fine rate reduced the CBR by 0.31. As presented in Table A3, the mean value of the policy fine rate was 1.74. The policy fine should have therefore reduced the CBR by 0.54, which was approximately 3.5% of the mean CBR of 15.50 in China during 1980–2010.

dynamic effects of fertility on income growth; and (iii) the positive effect of higher fertility may increase with the economic development level if more developed countries are more efficient in innovation and accumulating physical capital.

Inspired by these theoretical implications, this paper adopted an empirical strategy that could identify the long-run dynamic effects (and hence the long-run average effect) of a fertility change on income growth, and thus avoided the bias that arises from examining only short-run effects. Additionally, this paper made great efforts to address the endogeneity bias caused by omitted variables and reverse causation, which has been another important cause of inconsistency in the empirical literature. The long-run dynamic effects were estimated by a series of panel models, each of which regressed the income growth rate on the fertility rate that was lagged by different years. Endogeneity bias was then addressed by the fixed effects in the panel model and by the instrumental variables constructed from the intensity measures of birth control policies.

The empirical model was estimated using national data from 164 countries and provincial data from China. The estimated dynamic effects were consistent with the theoretical prediction that a higher fertility rate first reduces and then increases income growth. Subsample analyses confirmed the theoretical prediction that the positive effect of fertility would increase with the economic development level. Importantly, although the long-run average effect of fertility is significantly positive in high-income countries, middle-income countries, and take-off low-income countries, the effect is significantly negative in stagnant low-income countries. These findings lead to two major implications: (i) the secular fertility declines observed worldwide represent a strong force driving down global economic growth; and (ii) pro-natalist policies could increase long-run economic growth for most countries, except for those stagnant low-income countries in which anti-natalist policies are likely to be growth-enhancing.

Reference

Almon, Shirley. 1965. "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica* 33 (1):178-196.

Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household Bargaining and Excess Fertility: An Experimental Study in Zambia." *The American Economic Review* 104 (7):2210-2237.

Baochang, Gu, Wang Feng, Guo Zhigang, and Zhang Erli. 2007. "China's local and national fertility policies at the end of the twentieth century." *Population and Development Review* 33 (1):129-148.

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

Becker, Gary S, Kevin M Murphy, and Robert Tamura. 1990. "Human Capital, Fertility, and Economie Growth." *Journal of Political Economy* 98 (5):S12-S37.

Bingjing Li, and Hongliang Zhang. 2017. "Does population control lead to better child quality? Evidence from China's one-child policy enforcement." *Journal of Comparative Economics* 45 (2):246-260.

Bloom, David E, David Canning, Günther Fink, and Jocelyn E Finlay. 2009. "Fertility, female labor force participation, and the demographic dividend." *Journal of Economic growth* 14 (2):79-101.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87 (1):115-143.

Boserup, E. 1981. *Population and technological change: a study of long-term trends*. Chicago, Ill., United States: University of Chicago Press.

Brandt, Loren, and Thomas G Rawski. 2008. *China's great economic transformation*. Edited by Loren Brandt and Thomas G Rawski. Cambridge: Cambridge university press.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics* 90 (3):414-427.

Cass, David. 1965. "Optimum growth in an aggregative model of capital accumulation." *The Review of Economic Studies* 32 (3):233-240.

Cavalcanti, Tiago, Georgi Kocharkov, and Cezar Santos. 2020. "Family Planning and Development: Aggregate Effects of Contraceptive Use." *The Economic Journal*.

Chatterjee, Shoumitro, and Tom Vogl. 2018. "Escaping Malthus: Economic Growth and Fertility Change in the Developing World." *American Economic Review* 108 (6):1440-1467.

Coale, Ansley J. 1981. "Population trends, population policy, and population studies in China." *Population and Development Review* 7 (1):85-97.

Dasgupta, Partha. 1995. "The population problem: theory and evidence." *Journal of Economic Literature* 33 (4):1879-1902.

De Silva, Tiloka, and Silvana Tenreyro. 2017. "Population control policies and fertility convergence." *Journal of Economic Perspectives* 31 (4):205-28.

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.

Feng, Wang, Yong Cai, and Baochang Gu. 2013. "Population, policy, and politics: How will history judge China's one - child policy?" *Population and Development Review* 38:115-129.

Galor, Oded, and David N Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review*:374-387.

Goodkind, Daniel. 2017. "The astonishing population averted by China's birth restrictions: Estimates, nightmares, and reprogrammed ambitions." *Demography* 54 (4):1375-1400.

Headey, Derek D., and Andrew Hodge. 2009. "The Effect of Population Growth on Economic Growth: A Meta-

Regression Analysis of the Macroeconomic Literature." *Population and Development Review* 35 (2):221-248.

Herzer, Dierk, Holger Strulik, and Sebastian Vollmer. 2012. "The long-run determinants of fertility: one century of demographic change 1900–1999." *Journal of Economic Growth*:1-29.

Hongbin Li, and Junsen Zhang. 2007. "Do high birth rates hamper economic growth?" *The Review of Economics and Statistics* 89 (1):110-117.

Huang, Wei, Xiaoyan Lei, and Ang Sun. 2020. "Fertility Restrictions and Life-Cycle Outcomes: Evidence from the One-Child Policy in China." *The Review of Economics and Statistics* 0 (ja):1-41.

Huang, Wei, Xiaoyan Lei, and Yaohui Zhao. 2016. "One-child policy and the rise of man-made twins." *The Review of Economics and Statistics* 98 (3):467-476.

Jones, Charles I. 1995. "R & D-based models of economic growth." *Journal of political Economy*:759-784.

Junsen Zhang. 2017. "The evolution of China's one-child policy and its effects on family outcomes." *Journal of Economic Perspectives* 31 (1):141-60.

Kelley, Allen C. 1988. "Economic consequences of population change in the Third World." *Journal of Economic Literature* 26 (4):1685-1728.

Kelley, Allen C, and Robert M Schmidt. 1994. *Population and income change: Recent evidence*. Washington, D.C.: The World Bank.

Kremer, Michael. 1993. "Population growth and technological change: one million BC to 1990." *The Quarterly Journal of Economics* 108 (3):681-716.

Kuznets, Simon. 1967. "Population and economic growth." *Proceedings of the American philosophical Society* 111 (3):170-193.

Malthus, Thomas Robert. 1798. *An essay on the principle of population*. Cambridge: W. Pickering, 1986.

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

Paul Schultz. 2008. "Population Policies, Fertility, Women's Human Capital, and Child Quality." In *Handbook of development economics*, 3249-3303. Elsevier.

Robinson, Warren C., and John A. Ross. 2007. *The global family planning revolution: three decades of population policies and programs*. Washington, D.C.: World Bank.

Ross, John, and John Stover. 2001. "The Family Planning Program Effort Index: 1999 Cycle." *International Family Planning Perspectives* 27 (3): 119–129.

Simon, Julian. 1992. *Population and development in poor countries: selected essays.* Princeton, NJ: Princeton University Press.

Simon, Julian L. 1989. "On aggregate empirical studies relating population variables to economic development." *Population and Development Review*:323-332.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *The quarterly journal of economics* 70 (1):65-94.

Thornton, John. 2001. "Population growth and economic growth: long-run evidence from Latin America." *Southern Economic Journal*:464-468.

United Nations. 2015. World Population Policies Database. https://esa.un.org/PopPolicy/about_database.aspx: The United Nations.

Wei, Shang-Jin, and Xiaobo Zhang. 2011. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China." *Journal of Political Economy* 119 (3):511-564.

Online Appendix

A. Data Appendix

Table A1 contains the data sources and summary statistics of the variables used in the global analysis. Table A2 lists the 164 sample countries. Table A3 includes the data sources and summary statistics of the variables used in the China analysis. The growth rate of GDP per capita and the fertility measures are calculated as five-year moving averages to reduce the confounding effects of short-term fluctuations; the results are similar if the moving averages are not taken into consideration.

Table A1. Sources and Summary Statistics of the Global Data

Variable Name	Definition	Source	Mean
Growth rate of GDP per capita	Annual growth rate of real GDP per capita in 2011 US$	A	0.019
Total fertility rate (children per woman)	The average number of children that a woman would have over her childbearing years	B	4.05
Family planning program effort score (ranging from 0 to 300)	The Family Planning Program Effort Index developed by Ross and Stover (2001)	C	39.8
The percentage of women exposed to family planning messages (%)	The percentage of women who have been exposed to family planning messages on the radio, television, or newspapers, calculated based on the Demographic and Health Surveys	C	51.8
Funds for family planning per capita (US$)	Funds for family planning from both government and nongovernment sources, in 2005 US$	C	39.9
Five-year lagged GDP per capita (Log)	Five-year lagged GDP per capita in natural log, 2011 US$	A	8.58
Infant mortality rate (1/1000)	The number of deaths per 1,000 live births of children	B	56.6
The share of urban population (%)	Urban population as a percentage of total population	B	48.8
Years of schooling (year)	Years of schooling for individuals aged 25 or over	D	5.44

Note: 1. Data sources:

A: The Maddison Project Database 2018

B: World Development Indicators, the World Bank

C: The dataset of De Silva and Tenreyro (2017)

D: The dataset of Barro and Lee (2013)

2. All variables are in country level. All data are for the 164 sample countries during 1960–2016.

Table A2. Sample Countries

Afghanistan	Djibouti	Lao PDR	Romania
Albania	Dominica	Latvia	Russian Federation
Algeria	Dominican Republic	Lebanon	Rwanda
Angola	Ecuador	Lesotho	Sao Tome and Principe
Argentina	Egypt, Arab Rep.	Liberia	Saudi Arabia
Armenia	El Salvador	Libya	Senegal
Australia	Equatorial Guinea	Lithuania	Serbia
Austria	Estonia	Luxembourg	Seychelles
Azerbaijan	Ethiopia	Macedonia	Sierra Leone
Bahrain	Finland	Madagascar	Singapore
Bangladesh	France	Malawi	Slovak Republic
Barbados	Gabon	Malaysia	Slovenia
Belarus	Gambia	Mali	South Africa
Belgium	Georgia	Malta	Spain
Benin	Germany	Mauritania	Sri Lanka
Bolivia	Ghana	Mauritius	St. Lucia
Bosnia and Herzegovina	Greece	Mexico	Sudan
Botswana	Guatemala	Moldova	Swaziland
Brazil	Guinea	Mongolia	Sweden
Bulgaria	Guinea-Bissau	Montenegro	Switzerland
Burkina Faso	Haiti	Morocco	Syrian Arab Republic
Burundi	Honduras	Mozambique	Tajikistan
Cabo Verde	Hong Kong SAR, China	Myanmar	Tanzania
Cambodia	Hungary	Namibia	Thailand
Cameroon	Iceland	Nepal	Togo
Canada	India	Netherlands	Trinidad and Tobago
Central African Republic	Indonesia	New Zealand	Tunisia
Chad	Iran, Islamic Rep.	Nicaragua	Turkey
Chile	Iraq	Niger	Turkmenistan
China	Ireland	Nigeria	Uganda
Colombia	Israel	Norway	Ukraine
Comoros	Italy	Oman	United Kingdom
Congo, Dem. Rep.	Jamaica	Pakistan	United States
Congo, Rep.	Japan	Panama	Uruguay
Costa Rica	Jordan	Paraguay	Uzbekistan
Cote d'Ivoire	Kazakhstan	Peru	Venezuela, RB
Croatia	Kenya	Philippines	Vietnam
Cuba	Korea, Dem. People's Rep.	Poland	West Bank and Gaza
Cyprus	Korea, Rep.	Portugal	Yemen, Rep.
Czech Republic	Kuwait	Puerto Rico	Zambia
Denmark	Kyrgyz Republic	Qatar	Zimbabwe

Note: The 164 sample countries (or regions) for which the annual data on the growth rate of GDP per capita and total fertility rate are available.

Table A3. Sources and Summary Statistics of the China Provincial Data

Variable Name	Definition	Source	Mean
Growth rate of GDP per capita	Annual growth rate of real GDP per capita, in 2010 CNY	A, B	0.078
Crude birth rate (1/1000)	The annual number of births per thousand population	A, B	17.9
Policy fine rate (years of local household income)	The average monetary penalty rate for one unauthorized birth, in years of local household income	F	1.74
Excess fertility rate (%)	The percentage of non-agricultural Han mothers aged 15–49 who gave a higher order birth in 1981	D	4.24
Five-year lagged GDP per capita (Log)	Five-year lagged real GDP per capita in natural log, 2010 CNY	A, B	8.36
Share of labor with secondary education	Percentage of labor with middle and high school education (grades 7-12)	C	0.45
Share of population in urban areas	Percentage of the population living in urban areas	A, B	0.30
Crude death rate (1/1000)	The annual number of deaths per thousand population	A, B	6.41
Out-migration rate (%)	Out-migration as a percentage of the total population	A, B	0.34
Trade share in GDP	Trade (exports plus imports) as a percentage of GDP	A, B	0.25
Government spending share	Government spending as a percentage of GDP	A, B	0.15
Distance to port (100 km)	The distance from each province's centroid to the nearest port	G	5.15
Share of services in GDP	The contribution of services to GDP	A, B	0.29
Minority population share	Percentage of minorities in the population	D, E	0.10
Fertility preference (births)	The average total number of births of females aged 45–54 in 1981	D	5.43

Note: 1. Data sources:

A: China Compendium of Statistics 1949–2008
B: National Bureau of Statistics of China
C: China Population (and Employment) Statistics Yearbook (various years)
D: National Population Census of the PRC (decennial census)
E: The 1% Population Sample Survey (during the inter-censual years ending with 5)
F: The dataset of Ebenstein (2010)
G: China province Shapefile

2. All variables are in province level. All data are from 1980 to 2010 if not specified in the definition. The data for the youth dependency ratio, the old-age dependency ratio, education (before 1989), migration (after 2007), and minority population share are only available in five-year intervals, and continuous yearly data were obtained by linear interpolation.

B. Global Evidence Appendix

B1. Remaining fertility variation in a fixed-effects panel model

Including the location fixed effects in a panel model that regresses the income growth rate on fertility would eliminate all cross-sectional long-run fertility differences, and the identification would therefore only depend on annual fertility changes. If the time-series of the panel data is sufficiently long and if the fertility changes mostly occurred in the early years of the time-series, the fixed-effects panel model is capable of capturing the long-run effects; this can be most clearly illustrated by a standard DID model (see Footnote 7). However, for the 164 countries examined in this paper, fertility declines were continuous over time. Therefore, the fertility variation used to identify the pane model with country-fixed effects is a mix of fertility declines that occurred in different years. Since it takes about 20 years for a newborn to become an adult, the length of time that is defined as being "long-run" in a fertility-growth study should be a minimum of 20 years; in other words, only fertility declines that occurred 20 years ago should be used to capture long-run effects. Consequently, most of the fertility variations employed in the fixed-effects panel model are "short-run" fertility variations, and thus, short-run effects are what are mostly captured.

Table B1 illustrates this argument by using the five-decade fertility data from 1960–2009 for the 164 sample countries. For simplicity, the average yearly fertility declines were calculated for each decade across all countries and denoted as V^1, V^2, V^3, V^4, and V^5, respectively. As presented in Panel A of Table B1, significant yearly fertility declines occurred in every decade. Panel B present the current and lagged fertility declines used when estimating the fixed-effects panel model. For example, the fertility declines in the third decade (1980–1989) were derived from the first, second, and third decades, which are denoted as V^1_{-2} , V^2_{-1}, and V^3_0, respectively (the subscript denotes the lagged decades of the fertility decline). For a fixed-effects model based on these five decades of data, the share of the fertility variation that is lagged by 0, 1, 2, 3, 4, and 5 decades can be calculated approximately 34.3%, 29.8%, 20.1%, 18.8%, 4.0%, and 0%, respectively.[23] Therefore, a little more than one-third of the variation (i.e., 20.1% + 18.8% + 4.0% = 35.9%) can be used to capture the effect that is lagged by more than 20 years. If the time-series is shorter, the share of the lagged effects that are captured is smaller.

Table B1. Remaining Fertility Variation in Panel Model with Fixed Effects

	Decade				
	1960-1969	1970–1979	1980–1989	1990–1999	2000–2009
Panel A: Average yearly fertility declines					
	V^1 = -0.029	V^2 = -0.057	V^3 = -0.061	V^4 = -0.071	V^5 = -0.033
Panel B: Fertility declines used in the estimation in each decade					
	V^1_0	V^1_{-1}	V^1_{-2}	V^1_{-3}	V^1_{-4}
		V^2_0	V^2_{-1}	V^2_{-2}	V^2_{-3}
			V^3_0	V^3_{-1}	V^3_{-2}
				V^4_0	V^4_{-1}
					V^5_0
Panel C: Fertility declines used in the estimation in each decade, one decade lagged model					
		V^1_{-1}	V^1_{-2}	V^1_{-3}	V^1_{-4}
			V^2_{-1}	V^2_{-2}	V^2_{-3}
				V^3_{-1}	V^3_{-2}
					V^4_{-1}

Notes: Fertility declines listed in Panel A represent remaining fertility variations in a panel model with country fixed effects, calculated for each decade from 1960–2009 based on fertility data from 164 countries.

A natural way to identify the long-run lagged effects based on the fixed-effects panel model is to replace the TFR with the lags thereof. To show this, Panel C in Table B1 presents the yearly fertility declines used in the identification when the TFR is lagged

[23] For example, the share of the fertility variation that is lagged by one decade is calculated according to $(V^1_{-1}+V^2_{-1}+V^3_{-1}+V^4_{-1})/\sum V^i_{-j}=0.217/0.730=29.8\%$, in which the denominator is the sum of all values in Panel B; and this calculation uses the conditions of, for example, $V^1_{-1}=V^1=-0.029$.

by one decade. It can be calculated that now the share of the fertility variation that can be used to capture the effect lagged by more than 20 years increased to 54.6%. Panel C also suggests that the coefficient of, for example, the one-decade-lagged TFR in the fixed-effects panel model should be interpreted as the weighted average of the effects of a fertility change on the income growth *at least* one decade later, and the weighting diminishes with the lag length. This is because the fertility variations used in the estimation are those lagged 1–4 decades, and the share of the variation declined according to the lag length.

B2. Full results of the lagged effects

Table B2 presents the estimates for model (8) with a maximum lag length of 50 years. Column 1 presents the estimates based on all 164 countries. Columns 2, 3, and 4 focus on countries for which the 1960 GDP per capita ranked in the bottom, middle, or top one-third, respectively, of the 164 countries.

Table B2. Current and Lagged Effects of TFR on the Growth Rate of GDP Per Capita

	(1) Full sample		(2) Bottom 1/3		(3) Middle 1/3		(4) Top 1/3	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
TFR	-0.0032	(0.0006)	-0.0106	(0.0014)	-0.0054	(0.0016)	-0.0030	(0.0022)
L.TFR	-0.0035	(0.0006)	-0.0100	(0.0014)	-0.0054	(0.0016)	-0.0029	(0.0022)
L2.TFR	-0.0036	(0.0007)	-0.0099	(0.0014)	-0.0054	(0.0016)	-0.0030	(0.0022)
L3.TFR	-0.0036	(0.0007)	-0.0096	(0.0014)	-0.0053	(0.0016)	-0.0027	(0.0023)
L4.TFR	-0.0034	(0.0007)	-0.0093	(0.0015)	-0.0049	(0.0016)	-0.0019	(0.0024)
L5.TFR	-0.0030	(0.0007)	-0.0092	(0.0015)	-0.0041	(0.0016)	-0.0007	(0.0024)
L6.TFR	-0.0025	(0.0007)	-0.0092	(0.0016)	-0.0035	(0.0017)	0.0003	(0.0025)
L7.TFR	-0.0018	(0.0008)	-0.0093	(0.0016)	-0.0032	(0.0017)	0.0009	(0.0026)
L8.TFR	-0.0011	(0.0008)	-0.0089	(0.0016)	-0.0028	(0.0017)	0.0011	(0.0028)
L9.TFR	-0.0004	(0.0008)	-0.0079	(0.0016)	-0.0027	(0.0018)	0.0009	(0.0030)
L10.TFR	0.0001	(0.0008)	-0.0066	(0.0016)	-0.0029	(0.0018)	0.0004	(0.0031)
L11.TFR	0.0006	(0.0008)	-0.0051	(0.0017)	-0.0027	(0.0019)	0.0000	(0.0034)
L12.TFR	0.0013	(0.0008)	-0.0032	(0.0017)	-0.0024	(0.0020)	0.0004	(0.0035)
L13.TFR	0.0020	(0.0009)	-0.0016	(0.0017)	-0.0020	(0.0020)	0.0018	(0.0037)
L14.TFR	0.0029	(0.0009)	-0.0003	(0.0017)	-0.0014	(0.0021)	0.0040	(0.0039)
L15.TFR	0.0036	(0.0009)	0.0010	(0.0017)	-0.0004	(0.0021)	0.0068	(0.0039)

L16.TFR	0.0040	(0.0009)	0.0019	(0.0017)	0.0006	(0.0022)	0.0101	(0.0038)
L17.TFR	0.0042	(0.0009)	0.0025	(0.0018)	0.0018	(0.0022)	0.0134	(0.0037)
L18.TFR	0.0040	(0.0010)	0.0035	(0.0018)	0.0033	(0.0022)	0.0157	(0.0037)
L19.TFR	0.0042	(0.0010)	0.0050	(0.0019)	0.0050	(0.0022)	0.0180	(0.0038)
L20.TFR	0.0046	(0.0010)	0.0064	(0.0020)	0.0065	(0.0022)	0.0197	(0.0038)
L21.TFR	0.0051	(0.0010)	0.0081	(0.0021)	0.0078	(0.0022)	0.0205	(0.0039)
L22.TFR	0.0059	(0.0010)	0.0098	(0.0021)	0.0090	(0.0022)	0.0214	(0.0042)
L23.TFR	0.0068	(0.0011)	0.0112	(0.0021)	0.0099	(0.0021)	0.0228	(0.0043)
L24.TFR	0.0072	(0.0011)	0.0115	(0.0023)	0.0102	(0.0022)	0.0238	(0.0045)
L25.TFR	0.0077	(0.0012)	0.0114	(0.0025)	0.0103	(0.0023)	0.0241	(0.0049)
L26.TFR	0.0082	(0.0013)	0.0111	(0.0028)	0.0100	(0.0024)	0.0243	(0.0054)
L27.TFR	0.0082	(0.0014)	0.0102	(0.0031)	0.0092	(0.0025)	0.0242	(0.0062)
L28.TFR	0.0079	(0.0015)	0.0094	(0.0034)	0.0082	(0.0026)	0.0212	(0.0062)
L29.TFR	0.0076	(0.0015)	0.0084	(0.0036)	0.0080	(0.0025)	0.0179	(0.0061)
L30.TFR	0.0068	(0.0015)	0.0068	(0.0037)	0.0074	(0.0025)	0.0153	(0.0063)
L31.TFR	0.0055	(0.0014)	0.0051	(0.0034)	0.0065	(0.0023)	0.0116	(0.0061)
L32.TFR	0.0043	(0.0014)	0.0040	(0.0033)	0.0055	(0.0023)	0.0063	(0.0043)
L33.TFR	0.0031	(0.0014)	0.0024	(0.0031)	0.0039	(0.0023)	0.0045	(0.0044)
L34.TFR	0.0020	(0.0014)	0.0016	(0.0034)	0.0013	(0.0023)	0.0027	(0.0043)
L35.TFR	0.0013	(0.0015)	0.0022	(0.0040)	-0.0010	(0.0023)	0.0012	(0.0042)
L36.TFR	0.0010	(0.0017)	0.0038	(0.0044)	-0.0024	(0.0026)	0.0027	(0.0043)
L37.TFR	0.0002	(0.0017)	0.0043	(0.0046)	-0.0046	(0.0026)	0.0069	(0.0043)
L38.TFR	0.0001	(0.0018)	0.0057	(0.0044)	-0.0061	(0.0026)	0.0124	(0.0045)
L39.TFR	0.0002	(0.0020)	0.0078	(0.0044)	-0.0060	(0.0028)	0.0174	(0.0050)
L40.TFR	-0.0007	(0.0022)	0.0076	(0.0049)	-0.0061	(0.0031)	0.0186	(0.0052)
L41.TFR	-0.0013	(0.0025)	0.0102	(0.0062)	-0.0069	(0.0033)	0.0190	(0.0063)
L42.TFR	-0.0010	(0.0029)	0.0161	(0.0078)	-0.0060	(0.0037)	0.0172	(0.0071)
L43.TFR	-0.0023	(0.0034)	0.0198	(0.0094)	-0.0045	(0.0041)	0.0121	(0.0081)
L44.TFR	-0.0051	(0.0039)	0.0183	(0.0104)	-0.0051	(0.0049)	0.0076	(0.0084)
L45.TFR	-0.0066	(0.0044)	0.0145	(0.0097)	-0.0068	(0.0066)	0.0054	(0.0101)
L46.TFR	-0.0067	(0.0051)	0.0199	(0.0121)	-0.0075	(0.0085)	0.0072	(0.0113)
L47.TFR	-0.0058	(0.0054)	0.0229	(0.0149)	-0.0108	(0.0110)	0.0029	(0.0096)
L48.TFR	-0.0115	(0.0086)	0.0276	(0.0197)	-0.0179	(0.0149)	-0.0004	(0.0151)
L49.TFR	-0.0046	(0.0097)	0.0001	(0.0230)	-0.0239	(0.0189)	-0.0046	(0.0154)
L50.TFR	0.0115	(0.0165)	0.0120	(0.0317)	-0.0035	(0.0250)	0.0369	(0.0258)

Note: This table presents the estimates of model (8) based on the full sample or the subsamples as indicated in the header of each column. Standard errors clustered at country-level are in parentheses.

B3 Effects on the level of GDP per capita

The dependent variable in the main analysis is the growth rate of GDP per capita. In this robustness check, the dependent variable of model (8) was replaced with the log GDP per capita. The modified model further controlled for the lagged log GDP per capita to address the serial correlation of the level of GDP per capita. As presented in Figure B1, the estimated effect pattern is the same as that presented in Figure 2. I have also estimated this modified model by the Arellano-Bond System GMM dynamic panel estimation (Blundell and Bond 1998) and found a similar result.

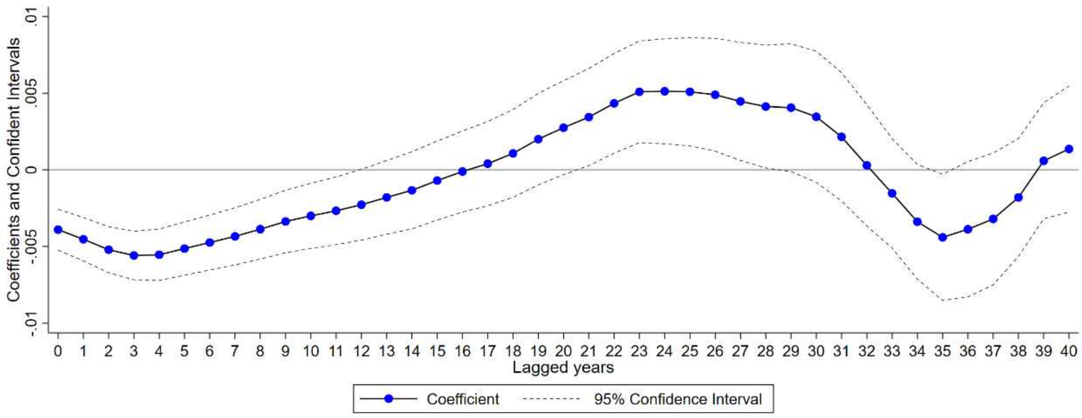

Figure B1. Current and Lagged Effects of Fertility on Log GDP per capita

Notes: This figure replicates Figure 2. The difference is that regressions generating these estimates using the log GDP per capita (instead of the growth rate) as the dependent variable.

B4. Robust to omitted variables

This appendix provides various evidence supporting that the effect pattern presented in Figure 2 is not primarily driven by omitted variables.

B4.1 *Excluding the most important time-varying control variables*

Figure B2 replicates Figure 2, but excludes the four time-varying control variables. Removing these control variables does not change the estimated effect pattern (i.e., the effect of a higher TFR is first negative and then positive and lasts for more than three decades), although the initial negative effect lasts for fewer years. Since removing these most important time-varying control variables does not change the estimated effect pattern, it is very unlikely that the effect pattern is primarily driven by other omitted variables.

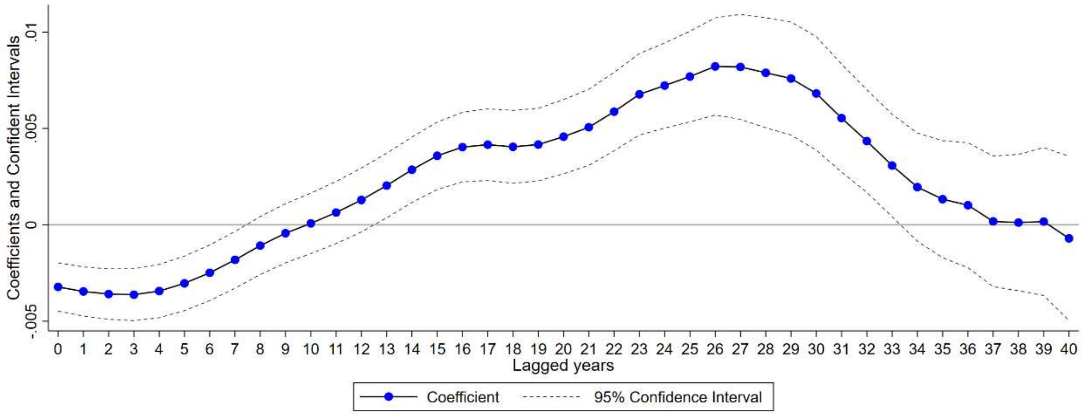

Figure B2. Robust to Excluding the Time-Varying Control Variables

Notes: This figure replicates Figure 2. The difference is that the regressions exclude the four time-varying control variables.

B4.2 *Including additional control variables*

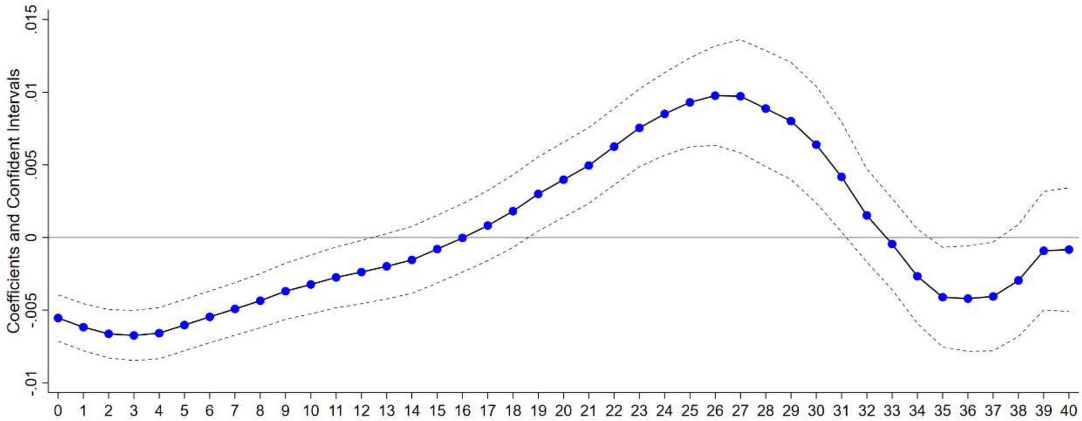

Figure B3. Robust to Including Additional Time-Varying Control Variables

Notes: This figure replicates Figure 2. The difference is that the regressions include five additional time-varying control variables.

Figure B3 replicates Figure 2, but includes five additional time-varying control variables: the crude mortality rate, life expectancy at birth, net immigration, trade share in GDP, and total natural resources rents as a share of GDP. All these additional controls are derived from the World Bank's World Development Indicators. Missing values are filled in by a linear interpolation. These control variables are excluded from the main analysis due to missing values and the concern of endogeneity. Including these five controls only has a very small effect on the estimates, suggesting again that the OLS estimates are robust to omitted variables.

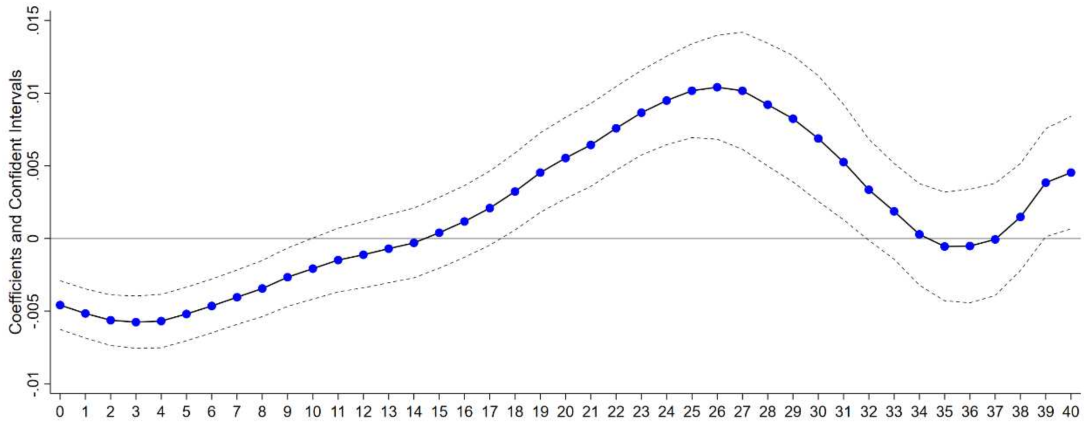

Figure B4. Robust to Including the Interactions between Time-Invariant Control Variables and Year Dummies

Notes: This figure replicates Figure 2. The difference is that the regressions additionally control for the interactions between a full set of year dummies and three time-invariant variables: landlocked, latitude, and official language.

Figure B4 replicates Figure 2, but additionally controls for the potential non-linear effects of three time-invariant variables: whether the country is landlocked, the latitude of the country's geographic center, and the first official language of the country. Although the linear effect of these time-invariant variables has been well controlled for by the country and year fixed effects, these variables could still bias the fertility estimates if their effects vary systematically over time. For example, countries using English as the official language may benefit disproportionally more from foreign technological improvements. This possibility can be examined by controlling for the interactions between a full set of year dummies and the time-invariant factors. The resulting estimates are very similar to those in Figure 2, indicating that the effects of these time-invariant factors have been well controlled for.

B4.3 *Controlling for country-specific time trends*

Figure B5 accounts for the potential serial correlation, which could be caused by omitted variables, by controlling for country-specific linear time trends. The resulting estimates have virtually no difference from those presented in Figure 2. Very similar estimates are obtained when controlling for country-specific quadratic time trends.

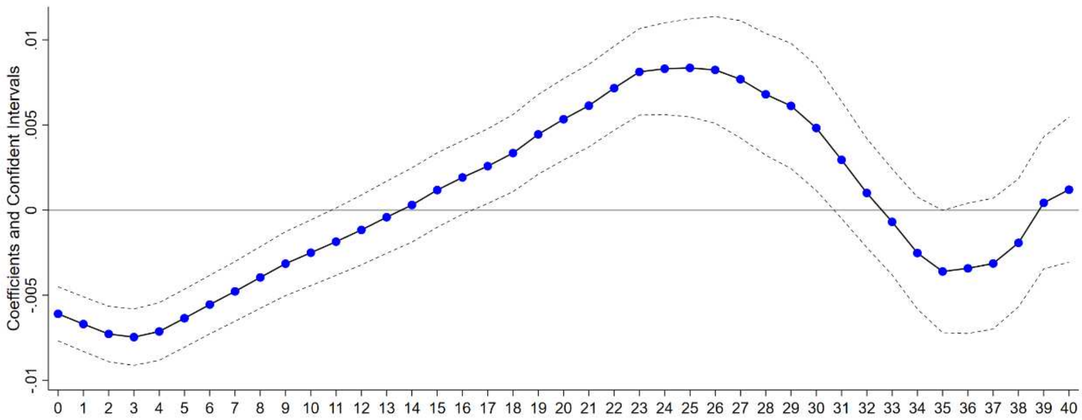

Figure B5. Robust to Country-Specific Time Trends

Notes: This figure replicates Figure 2. The difference is that the regressions control for country-specific linear time trends.

B4.4 *Clustering the error term at the region-year level*

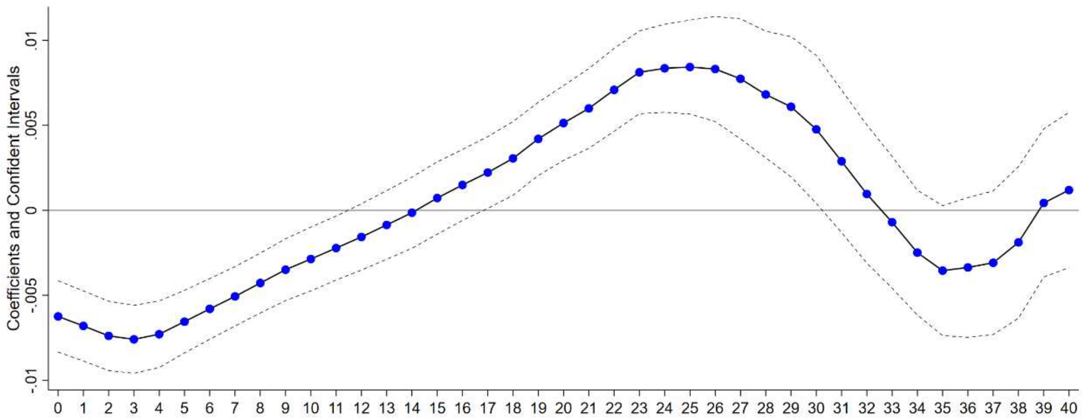

Figure B6. Robust to Clustering the Error Term at the Region-Year Level

Notes: This figure replicates Figure 2. The difference is that the error term in each regression is clustered at the region-year level.

Figure B6 accounts for the potential serial and spatial correlations, which could be caused by omitted variables, by clustering the error term at the region-year level. The regions used in the clustering refers to Asia, Africa, Europe, the Americas, and Oceania. The resulting confidence intervals are very similar to those presented in Figure 2. Similar results are found when clustering the error term at the country level, which accounts for within-country serial correlation but not cross-country spatial correlation.

B4.5 *Using the control variables that are not lagged*

The baseline estimation lags the control variables by the same years as the TFR. This is because if the control variables are not lagged, the model actually controls for the future values of the control variables (relative to the lagged TFR). If there is any correlation between the future values of control variables and the lagged TFR, it is most likely that the lagged TFR is the cause. In this case, controlling for the current values of these control variables could partly account for the true effect of the TFR and thus bias the TFR estimate towards zero. Figure B7 estimates a version of the baseline model where the control variables are not lagged. Although the estimated effect pattern is the same, the estimated marginal effects are indeed smaller.

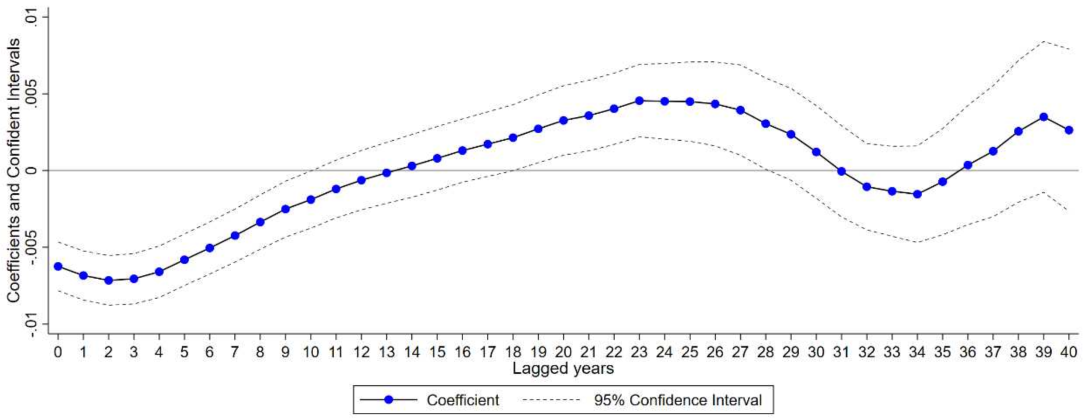

Figure B7. Robust to Using the Control Variables That are not Lagged

Notes: This figure replicates Figure 2. The difference is that control variables are not lagged.

B5 The exclusion restriction

The 2SLS estimation using the three intensity measures as the IVs depends on the assumption that the exclusion restrictions of the IVs are satisfied. This assumption can be tested by investigating whether these intensity measures are correlated with preexisting growth trends. To do this, I regress each intensity measure in the next one, three, and five years, respectively, on the current growth rate of GDP per capita. All regressions include the country and year fixed effects and the four time-varying control variables. If the association is significant, then the endogeneity problem is worthy of concern. As reported in Table B3, all estimates are statistically insignificant and extremely small (relative to the mean values of the intensity measures reported in Table A1), supporting the exogeneity assumption.

Table B3. Predictive Power of Current Income Growth Rate on Future Intensity Measures of Family Planning Programs

	Dependent variable: the intensity measures lagged by:								
	Program effort score			Message exposure			Funds per capita		
	L1	L3	L5	L1	L3	L5	L1	L3	L5
Growth rate of GDP per capita	-0.17 (0.23)	-0.15 (0.26)	-0.07 (0.21)	0.06 (0.70)	1.06 (0.99)	0.39 (1.01)	-0.69 (0.85)	-0.55 (0.69)	-0.78 (0.54)
Four control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.83	0.83	0.83	0.77	0.77	0.77	0.86	0.86	0.87

Notes: This table regresses each of the intensity measures in the next one (L1), three (L3), and five (L5) years on the current growth rate of GDP per capita, respectively. All regressions include country and year fixed effects and the four time-varying control variables. The standard errors (in parentheses) account for arbitrary heteroskedasticity. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

C. China Evidence Appendix

C1. Spatial and temporal variation in the policy fine rate

Figure C1 presents the monetary penalty rate for one unauthorized birth in each province for the period of 1979–2000. The data were derived from Ebenstein (2010). The fine rate is measured in years of local household income.

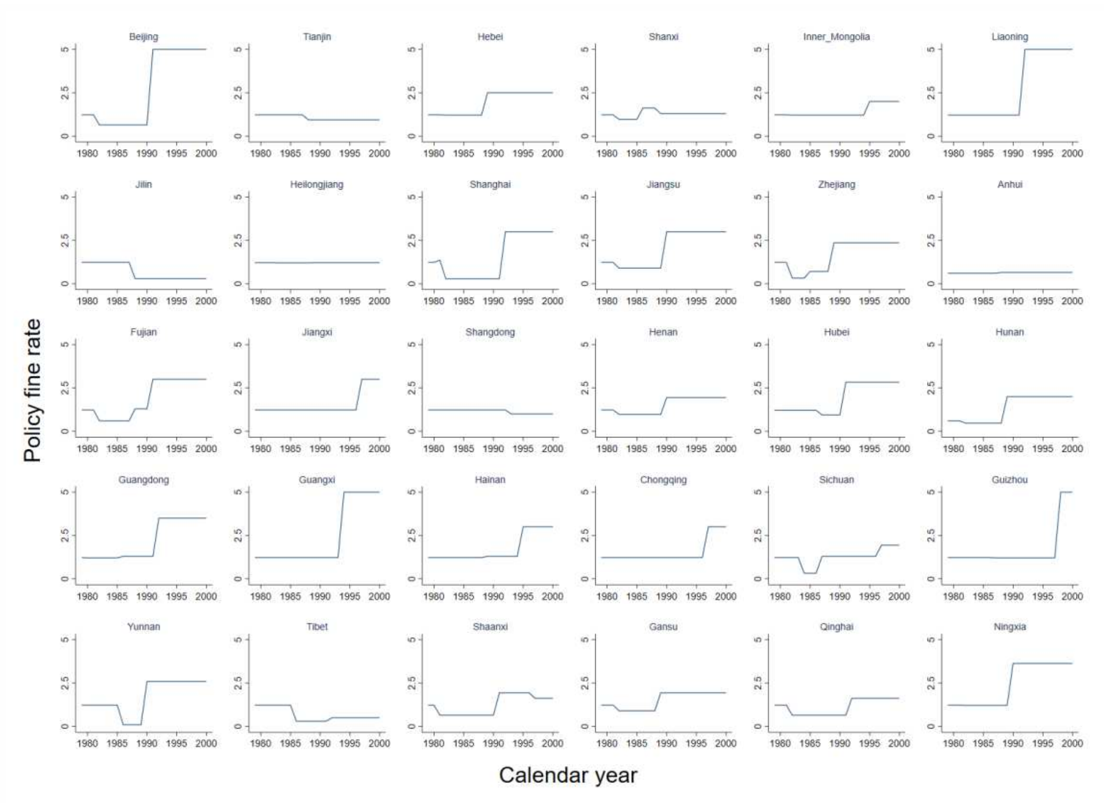

Figure C1. One-Child Policy Violation Fine Rates, 1979–2000

Data sources: Ebenstein (2010)
Notes: This figure plots the monetary penalty rate (in years of local household income) for one unauthorized birth from 1979 to 2000 in each province.

C2. Robustness of the OLS estimates

This appendix examines the robustness of the OLS estimates presented in Figure 5. Figure C2 excludes the seven time-varying control variables; Figure C3 controls for the three concurrent events (see Footnote 17); Figure C4 controls for province-specific linear time trends; and Figure C5 clusters the error term at the province level, using the bootstraps resampling procedure suggested by Cameron et al. (2008) to address the downward bias of small number of clusters (i.e., 27 provinces). All these robustness checks generate very similar estimates.

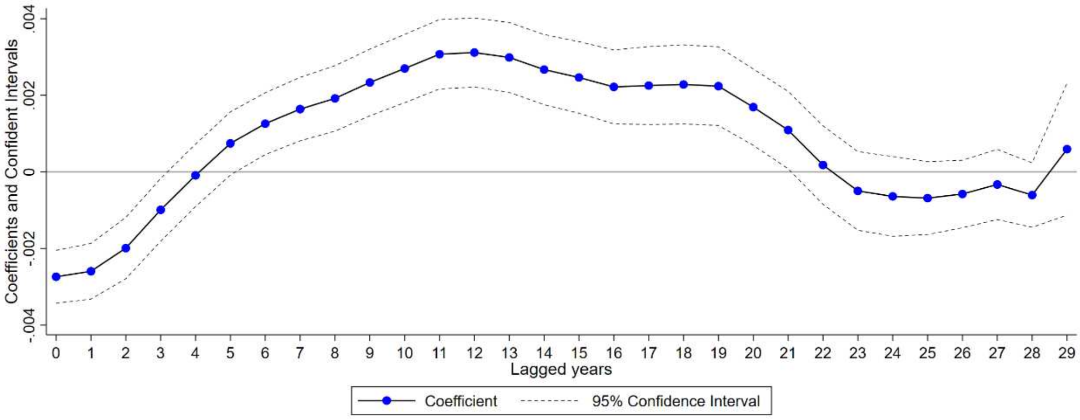

Figure C2. Robust to Excluding the Seven Time-Varying Control Variables

Notes: This figure replicates Figure 5. The only difference is that all regressions exclude the seven time-varying control variables.

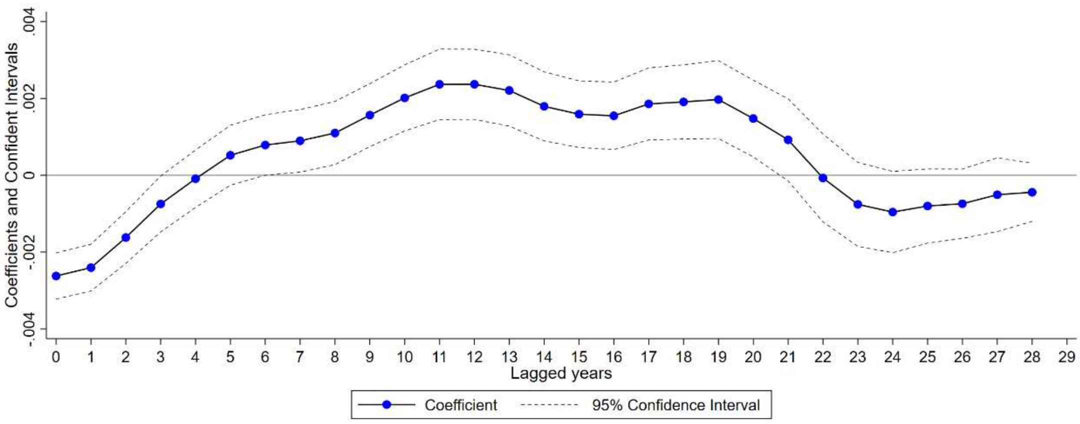

Figure C3. Robust to Controlling for Concurrent Events

Notes: This figure replicates Figure 5. The only difference is that all regressions additionally control for the five indicators of concurrent events detailed in Footnote 17.

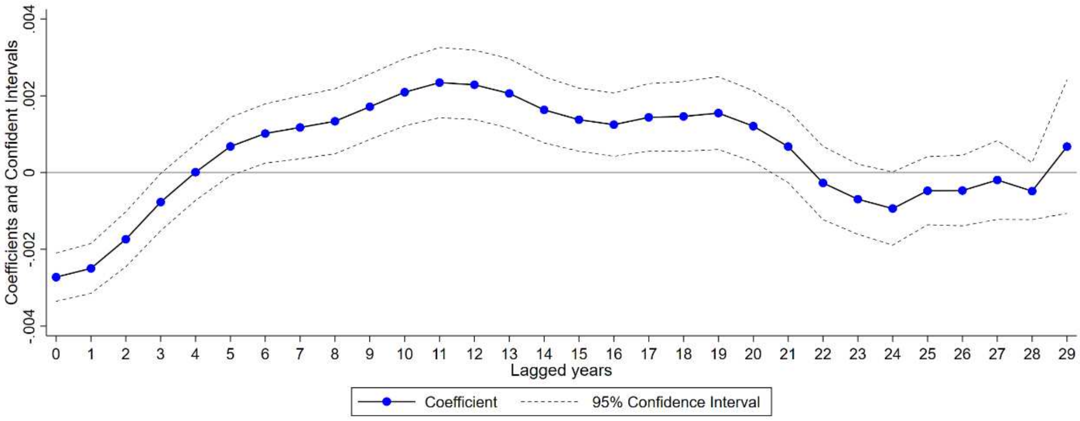

Figure C4. Robust to Controlling for Province-Specific Linear Time Trends

Notes: This figure replicates Figure 5. The difference is that all regressions control for province-specific linear time trends.

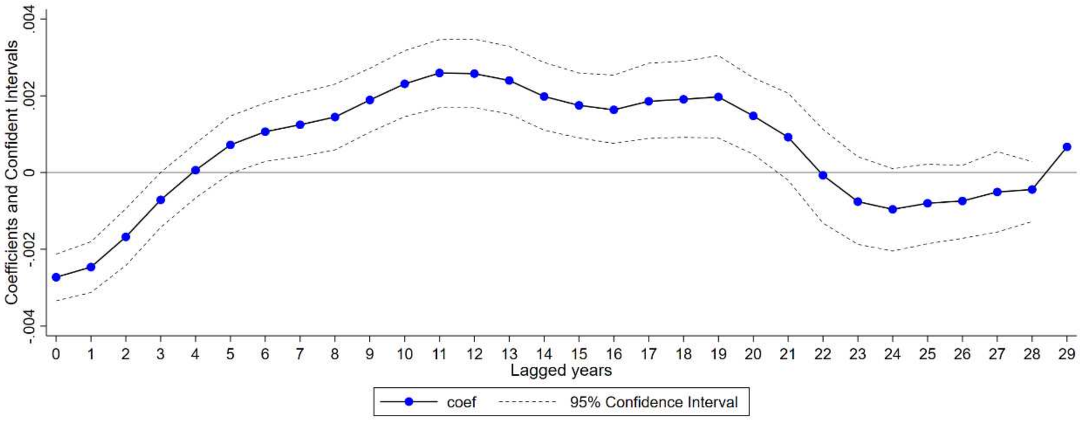

Figure C5. Robust to Clustering the Error Term at the Province Level

Notes: This figure replicates Figure 5. The difference is that the error terms are clustered at the province-level, based on the bootstraps resampling procedure suggested by Cameron et al. (2008). The confidence intervals cannot be calculated in the bootstraps procedure when the lag length is 29 years because of a too small sample size.

C3. First-stage estimates of the 2SLS estimation

Table C1 reports the first-stage estimates of the 2SLS estimation of model (10). Specifically, column 1 regresses the crude birth rate on the policy fine rate in a panel model that includes province and year fixed effects and the seven time-varying control variables. The estimates confirm that there is a significantly negative association between the policy fine rate and the crude birth rate. The remaining columns contain robustness tests. Column 2 excludes the seven time-varying control variables, column 3 controls for concurrent events using the indicates detailed in Footnote 17, column 4 controls for province-specific linear time trends, and column 5 clusters the error term at the province level using the bootstrap procedure. Estimates from these robustness checks are similar to that in column 1, and *t*-tests did not find statistically significant differences between them.

Table C1. First-Stage Regression Results

	Dependent variable: The crude birth rate				
	(1)	(2)	(3)	(4)	(5)
Policy fine rate (years of household income)	-0.31*** (0.11)	-0.46*** (0.10)	-0.38*** (0.11)	-0.34*** (0.11)	-0.34*** (0.11)
Seven time-varying control variables	Yes		Yes	Yes	Yes
Indicates of concurrent events			Yes	Yes	Yes
Province time trend				Yes	Yes
Clustering stander error					Yes
Province FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	781	781	781	781	781
R-squared	0.884	0.888	0.866	0.886	0.886

Notes: This table regresses the crude birth rate on the policy fine rate in a panel model. All columns include the province and year fixed effects. Column 1 includes the seven control variables, column 2 excludes the seven control variables, column 3 additionally controls for the concurrent events, column 4 controls for province-specific time trends, and column 5 clusters the error term at the province level. The standard errors (in parentheses) account for arbitrary heteroskedasticity. Significance levels are *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

C4. Robustness of the 2SLS estimates

This appendix examines the robustness of the 2SLS estimates in Figure 6. Figure C6 excludes the seven time-varying control variables; Figure C7 controls for the reform and opening-up in 1978 and the tax system reform in 1994 (see Footnote 17); Figure C8 controls for province-specific linear trends; Figure C9 clusters the error term at the province level, using the bootstraps resampling procedure; and Figure C10 allows for an additional lag in the policy fine rate to reflect the lag in the translation of the policy fine rate change to fertility change. All these robustness checks lead to estimates comparable to those presented in Figure 6.

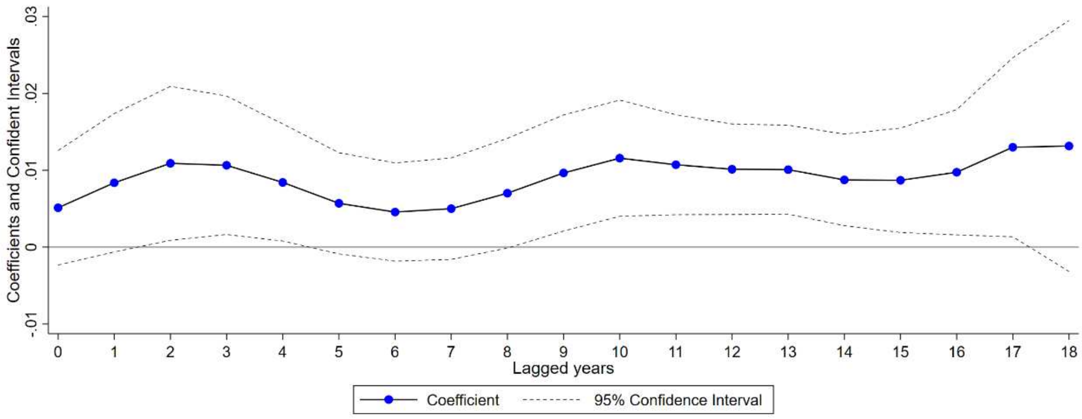

Figure C6. Robust to Excluding the Seven Time-Varying Control Variables

Notes: This figure replicates Figure 6. The only difference is that all regressions exclude the seven time-varying control variables.

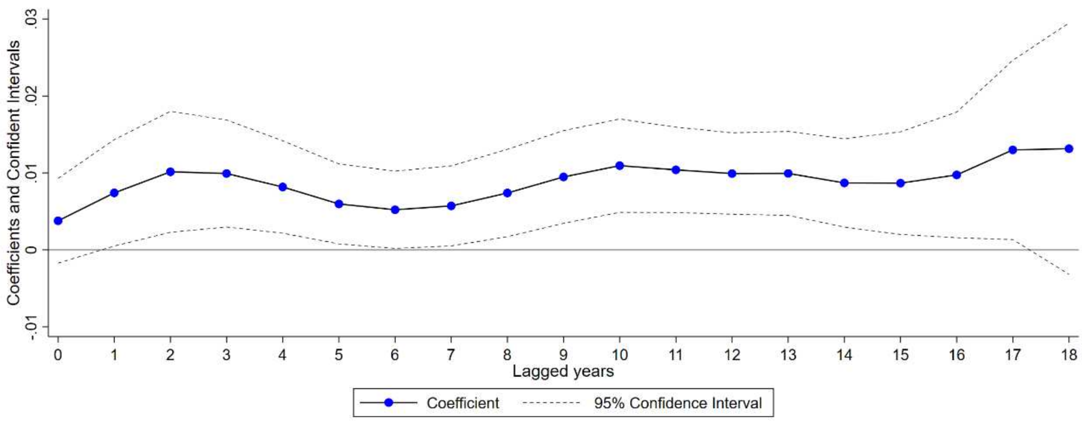

Figure C7. Robust to Controlling for Concurrent Events

Notes: This figure replicates Figure 6. The only difference is that all regressions additionally control for the reform and opening-up in 1978 and the tax system reform in 1994.

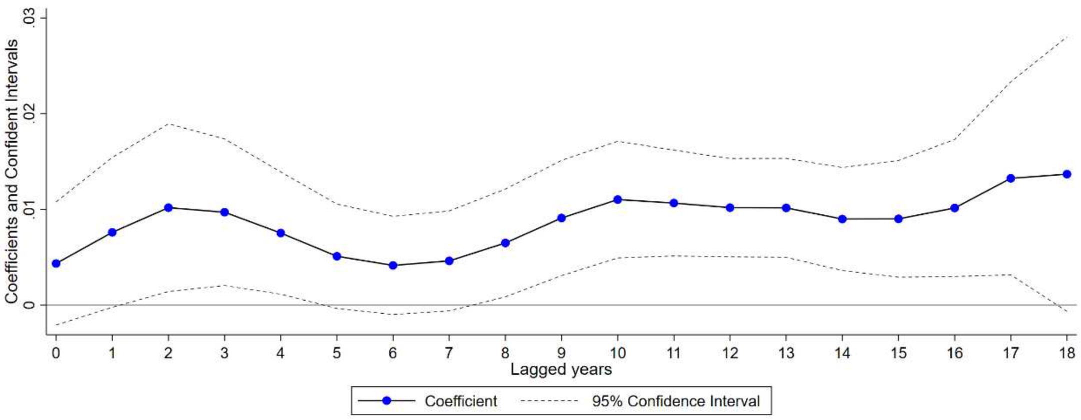

Figure C8. Robust to Controlling for Province-Specific Linear Time Trends

Notes: This figure replicates Figure 6. The difference is that all regressions control for province-specific linear time trends.

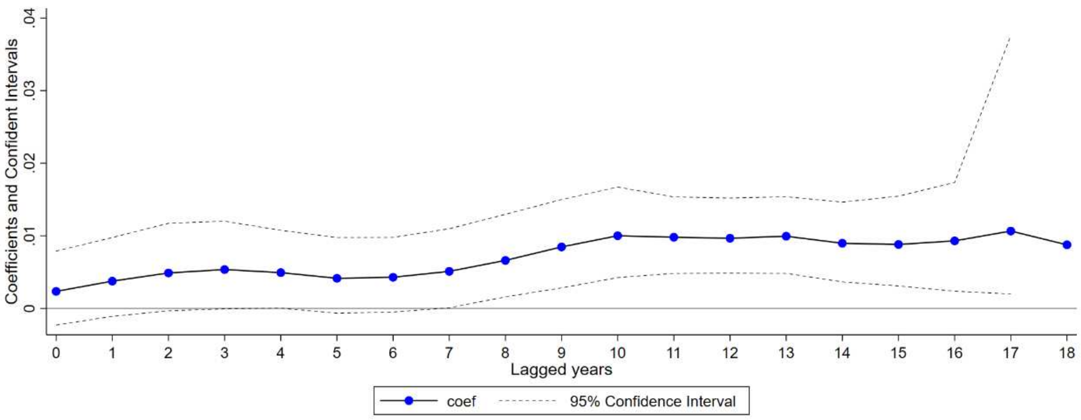

Figure C9. Robust to Clustering the Error Term at the Province Level

Notes: This figure replicates Figure 6. The difference is that the error terms are clustered at the province-level, based on the bootstraps resampling procedure suggested by Cameron et al. (2008).

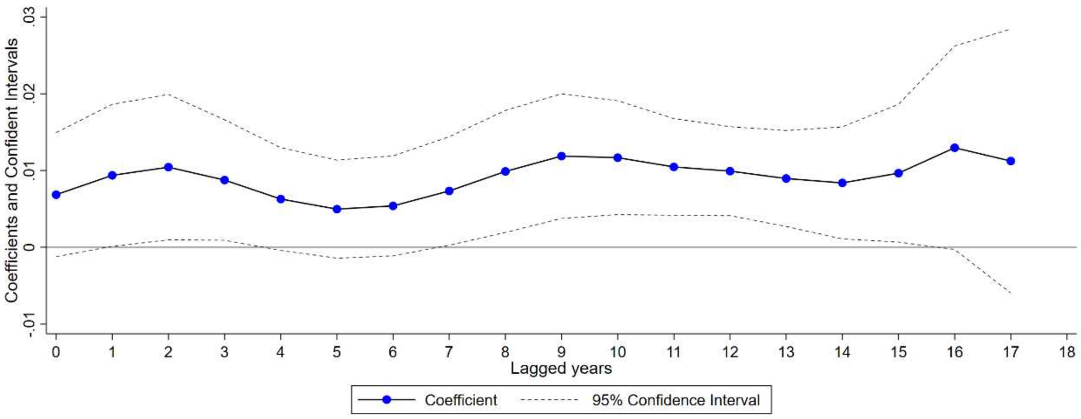

Figure C10. Robust to Allowing for an Additional Lag in the Policy Fine Rate

Notes: This figure replicates Figure 6. The only difference is that the $Fine_{p(t-s)}$ in the first-stage regression is replaced by $Fine_{p(t-s-1)}$ to allow a lag in the translation of the policy fine rate change to fertility change.

C5. Exogeneity of the policy fine rate

A critical assumption of using the policy fine rate as the IV is that it is exogenous to income growth. This appendix presents three pieces of evidence supporting this assumption.

C6.1 *The predictive power of preexisting income levels or growth rates*

A major endogeneity concern is that the policy fine rate could be determined by income levels or growth rates. To relieve this concern, Table C2 regresses the policy fine rates in the next one, three, and five years on the current growth rate (or level) of GDP per capita, respectively, in columns 1a, 2a, and 3a (columns 1b, 2b, and 3b.) All estimates are small and with a p-value larger than 0.1, suggesting that the policy fine rate is not determined by preexisting income growth rates or levels. Although it is infeasible to directly examine whether the policy fine rate is determined by expectations of future incomes, it seems reasonable to believe that if the policy fine rates were not set based on the readily available information of past incomes, they were even less likely to be set based on the uncertain predictions of future incomes.

Table C2. Predictive Power of Preexisting Incomes on the Policy Fine Rates

	Dependent variable: One-child policy fine rate					
	1-year later		3-year later		5-year later	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Growth rate of GDP per capita (%)	-0.002		0.015		0.025	
	(0.029)		(0.026)		(0.026)	
Log GDP per capita		0.278		0.193		-0.034
		(0.707)		(0.817)		(0.831)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year trends	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.69	0.69	0.70	0.70	0.73	0.73

Notes: This table regresses the policy fine rates in the next one, three, and five years on the current growth rate (or level) of GDP per capita, respectively, in columns 1a, 2a, and 3a (columns 1b, 2b, and 3b.) All regressions include the province and year fixed effects as well as the province-specific linear year trends. The standard errors (in parentheses) are clustered at the province level. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

C6.2 *Correlations with time-varying income determinants*

Another endogeneity concern is that the policy fine rate may be correlated with omitted determinants of income growth. This concern has been substantially reduced by the province and year fixed effects and the various time-varying control variables included. Therefore, the remaining concern pertains only to the omitted province-specific, time-varying factors. Although it is impossible to examine the correlation with unobservable factors, this concern can be reduced if the policy fine rate is not correlated with even the most important observable factors.

Table C3. *P*-values for the Correlations of the Policy Fine Rates with Nine Growth Determinants

	Dependent variable: One-child policy fine rate			
	(1) 1-year later	(2) 3-year later	(3) 5-year later	(4) First difference
Share of labor with secondary education	0.19	0.12	0.31	0.19
Share of urban population	0.36	0.49	0.76	0.36
Crude death rate	0.13	0.20	0.10	0.13
Out-migration rate	0.16	0.18	0.35	0.16
Trade share in GDP	0.24	0.20	0.18	0.24
Government spending share	0.41	0.41	0.26	0.41
Trade-to-GDP ratio×1978 dummy	0.24	0.20	0.18	0.24
Distance to the nearest port ×1978 dummy	0.57	0.54	0.51	
Government spending share×1994 dummy	0.52	0.47	0.42	0.52

Notes: Columns 1, 2, and 3 regress the policy fine rates in the next one, three, and five years, respectively, on the current value of each of the nine growth determinants. Column 4 regresses the changes in the policy fine rate on changes in each of the growth determinants (except for the Distance to the nearest port ×1978 dummy, which has not enough variation for the estimation.) All regressions include the province and year fixed effects and the province-specific linear trends. All values reported are p-values. The standard errors used for calculating the p-values are clustered at the province level.

Table C3 examines the correlations between the policy fine rate and a set of nine time-varying growth determinants.[24] Specifically, I regressed the policy fine rates in the next one, three, and five years, respectively, on each of the nine determinants in a

[24] These are the control variables used in the main regressions or robustness checks. The 5-year lagged GDP per capita is not examined here because it has been examined in Table C2. The indicators for joining the WTO in 2001 are also not examined because the policy fine data used are during 1980–2000.

panel model with province and year fixed effects (columns 1, 2, and 3 of Table C3). I also examined whether changes in these determinants are correlated with changes in the policy fine rate in a similar model setting (column 4). None of the p-values associated with the coefficients of these nine variables are smaller than 0.1, suggesting no significant correlation of them with the policy fine rate. I have also examined the joint significance of all or subsets of these variables and still found no significant association.

C6.2 *Effects of the lead of the policy fine rate*

To the extent that the policy fine rate captures the impact of the OCP's strictness on fertility rather than differential trends across provinces (which could be caused by omitted variables), the future policy fine rate should not predict current fertility and income growth. Table C4 examines the effects of the five-year lead policy fine rate on income growth and fertility by including it as a control variable in the first- and second-stage regressions of the 2SLS estimation. To facilitate the comparison, column 1 of Table C4 lists the baseline estimates that were presented in Figure 6. For brevity, the table only reports the estimates from regressions in which the CBR is lagged by 5 or 10 years; the findings are similar for other lags.

Panel A presents the effect of the five-year lead policy fine rate on the income growth rate. The estimated coefficient for the five-year lead policy fine rate is extremely small and statistically insignificant. In addition, including the five-year lead policy fine rate as a control variable has a very small effect on the estimated effect of the CBR. Panel B presents the effect of the five-year lead policy fine rate on fertility. Similarly, the five-year lead policy fine rate has no significant effect on fertility. Therefore, the evidence does not support the concern that the policy fine rate captures the impact of differential trends across provinces.

Table C4. Effects of the Lead of the Policy Fine Rate on Income Growth and Fertility

	Baseline		Controlling for the 5-year lead fine rate	
	(1a)	(1b)	(2a)	(2b)
	Panel A: The second-stage estimates (Dependent variable: the growth rate of GDP per capita)			
5-year lagged CBR	0.004**		0.004**	
	(0.002)		(0.002)	
10-year lagged CBR		0.01***		0.012***
		(0.002)		(0.004)
5-year lead policy fine rate			0.0002	0.0008
			(0.001)	(0.003)
	Panel B: The first-stage estimates (Dependent variable: 5-year and 10-year lagged CBR in columns a and b, respectively)			
5-year lagged policy fine rate	-0.47***		-0.54***	
	(0.11)		(0.13)	
10-year lagged policy fine rate		-0.54***		-0.59***
		(0.13)		(0.15)
5-year lead policy fine rate			-0.01	-0.02
			(0.19)	(0.29)
Province FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R-squared	0.86	0.82	0.81	0.77

Notes: This table examines the effects of the five-year lead policy fine rate on income growth and fertility by including it as a control variable in the 2SLS estimation of model (10). Panel B presents the first-stage estimates, and Panel A presents the second-stage estimates. The standard errors (in parentheses) are clustered at the province level. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

D. Causal Evidence Based on Local Policy Violations

An alternative intensity measure of the OCP is the extent of local OCP violation. Based on microdata from the 1982 Chinese Population Census, Bingjing Li and Hongliang Zhang (2017) constructed the excess fertility rate (EFR) as a measure of local violations of the OCP. The EFR was constructed as the percentage of Han Chinese mothers aged 15–49 years who gave a higher order birth in 1981. They found substantial regional differences in the EFR and used it as an exogenous source of variation in fertility to examine the causal effects of child quantity on child quality. Using a similarly constructed EFR, Junsen Zhang (2017) examined the effect of the OCP on marital status, labor supply, and migration. Since only two waves of the population censuses (in 1982 and 1990) contained sufficient information to construct a provincial EFR, the EFR cannot be used as an IV to identify the dynamic causal effects.[25] Instead, this appendix uses the EFR as an intensity measure in a DID model to estimate the long-run average causal effect of a decline in fertility.

D1. The excess fertility rate

Regional differences in such factors as implementation methods and work styles could lead to differential local violation of the OCP. This appendix follows Bingjing Li and Hongliang Zhang (2017) to construct the EFR as a measure of local violation of the OCP, using the microdata from the 1982 and 1990 Chinse Population Censuses, which contained information for 1981 and 1989, respectively. The EFR is constructed as follows:

$$EFR_{p,t} = \frac{\sum_j \left(Birth_{pjt} \times 1\left(NSC_{pjt} \geq 2 \right) \right)}{\sum_j 1\left(NSC_{pjt} \geq 1 \right) - \sum_j \left(Birth_{pjt} \times 1\left(NSC_{pjt} = 1 \right) \right)} \times 100 \quad , \tag{11}$$

[25] There were three Censuses during the sample period, but the publicly available microdata from the 2000 census do not contain sufficient geographic information to construct the provincial EFR.

where $Birth_{pjt}$ is an indicate of whether woman j in province p gave a birth in year t (either 1981 or 1989), and NSC_{pjt} denotes her number of surviving children for woman j by the end of year t. I calculate the $EFR_{p,t}$ for all *Han* Chinese women aged 15–49 from *non-agricultural* households. Thus, the $EFR_{p,t}$ measures the percentage of non-agricultural Han mothers (i.e., those with at least one surviving child) aged 15–49 who gave a higher order birth in year t.[26] This construction is slightly different from that of Bingjing Li and Hongliang Zhang (2017); their construction focused on all Han Chinese mothers (from both agricultural and non-agricultural households) aged 25–44 (instead of 15–49). Focusing on non-agricultural Chinese mothers helped to avoid a potential bias due to pre-existing correlations between income growth and the share of rural residents (couples with an agricultural *hukou* were allowed to have a second child if the first was a girl, see Footnote 13). A robustness check (row 3 of Table D3) shows that using the EFR constructed for both agricultural and non-agricultural mothers leads to a comparable result.

The EFR would be 0 if the OCP was strictly enforced with no violations, and a larger EFR value corresponds to more relaxed enforcement. As presented in Figure D1, the EFR indicates that the OCP was not perfectly enforced. For example, the 1981 EFR for non-agricultural Han Chinese women ranged from 0.92–7.89 across provinces, with a mean of 4.24 and a standard deviation of 1.60. Figure D1 also shows that the EFR is positively and significantly correlated with the crude birth rate.

[26] I used the number of surviving children in mid-1982 to proxy for that in end-1981, which was not available from the census.

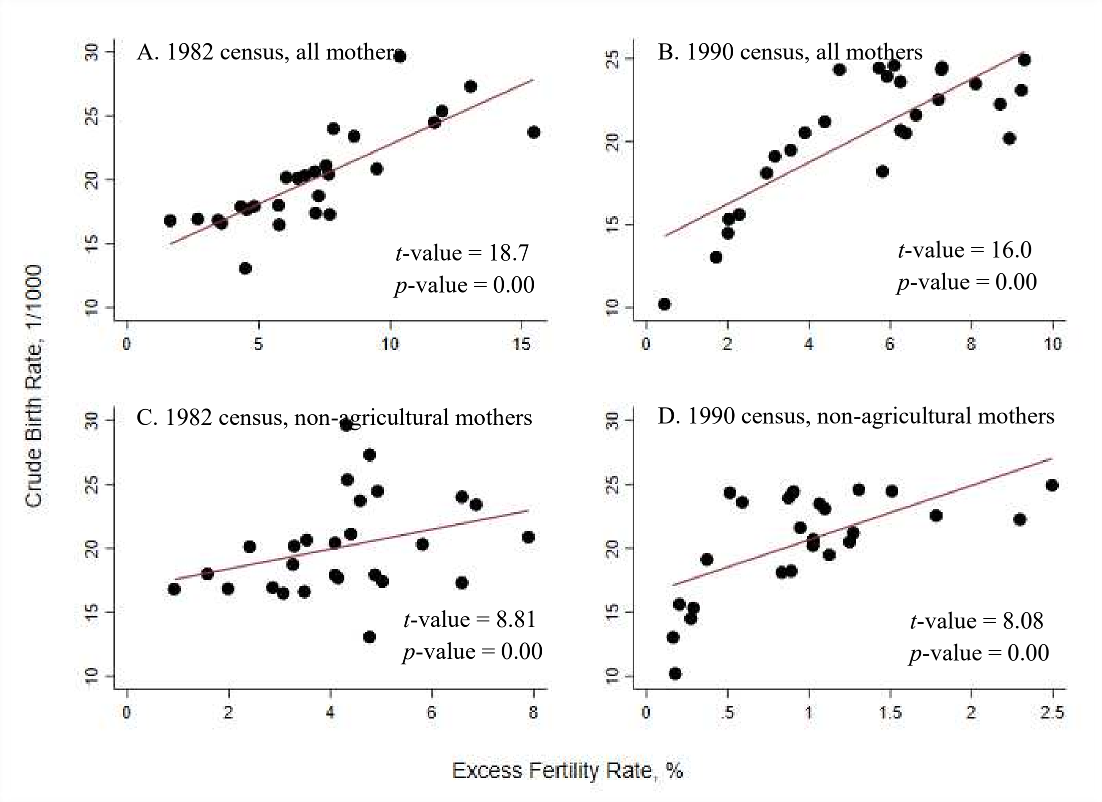

Figure D1. Correlation between the CBR and the EFR Calculated based on the Population Censuses 1982 and 1990

Notes: The correlations are plotted for the 27 sample provinces in China. The EFR in Panel A (Panel B) is calculated for all Han Chinese mothers using 1982 (1990) Census data, while the EFR in Panel C (Panel D) is calculated for non-agricultural Han Chinese mothers using 1982 (1990) Census data.

D2. The DID estimates

This subsection employs variations from both the timing of the OCP and the local violation thereof in a reduced-form model to examine the long-run average effect of fertility change on income growth. The estimation strategy is to compare the relative changes in economic growth between provinces with high violation and low violation of the OCP in the post-OCP period to that in the pre-OCP period. The estimation equation is written as follows:

$$y_{pt} = \nu_p + \tau_t + \alpha_1 EFR_{p,1981} \times post_t + Z_{pt}\lambda + \vartheta_{pt} \quad , \tag{12}$$

where $EFR_{p,1981}$ is the excess fertility rate of province p in 1981, $post_t$ is an indicator variable that equals one for the periods after and including 1980, and ϑ_{pt} is the error term. Other variables are defined the same as in the main text. Coefficient α_1 captures the additional income growth experienced after the OCP by provinces with higher OCP violation rates. Since the OCP caused a long-run fertility difference between provinces with different EFRs, coefficient α_1 reflects the long-run average effect of a change in fertlity on income growth.

Table D1. Difference-in-Differences Estimates of the Effect of OCP Violation on the Growth Rate of GDP per capita

	Low Violation	High Violation	Difference
Panel A: Experiment of Interest (growth rate of GDP per capita)			
1960-1979 average	0.034	0.027	-0.007**
			(0.003)
1980-2010 average	0.083	0.089	0.006**
			(0.002)
Difference-in-differences			**0.013*** **
			(0.004)
Panel B: Control Experiment (growth rate of GDP per capita)			
1960-1969 average	0.011	0.003	-0.008
			(0.005)
1970-1979 average	0.057	0.049	-0.008
			(0.005)
Difference-in-differences			**-0.0005**
			(0.007)

Notes: This table disaggregates the sample provinces into two approximately equal-sized groups according to their 1981 EFR and compares the growth rates between these two groups in different periods. Standard errors are in parentheses. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

This estimation strategy can most clearly be illustrated by the simplified DID estimates presented in Panel A of Table D1. The table disaggregates the sample provinces into two groups that approximately equal in size—the "high-violation" group and the "low-violation" group—according to the 1981 EFR. It then compares the average growth rates of GDP per capita during 1960–1979 to that during 1980–2010 across the two groups of provinces. The DID estimates presented in the final column indicate that after 1980, the high-violation provinces experienced a growth rate that was 1.3 percentage points higher, with a standard error of 0.4. Because the EFR is strongly and positively correlated with the CBR (see Figure D1), this DID estimate suggests that provinces with higher fertility experienced faster income growth.

This estimation strategy is based on a parallel-trends assumption that in the absence of the OCP, changes in the income growth rate would not have been systematically different across the low- and high-violation provinces. This assumption is supported by two pieces of evidence. Panel B in Table D1 shows that prior to 1980, there was no significant difference between the changes in the income growth rate of low- and high-violation provinces. Specifically, I compare the average growth rates for 1960–1969 and that for 1970–1979 across low- and high-violation provinces and find a DID estimate close to zero (-0.05 percentage points). In addition, the next subsection will estimate a flexible version of model (12) that includes interactions between the 1981 EFR and a full set of year dummies. The estimation found that prior to the OCP, the effects of the EFR on income growth were all small and statistically insignificant, but after the policy was implemented, the EFR effects were all positive, noticeably larger, and mostly statistically significant.

The estimates for model (12) are presented in Table D2. The estimations were based on 1970–2010 data for the 27 sample provinces.[27] The baseline estimates presented in column 1 show that a one-percentage-point increase in the EFR raised the growth rate of GDP per capita by 0.62 percentage points, and the effect was statistically significant at the 1% level. Because the EFR is strongly and positively correlated with the CBR, this estimate suggests that the long-run average effect of higher fertility rates on the aggregate income growth rate is significantly positive.

[27] The data prior to 1970 were not used in this estimation because of the concern that China provincial data prior to 1970 might be unreliable. A similar result was found when the data were extended back to 1960.

Table D2. Effect of Local OCP Violation on Long-Run Income Growth

	Dependent variable: Growth rate of GDP per capita				
	(1)	(2)	(3)	(4)	(5)
$EFR_{1981} \times dummy_{1980}$	0.0062***	0.0060***	0.0049***	0.0047***	
	(0.0011)	(0.0012)	(0.0012)	(0.0012)	
$CBR_t \times dummy_{1980}$ (IV: $EFR_{1981} \times dummy_{1980}$)					0.0097***
					(0.0032)
Seven time-varying controls		Yes	Yes	Yes	Yes
Fertility preferences × all year dummy			Yes	Yes	Yes
Controls for the reform and opening-up in 1978					
Trade share in GDP $\times dummy_{1978}$				Yes	Yes
Distance to port $\times dummy_{1978}$				Yes	Yes
Control for the tax system reform in 1994					
Government spending share $\times dummy_{1994}$				Yes	Yes
Controls for joining the World Trade Organization in 2001					
Trade share in GDP $\times dummy_{2001}$				Yes	Yes
Share of services in GDP $\times dummy_{2001}$				Yes	Yes
Province fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
First-stage F-statistics					26.7
Observations	1,107	1,107	1,107	1,107	1,107
R-squared	0.519	0.598	0.662	0.679	0.328

Notes: Column 1 presents the baseline OLS estimate of model (12). Columns 2–4 provide robustness checks that increasing include more sets of control variables. Column 5 contains the 2SLS estimates of a modified version of model (12) that replaces $EFR_{p,1981} \times post_t$ by $CBR_{p,t} \times post_t$. The standard errors (in parentheses) account for arbitrary heteroskedasticity. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

Robustness tests are presented in columns 2–4 of Table D2 and in subsection D4. Column 2 of Table D2 includes the seven time-varying control variables; column 3 further controls for preexisting fertility preferences, which are measured by the interactions between a full set of year dummies and the average total number of births of females aged 45–54 in 1981 (calculated from the microdata from the 1982 Census); column 4 further controls for three events that had the potential to confound the estimated effects: the reform and opening-up in 1978, tax system reform in 1994, and joining the World Trade Organization (WTO) in 2001 (see Footnote 17 for details). Including these control variables leads to slightly smaller estimates, but t-tests found no significant difference from the baseline estimate reported in column 1. In addition, Section D4 shows that the finding is robust to alternative EFR constructions, subsamples, and series correlation.

To obtain the marginal effect of fertility, column 5 of Table D2 provides the 2SLS estimate of a modified version of model (12) that replaces $EFR_{p,1981} \times post_t$ with $CBR_{p,t} \times post_t$. The first-stage regression of the 2SLS estimation is

$$CBR_{p,t} \times post_t = \nu_p + \tau_t + \beta EFR_{p,1981} \times post_t + Z_{pt}\lambda + \mu_{it} \quad . \tag{13}$$

The 2SLS estimate suggests that a one-unit increase in the CBR increased the average growth rate of GDP per capita for 1980–2010 by 0.97 percentage points. Since the average CBR during this period was 15.4, the 2SLS estimate suggests that a 1% increase in fertility leads to an income growth rate that is 0.15 percentage points higher.

D3. Exogeneity of the EFR

A crucial assumption of identifying the causal effect by equation (12) is that provinces with different EFRs would have the sample growth trends if without the OCP. If this assumption is true, the EFR should have no effect on income growth prior to the OCP. As such, this assumption can be tested by estimating the following flexible version of equation (12) that includes the interactions between the 1981 EFR and a full set of year dummies:

$$y_{pt} = \nu_p + \tau_t + \sum_{j=1971}^{2010} \alpha_j EFR_{p,1981} \times dummy_j + Z_{pt}\lambda + \vartheta_{pt} \quad , \tag{14}$$

where $dummy_j$ equals 1 in year j. The estimated vector of α_js reveals the correlation between the EFR and the growth rate in each year. If the EFR was not correlated with growth trends prior to the OCP, then the estimated α_js would be expected to be close to zero for the years before the OCP was implemented.

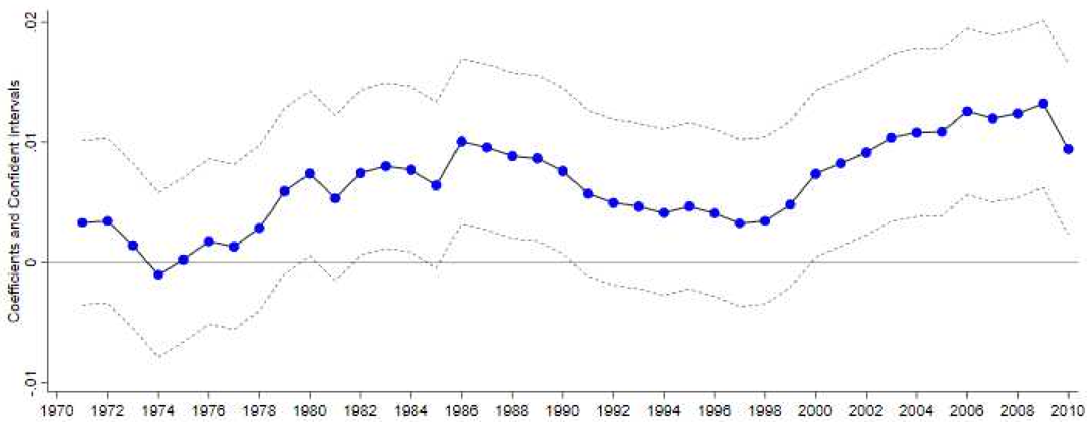

Figure D2. Flexible Estimates of the Relationship between the EFR and the Income Growth Rate

Note: Each dot on the solid line is the point estimate of a_js from equation (14), and the broken lines indicate the corresponding 95% confidence intervals. The equation is estimated using annual data from 1970 to 2010 for the 27 Chinese provinces. The estimation controls for the province and year fixed effects as well as the seven time-varying control variables.

Equation (14) is estimated based on the 1970–2010 data for the 27 sample provinces. Figure D2 plots the point estimates of α_js (dots on the solid line) and the corresponding 95% confidence intervals (the broken lines). During the 1971–1979 period, the estimated coefficients are all small and statistically insignificant, which supports the assumption that the EFR was not correlated with growth trends prior to the OCP. The figure also shows that the coefficients after 1980 are much larger and mainly statistically significant after 1980, which suggests that provinces with higher OCP violations experienced faster income growth after 1980.

D4. Further robustness tests

Table D3 provides five additional robustness tests for the baseline estimates reported in column 1 of Table D2. All robustness tests have the same model setting as the baseline estimation, except for the one specified in each test. For simplicity, only the estimated coefficient of the EFR is reported. To facilitate comparison, row 1 replicates the baseline estimate.

Table D3. Robustness Tests of the Effect of the EFR on Income Growth

	Coefficient of interest	Standard error
(1) The baseline estimate from column 1 of Table D2	0.0062***	(0.0011)
(2) The EFR calculated from the 1990 Population Census	0.0027***	(0.0008)
(3) The 1981 EFR for both agricultural and non-agricultural mothers	0.0022***	(0.0007)
(4) Excluding provinces with minority population share higher than 10%	0.0051***	(0.0011)
(5) Controlling for province-specific time trends	0.0063***	(0.0011)
(6) Clustering the standard errors at the province level	0.0062***	(0.0021)

Notes: All robustness tests have the same model setting as that in column 1 of Table D2, except for the one specified in each test. The standard errors (in parentheses) account for arbitrary heteroskedasticity (and clustered at the province level in row 6). Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

Row 2 uses the EFR calculated from the 1990 Census instead of that from the 1982 Census. The estimated effect of the EFR is still positive and statistically significant, but it is smaller, potentially because the 1990 EFR captures the average effect over a shorter period (1990–2010). Row 3 uses the 1981 EFR calculated for both agricultural and non-agricultural mothers (recall that the baseline analysis only used the 1981 EFR calculated for the non-agricultural mothers) and presents a smaller marginal effect. This finding is reasonable because the mean value of the 1981 EFR calculated in this way is approximately two-times larger. Rows 4 excludes the five provinces with a minority population share that is greater than 10% in order to further address the concern that minority provinces might have different growth trends from other provinces (recall that the EFR is only calculated for Han Chinese mothers). The estimated effect is slightly smaller, but there is no statistically significant difference from row 1. Rows 5 and 6 examine the robustness to series correlation by controlling for province-specific time trends and clustering the standard errors at the province level, respectively. The resulting estimates are very close to the baseline estimate.

E. Causal Evidence based on the Minority Population Share

A third intensity measure of the OCP is the ethnic minority population share (MPS). Recall that ethnic minorities were subjected to less-strict birth control measures during the OCP. The MPS was used as an IV for fertility by Hongbin Li and Junsen Zhang (2007) when they examined the effect of fertility on income growth. As detailed below, however, the province-level MPS is endogenous in the sense that even after controlling for the province and year fixed effects and various time-varying factors, it is strongly correlated with preexisting growth trends. As such, the main analysis of this paper does not depend on this intensity measure. Nevertheless, comparable dynamic effects are found when using the MPS as the IV for fertility.

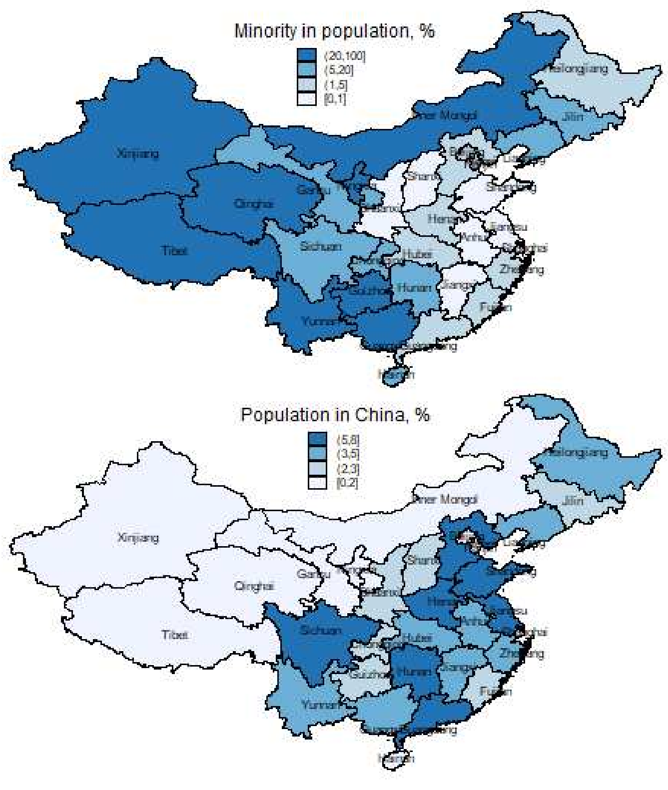

Figure E1. Percentage of minorities in each province (upper panel) and percentage of each province's population in China (lower panel), 1980–2010

Notes: The figure only shows the 31 mainland Chinese provinces. See data sources from Table A3.

The endogeneity of the MPS is a concern because of the fact that minorities only comprised a small share (about 10%) of the Chinese population and most minorities live in non-presentative western provinces. Figure E1 shows that minorities mainly live in the seven western provinces, which together contain less than 10% of the Chinese population but cover more than half of China's territory. It is difficult to believe that the western minority provinces, which have significantly lower populations and economic densities, could experience the same growth trends as other provinces.

Table E1. Predictive Power of the Current Income Growth Rate on the Future Minority Population Share

	Dependent variable: The minority population share		
	1-year later	3-year later	5-year later
	(1)	(2)	(3)
Growth rate of GDP per capita	-0.74***	-0.87***	-0.68***
	(0.023)	(0.022)	(0.020)
7 control variables	Yes	Yes	Yes
Province FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R-squared	0.99	0.99	0.99

Notes: This table examines whether the current income growth rate has predictive power on the minority population shares in the next one, three, and five years, respectively. All regressions include province and year fixed effects and the seven time-varying control variables. The standard errors (in parentheses) are clustered at the province level. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

This concern can be confirmed by examining the predictive power of the current income growth rate on the future MPS. If the MPS is exogenous to income growth, the current income growth rate should have no predictive power on the future MPS, conditional on province and year fixed effects. I regress the MPS in the next one, three, and five years on the current growth rate of GDP per capita, respectively, in columns 1, 2, and 3 of Table E1. All regressions include province and year fixed effects and the seven time-varying control variables. All regressions are based on the 1980–2010 data for the 27 provinces. Details of the MPS data are presented in Table A3. The estimates are all large and statistically significant at the 1% level, which suggests that the MPS is endogenous.

Nevertheless, I still provide the 2SLS estimates of model (10) that uses the MPS as the IV for the CBR. As presented in Figure E2, consistent with the 2SLS estimates based on the policy fine rate, the effect of higher fertility rates is statistically insignificant when the lag length is smaller than 3 years, but becomes significantly positive and much larger after that. The figure only presents the estimates up to a 16-year lag length, because the following estimates are all statistically insignificant and unreasonably large, possibly due to the endogeneity bias.

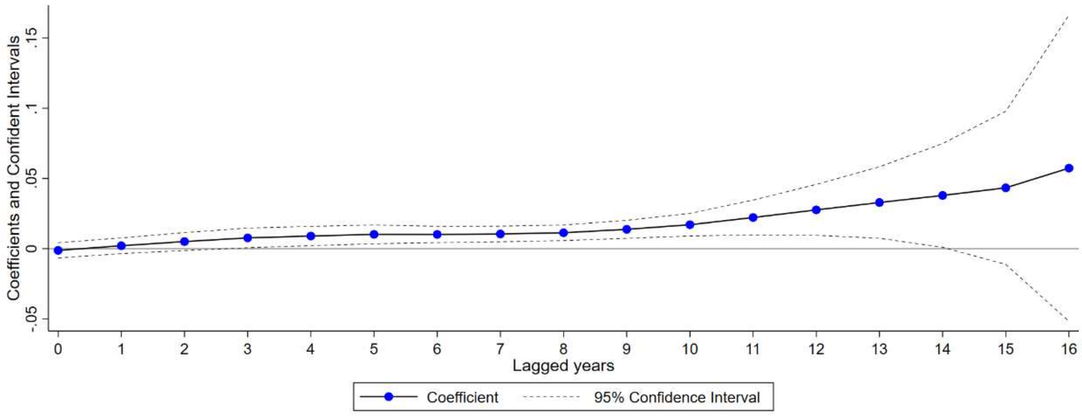

Figure E2. Current and Lagged Effects of Fertility on Income Growth in China (2SLS, Using the minority population share as the IV)

Notes: The figure presents the 2SLS estimates of model (10) that uses the minority population share as the IV. The x-axis indicates the lagged years of the CBR. Each dot on the solid line is the point estimate of the coefficient, and the broken lines indicate the corresponding 95% confidence intervals.

It worth to note that Hongbin Li and Junsen Zhang (2007) also used the MPS as an IV to estimate the causal effect of fertility on income growth in China. Specifically, depending on China provincial data from 1978 to 1998, they estimated the *current* (instead of the lagged) effect of the CBR on the growth rate of GDP per capita in a fixed-effects panel model that used the MPS as an IV for the CBR. They found a negative effect of fertility on income growth. Because their model only used the current CBR as the explanatory variable, as illustrated in Appendix B1 of this paper, what they estimated was mostly the short-run effect. I replicated their estimation using the data during their sample period and found a similarly negative short-run effect of higher fertility on income growth. Because the MPS is likely endogenous, however, the IV estimates based on it may be biased.