Secular Fertility Declines Hinder Long-Run Economic Growth

Huang, Kaixing

Peking University

3 November 2018

Secular Fertility Declines Hinder Long-Run Economic Growth

HUANG KAIXING

Peking University, China, kaixing.huang@pku.edu.cn

To address the endogeneity bias of early cross-country studies of the effect of fertility on economic growth, later studies adopt panel models with fixed effects. However, the fixed effects eliminate long-term fertility differences, and thus the estimation captures mainly the short-term effect. To capture the long-term effect while addressing endogeneity, this article estimates a long-term lagged panel model using one-off fertility shocks. Based on the data from 138 countries from 1960 to 2016, this article found that an increase in fertility first reduces and then increases economic growth, and the long-term average effect is significantly positive. Comparable results are obtained when focusing on countries in different development levels or using within-country variation from China's one-child policy. This finding explains why previous studies, which differently capture the long-term effect of fertility, obtain mixed results. More importantly, it suggests that anti-natalist policies prevalent in the developing world may hinder long-run economic growth.

Keywords: Anti-natalist policies, secular fertility declines, long-run economic growth

JEL: J13, O47, N30

1. Introduction

It is crucial not only for demographers and economists but also for policymakers to investigate the impact that declines in secular fertility may have on long-run economic growth. 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 in 1960 to 2.41 in 2018. Most of the global fertility declines came from middle- and low-income countries, but during this period, fertility rates in high-income countries were nearly halved, from 3.0 to 1.6. Although many high-income countries adopt pro-natalist policies to increase their birth rates, most low- and middle-income countries still adopt anti-natalist policies to curb fertility (United Nations 2015). In sharp contrast to the worldwide adoption of pro-natalist or anti-natalist policies, almost all literature reviews find that there is an ambiguous effect of fertility on economic growth (e.g., Simon 1992, Kelley and Schmidt 1994, Schultz 2008, Headey and Hodge 2009). Without a robust estimate of the causal effect of fertility on economic growth, it is difficult for policymakers to evaluate the costs and benefits of the population policies adopted in their country.

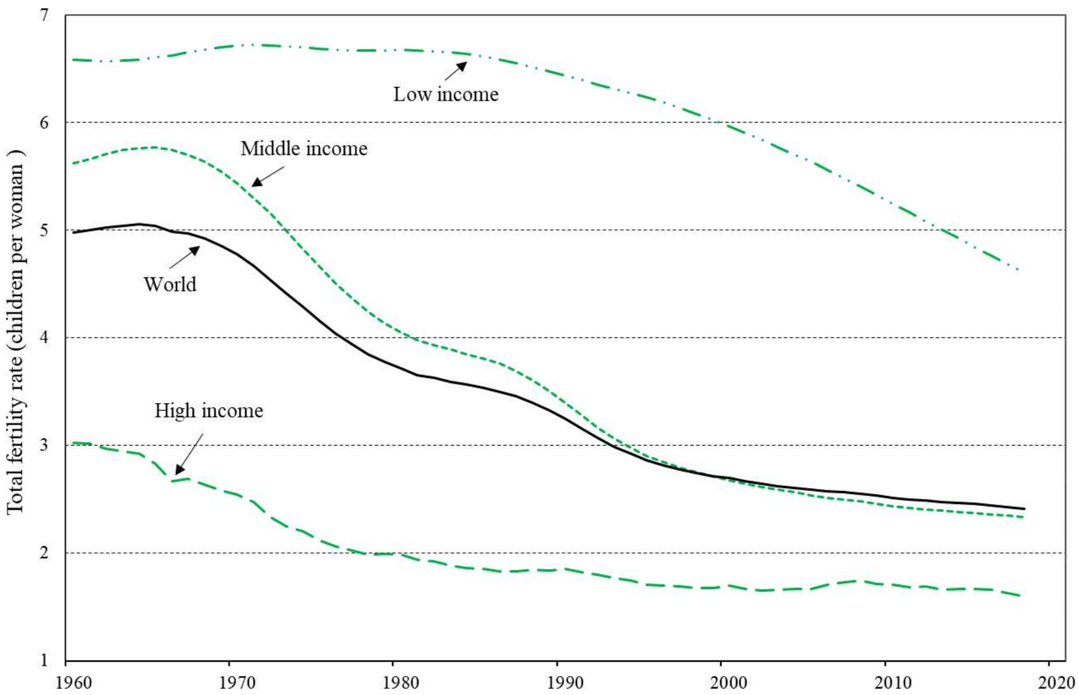

Figure 1. Fertility Trends Over the Past Six Decades

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

Table 1 summarizes the existing macro-level studies on the fertility–income nexus, finding that evidence of the *causal* effect of fertility on income growth is surprisingly scarce.[1] Numerous early studies have examined the association between fertility and income growth, but most of them were based on cross-sectional regression models that did not adequately address endogeneity bias (Simon 1989, Kelley and Schmidt 1994). These early studies generally found no robust association between fertility and income growth. For example, Julian Simon (1992, p.ix) summarized that "the most important 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." More recent studies have shifted their attention to the (reversed) causal effect of income on fertility, generally finding that long-run economic growth reduced fertility. Possibly because of the difficulty of addressing the endogeneity of fertility, few studies have estimated the causal effect of fertility on income growth using instrumental variables (IV) estimation, and these studies also found mixed results. For example, Bloom and Williamson (1998) used several IVs (e.g., population policy and religious composition) and found a positive effect of population growth on income growth, and Li and Zhang (2007) employed an intensity measure of China's one-child policy as the IV and found a negative effect of fertility on income growth.

Table 1. A Summary of Empirical Studies on the Fertility–Income Nexus

Fertility $\Leftrightarrow$ Income	Abundant but mixed evidence: see the surveys from, e.g., Simon (1992), Kelley and Schmidt (1994), Paul Schultz (2008), and Headey and Hodge (2009).
Income $\Rightarrow$ Fertility	Negative long-run causal effects: e.g., Eckstein et al. (1999), Herzer et al. (2012), Chatterjee and Vogl (2018).
Fertility $\Rightarrow$ Income	Scarce and mixed evidence: Bloom and Williamson (1998) found a positive effect, but Li and Zhang (2007) found a negative effect.

Note: A $\Leftrightarrow$ B denotes an association between A and B, and A $\Rightarrow$ B denotes a causal effect of A on B.

[1] This table excludes the following three branches of studies on the fertility–income nexus. First, it excludes micro-level studies because they are incapable of capturing the economy-wide spillovers associated with population growth (Dasgupta 1995). Second, it excludes studies based on data from premodern times because the fertility–income association may change from premodern times to modern times (Galor 2011). Third, it excludes pure time sereis studies because it is difficult to interpret the Granger causality as the true causality because of the concern of post hoc fallacy. Note that there exist excellent studies providing family-level evidence (e.g., Black et al. 2005, Ashraf et al. 2014), using data from pre-modern times (e.g., Lee and Anderson 2002, Ashraf and Galor 2011), and adopting time series estimation methods (e.g., Herzer et al. 2012, Hafner and Mayer-Foulkes 2013).

The current article argues that the failure to capture the long-term lagged effects of fertility is a major reason for why existing studies have not found a robust effect of fertility on income growth. Individuals interact differently with the economy over different stages of the life cycle; thus, the effect of a change in the fertility rate may last for decades and differ substantially in the short term and long term. A larger newborn cohort could reduce short-run economic growth by reducing parental labor supply (Galor and Weil 1996) and could hinder long-run economic growth by reducing human capital investment (Becker et al. 1990) and diluting per capita physical capital (Solow 1956). A larger newborn cohort may also promote short-run economic growth by generating higher demands and more pressures on the economy (Boserup 1981) and may enhance long-run economic growth by providing a larger working age cohort (Bloom et al. 2009) and more potential innovators (Romer 1990). Therefore, a relevant estimate of the effect of fertility on income growth should capture both the short- and long-term effects. Studies that have attempted to identify the causal effect of fertility usually depend on panel models that use fixed effects to account for confounding factors. The present article illustrates that because the fixed effects eliminate most long-term cross-sectional fertility differences, what is captured in the model is mainly the short-term effect.[2]

To complement the literature, the current article adopts two estimation strategies that can capture the long-term lagged effects of fertility. The first strategy is to estimate a series of panel models, each of which regresses the current income growth rate on the fertility rate lagged by different years, ranging from zero to decades. The potential endogeneity bias is then addressed by fixed effects and plausibly exogenous IVs constructed from birth control policies. The fertility coefficients estimated from this strategy reveal the dynamic effects of a fertility change on income growth over the life cycle of individuals, and the average of these estimates reflects the long-term average effect of fertility. The second strategy follows the logic of the standard difference-in-differences (DID) model to capture the long-term lagged effects by employing one-off fertility shocks from exogenous events. Specifically, I employ plausible exogenous fertility shocks from the global epidemic disease interventions around 1940 and from the global family planning campaigns around the mid-1960s. This strategy captures the

[2] Although cross-sectional studies have the potential to capture the long-term effect by comparing cross-sectional long-term differences in fertility and economic growth, the resulting estimates could be subject to substantial endogeneity bias.

long-term lagged effect of fertility shock over the post-shock periods.

I applied these two estimation strategies to national data from 138 countries and to provincial data from China. Based on the first strategy, both the ordinary least squares (OLS) estimates and two-stage least squares (2SLS) estimates suggest that although the short-term effect of higher fertility on income growth is negative or statistically insignificant, the long-term average effect is significantly positive. These results apply to low-, middle-, and high-income countries although the long-term effect is much more positive in high-income countries. The effect estimated based on national data is comparable to that based on China provincial data. The most credible 2SLS estimate based on the first strategy suggests that a 10% increase in fertility would raise the long-term GDP per capita by 23.8%. Evidence based on the second strategy supports the finding that the long-term average effect of a higher fertility rate is significantly positive. The most credible 2SLS estimate based on the second strategy indicates that a 10% increase in fertility would raise the long-term GDP per capita by 23.5%.

The findings of the current article have three important implications. First, the results imply that secular fertility declines represent a strong force driving down long-run economic growth. This finding is in sharp contrast with the conventional view that there is a virtuous cycle between fertility decline and income growth. The observation that growth spurts are often associated with demographic transitions had led to the hypothesis that fertility decline promotes economic growth. [3] Combining this hypothesis with the fact that higher income leads to lower fertility, a virtuous cycle emerges: growth of income per capita leads to reduced fertility, which in turn causes income growth to rise further, which leads to a further decline of fertility, and so forth. This virtuous cycle has been stressed so much in development economics. The findings of this article, however, do not support the existence of this cycle. Instead, it suggests the existence of a long-run equilibrium between fertility and growth: higher income leads to lower fertility, which in turn reduces income growth.

Second, the present article provides strong evidence supporting the scale effect prediction of R&D-based growth models—faster population growth increases the long-

[3] However, there are good reasons to believe that this association may actually reflect that income growth reduces fertility. For example, it has been proposed that rising income increases the opportunity cost of fertility for women (Galor and Weil 1996), that technological progress raises the importance of human capital relative to raw labour (Galor and Weil 2000), and that higher income reduces the needs of old-age security from children (Strulik 2003).

run economic growth. First-generation models of R&D-based growth (Romer 1990, Grossman and Helpman 1991, Aghion and Howitt 1992) predict a strong scale effect in which a larger population leads to faster economic growth. Later contributions eliminate the strong scale effect by assuming either lower intertemporal knowledge spillovers (Jones 1995), increasing the difficulty of R&D (Kortum 1997, Segerstrom 1998), or a diluting effect of product proliferation (Peretto 1998, Howitt 1999). Nevertheless, even in these frameworks, a weak scale effect is still present—faster population growth increases long-run economic growth. Although the scale effect prediction is naturally derived from R&D-based growth models, empirical evidence supporting this prediction is scarce. The present article shows that once adopting an estimation method that can capture the long-term lagged effects of population growth, strong evidence for the scale effect can be found. Therefore, in line with the seminal work by Kremer (1993) and Jones (1995), the current article contributes to establishing a positive link between population, technological change, and economic growth in the long run.

Finally, the findings of this article are useful when evaluating the economic impact of family planning programs. The Malthusian fears that fast population growth hinders income growth have led many developing countries to adopt family planning programs (Coale and Hoover 1958, Ehrlich 1968). The number of countries that adopted family planning programs reached 95 by 1976 and increased to 160 by 2013 (United Nations 2015), and birth control efforts are still in progress in the developing world (Kuang and Brodsky 2016, Stover and Sonneveldt 2017). In recent years, many studies have evaluated the impact of family planning interventions on development and well-being (e.g., Ashraf et al. 2014, Cavalcanti et al. 2020). However, evidence on the impact of family planning programs on income growth largely comes from *family-level* studies, which generally have found that families with fewer children have a higher per capita income. Micro-level studies are, however, incapable of capturing the positive spillovers associated with population growth because these effects are economy wide (Dasgupta 1995). In contrast to the promise of family planning programs,[4] the current article finds strong macro-level evidence that birth control efforts could significantly hinder long-run economic growth.

[4] The rationale for family planning programs has *extended* from promoting per capita income to women's empowerment and reproductive health and rights since the fifth international population conference held in Cairo in 1994 (Cleland et al. 2006).

2. Motivating Theory and Estimating Framework

2.1 Motivating Theory

To frame the empirical analysis, I characterize the effect of a change in the fertility rate over the life cycle of a newborn cohort in the closed-economy neoclassical (Solow-Swan) growth model. Economy i has the constant returns to scale aggregate production function:

$$Y_{it} = A_{it} H_{it}^{\alpha} K_{it}^{\beta} L_{it}^{1-\alpha-\beta}, \tag{1}$$

where $\alpha + \beta < 1$, A_{it} denotes total factor productivity (TFP), K_{it} is physical capital, L_{it} is the supply of land, and H_{it} is the effective units of labor given by $H_{it} = h_{it} N_{it}$, where N_{it} is the total employment and h_{it} is the human capital per worker. Without loss of generality, I normalize $L_{it} = L_i = 1$ for all i and t. Thus, the output per worker is

$$y_{it} = \frac{Y_{it}}{N_{it}} = \frac{A_{it} H_{it}^{\alpha} K_{it}^{\beta}}{N_{it}}. \tag{2}$$

Based on the standard assumptions of the Solow-Swan model, the growth rate of the output per worker around the steady state can be written as

$$g_{it}^{y} = -\beta^{*} \left(\ln y_{it} - \ln y_i^{*} \right), \tag{3}$$

where y_i^{*} denotes the steady-state output per worker in country i, and β^{*} measures the speed of convergence. Equation (3) will be familiar to anyone who has read an advanced macroeconomics textbook (e.g., Barro and Sala-i-Martin 2003). It is also consistent with the empirical growth literature, especially that studies focusing on conditional convergence (e.g., Barro 1991, Mankiw et al. 1992, Sachs et al. 1995). As highlighted by the demographic-dividend literature (e.g., Bloom and Williamson 1998), *per capita* output can be written as:

$$\tilde{y}_{it} = \frac{Y_{it}}{P_{it}} = \frac{Y_{it}}{N_{it}} \frac{N_{it}}{P_{it}} = y_{it} \frac{N_{it}}{P_{it}}, \tag{4}$$

where P_{it} is the total population. This expression can be converted into the growth rate of per capita output:

$$g_{it}^{\tilde{y}} = g_{it}^{y} + g_{it}^{N} - g_{it}^{P}, \tag{5}$$

where g_{it}^{y} is given by equation (3) and g_{it}^{N} and g_{it}^{P} are the growth rates of the total employment and population, respectively.

To characterize the dynamic effects of a change in fertility over the lifecycle of the newborn cohort, the model focuses on a time horizon from the birth ($t = 0$) to the death ($t = T$) of a single cohort. For simplicity, the model assumes that the economy contains only two cohorts—the newborn cohort and the parent cohort—and that the mortality rate is zero. Thus, the total population is kept constant during $t \in (0,T)$: $P_i = \bar{N}_i (1+e_i)$, where $\bar{N}_i$ denotes the parent cohort population and e_i is the fertility rate. Thus, the fertility rate in the model is defined as the births per adult in a single year ($t = 0$).

To capture the time-varying effects of e_i over the life cycle in a reduced-form manner, I assume the following relationships:

$$N_{it} = \bar{N}_i (1+e_i)^{\lambda_t}, \tag{6}$$

$$A_{it} = \bar{A}_i N_{it}^{\gamma} = \bar{A}_i \bar{N}_i^{\gamma} (1+e_i)^{\gamma\lambda_t}, \tag{7}$$

$$h_{it} = \bar{h}_i (1+e_i)^{\eta_t}, \tag{8}$$

$$H_{it} = h_{it} N_{it} = \bar{h}_i \bar{N}_i (1+e_i)^{\eta_t + \lambda_t}, \tag{9}$$

$$K_{it} = \bar{K}_i (1+e_i)^{\delta_t}, \tag{10}$$

where $\bar{A}_i$, $\bar{h}_i$, and $\bar{K}_i$ are the base-year TFP, human capital, and total physical capital, respectively. Equation (6) characterizes the time-varying effects of fertility on the total labor supply (Galor and Weil 1996, Bloom et al. 2009). Newborns reduce parents' labor supply because childrearing is time-consuming, but the time parents spend on each child may decline as the child matures. The newborn cohort eventually contributes to the labor force when reaching working age. Therefore, I assume that the elasticity coefficient λ_t increases from *negative* to *positive* over the life cycle. Equation (7) follows the R&D-based growth models to assume $\gamma > 0$, which means that a higher labor supply (and thus more potential innovators) leads to faster technological progress (e.g., Romer 1990, Aghion and Howitt 1992). Thus, fertility has time-varying effects on TFP through the labor supply (captured by $\gamma\lambda_t$). Equation (8) follows the quality–quantity trade-off theory (e.g., Becker et al. 1990, Galor 2005) to assume that higher fertility leads to lower human capital investment in each child. For simplicity, I assume $\eta_t = 0$ before the new cohort enters the labor force and $\eta_t < 0$ after that. Equation (9) characterizes the effect of fertility on the effective labor supply through the human capital and labor supply ($\eta_t + \lambda_t$). Finally, equation (10) captures the potentially positive effect of fertility on physical capital accumulation ($\delta_t > 0$). Numerous studies (e.g., Solow 1956, Cass 1965, Boserup 1981, Simon 1992) have argued that the higher demand and pressure on the economy arising from population growth might promote physical capital accumulation. To the extent that a newborn's demand and, thus, pressure on the economy increase with age, the elasticity coefficient δ_t could increase over time as well.

To sum up, equations (6)–(10) characterize the time-varying effects of fertility through four channels: labor supply, TFP, human capital, and physical capital. In addition, $g_{it}^N - g_{it}^P$ in equation (5) characterizes the fifth channel by the dilution effect of population growth on per capita output (Malthus 1798, Solow 1956). Substituting (6) –(10) into (2) and taking logs, I obtain the following log-linear relationship between per worker output and fertility:

$$\ln y_{it} \approx D_i + \left(\alpha\eta_t + (\gamma + \alpha - 1)\lambda_t + \beta\delta_t\right)e_i \ , \tag{11}$$

where the time-invariant term $D_i = \ln \bar{A}_i \bar{N}_i^{\gamma} + \alpha \ln \bar{h}_i \bar{N}_i + \beta \ln \bar{K}_i - \ln \bar{N}_i$, and I applied the approximation $e_i \approx \ln(1+e_i)$.[5] I combine (11) with (3) and (5) to obtain a relationship between the growth rate of per capita output and fertility:

$$g_{it}^{\tilde{y}} \approx \beta^* \ln y_i^* - \beta^* \ln y_{it} + \lambda_t^{'} e_i \approx \beta^* \left(\ln y_i^* - D_i\right) + \psi_t e_i \ , \tag{12}$$

where $\psi_t = \lambda_t^{'} + (1-\gamma-\alpha)\beta^*\lambda_t - \beta^*\beta\delta_t - \beta^*\alpha\eta_t$ is the time-varying coefficient of fertility. Note that (12) uses the conditions $g_{it}^P = 0$ (i.e., constant population during $0 < t < T$) and $g_{it}^N \approx \lambda_t^{'} e_t$ (which is derived from (6), and $\lambda_t^{'}$ denotes the first derivative of λ_t). Equation (12) highlights the complexity and time-varying nature of the effect of a change in fertility rate on the per capita income growth over the life cycle of the newborns. The effect depends not only on constant elasticity coefficients (α , β , and γ), but also on the time-varying elasticity coefficients (λ_t , δ_t , and η_t), the first derivative of λ_t ($\lambda_t^{'}$), and the speed of convergence (β^*). The complex combination of these coefficients implies that the effect of fertility on income growth is theoretically ambiguous and most likely time varying. The next subsection will convert this theoretical framework into an estimation framework that can be used with real data.

[5] Note that the fertility rate is a very small value because it is defined in the model as the births per adult that occur in a given year.

2.2 Estimating Framework

The above theoretical model focuses on a single newborn cohort, but in reality, new cohorts emerge continuously. To transform (12) into a regression equation that can be estimated by the real data, I introduce a new time dimension to it to denote the birth year of each cohort relative to the current time. Specifically, when adding an error term, potential covariates, and the new time dimension to (12), I obtain the following estimation equation:

$$g_{it}^{\tilde{y}} = \nu_i + \tau_t + \sum_{s=0}^{T} \psi_s TFR_{i(t-s)} + Z_{it}\mu + \varepsilon_{it} \ , \tag{13}$$

where $g_{it}^{\tilde{y}}$ is the annual growth rate of GDP per capita in year t and country i ; v_i denotes the country-fixed effects that are used to account for the confounding effects of country-specific, time-invariant determinants of income growth, i.e., $\beta^*(\ln y_i^* - D_i)$; τ_t is year fixed effects used to account for the confounding effects of time-varying factors common across countries; Z_{it} denotes a vector of control variables; and ε_{it} is the error term. The new time dimension s denotes the cohort born s years prior to the current time t . The key explanatory variable, $TFR_{i(t-s)}$, is TFR s years prior to the current year. Thus, the coefficient $\psi_s = \lambda_s^{'} + (1-\gamma-\alpha)\beta^*\lambda_s - \beta^*\beta\delta_s - \beta^*\alpha\eta_s$ captures the s-year lagged effect of fertility on income growth. In this distributed-lag model setting, the changes of ψ_s over $0 \le s \le T$ reflect the differential effects of a change in fertility rate on income growth over the life cycle of the newborn cohort, and the average of ψ_s captures the long-term average effect of fertility.

A major advantage of (13) compared with the standard fixed effects panel model used in previous studies is in capturing the long-term lagged effects of fertility. Previous fixed effects panel regressions usually include only the current fertility or fertility lagged by several years (usually one or five years), thus capturing only the short-term effects of fertility. This is because the fixed effects eliminate long-term cross-sectional fertility differences; thus, the identification depends mainly on short-run interannual fertility changes. As illustrated in Appendix B1, even for a panel of fertility data covering five decades, most of the fertility changes used in the identification of the fixed effects panel model are short-term fertility changes. By using long-term lagged fertility rates as explanatory variables, (13) provides a natural way to identify the long-term lagged effects of fertility in a fixed-effects panel model.[6] An alternative way to capture the long-term lagged effect is to estimate the effect of one-off fertility shocks (occurring at the early stages of a long sample period), following the logic of a standard DID model.[7] This alternative approach will be detailed later.

However, it is infeasible to directly estimate model (13) because of serious collinearity issues: fertility rates in successive years are highly correlated with one another.[8] Unless collinearity is adequately addressed, the estimate of ψ_s could be imprecise and have incorrect signs. A standard method for addressing collinearity is to use a restricted least squares estimator that depends 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 of the effects of fertility is too complicated to be characterized by a tractable functional form.[9] Because imposing incorrect restrictions on parameters can lead to additional biases, I do not seek to solve the collinearity problem by using restricted least squares estimators.

Instead, I transform (13) into a series of estimating equations, each of which only includes one of the lag terms of TFR, with a lag length ranged from 0 to T years:

$$g_{it}^{\tilde{y}} = \nu_i + \tau_t + \psi_s TFR_{i(t-s)} + Z_{it}\mu + \varepsilon_{it}^{s}\ , \quad s = 0,1,2,\cdots,T\ . \tag{14}$$

To the extent that nearby lags are correlated with one another, the coefficient of the included lag, ψ_s, captures the effect of the "omitted" nearby lags. The estimate of ψ_s in model (14) can be seen as the weighted average of the effects of the included lag and

---

[6] The current article is not the first to adopt a lagged model setting to estimate the long-term lagged effect of fertility. Barlow (1994) made the first attempt to separate the long-term and short-term effects of fertility by including both long-term lagged fertility and current fertility in a single regression. Based on data from 86 countries, he found that per capita income growth is negatively related to current fertility and positively related to 17-year lagged fertility. However, no study has used this estimation strategy after Barlow's paper. Note that studies that have lagged the fertility rate by one or a few years do not fulfill this criterion because the positive effect of fertility takes longer to express.

[7] 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 occurred in the tenth year in half of the sample countries. A standard DID model that compares the income growth before and after the fertility decline across the 100 countries could capture the long-term average effect of the fertility decline over the 50 years following it.

[8] If $TFR_{i(t-s)}$ follows a pattern over time, then $TFR_{i(t-s+1)}$ will follow a similar pattern, thus causing $TFR_{i(t-s)}$ and $TFR_{i(t-s+1)}$ to be strongly correlated. The data from the 138 countries examined in this article show that the correlation between the current TFR and TFR lagged by 3, 5, and 10 years are 0.99, 0.98, and 0.96, respectively.

[9] The lagged effects of fertility are more complicated than the lagged effects of other factors, such as fiscal policies, because an individual's interaction with the economy may last for decades and change in a nonlinear manner over time.

the omitted nearby lags, and the weighting is the strength of the correlation. It is well-known that this kind of pure serial correlation does not cause bias in the regression coefficient estimates (Greene 2010, p.903), but it tends to bias the estimated variances of the regression coefficients. For this reason, I report autocorrelation-consistent standard errors (e.g., the Newey–West standard errors) throughout this article. Note that the simple average of ψ_s (over $s = 0,1,2,\cdots,T$) captures the long-term average effect of a change in fertility rate on income growth.

Although it is good at capturing the long-term lagged effects of fertility, model (14) does not adequately address the endogeneity bias from the omitted variables and reverse causality. The model includes country-fixed effects to account for the confounding effects of country-specific time-invariant factors and includes year fixed effects to account for the confounding effects of annual shocks common to all countries. However, these fixed effects cannot address the potential bias caused by omitted *country-specific time-varying* factors. The model also reduces the potential bias from reverse causality thanks to its lagged model design. However, reverse causality is not fully addressed because fertility could be affected by expectations of future income growth. Considering that income growth generally has a negative effect on fertility in modern times (Eckstein et al. 1999, Herzer et al. 2012, Chatterjee and Vogl 2018), the estimate of ψ_s is expected to be downwardly biased by reverse causality.

I have adopted two approaches to address the remaining endogeneity. The first is to find an excluded IV for TFR and then conduct the 2SLS estimation of model (14). Note that a suitable IV must have sufficient time variation because TFR in the model is lagged by various years. This kind of IV is only available when using Chinese provincial data for the estimation. The second approach is to employ one-off fertility shocks to estimate a version of the model (14) that includes only the current TFR. Because only the current TFR is included in the modified model, it is not difficult to construct excluded IVs for TFR from one-off fertility shocks when using country-level data. As mentioned above and detailed in Appendix B1, if the fertility shock occurred at the early stages of a long sample period, this approach can also capture the long-term lagged effects of fertility. As presented in the following sections, both the OLS and 2SLS estimates of model (14) (and its variation) suggest that higher fertility rates significantly increase long-term average income growth, even though the 2SLS estimates suggest a much larger effect than the OLS estimates.

3. Global Evidence

This section estimates model (14) based on data from 138 countries. The estimation finds that a higher fertility rate first reduces and then increases income growth, and the long-term average effect is significantly positive. This finding is robust when adjusting for serial correlation, including various control variables, and when focusing on countries with different income levels. Estimates based on plausibly exogenous fertility shocks confirm this finding. The data sources and summary statistics of all variables are presented in Appendix Table A1.

3.1 Long-term Average Effect of Fertility

Figure 2 presents the OLS estimate of ψ_s (the solid line) from model (14) and the corresponding 95% confidence intervals (broken lines). The confidence intervals are calculated based on the Newey–West standard errors, which are robust to autocorrelation and heteroskedasticity. The estimation uses annual data from 138 countries (listed in Table A2), where GDP per capita and TFR are available continuously from 1960 to 2016. The GDP data (in 2011 USD) are derived from the Maddison Project Database 2018, and TFR data are derived from the World Development Indicators. The estimation controls for four of the most important determinants of income or fertility: the five-year lagged log GDP per capita (used to capture the effect of economic convergence), years of total schooling (used to capture the effect of human capital), infant mortality, and life expectancy. Other control variables are set aside for robustness checks. All control variables are lagged by the same years as TFR in each regression to avoid overcontrol bias.[10] The coefficients of the current TFR and TFR lagged up to 50 years are estimated, but the figure only reports the coefficients up to 40 lagged years because the remaining estimates are statistically insignificant and have very wide confidence intervals (complete results are reported in Appendix Table B1).

[10] If the control variables are not lagged, which means that they are in the same time period as the dependent variable, the control variables could partially account for the true effect of the lagged fertility. Specifically, if there are any correlations between the future values of the control variables and the past TFR, it is most likely that the past TFR is the cause of the correlation. In this case, these (current) control variables are the channel variables through which the lagged TFR affects income growth, and controlling for these channel variables would account for the true effect of the lagged fertility.

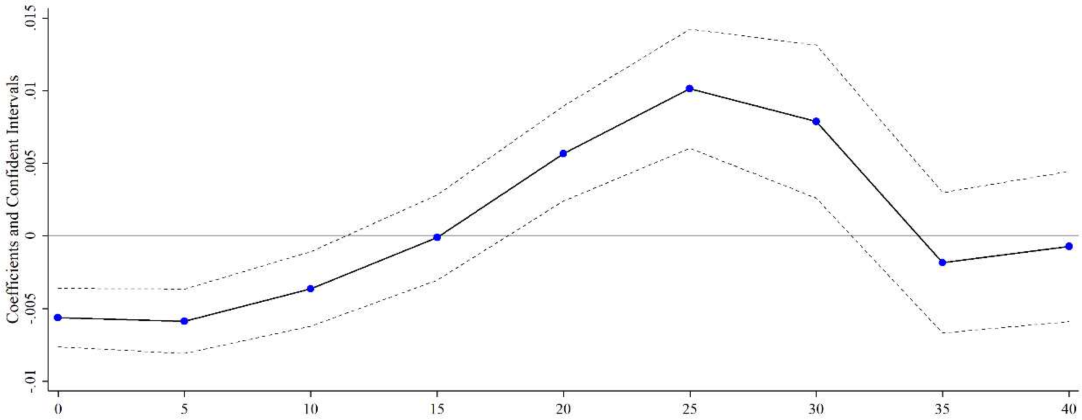

Figure 2. Current and Lagged Effects of the Fertility Rate on the Growth Rate of GDP per capita

Notes: The figure presents the estimated TFR coefficients of model (14) using data from 138 countries from 1960 to 2016. Each dot on the solid line is the point estimate of the coefficient of TFR lagged by the year indicated by the x-axis, and the broken lines indicate the corresponding 95% confidence intervals that are calculated based on the Newey–West standard errors.

Figure 2 shows that a higher fertility rate first reduces and then increases income growth and that the effect lasts for more than three decades. Specifically, the initially negative effect turns positive after 15 lagged years, and the positive effect peaks at 25 lagged years, after which it declines and becomes statistically insignificant after 33 lagged years. This finding highlights the importance of estimating the long-term *average* effect: studies focusing only on the short-term effect tend to overestimate the negative effect, and studies focusing only on the long-term effect tend to overestimate the positive effect. It can be calculated that the average of the estimates of ψ_s over 0 to 40 lagged years is 0.12%, with a 95% confidence interval (0.03, 0.21). Therefore, a one-unit increase in TFR would significantly raise the long-term average growth rate of GDP per capita by 0.12%. The accumulated effect calculated based on this estimate suggests that a one-unit increase in TFR would raise the GDP per capita by 5.0% for an average sample country over four decades.[11] Because the average TFR had declined by 2.53 from 1960 to 2016, this estimate implies that secular fertility declines reduced GDP per capita by 9.1% (5.0%*2.53*40/56) over four decades. Note that, however, the OLS estimates tend to underestimate the positive effect of fertility because of the downward bias from reverse causality.

The effect pattern presented in Figure 2 can be explained by the mechanisms

[11] The accumulative effect is approximately calculated according to $\prod_{0\le s\le 40}(1+\psi_s)-1$.

proposed in the motivating theory: (1) the initial (0–15 lagged years) negative effect of higher fertility could be driven by the negative dilution effect of population growth on per capita income and the negative effect from the time cost of childrearing; (2) the negative effect gradually disappears and becomes positive, possibly because the time cost of childrearing declines with the lag length, while the stimulating effect on physical capital formation increases with it; (3) the positive effect increases further possibly because the new cohort directly contributes to the labor force and innovations; (4) the effect eventually peaks (at 25 lagged years) and then gradually declines, eventually reaching zero, possibly because 25 years old is approximately the reproductive age of individuals (and thus a new cycle begins). Figure 2 also suggests that the negative effect of higher fertility on human capital formation (i.e., quality–quantity trade-off) does not dominate the lagged effect.[12]

3.2 Robustness Checks

Table 2 examines the robustness of the baseline results (presented in Figure 2) to autocorrelation, omitted variables, and income levels. To facilitate a comparison, column 1 replicates the baseline estimates. All robustness checks have the same model setting as the baseline estimation, except for the one specified in each check. Additional robustness checks are presented in Appendix B3.

Columns 2–4 of Table 2 show that the estimated standard errors are robust to autocorrelation. As detailed before, the estimated variances of the coefficients from model (14) are likely downwardly biased by autocorrelation. This is why the baseline estimation reports the Newey–West standard errors that are robust to autocorrelation. Column 2 shows the extent of the bias by presenting the OLS estimates of the standard errors that do not adjust for autocorrelation. The OLS standard errors are only slightly smaller than the Newey–West standard errors (reported in column 1), suggesting no substantial bias from autocorrelation. This finding is not surprising when considering that the year fixed effects included in the model could have accounted for most of the trends in fertility. Column 3 clusters the error terms at the country level to adjust for within-country correlation of the error terms caused for any reason, instead of only by

[12] This finding is in line with the existing empirical evidence. Although early studies (e.g., Leibowitz 1974, Hanushek 1992) found evidence supporting the quality–quantity trade off, later studies using the incidence of twinning or the sibling sex composition as instruments have usually found mixed evidence (Angrist and Evans 1998, Black et al. 2005, Cáceres-Delpiano 2006, Lee 2008). See Schultz (2007) for a review of this literature.

autocorrelation (Abadie et al. 2017). The resulting standard errors are larger, but all the corresponding estimates are statistically significant, at least at the 10% level. Note that clustering the error terms at the country level is not a preferred method for addressing autocorrelation here because it also accounts for the error correlations caused by omitted channel variables. [13] Column 4 directly accounts for autocorrelation by controlling for linear and quadratic country-specific time trends while also finding comparable results.

Columns 5 and 6 show that the estimates are robust to omitted variables. The baseline estimation only controls for four time-varying determinants of fertility and growth. Column 5 controls for four additional time-varying factors: GDP per capita as a ratio of US GDP per capita (which is another control of income convergence), the share of urban population, net international migration, and the share of natural resource rents in GDP. Controlling for these factors has only a negligible effect on the estimates, and the long-term average effect calculated (presented in the last row) is identical to the baseline estimate. I have also tried to control for other (less relevant) time-varying factors, such as temperature, inflation, and foreign direct investment and found very similar results. Note that all time-invariant factors are excluded from the model because they have been well controlled for by the included country-fixed effects. However, a potential concern is that time-invariant factors may have time-varying effects on income growth, which cannot be controlled for by fixed effects. To address this concern, column 6 controls for the potential time-varying effects of three important time-invariant variables (the dummy of landlocked, first official language, and average index of political stability) by including the interactions of each of them with a full set of year dummies. [14] The resulting estimates are slightly larger but have no statistically significant difference from the baseline estimates.

[13] For example, lagged TFR could affect future income growth by affecting future education, so the model intentionally excludes future (nonlagged) education as a control variable to avoid overcontrol bias (see Footnote 10 for details). The "omitted" future education naturally leads to a within-country error correlation given the lagged model setting of (14). Adjusting for this kind of error correlation could lead to upward-biased standard errors.

[14] The index of political stability is a time-varying variable. However, because the data are only sparsely available mainly after 1990s for most countries, I have chosen to control for it by using its mean value over all available observations.

Table 2. Robust to Autocorrelation, Omitted Variables, and Income Levels

Lags	(1) Baseline (Newey–West)	Autocorrelation (2) OLS	(3) Cluster at country level	(4) Control for time trends	Omitted variables (5) Additional time-varying controls	(6) Additional time -invariant controls	Income levels (7) Low -income	(8) Middle -income	(9) High -income
L0	-0.0056***	-0.0056***	-0.0056***	-0.0053***	-0.0069***	-0.0069***	-0.0074***	-0.0043**	-0.0056*
	(0.0010)	(0.0007)	(0.0020)	(0.0007)	(0.0010)	(0.0010)	(0.0016)	(0.0018)	(0.0031)
L5	-0.0058***	-0.0058***	-0.0058***	-0.0056***	-0.0057***	-0.0058***	-0.0057***	-0.0067***	-0.0040
	(0.0011)	(0.0008)	(0.0019)	(0.0008)	(0.0011)	(0.0011)	(0.0019)	(0.0019)	(0.0036)
L10	-0.0036***	-0.0036***	-0.0036*	-0.0033***	-0.0032**	-0.0030**	-0.0018	-0.0083***	-0.0048
	(0.0013)	(0.0009)	(0.0019)	(0.0009)	(0.0013)	(0.0014)	(0.0020)	(0.0021)	(0.0050)
L15	-0.0001	-0.0001	-0.0001	0.0003	0.0004	-0.0000	0.0036*	-0.0045*	0.0032
	(0.0014)	(0.0011)	(0.0024)	(0.0011)	(0.0015)	(0.0016)	(0.0021)	(0.0025)	(0.0072)
L20	0.0056***	0.0056***	0.0056*	0.0058***	0.0060***	0.0064***	0.0107***	0.0042	0.0222***
	(0.0016)	(0.0012)	(0.0033)	(0.0012)	(0.0016)	(0.0017)	(0.0026)	(0.0027)	(0.0075)
L25	0.0101***	0.0101***	0.0101**	0.0100***	0.0099***	0.0105***	0.0137***	0.0181***	0.0294***
	(0.0020)	(0.0015)	(0.0042)	(0.0015)	(0.0020)	(0.0021)	(0.0036)	(0.0028)	(0.0092)
L30	0.0078***	0.0078***	0.0078**	0.0079***	0.0081***	0.0082***	0.0099*	0.0148***	0.0268**
	(0.0026)	(0.0021)	(0.0033)	(0.0021)	(0.0026)	(0.0026)	(0.0050)	(0.0024)	(0.0108)
L35	-0.0018	-0.0018	-0.0018	-0.0019	-0.0009	-0.0003	-0.0023	0.0027	0.0235***
	(0.0024)	(0.0019)	(0.0040)	(0.0018)	(0.0024)	(0.0026)	(0.0050)	(0.0022)	(0.0076)
L40	-0.0007	-0.0007	-0.0007	-0.0005	0.0008	0.0026	0.0065	-0.0030	0.0252***
	(0.0026)	(0.0021)	(0.0048)	(0.0020)	(0.0025)	(0.0027)	(0.0060)	(0.0026)	(0.0079)
Average (L0–L40)	0.0012***	0.0012***	0.0012**	0.0014***	0.0012***	0.0013***	0.0033***	0.0019***	0.0096***
	(0.0004)	(0.0004)	(0.0006)	(0.0004)	0.0004	0.0005	0.0007	0.0007	0.0012

Notes: This table reports the estimates of the lagged TFR from model (14), with the lag length denoted in the first column. Column 1 replicates the baseline estimates reported in Figure 2. Columns 2–4 examine the robustness to autocorrelation, columns 5 and 6 the robustness to control variables, and columns 7–9 the robustness to the income levels of the sample countries. The last row reports the long-term average effect of TFR lagged by 0 to 40 years. Robust standard errors are reported in parentheses. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

Columns 7–9 separately estimate the model for low-, middle-, and high-income countries. To avoid using a subjective definition of income levels, I classify the 138 sample countries into three equal-sized groups based on the ranking of their 1960 GDP per capita.[15] The estimations find the same effect pattern for each group of countries: higher fertility first reduces and then increases per capita income, and the long-term average effect is significantly positive. As expected, the estimated positive effect is much larger in high-income countries than in low- and middle-income countries.[16] These subsample regressions also provide a way to verify the existence of downward bias from reverse causality. Because fertility declines with income, the reverse causality bias should be larger when pooling countries from different income groups. This is because when using the pooled data, more fertility variation comes from cross-country fertility differences caused by income differences. If the effects that are estimated separately for each income group are all larger than the pooled baseline estimate, this can be taken as evidence of reverse causality bias. The estimated long-term average effects are 0.33%, 0.19%, and 0.96%, respectively, for low-, middle-, and high-income countries, which are all much larger than the baseline estimate using data from all countries (i.e., 0.12%).

Appendix B3 presents the three groups of additional robustness checks. First, I adopt an alternative fertility measure, the crude birth rate, in place of TFR. Second, I exclude sample countries with a 1960 population smaller than 2 million (40 countries) or larger than 50 million (12 countries) to determine whether the findings are mainly driven by these relatively small or large countries. Finally, I exclude countries from America, Asia, Europe, and Sub-Saharan Africa one region each time to check if the findings are sensitive to countries from a specific region. All of these robustness checks find estimates comparable to the baseline estimates.

[15] I have also used the classification of low-, middle-, and high-income countries provided by the World Bank and found comparable results.

[16] The relatively abundant physical and human capital stocks in high-income countries imply smaller negative effects of fertility through physical capital dilution and quantity–quality trade off. In addition, as usually modelled in theoretical studies (e.g., d'Albis 2007, Boucekkine et al. 2002), the marginal positive effect of fertility could be higher when the fertility level is lower, which is the case regarding high-income countries.

3.3 Evidence from Global Epidemic Disease Interventions

A major concern of the above estimation is endogeneity bias, especially bias from reverse causality. To address this concern, this subsection adopts an alternative estimation strategy employing plausibly exogenous one-off fertility shocks to identify the long-term causal effect. As mentioned before, the coefficient of the current fertility in a fixed effects panel model mainly captures the short-term effect because the fixed effects eliminate most of the long-term fertility variation. However, if the identification is based on one-off fertility shocks, long-term effects can be captured (see Appendix B1 for a detailed discussion). Specifically, this can be done by constructing IVs from exogenous fertility shocks for the 2SLS estimation of the following fixed effects panel model:

$$\ln y_{it} = \nu_i + \tau_t + \psi TFR_{it} + Z_{it}\mu + \vartheta_{it} \ , \tag{15}$$

which regresses log GDP per capita, $\ln y_{it}$, on the current fertility rate. The first stage of the 2SLS estimation is given by

$$TFR_{it} = \nu_i + \tau_t + \chi ShockIV_{it} + Z_{it}\gamma + \xi_{it} \ , \tag{16}$$

where $ShockIV_{it}$ is the IV constructed from plausibly one-off fertility shocks, χ is a coefficient, and ξ_{it} is the error term. The identification of the 2SLS estimation depends on comparing the relative changes in income between countries with high and low intensities of fertility shocks in the post-shock period to that in the pre-shock period. As long as the fertility shock is approximately a one-off and the post-shock period is long enough, the long-term average effect of fertility can be captured by the coefficient ψ. Note that the log GDP per capita (instead of the growth rate) is an appropriate dependent variable in this estimation framework; this is because depending on one-off fertility shocks means insufficient fertility variation that can be used to identify the time-varying effects on the growth rate.[17]

I first estimate model (15) using an IV constructed from the global epidemic disease interventions that occurred around 1940. As detailed in Acemoglu and Johnson (2007),

[17] In addition, it is not a good idea to estimate the effect on the long-term *average* growth rate in this estimation framework because of the concern of a downward bias. This is because, as shown in Figure 2, the initial effects of fertility on the growth rates are negative while the lagged effects are positive. Because income is generally an increasing trend, a 1% initial income change should be smaller than a 1% later income change. As such, the estimated effect on the average growth rate underestimates the true effect on income growth.

a series of global drug and chemical innovations around the year 1940 dramatically reduced mortality from epidemic diseases across the world. Because mortality and fertility are strongly correlated (Olsen and Wolpin 1983, Angeles 2010), this event also generated a shock to fertility. I follow Acemoglu and Johnson (2007) in constructing an IV for fertility based on the mortality rate of each country from 15 of the most important infectious diseases prior to 1940 and the global intervention dates for each specific disease. Specifically, Acemoglu and Johnson (2007) constructed a predicted mortality instrument based on

$$M_{it}^{I} = \sum_{d \in D} \left[(1 - I_{dt}) M_{di40} + I_{dt} M_{dFt} \right] , \tag{17}$$

where I_{dt} is a dummy for intervention for disease d ($I_{dt} = 1$ after the intervention), M_{di40} is the preintervention mortality from disease d in country i, and M_{dFt} is the mortality rate from disease d at the health frontier of the world at time t. The authors argued for the exogeneity of the predicted mortality instrument based on the fact that the global intervention dates for each disease (I_{dt}) are exogenous to individual countries.

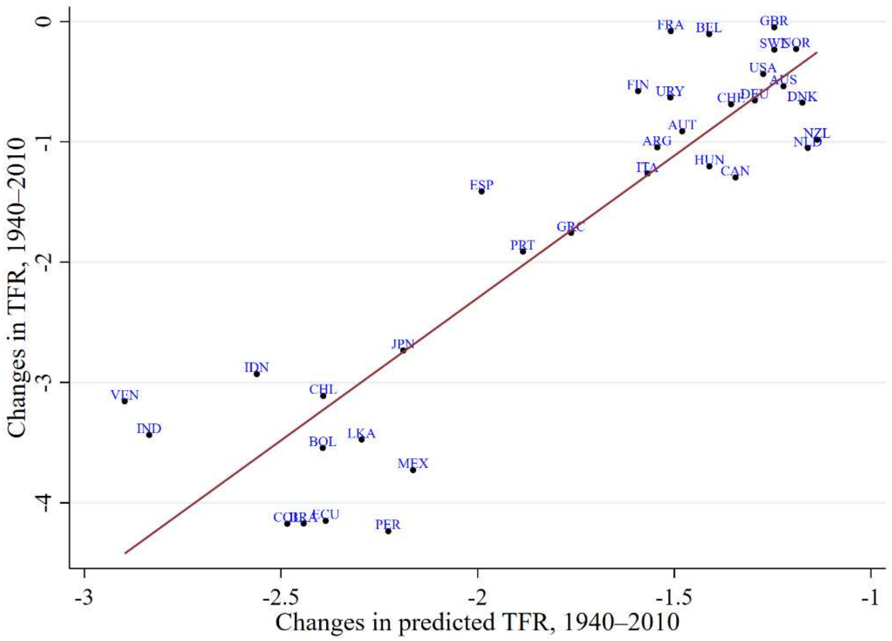

Figure 3. Changes in TFR and predicted TFR

Notes: The figure presents 34 countries where the data for TFR, GDP per capita, and predicted mortality were available from 1900 to 2010.

It may not be a good idea to directly use predicted mortality as the IV for the fertility rates when examining the effect of fertility on income because the predicted mortality could affect income through channels that are uncorrelated with fertility. To address this concern, in two steps, I construct an IV for fertility based on the predicted mortality. First, I use predicted mortality as an IV for life expectancy to obtain the 2SLS estimates of the following:

$$TFR_{it} = \nu_i + \tau_t + \kappa Lifee_{it} + Z_{it}\rho + \pi_{it} \ , \tag{18}$$

where $Lifee_{it}$ denotes life expectancy (instrumented by predicted mortality), κ is a coefficient, and π_{it} is an error term. Note that the 2SLS estimate of κ captures the effect of predicted mortality on TFR *only* through the channel of life expectancy. Second, I predict TFR from the 2SLS estimates of (18) and then use the predicted TFR ($pTFR_{it}$)as the IV for TFR. By construction, $pTFR_{it}$ could affect income only (directly or indirectly) through fertility.[18] Figure 3 shows that the changes in TFR from 1940 to 2010 are strongly correlated with the changes in the predicted TFR over the same period.

Table 3. Causal Effect of TFR on Log GDP per Capita

	(1) IV = Predicted mortality	(2) IV = Predicted TFR	(3) IV = Program starting year
Panel A. *Second stage* (Dependent variable: log GDP per capita)			
TFR (instrumented)	0.298*** (0.093)	0.168*** (0.042)	0.574*** (0.160)
Panel B. *First stage* (Dependent variable: TFR)			
Predicted mortality	1.06*** (0.21)		
Predicted TFR		0.99*** (0.09)	
Program dummy			-0.26*** (0.06)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
First-stage *F*	24.7	110	19.4
Observations	408	408	1683
Countries	34	34	31

Notes: The table presents the 2SLS estimates based on Equations (15) and (16). The IVs are predicted mortality, predicted TFR, and the dummy of the family planning programs starting year in columns (1), (2), and (3), respectively. Robust standard errors are reported in parentheses. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

Columns 1 and 2 of Table 3 present the 2SLS estimates of model (15) using

[18] Departing from the standard 2SLS regressions, I use a generated IV. According to Wooldridge (2010, p.124), parameter estimates in 2SLS regressions with generated instruments are asymptotically distributed, as in standard 2SLS regressions.

predicted mortality and predicted fertility as IVs, respectively. The estimations are based on 10-year interval data from 1900 to 2010 (instead of the continuous annual data from 1960 to 2016) because the predicted mortality data constructed by Acemoglu and Johnson (2007) are available in 10-year intervals since 1900. The data on GDP per capita are still derived from the Maddison Project Database 2018, but the data on TFR are derived from Roser (2014) (instead of from the World Development Indicators, where the data prior to 1960 are unavailable). There are 34 countries (see Figure 3) where the data for TFR, GDP per capita, and predicted mortality are available for this period. Employing the early data implies that most of the usually used control variables are unavailable, so these estimations control only for the country and year fixed effects.[19]

The first-stage estimates in columns 1 and 2 (Panel B) indicate that both predicted mortality and predicted fertility have a significantly positive effect on TFR, confirming that they are likely strong IVs for TFR. The second-stage estimates in columns 1 and 2 (Panel A) suggest that a one-unit increase in TFR raises the long-term GDP per capita by 29.8% and 16.8%, respectively, when using predicted mortality and predicted fertility as IVs. As detailed above, the estimated effect based on the predicted fertility IV is more credible. However, no matter which IVs are being used, the 2SLS estimates are much larger than the OLS estimate from the last section (which suggests a 5.0% long-term effect). This finding confirms that the OLS estimate is indeed substantially downward biased, potentially because of reverse causality. Recall that these 2SLS estimates reflect the long-term effect of fertility on income only when the fertility change from the global disease interventions can be approximately seen as a one-off fertility shock. To the extent that the interventions have long-term lagged effects on fertility, the 2SLS estimates still tend to underestimate the long-term effects of fertility.

The above 2SLS estimations are based on the assumption that predicted mortality (and, thus, predicted fertility constructed from it) is exogenous. The exogeneity of predicted mortality has been thoroughly examined in Acemoglu and Johnson (2007). Similar to their examinations, Appendix B4 presents two pieces of evidence supporting that predicted mortality and predicted fertility are likely exogenous. First, columns 1 and 2 of Table B2 show that the leads of predicted mortality and predicted fertility have

[19] The early data for some channel variables, such as education, are available, but controlling for channel variables could lead to overcontrol bias.

no significant effect on TFR. Because the global disease interventions did not start before 1940, the leads of the IVs should have no effect on TFR. If the leads are correlated with TFR, the correlation must be driven by preexisting trends, which implies endogeneity. Second, columns 3–6 of Table B2 show that the changes in TFR and GDP per capita prior to the global interventions (from 1930 to 1940) have no predictive power on the preintervention changes in predicted mortality and predicted fertility (from 1940 to 2010). If prior-intervention changes in fertility or income have a significant effect on post-intervention changes in the IVs, the IVs are endogenous in the sense of reverse causality. The findings are the same when examining the correlation using data from different pre- and post-intervention periods.

3.4 Evidence from the Global Family Planning Campaigns

A potential concern of the above 2SLS estimates is that because global disease interventions only indirectly affect fertility through mortality, it may take a long time (possibly decades) for fertility to respond to the mortality changes from the disease interventions (Angeles 2010). If this is true, the above 2SLS estimates may still have underestimated the long-term effect of fertility. For this reason, this subsection employs fertility changes from a second event that directly affected fertility: the global family planning campaigns.

Starting around the mid-1960s, increasing concerns over the unprecedented levels of population growth in the developing world have led many developing countries to adopt family planning programs (Robinson and Ross 2007). De Silva and Tenreyro (2017) found strong evidence that national family planning programs significantly reduced the fertility rates in developing countries. I use the dummy of the national family planning program starting year (equals one for all years after the starting year) as the IV for fertility for the 2SLS estimation of model (15). As compiled by De Silva and Tenreyro (2017) from various sources, the exact starting year of state-led family planning programs is available for 31 developing countries. As presented in Figure 4, most of the 31 countries started their family planning programs in the time period of 1960–1970, which is consistent with the timing of the global family campaigns. Note that although several intensity measures of family planning programs are available, the estimation of this article does not depend on these intensity measures because of data limitations and endogeneity concerns.[20]

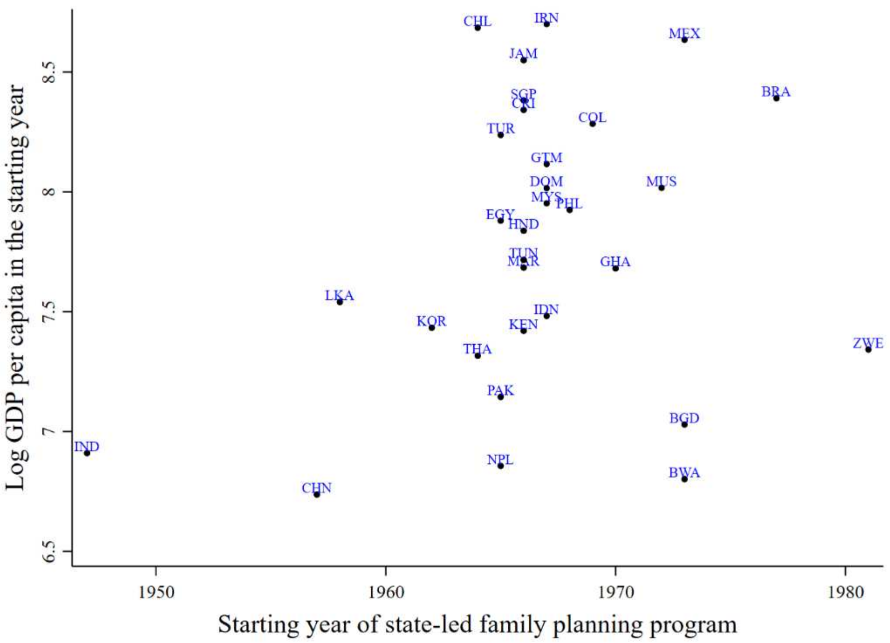

Figure 4. The Starting Year of State-Led Family Planning Programs and Log GDP per Capita in the Starting Year

The 2SLS estimates using the program starting year as the IV are presented in column 3 of Table 3. The estimation depends on continuous annual data from 1960 to 2016 for the 31 sample countries and controls for five potential confounding factors.[21] The first-stage estimates suggest that family planning programs significantly reduced fertility by 0.26 children per woman. The second-stage estimates suggest that a one-unit increase in TFR raised the long-term GDP per capita by 57.4%. As expected, the estimated effect is much larger when using the program starting year instrument than when using the predicted fertility instrument. This could reflect the fact that family planning programs directly (thus promptly) affected fertility, while disease interventions mainly indirectly affected fertility by affecting the size of childbearing

[20] There are three intensity measures of family planning programs: the effort score, the percentage of women exposed to family planning messages, and the funds for family planning per capita (De Silva and Tenreyro 2017). However, all these measures are not suitable for constructing IVs for fertility (when the target is to identify the long-term effect) because the data are only sparsely available over narrow (post-program) periods. For example, data on the percentage of women exposed to family planning messages are only available for several years from 1993 to 2013. More importantly, as argued by Bloom et al. (2009), these intensity measures are most likely endogenous to countries' economic performances.

[21] The five control variables are infant mortality, life expectancy, the share of urban population, net international migration, and the share of natural resource rents in GDP. Other (contemporaneous) control variables are not included because of the concern of overcontrol bias. For example, years of total schooling is not included because fertility could affect the long-term income growth by affecting education.

cohorts and parents' preferences regarding fertility, which may take decades to see the results of. Because the average TFR for the study sample is 4.1, this estimate suggests that a 10% increase in TFR could raise the long-term GDP per capita by 23.5%.

The 2SLS estimation is based on the assumption that—conditional on country and year fixed effects—the starting year of the program is exogenous to income growth. I found four pieces of evidence supporting this assumption. First, the consistency of the program starting years with the timing of the global family planning campaigns (see Figure 4) suggests that the program starting year is primarily determined by exogenous global family planning campaigns. Second, as presented in Figure 4, there is no obvious correlation between the program starting year and log GDP per capita in that year, suggesting that the income level of the country is not a determinant of the staring year. Third, the starting year is not determined by the income growth rates prior to it. As presented in Appendix Table B3 (columns 2–4), the average growth rate of GDP per capita 3, 5, or 10 years prior to the starting year has no effect on the starting year. Fourth, as presented in column 1 of Table B3, the 1-, 5-, and 10-year leads of the starting year have no significant effect on TFR. This finding confirms that the effect of the starting year on fertility indeed comes from family planning programs instead of from any omitted factors correlated with the starting year.

4. Evidence from China's One-Child Policy

This section depends on data from China's one-child policy (OCP) to estimate model (14). 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. From 1979–2010, the OCP generally allowed each couple to only have one child but had several exemption rules.[22] Residents who violated the OCP faced a stiff fine. Ebenstein (2010) collected province-level OCP violation fine rates (measured in times of local yearly household income) in China from 1979–2000, finding substantial cross-province and temporal variations in the fine rate (see Figure C1). Many studies have used the policy fine rate as an IV or 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

[22] 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).

(Huang et al. 2016), or various micro-level individual outcomes (Huang et al. 2020). Appendix Table C1 estimates that the policy fine rate had a significantly negative and robust effect on fertility.

I use the policy fine rate lagged by s years ($Fine_{p(t-s)}$) as the IV for the fertility rate lagged by the same years ($CBR_{p(t-s)}$) in the 2SLS estimation of a version of model (14):

$$g_{pt}^{\tilde{y}} = \nu_p + \tau_t + \psi_s CBR_{p(t-s)} + Z_{p(t-s)}\lambda + \varepsilon_{pt}^{s}, \quad s = 0,1,2,\cdots,T \quad . \tag{19}$$

The key explanatory variable, $CBR_{p(t-s)}$, is the crude birth rate (CBR) in province p lagged by s years. Here, fertility is measured by CBR instead of TFR because province-level TFR data are not available for China.[23] All other variables are the same as defined before but at the province level. The estimation depends on annual data from 1980 to 2010 for 27 of the 31 mainland Chinese provincial districts that enforced the OCP.[24] The data sources and summary statistics of all variables are presented in Table A3.

There are two alternative intensity measures of OCP used as IVs for fertility in the literature. The first 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 second is the ethnic minority population share (because ethnic minorities are subjected to less-strict birth control measures, see Footnote 22), which has been used to examine the effect of fertility on income growth (Li and Zhang 2007). The main analysis of this article does not depend on these two intensity measures because the data on excess fertility rate are only available for two census years, while the ethnic minority population share is subject to endogeneity concerns. However, as presented in Appendices D and E, analyses based on these two intensity measures also find that the long-term effect of fertility on income growth is significantly positive.

[23] CBR is defined as the annual number of births per thousand *population*, while TFR is defined as the average number of children that a woman would have over her childbearing years. Therefore, TFR is a better measure of current fertility than CBR because it is not affected by the age distribution of the population. Note that using CBR as the fertility measure in the global analysis yielded a comparable result (see Figure B1).

[24] 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 and Sichuan in 1988 and 1997, respectively. The year of 1979 is excluded because the OCP was implemented at the end of 1979, thus having no effect on 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.

The 2SLS estimation of (19) is based on the assumption that changes in the policy fine rates are exogenous. Studies using the policy fine rate as an IV have usually argued that changes in the province-level policy fine rate are determined by local-specific factors, such as the new provincial governors' preferences, which are not systematically correlated with the outcomes of interest. This argument is supported by Figure C1: no obvious common patterns are found for the timing, magnitude, or direction of the changes in the policy fine rate. A series of tests presented in Appendix C3 also supports the exogeneity assumption. First, Table C2 shows that prior income levels or growth rates have no predictive power on the current policy fine rate. Second, Table C3 shows that the policy fine rate is uncorrelated with various important determinants of income growth. Third, Table C4 shows that the lead of the policy fine rate is not correlated with current fertility and income growth.

The estimation further addresses endogeneity concerns by including various control variables. First, it controls for five time-varying determinants of fertility and growth: five-year lagged log GDP per capita, share of labor with secondary education, crude death rate, net migration rate, and share of urban population. All these controls are lagged by the same years as the CBR to avoid overcontrol bias. Second, the estimation controls for the indicators of two important events that may confound the effects: the tax system reform in 1994 and joining the World Trade Organization (WTO) in 2001. Specifically, the estimation controls for the interactions between the timing and intensity of each event.[25] China reformed its tax system in 1994 to triple the central government's share of revenues in GDP from 3% to 9% (Brandt and Rawski 2008, pp. 431–440). I control for this event by the interaction between the 1994 dummy and government spending share of GDP. China joined the WTO in 2001, which dramatically increased its international trade and liberalized its service sectors (Brandt and Rawski 2008, pp.657–659). I control for this event by using the interactions between the 2001 dummy and trade-to-GDP ratio and the contribution of services to the GDP, respectively. Finally, the model also controls for province-specific time trends.

[25] Another important event, the reform and opening up in 1978, is not controlled for because it occurred before the OCP. I have attempted to control for this event by the interactions between a full set of year dummies and two intensity measures of this event (the trade-to-GDP ratio and distance to the nearest port) and find a similar result.

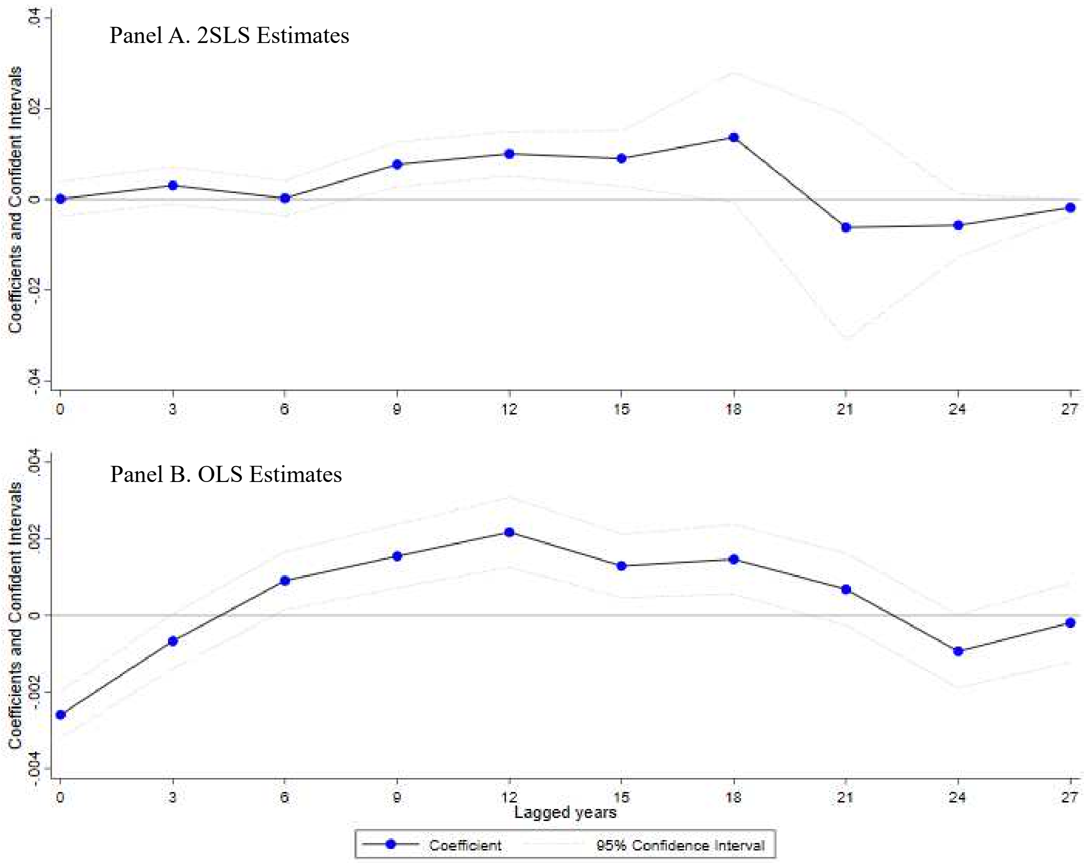

Figure 5. Current and Lagged Effects of Fertility on Income Growth in China

Notes: The figure presents the 2SLS estimates (Panel A) and OLS estimates (Panel B) of model (19). 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.

Panels A and B of Figure 5 present the 2SLS and OLS estimates, respectively, of model (19).[26] The 2SLS estimates show a similar effect pattern as the OLS estimates: the effect of a higher fertility rate first increases and then declines (although the 2SLS estimates of the initial effects are close to zero). As expected, the 2SLS estimates are much more positive than the OLS estimates. For example, when the lag length is 12 years, the OLS estimate suggests that a one-unit increase in CBR raises the growth rate of GDP per capita by 0.23%, but the 2SLS estimate suggests that the effect is as large as 0.99%. The average OLS estimate of ψ_s over 0 to 20 lagged years (the last year with a significant effect) is 0.07%, while the average 2SLS estimate over the same period is as high as 0.66%. When calculating the accumulated effect over 20 years, the OLS estimates suggest that a one-unit increase in CBR will raise GDP per capita by only

[26] Panel A of the figure only presents the second-stage estimates, and the first-stage estimates are reported in Table C1, which shows that the policy fine rate has a significantly negative and robust effect on CBR.

1.5%, while the 2SLS estimates suggest an accumulated effect of 14.1%.[27]

To make a comparison with the 2SLS estimates based on country-level data, I transform the marginal effect of CBR in China to the marginal effect of TFR. The country-average CBR and TFR in China from 1980 to 2010 were 16.9 and 1.99, respectively. As such, according to the 2SLS estimate, a one-unit increase in TFR would raise the long-term GDP per capita by 119.7% (i.e., 14.1*16.9/1.99). Thus, a 10% increase in TFR from its mean (1.99) raises the long-term GDP per capita in China by 23.8%. This estimate is close to the most credible 2SLS estimate obtained based on the timing of national family planning programs, which suggests that a 10% increase in TFR raises long-term GDP per capita by 23.5%.

Table 4 presents the robustness checks for the 2SLS estimates. For simplicity, the table only reports the estimates for contemporaneous CBR and the lagged CBR in five-year intervals, and the full results are reported in Table C5. Column 1 repeats the baseline 2SLS estimates reported in Figure 5. Recall that the baseline estimation controls for five time-varying factors, indicators of two concurrent events, and province-specific time trends. I test the robustness to these controls by excluding all of them (column 2), including only the five time-varying controls (column 3), and including only indicators of the two events (column 4). These tests produce results similar to the baseline estimates in column 1, suggesting that the estimates are not sensitive to omitted variables. Column 5 clusters the error term at the province level, based on the bootstrap procedure suggested by Cameron et al. (2008), given the small number of clusters. The clustered standard errors are slightly larger, but the significance levels are unaffected. Column 6 uses the one-additional-year lagged policy fine rate $Fine_{p(t-s-1)}$ as the IV for $CBR_{p(t-s)}$ to allow a lag for the translation of the policy fine rate change to fertility change. Doing so leads to a slightly larger estimated effect than that from the baseline estimation.

[27] The OLS estimates based on the data from China indicate shorter initial negative effects than the OLS estimates based on global data (Figure 2). This may reflect the fact that China implemented the most coercive birth control policy in the world. If more fertility declines were from birth control policies in China than in other countries, the OLS estimates based on the data from China tend to subject to less bias from reverse causality.

Table 4. Robustness Checks of the 2SLS Estimates

Lag length	(1) All controls	(2) No controls	(3) Five controls	(4) Five controls and two events	(5) Clustered at the province level	(6) Lagged policy fine rate
L0	0.0001	0.0023	0.0051	0.0004	0.0001	0.0017
	(0.0019)	(0.0024)	(0.0038)	(0.0020)	(0.0021)	(0.0021)
L5	0.0001	0.0041	0.0056*	0.0008	0.0001	0.0001
	(0.0019)	(0.0025)	(0.0033)	(0.0021)	(0.0021)	(0.0020)
L10	0.0102***	0.0100***	0.0115***	0.0104***	0.0102***	0.0105***
	(0.0027)	(0.0029)	(0.0038)	(0.0030)	(0.0029)	(0.0030)
L15	0.0090***	0.0088***	0.0086**	0.0086**	0.0090***	0.0096**
	(0.0031)	(0.0028)	(0.0034)	(0.0034)	(0.0042)	(0.0045)
L20	0.0050	0.0258	0.0031	0.0031	0.0050	-0.0039
	(0.0159)	(0.2116)	(0.0179)	(0.0179)	(0.0051)	(0.0106)
L25	-0.0062	-0.0078	-0.0059	-0.0059	-0.0062	-0.0095
	(0.0043)	(0.0073)	(0.0050)	(0.0050)	(0.0037)	(0.0130)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
5 controls	Yes	No	Yes	Yes	Yes	Yes
2 events	Yes	No	No	Yes	Yes	Yes
Time trends	Yes	No	No	No	Yes	Yes
Clustered standard error	No	No	No	No	Yes	No
Additional policy fine lag	No	No	No	No	No	Yes

Notes: Column 1 repeats the 2SLS estimates of Figure 5. The remaining columns test the robustness to the control variables (columns 2, 3, and 4), clustered standard errors (column 5), and an additional lag year of the policy fine rate (column 6). Robust standard errors are reported in parentheses. For simplicity, the table only reports the estimates in five-year intervals, and the full results are reported in Table C5. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

5. Concluding Remarks

This article has revisited the causal effect of fertility on long-run economic growth. Existing empirical studies on this topic generally find mixed results. This article highlights that failing to capture the long-term lagged effects of fertility is a crucial reason for the mixed findings found in the literature. When adopting estimation methods that can capture the long-term lagged effects of fertility, both the OLS and 2SLS estimates indicate a significantly positive long-term average effect of fertility on income growth. This result can be found separately for low-, middle-, and high-income countries, and when using province-level data from China.

This finding has important implications for economists and policy makers. Although the R&D-based growth models unambiguously predict the scale effect that faster population growth raises long-run economic growth, existing empirical studies find little evidence supporting this prediction. In contrast, development economists usually hypothesize a virtuous cycle between fertility decline and income growth based on the assumption that lower fertility promotes income growth (and the fact that higher income reduces fertility). This article provides strong evidence supporting the scale effect prediction of R&D-based growth theories, thus questioning the existence of the virtuous cycle between fertility decline and income growth. Although fertility has dramatically declined in almost all countries over the past few decades, many developing countries still adopt family planning programs to curb their population growth, partly motivated by the conventional view that fast population growth hinders long-run economic growth. However, the findings of this article suggest that fertility declines caused by family planning programs could significantly reduce these countries' long-run economic growth.

Reference

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. When should you adjust standard errors for clustering? : National Bureau of Economic Research.

Acemoglu, Daron, and Simon Johnson. 2007. "Disease and development: The effect of life expectancy on economic growth." *Journal of Political Economy* 115 (6):925-985.

Aghion, Philippe, and Peter Howitt. 1992. "A model of growth through creative destruction." *Econometrica*:323-351.

Almon, Shirley. 1965. "The distributed lag between capital appropriations and expenditures." *Econometrica* 33 (1):178-196.

Angeles, Luis. 2010. "Demographic transitions: analyzing the effects of mortality on fertility." *Journal of Population Economics* 23 (1):99-120.

Angrist, Joshua D, and William N Evans. 1998. "Children and their parents' labor supply: evidence from exogenous variation in family size." *American Economic Review* 88 (3):450-477.

Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household bargaining and excess fertility: An experimental study in Zambia." *American Economic Review* 104 (7):2210-2237.

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and stagnation in the Malthusian epoch." *American Economic Review* 101 (5):2003-41.

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.

Barlow, Robin. 1994. "Population growth and economic growth: Some more correlations." *Population and Development Review*:153-165.

Barro, Robert J. 1991. "Economic growth in a cross section of countries." *The Quarterly Journal of Economics* 106 (2):407-443.

Barro, Robert J., and Xavier Sala-i-Martin. 2003. *Economic growth*. Second Edition ed. Cambridge, Massachusetts: The MIT Press.

Becker, Gary S, Kevin M Murphy, and Robert Tamura. 1990. "Human capital, eertility, 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.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "The more the merrier? The effect of family size and birth order on children's education." *The Quarterly Journal of Economics* 120 (2):669-700.

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.

Bloom, David E, and Jeffrey G Williamson. 1998. "Demographic transitions and economic miracles in emerging Asia." *The World Bank Economic Review* 12 (3):419-455.

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

Boucekkine, Raouf, David De la Croix, and Omar Licandro. 2002. "Vintage human capital, demographic trends, and endogenous growth." *Journal of Economic Theory* 104 (2):340-375.

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

Cáceres-Delpiano, Julio. 2006. "The impacts of family size on investment in child quality." *The Journal of Human Resources* 41 (4):738-754.

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.

Cleland, John, Stan Bernstein, Alex Ezeh, Anibal Faundes, Anna Glasier, and Jolene Innis. 2006. "Family planning: The unfinished agenda." *The Lancet* 368 (9549):1810-1827.

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

Coale, Ansley Johnson, and Edgar M Hoover. 1958. *Population growth and economic development in low income countries: A case study of India's prospects*. Princeton: Princeton University Press.

d'Albis, Hippolyte. 2007. "Demographic structure and capital accumulation." *Journal of Economic Theory* 132 (1):411-434.

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.

Eckstein, Zvi, Pedro Mira, and Kenneth I. Wolpin. 1999. "A quantitative analysis of Swedish fertility dynamics: 1751–1990." *Review of Economic Dynamics* 2 (1):137-165.

Ehrlich, Paul R. 1968. *The population bomb*. New York: Ballentine.

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, O. 2005. "The demographic transition and the emergence of sustained economic growth." *Journal of the European Economic Association* 3 (2-3):494-504.

Galor, Oded. 2011. *Unified growth theory*: Princeton University Press.

Galor, Oded, and David N Weil. 1996. "The gender gap, eertility, and growth." *American Economic Review*:374-387.

Galor, Oded, and David N Weil. 2000. "Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond." *American Economic Review*:806-828.

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

Greene, William H. 2010. *Econometric analysis*. 7 ed. New Jersey: Prentice Hall.

Grossman, Gene M, and Elhanan Helpman. 1991. "Quality ladders in the theory of growth." *The Review of Economic Studies* 58 (1):43-61.

Hafner, Kurt A., and David Mayer-Foulkes. 2013. "Fertility, economic growth, and human development causal determinants of the developed lifestyle." *Journal of Macroeconomics* 38:107-120.

Hanushek, Eric A. 1992. "The trade-off between child quantity and quality." *Journal of political economy* 100 (1):84-117.

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.

Howitt, Peter. 1999. "Steady endogenous growth with population and R&D inputs growing." *Journal of Political Economy* 107 (4):715-730.

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

Kortum, Samuel S. 1997. "Research, patenting, and technological change." *Econometrica*:1389-1419.

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

Kuang, Bernice, and Isabel Brodsky. 2016. "Global trends in family planning programs, 1999–2014." *International Perspectives on Sexual and Reproductive Health* 42 (1):33-44.

Lee, Jungmin. 2008. "Sibling size and investment in children's education: An Asian instrument." *Journal of Population Economics* 21 (4):855-875.

Lee, Ronald, and Michael Anderson. 2002. "Malthus in state space: Macro economic-demographic relations in English history, 1540 to 1870." *Journal of Population Economics* 15 (2):195-220.

Leibowitz, Arleen. 1974. "Home investments in children." *Journal of Political Economy* 82 (2, Part 2):S111-S131.

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

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

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A contribution to the empirics of economic growth." *The Quarterly Journal of Economics* 107 (2):407-437.

Mayer, Thierry, and Soledad Zignago. 2011. "Notes on CEPII's distances measures: The GeoDist database." *CEPII Working Paper 2011-25*.

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

Olsen, Randall J., and Kenneth I. Wolpin. 1983. "The Impact of Exogenous Child Mortality on Fertility: A Waiting Time Regression with Dynamic Regressors." *Econometrica* 51 (3):731-749.

Peretto, Pietro F. 1998. "Technological change and population growth." *Journal of Economic Growth* 3 (4):283-311.

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.
Romer, Paul M. 1990. "Endogenous technological change." *Journal of Political Economy*:S71-S102.
Roser, Max. 2014. Fertility Rate. https://ourworldindata.org/fertility-rate: OurWorldInData.
Sachs, Jeffrey D., Andrew Warner, Anders Åslund, and Stanley Fischer. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1995 (1):1-118.
Schultz, Paul. 2008. "Population policies, fertility, women's human capital, and child quality." In *Handbook of Development Economics*, 3249-3303. Elsevier.
Segerstrom, Paul S. 1998. "Endogenous growth without scale effects." *American Economic Review*:1290-1310.
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.
Stover, John, and Emily Sonneveldt. 2017. "Progress toward the goals of FP2020." *Studies in Family Planning* 48 (1):83-88.
Strulik, Holger. 2003. "Mortality, the trade-off between child quality and quantity, and demo-economic development." *Metroeconomica* 54 (4):499-520.
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.
Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT press.

Online Appendix

A. Data Appendix

Table A1 presents the data sources and summary statistics of all the variables used in the country-level analyses. Table A2 lists the 138 sample countries. Table A3 presents the data sources and summary statistics of all the variables used in the China province-level analyses. The growth rate of GDP per capita and fertility rates are calculated as five-year moving averages to reduce the confounding effects of short-term fluctuations; similar results are obtained if the moving averages are not taken into consideration.

Table A1. Sources and Summary Statistics of Country-level Data

Variable Name	Definition	Source	Mean
Growth rate of GDP per capita	Annual growth rate of real GDP per capita in 2011 USD	A	0.019
Total fertility rate	The average number of children a woman would have over childbearing years	B, E	4.1
Five-year lagged GDP per capita	Five-year lagged GDP per capita in natural log, 2011 USD	A	8.5
Infant mortality rate	The number of deaths per 1,000 live births	B	58.5
Life expectancy	Life expectancy at birth, year	B	62.6
Years of schooling	Years of total schooling for individuals aged 25 or over	C	5.1
Income ratio to US	GDP per capita as ratio of US GDP per capita	A	0.3
The share of urban population	Urban population as a percentage of total population, %	B	47.7
Migration	Net international out-migration, 1,000	B	801
Resource rents	The share of natural resource rents in GDP, %	B	6.6
Landlocked	The dummy of landlocked	F	0.2
Official language	The first official language	F	--
Political stability	Average of the political stability index	B	-0.2
Program staring year	Starting year of national family planning program	D	1966
Predicted mortality	The 1940 predicted mortality rate, 100 individuals per year	G	0.46

Note: 1. Data sources:

A: The Maddison Project Database 2018

B: World Development Indicators, the World Bank

C: Barro and Lee (2013)

D: De Silva and Tenreyro (2017)

E: Roser (2014), for data before 1960

F: Mayer and Zignago (2011)

G: Acemoglu and Johnson (2007)

2. All variables are at the country level. All data are for the 138 sample countries in the period of 1960–2016.

Table A2. Sample Countries

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

Note: This table lists the 138 sample countries (or regions) for which the data on the growth rate of GDP per capita and TFR are available for each year from 1960 to 2016.

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	A, B	0.078
Crude birth rate	The annual number of births per thousand population	A, B	17.9
Five-year lagged GDP per capita	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 the population in urban areas	Percentage of the population living in urban areas	A, B	0.30
Crude death rate	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
Distance to port	The distance from each province's centroid to the nearest port, 100 km	D	5.15
Share of services in GDP	The contribution of services to GDP	A, B	0.29
Policy fine rate	The average monetary penalty rate for one unauthorized birth, in years of local household income	E	1.74
Minority population share	Percentage of minorities in the population	F, G	0.10

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: China province Shapefile
E: The dataset of Ebenstein (2010)
F: National Population Census of the PRC (decennial census)
G: The 1% Population Sample Survey (during the inter-censual years ending with 5)

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

B. Global Evidence Appendix

B1. Fertility Variation in the Fixed Effects Panel Model

Because fixed effects eliminate all cross-sectional long-term fertility differences, the identification of the fertility coefficient in a fixed-effects panel model depends mainly on interannual fertility changes. Considering that fertility declines continuously over time, most of the fertility changes used in the identification are short-term fertility changes. To see this, assume that we regress the income growth rate on the fertility rate in a fixed effects panel model by using a panel of data covering five decades. As an illustration, the following matrix presents the contemporaneous and lagged fertility changes used in the estimation:

$$V = \begin{pmatrix} v_0^1 & v_1^1 & v_2^1 & v_3^1 & v_4^1 \\ & v_0^2 & v_1^2 & v_2^2 & v_3^2 \\ & & v_0^3 & v_1^3 & v_2^3 \\ & & & v_0^4 & v_1^4 \\ & & & & v_0^5 \end{pmatrix},$$

where v_i^j denotes the fertility change that occurred in decade j and affected income growth i decades later. For example, v_0^2 denotes the fertility change that occurred in the second decade and affected the income growth in the same decade, and v_2^1 denotes the fertility change that occurred in the first decade and affected the income growth two decades later. By assuming that fertility declines evenly over time, it can be calculated that only 40% of the fertility changes used in the identification occurred two decades ago, that is, $\left(v_2^1 + v_2^2 + v_2^3 + v_3^1 + v_3^2 + v_4^1\right) \Big/ \sum v_i^j = 6/15 = 40\%$. Because it takes about two decades for a newborn to grow up to be an adult, it is appropriate to define "long-term" in this study as longer than two decades. Therefore, even for a panel of data covering five decades, a large share (60%) of the fertility variation used in the identification reflects "short-term" fertility changes. Note that if the time span of the panel data is shorter than 20 years, no long-term effects of fertility can be captured in the fixed-effects panel model.

Two ways of capturing the long-term lagged effects of fertility in a fixed effects panel model have emerged from this illustration. The first is to use lagged fertility as the explanatory variable, which is the primary estimation strategy used in this article.

For example, when fertility is lagged by two decades, the matrix V turns out to be:

$$V_{L2} = \begin{pmatrix} 0 & 0 & v_2^1 & v_3^1 & v_4^1 \\ & 0 & 0 & v_2^2 & v_3^2 \\ & & 0 & 0 & v_2^3 \\ & & & 0 & 0 \\ & & & & 0 \end{pmatrix},$$

which shows that all of the remaining fertility changes occurred two decades ago. Intuitively, by estimating the effects of the fertility rate lagged by different years, we can reveal both the short-term and long-term effects of fertility.

The second way to identify the long-term effect of fertility in a fixed-effects panel model is to employ one-off fertility shocks occurring in the early stages of the sample period. For example, if there is only a one-off fertility shock that occurred in the first decade, matrix V would become:

$$V_{shock} = \begin{pmatrix} v_0^1 & v_1^1 & v_2^1 & v_3^1 & v_4^1 \\ & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 \\ & & & 0 & 0 \\ & & & & 0 \end{pmatrix}.$$

In this case, the fixed effects panel model captures the lagged effects of the fertility shock over the four decades following the shock. This identification strategy follows the same logic as the standard DID model, which is usually used to identify the long-term effect of a one-off treatment. In practice, one-off fertility shocks can be introduced to the model by IVs constructed from exogenous events. However, to the extent that the event employed has lagged effects on fertility, this approach tends to capture more of the short-term effect.

B2. Full Results of Column 1 in Table 2

Table B1 presents the full estimates of the fertility coefficients from model (14), with a maximum lag length of 50 years.

Table B1. Full Results of Column 1 in Table 2

Lags	Coef.	Se	**Lags**	Se	Lags	**Se**	Coef.	Se
L0	-0.0056	0.0010	**L17**	0.0020	0.0016	**L34**	-0.0009	0.0023
L1	-0.0061	0.0010	**L18**	0.0030	0.0016	**L35**	-0.0018	0.0025
L2	-0.0064	0.0011	**L19**	0.0044	0.0017	**L36**	-0.0021	0.0026
L3	-0.0064	0.0011	**L20**	0.0057	0.0017	**L37**	-0.0025	0.0026
L4	-0.0063	0.0011	**L21**	0.0067	0.0017	**L38**	-0.0021	0.0024
L5	-0.0059	0.0011	**L22**	0.0078	0.0017	**L39**	-0.0007	0.0024
L6	-0.0054	0.0012	**L23**	0.0089	0.0018	**L40**	-0.0007	0.0026
L7	-0.0049	0.0012	**L24**	0.0095	0.0019	**L41**	-0.0009	0.0031
L8	-0.0045	0.0012	**L25**	0.0101	0.0021	**L42**	-0.0008	0.0037
L9	-0.0040	0.0013	**L26**	0.0109	0.0023	**L43**	-0.0025	0.0042
L10	-0.0036	0.0013	**L27**	0.0110	0.0026	**L44**	-0.0056	0.0043
L11	-0.0032	0.0014	**L28**	0.0104	0.0027	**L45**	-0.0062	0.0047
L12	-0.0026	0.0014	**L29**	0.0096	0.0028	**L46**	-0.0057	0.0049
L13	-0.0018	0.0014	**L30**	0.0079	0.0027	**L47**	-0.0069	0.0056
L14	-0.0010	0.0015	**L31**	0.0056	0.0024	**L48**	-0.0046	0.0066
L15	-0.0001	0.0015	**L32**	0.0032	0.0022	**L49**	0.0040	0.0094
L16	0.0009	0.0015	**L33**	0.0010	0.0021	**L50**	0.0286	0.0164

B3 Additional Robustness Checks

This appendix presents three additional robustness checks for the baseline estimates in Figure 2. First, as presented in Figure B1, I use an alternative fertility measure—CBR—to replace TFR and find estimates very similar to the baseline estimates (although in different units). The main analysis does not measure fertility by CBR because it is sensitive to the age distribution of the population (see Footnote 24). Figure B2 excludes sample countries with a 1960 population smaller than 2 million (40 countries), and Figure B3 excludes sample countries with a 1960 population larger than 50 million (12 countries). The resulting estimates are quite similar to the baseline estimates, suggesting that the findings are not sensitive to relatively small or large countries. Figure B4 excludes sample countries from America, Asia, Europe, and Sub-Saharan Africa, respectively, showing that the baseline estimates are not mainly driven by countries from a specific region.

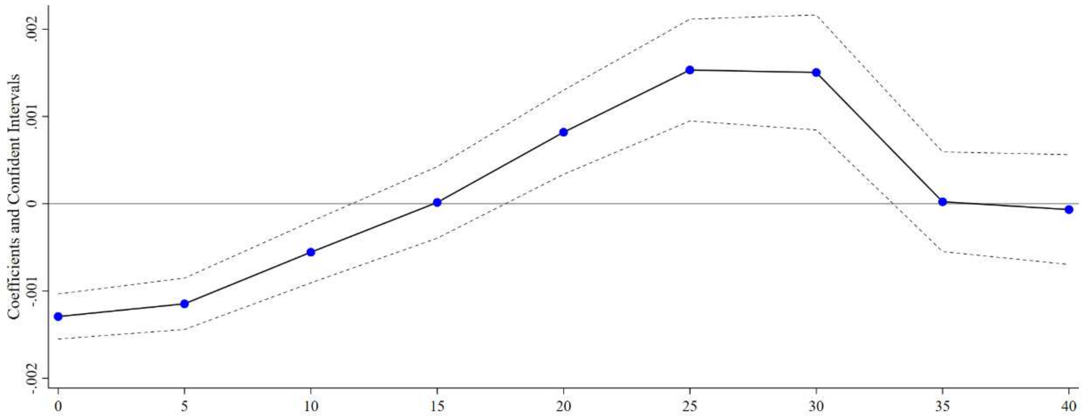

Figure B1. Measuring Fertility by Crude Birth Rate

Note: This figure replicates Figure 2. The only difference is that this figure measures fertility by the crude birth rate instead of by TFR.

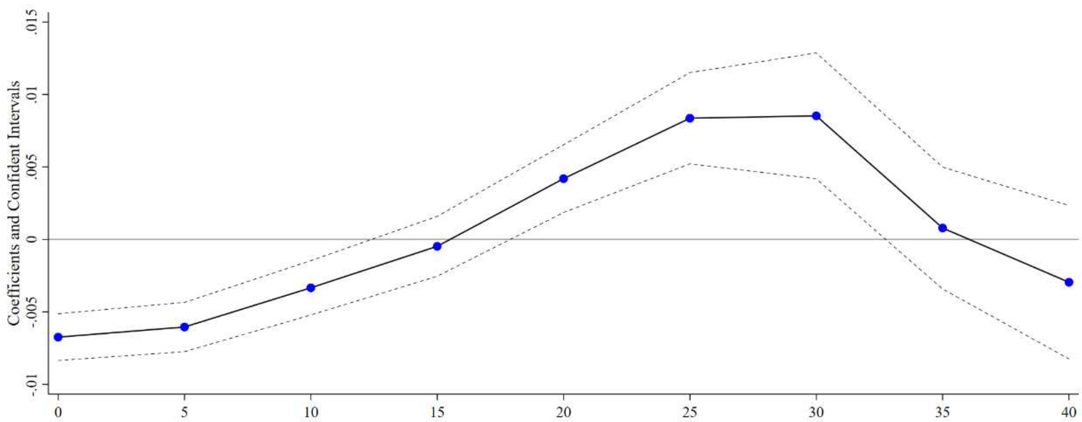

Figure B2. Excluding Small Countries

Note: This figure replicates Figure 2. The only difference is that countries with a 1960 population smaller than 2 million were excluded.

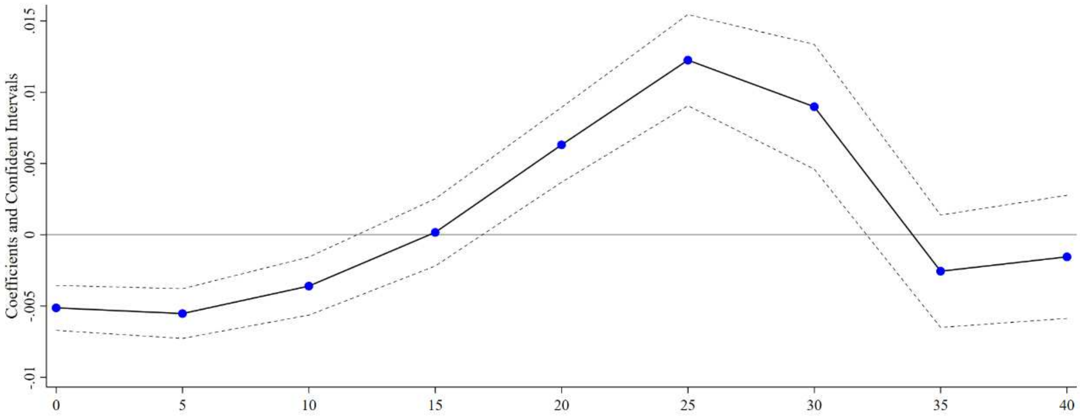

Figure B3. Excluding Large Countries

Note: This figure replicates Figure 2. The only difference is that countries with a 1960 population larger than 50 million were excluded.

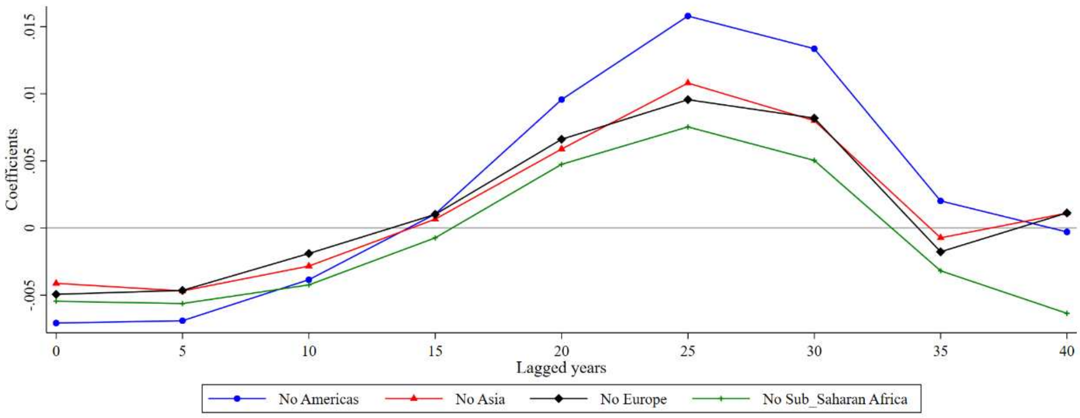

Figure B4. Excluding Countries from each Region

Note: This figure replicates Figure 2. The only difference is that countries from America, Asia, Europe, and Sub-Saharan Africa, respectively, were excluded.

B4 The Exclusion Restriction

Table B2 provides evidence to support the exogeneity of predicted mortality and predicted fertility. Column 1 regresses TFR on current predicted mortality, lag predicted mortality, and lead predicted mortality. Note that because the data used in the estimation are in 10-year intervals, one lag (lead) corresponds to 10 years. It finds that the lead predicted mortality has no effect on TFR, confirming that disease interventions—instead of preexisting trends—affect TFR. In addition, lag predicted mortality has a significant effect on TFR, reflecting the fact that it takes time for fertility to respond to disease interventions. Similarly, column 2 finds that lead predicted fertility has no effect on TFR, while lag predicted fertility has a significant effect. Columns 3 and 4 regress the changes in predicted mortality from 1940 to 2010 on changes in log GDP per capita and TFR from 1930 to 1940, respectively. These regressions indicate no effect of preintervention changes in income and TFR on post-intervention changes in predicted mortality, confirming the exogeneity of the predicted mortality instrument. Similarly, columns 5 and 6 find no effect of preintervention income and TFR on the predicted fertility instrument.

Table B2. Exogeneity of Predicted Mortality and Predicted Fertility

	(1)	(2)	(3)	(4)	(5)	(6)
	TFR	TFR	Changes in predicted mortality (1940–2010)	Changes in predicted mortality (1940–2010)	Changes in predicted fertility (1940–2010)	Changes in predicted fertility (1940–2010)
Predicted mortality	0.27					
	(0.19)					
Lag predicted mortality	0.70***					
	(0.15)					
Lead predicted mortality	0.14					
	(0.15)					
Predicted fertility		0.32				
		(0.47)				
Lag predicted fertility		1.37***				
		(0.36)				
Lead predicted fertility		-0.59				
		(0.38)				
Changes in log GDP per capita (1930–1940)			0.03		-0.56	
			(0.32)		(0.59)	
Changes in TFR (1930–1940)				0.19		0.38
				(0.14)		(0.27)
Country FE	Yes	Yes				
Year FE	Yes	Yes				

Observations	340	340	34	34	34	34

Notes: Column 1 regresses TFR on current predicted mortality, lag predicted mortality, and lead predicted mortality. Column 2 regresses TFR on current predicted fertility, lag predicted fertility, and lead predicted fertility. Columns 3 and 4 regress post-intervention changes in predicted mortality (1940–2010) on preintervention changes in log GDP per capita and TFR (1930–1940), respectively. Columns 5 and 6 regress post-intervention changes in predicted fertility (1940–2010) on preintervention changes in log GDP per capita and TFR (1930–1940), respectively. Robust standard errors are reported in parentheses. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

Table B3 presents evidence supporting the exogeneity of the starting year of national family planning programs. Column 1 regresses TFR on the starting year dummy, and 1-, 5-, and 10-year leads of the starting year dummy. It finds that the leads have no significant effect on fertility, confirming that it is the family planning programs instead of pre-existing trends that affect TFR. Columns 2–4 regress the starting year on the average growth rate of GDP per capita 3, 5, and 10 years prior to the starting year, respectively. The regressions show that the growth rates prior to the starting year have no significant effect on the starting year, which eases the concern that the starting year could be determined by the economic performance of the country examined.

Table B3. Exogeneity of the Starting Year of Family Planning Programs

	(1) TFR	(2) Starting year	(3) Starting year	(4) Starting year
Starting year	-0.17*** (0.08)			
1-year lead starting year	0.02 (0.10)			
5-year lead starting year	-0.10 (0.07)			
10-year lead starting year	0.03 (0.10)			
Average growth rate 3 years prior to the starting year		-6.94 (34.4)		
Average growth rate 5 years prior to the starting year			0.59 (33.7)	
Average growth rate 10 years prior to the starting year				-12.3 (36.4)
Country FE	Yes			
Year FE	Yes			
Observations	1373	31	31	31

Notes: Column 1 regresses TFR on the 1-, 5-, and 10-year leads of the family planning program starting year dummy. Columns 2–4 regress the starting year on the average growth rate of GDP per capita 3, 5, and 10 years prior to the starting year, respectively. Robust standard errors are reported in parentheses. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

C. China Evidence Appendix

C1. Spatial and Temporal Variations in the Policy Fine Rate

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

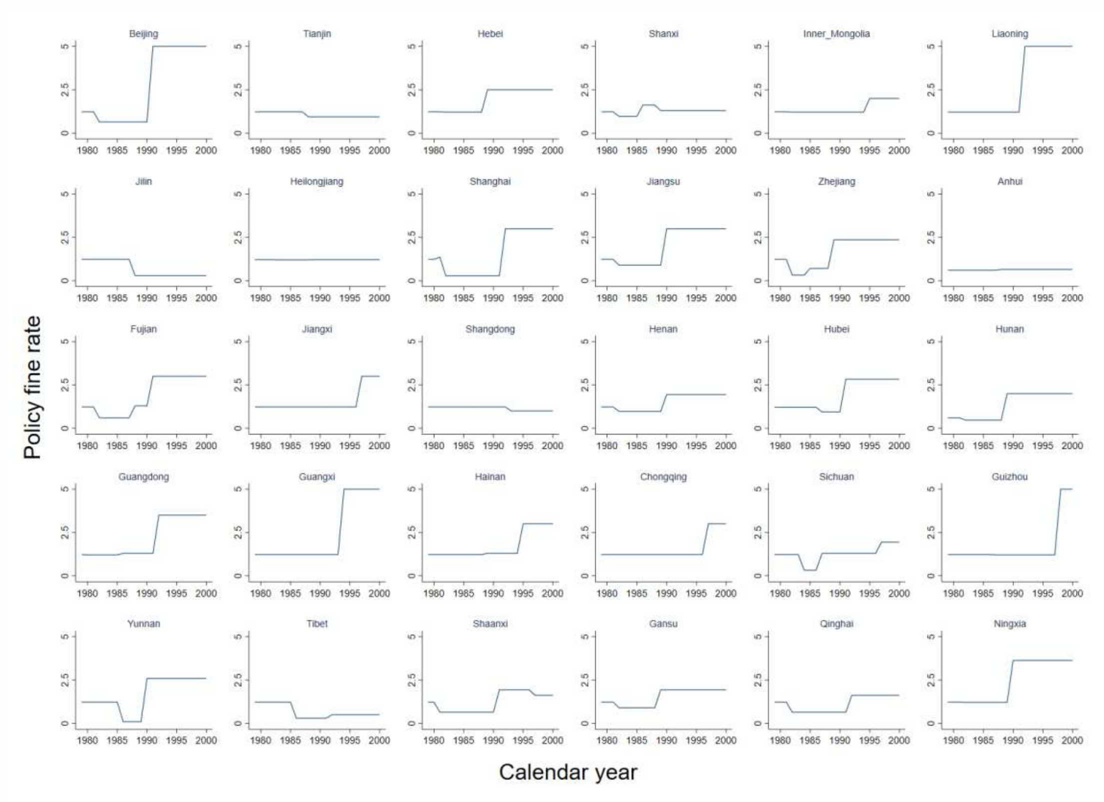

Figure C1. One-Child Policy Violation Fine Rates, 1979–2000

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. First-Stage Estimates of the 2SLS Estimation

Table C1 reports the first-stage estimates of the 2SLS estimation of model (19). Column 1 regresses the CBR on the policy fine rate in a panel model that includes province and year fixed effects. The estimates suggest that a one-unit increase in the policy fine rate reduces CBR by 0.42, and this effect is statistically significant at the 1% level. This finding is consistent with previous studies showing that the OCP significantly reduced fertility in China.[28] The remaining columns provide robustness tests. Column 2 includes five time-varying control variables (five-year lagged GDP per capita, share of labor with secondary education, share of urban population, crude death rate, and out-migration rate); column 3 controls for the three indicators of two concurrent events (as detailed in the main text) and province-specific linear time trends; and column 4 clusters the error term at the province level using a bootstrap procedure. Estimates from these robustness checks are very similar to those in column 1, suggesting that the policy fine rate has a robust effect on CBR.

Table C1. First-Stage Regression Results

	Dependent variable: The crude birth rate			
	(1)	(2)	(3)	(4)
Policy fine rate (years of household income)	-0.42*** (0.11)	-0.43*** (0.11)	-0.44*** (0.11)	-0.44*** (0.11)
Time-varying control variables		Yes	Yes	Yes
Indicators of concurrent events			Yes	Yes
Province time trends			Yes	Yes
Clustered stander error				Yes
Province FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	781	781	781	781
R-squared	0.870	0.886	0.887	0.887

Notes: The standard errors (in parentheses) account for arbitrary heteroskedasticity. Significance levels are *** p<0.01, ** p<0.05, * p<0.1.

[28] 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.

C3. Exogeneity of the Policy Fine Rate

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

C3.1 *The predictive power of preexisting income levels and growth rates*

A major endogeneity concern is that the policy fine rate could be determined by income levels or growth rates. To ease 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 in columns 1a, 2a, and 3a (columns 1b, 2b, and 3b), respectively. All estimates are small and have a p-value larger than 0.1, suggesting that the policy fine rate is not determined by prior income growth rates or levels. Although it is infeasible to directly examine whether the policy fine rate is determined by expectations of future income, it seems reasonable to believe that if the policy fine rates were not set based on the readily available information of past income, they were even less likely to be set based on the uncertain predictions of future income.

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

	Dependent variable: One-child policy fine rate					
	1 year later		3 years later		5 years 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 in columns 1a, 2a, and 3a (columns 1b, 2b, and 3b), respectively. 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.

C3.2 *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 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 eased 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 the Nine Growth Determinants

	Dependent variable: One-child policy fine rate			
	(1) 1 year later	(2) 3 years later	(3) 5 years 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

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 six growth determinants. Column 4 regresses the changes in the policy fine rate on the changes in each of the growth determinants. 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 time-varying growth determinants. Specifically, I regress the policy fine rates in the next one, three, and five years, respectively, on each of the six determinants in a panel model with province and year fixed effects (columns 1, 2, and 3 of Table C3). I also examine 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 variables are smaller than 0.1, suggesting no significant correlation with the policy fine rate. I have also examined the joint significance of all or subsets of these variables and still find no significant association.

C3.3 *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 this comparison, column 1 of Table C4 lists the baseline 2SLS estimates presented in Figure 5. 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 no 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	0.001 (0.001)		0.001 (0.001)	
10-year lagged CBR		0.01*** (0.003)		0.01*** (0.004)
5-year lead policy fine rate			-0.001 (0.001)	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	-0.50*** (0.11)		-0.55*** (0.12)	
10-year lagged policy fine rate		-0.56*** (0.13)		-0.55*** (0.15)
5-year lead policy fine rate			-0.22 (0.19)	-0.46 (0.33)
Province FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R-squared	0.86	0.82	0.81	0.85

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 (19). 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.

C4. Full Results of Table 4

Table C5. Various Robustness Checks of the 2SLS Estimates (Full results of Table 4)

Lags	(1a) Coefficient	(1b) Se	(2a) Coefficient	(2b) Se	(3a) Coefficient	(3b) Se	(4a) Coefficient	(4b) Se	(5a) Coefficient	(5b) Se	(6a) Coefficient	(6b) Se
0	0.0001	0.0020	0.0023	0.0024	0.0051	0.0038	0.0005	0.0020	0.0001	0.0021	0.0017	0.0021
1	0.0019	0.0021	0.0038	0.0027	0.0084	0.0046	0.0024	0.0023	0.0019	0.0021	0.0031	0.0023
2	0.0031	0.0022	0.0049	0.0028	0.0109	0.0051	0.0037	0.0024	0.0031	0.0022	0.0031	0.0022
3	0.0031	0.0021	0.0054	0.0028	0.0106	0.0046	0.0038	0.0022	0.0031	0.0021	0.0022	0.0021
4	0.0017	0.0020	0.0049	0.0027	0.0084	0.0039	0.0025	0.0022	0.0017	0.0020	0.0007	0.0021
5	0.0002	0.0020	0.0041	0.0026	0.0057	0.0034	0.0009	0.0021	0.0002	0.0021	0.0001	0.0021
6	0.0003	0.0020	0.0043	0.0026	0.0046	0.0033	0.0009	0.0021	0.0003	0.0020	0.0014	0.0021
7	0.0019	0.0020	0.0051	0.0027	0.0050	0.0034	0.0026	0.0022	0.0019	0.0020	0.0042	0.0024
8	0.0045	0.0023	0.0066	0.0028	0.0070	0.0036	0.0051	0.0025	0.0045	0.0023	0.0073	0.0028
9	0.0077	0.0025	0.0085	0.0029	0.0096	0.0039	0.0081	0.0028	0.0077	0.0025	0.0099	0.0030
10	0.0102	0.0027	0.0100	0.0029	0.0116	0.0039	0.0104	0.0030	0.0102	0.0029	0.0105	0.0030
11	0.0105	0.0026	0.0098	0.0026	0.0107	0.0033	0.0104	0.0028	0.0105	0.0026	0.0102	0.0028
12	0.0100	0.0025	0.0097	0.0024	0.0101	0.0030	0.0099	0.0027	0.0100	0.0025	0.0098	0.0027
13	0.0101	0.0026	0.0099	0.0025	0.0101	0.0030	0.0099	0.0028	0.0101	0.0026	0.0089	0.0030
14	0.0090	0.0027	0.0090	0.0026	0.0087	0.0030	0.0087	0.0029	0.0090	0.0027	0.0084	0.0036
15	0.0090	0.0031	0.0088	0.0029	0.0087	0.0035	0.0087	0.0034	0.0090	0.0042	0.0096	0.0045
16	0.0102	0.0037	0.0093	0.0034	0.0097	0.0042	0.0097	0.0042	0.0102	0.0037	0.0131	0.0069
17	0.0132	0.0051	0.0107	0.0044	0.0130	0.0059	0.0130	0.0059	0.0132	0.0051	0.0112	0.0088
18	0.0137	0.0073	0.0088	0.0058	0.0132	0.0083	0.0132	0.0083	0.0137	0.0073	0.0136	0.0333
19	0.0133	0.0129	0.0061	0.0123	0.0113	0.0150	0.0113	0.0150	0.0133	0.0129	-0.0176	0.0964
20	0.0050	0.0159	0.0258	0.2116	0.0031	0.0180	0.0031	0.0180	0.0050	0.0051	-0.0039	0.0106
21	-0.0062	0.0127	-0.0020	0.0290	0.0003	0.0089	0.0003	0.0089	-0.0062	0.0127	-0.0154	0.0204
22	-0.0100	0.0077	-0.0061	0.0111	-0.0040	0.0060	-0.0040	0.0060	-0.0100	0.0077	-0.0185	0.0227
23	-0.0093	0.0054	-0.0058	0.0072	-0.0045	0.0050	-0.0045	0.0050	-0.0093	0.0054	-0.0141	0.0152
24	-0.0057	0.0035	-0.0040	0.0051	-0.0037	0.0037	-0.0037	0.0037	-0.0057	0.0035	-0.0100	0.0088
25	-0.0062	0.0044	-0.0079	0.0073	-0.0060	0.0050	-0.0060	0.0050	-0.0062	0.0037	-0.0096	0.0130
26	-0.0029	0.0023	-0.0054	0.0057	-0.0028	0.0026	-0.0028	0.0026	-0.0029	0.0023	-0.0058	0.0058
27	-0.0018	0.0009	-0.0035	0.0020	-0.0018	0.0010	-0.0018	0.0010	-0.0018	0.0009	0.0008	0.0024
28	-0.0001	0.0009	0.0000	0.0013	-0.0002	0.0009	-0.0002	0.0009	-0.0001	0.0009	-0.0073	0.0051

Notes: This table presents the full results of Table 4. Columns 1 repeats the 2SLS estimation of Figure 5, column 2 excludes all control variables, column 3 includes the five controls, column 4 includes the indicators of the two events, column 5 clusters the standard error at province level, and column 6 uses the policy fine rate lagged by an additional year.

D. Evidence from 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.[29] Instead, this appendix uses the EFR as an intensity measure in a DID model to estimate the long-term 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:

$$EFR_{p,t} = \frac{\sum_j \left(Birth_{pjt} \times 1\left(NSC_{pjt} \geq 2\right)\right)}{\sum_j 1\left(NSC_{pjt} \geq 1\right) - \sum_j \left(Birth_{pjt} \times 1\left(NSC_{pjt} = 1\right)\right)} \times 100 \ , \tag{20}$$

where $Birth_{pjt}$ is an indicator 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

[29] 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.

who gave a higher order birth in year t.[30] 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 22). 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.

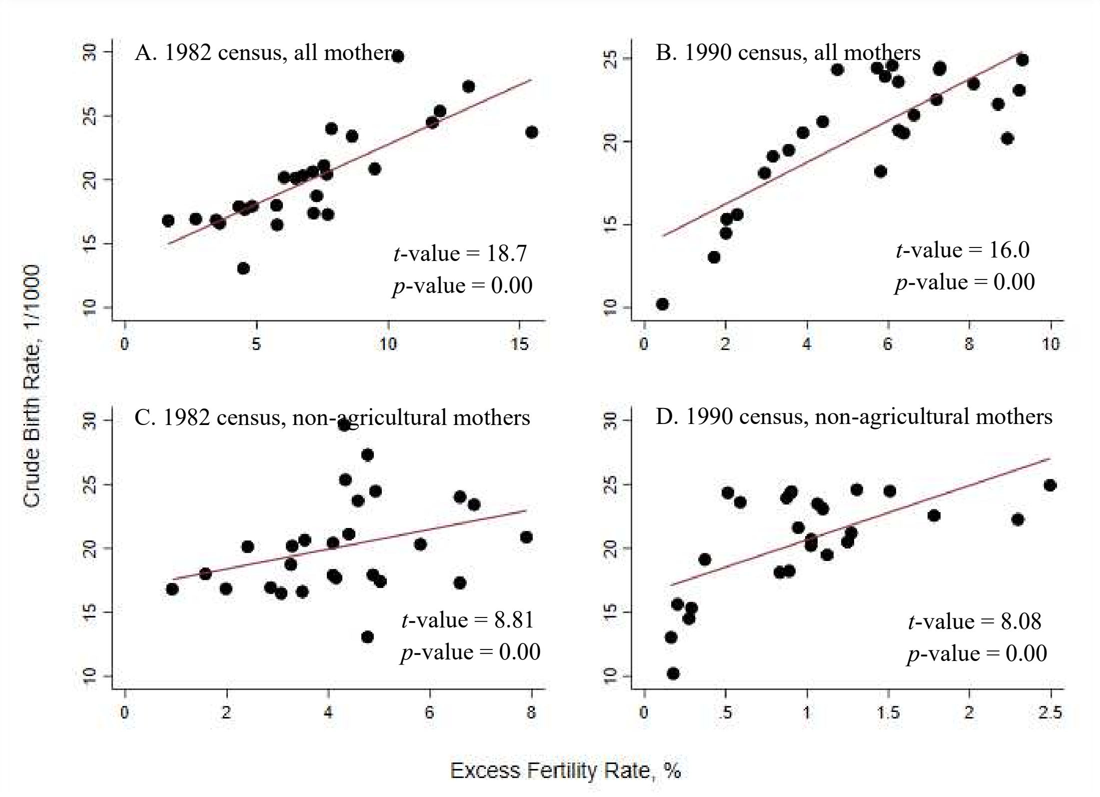

Figure D1. Correlation between the CBR and the EFR Calculated based on the Population Censuses 1982 and 1990

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.

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

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

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.

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-term 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:

$$y_{pt} = \nu_p + \tau_t + \alpha_1 EFR_{p,1981} \times post_t + Z_{pt}\lambda + \vartheta_{pt} \quad , \tag{21}$$

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-term average effect of a change in fertlity on income growth.

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.

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 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 (21) 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 (21) are presented in Table D2. The estimations were based on 1970–2010 data for the 27 sample provinces.[31] The baseline estimates presented in

[31] 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.

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-term average effect of higher fertility rates on the aggregate income growth rate is significantly positive.

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)
Five 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 (21). 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 (21) 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 section D4. Column 2 of Table D2 includes the five 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 the main text 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 (21) that replaces $EFR_{p,1981} \times post_t$ with $CBR_{p,t} \times post_t$. The first-stage regression of the 2SLS estimation is

$$CBR_{p,t} \times post_t = \nu_p + \tau_t + \beta EFR_{p,1981} \times post_t + Z_{pt}\lambda + \mu_{it} \quad . \tag{22}$$

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 (21) 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 (21) that includes the interactions between the 1981 EFR and a full set of year dummies:

$$y_{pt} = \nu_p + \tau_t + \sum_{j=1971}^{2010} \alpha_j EFR_{p,1981} \times dummy_j + Z_{pt}\lambda + \vartheta_{pt} \quad , \tag{23}$$

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.

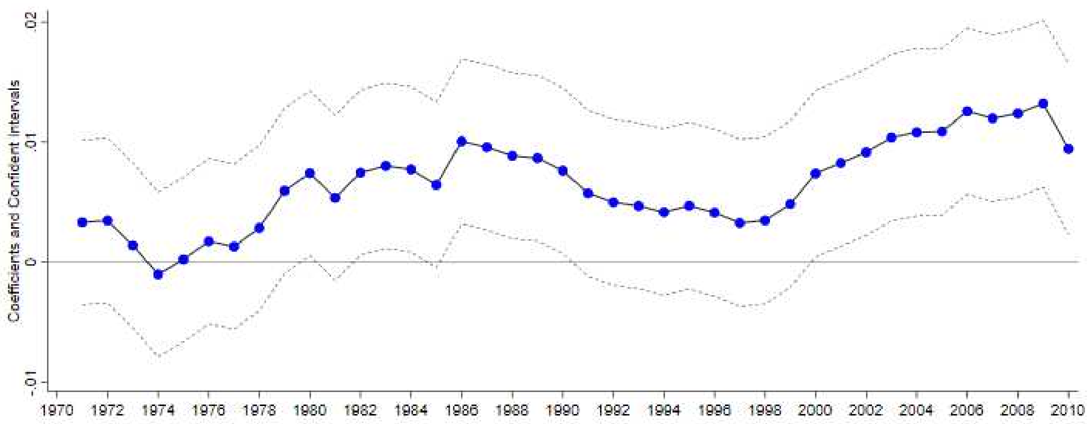

Figure D2. Flexible Estimates of the Relationship between the EFR and the Income Growth Rate

Note: Each dot on the solid line is the point estimate of α_js from equation (23), 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 five time-varying control variables.

Equation (23) 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.

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.

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.

E. Evidence from 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 Li and 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 article does not depend on this intensity measure. Nevertheless, comparable dynamic effects are found when using the MPS as the IV for fertility.

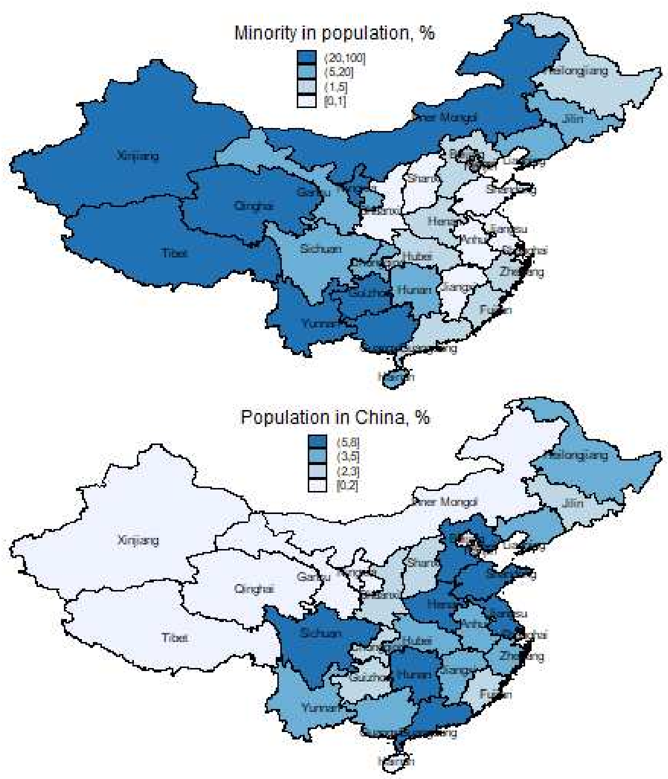

Figure E1. Percentage of minorities in each province (upper panel) and percentage of each province's population in China (lower panel), 1980–2010

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 five 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 five 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.

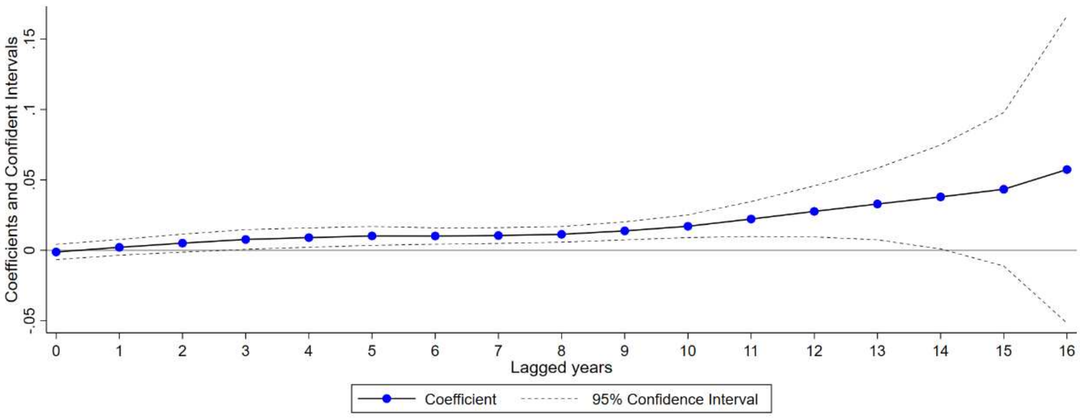

Figure E2. Current and Lagged Effects of Fertility on Income Growth in China (2SLS, Using the minority population share as the IV)

Notes: The figure presents the 2SLS estimates of model (19) 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.

Nevertheless, I still provide the 2SLS estimates of model (19) that uses the MPS as the IV for 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 small, but becomes significantly positive and much larger later on. 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.

It worth to note that Li and 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 uses 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 article, what they estimated is mostly the short-term effect. I replicated their estimation using the data during their sample period and found a similarly negative short-term effect of higher fertility on income growth. Because the MPS is likely endogenous, however, the IV estimates based on it may be biased.