SCHOOL OF ECONOMICS AND MANAGEMENT

Master's Programme in Economics

The Impact of Women's Education on Fertility Rate:

Evidence from China

Author: Duyun Lei

NEKP01

Master's Thesis (15 credits)

June 2023

Supervisor: Roel van Veldhuizen

Abstract

In recent years, the continuous decline in China's fertility rate has elicited a sense of crisis. The fertility trend appears to correlate with the rising educational attainment among women. Societal expectations and workplace dynamics create a challenging environment for women planning to start a family. As more Chinese women pursue higher education and professional careers, they may delay or even forgo childbirth. This thesis explores the effect of women's education on fertility rates in China, drawing upon recent two-year panel data from the China Family Panel Studies. Through the application of Pooled Ordinary Least Squares (OLS), Fixed Effects (FE) regression models, and Instrumental Variable (IV), Fixed Effects (FE) regression models, a discernible negative relationship emerges between a woman's years of schooling and fertility rates. This correlation will mitigate when including control variables age, marital status, working hours and income. These findings underscore the significant role of female education in influencing fertility trends. Despite existing laws, pregnant women and those considering motherhood face discrimination and inadequate support at work, implying the necessity of incorporating these dynamics into future policy decisions. Nevertheless, the study is constrained by its limited timeframe, necessitating further research to delve deeper into long-term trends and unveil the other potential factors influencing fertility rates.

Keywords: Women's education, Fertility rate, China, Fixed Effects Model, Instrument variable estimation

Acknowledgments

I wish to express my deepest gratitude to my supervisor Roel van Veldhuizen, whose guidance, support and expertise have been invaluable throughout this journey. Your patience and encouragement have been a great source of strength for me.

I also wish to acknowledge the faculty and administration of Lund University, who provided an enriching academic environment and resources crucial for completing this thesis.

Last but not least, I am deeply thankful for the unwavering support of my family. Your belief in me and my potential has been my greatest motivation. Thank you for standing by me throughout this journey.

Table of Contents

1. Introduction

As fertility rates decline worldwide, the global demographic shift brings us closer to closing the demographic dividend window. This also signifies that China is on the verge of entering an aging society. Chinese policy has evolved from the famous one-child policy to the current government's encouraging stance on childbirth. Besides the well-known policy changes, socioeconomic factors such as education and urbanization are believed to be essential reasons for the fertility decline (Lavely & Freedman, 1990). White and Parish (1984) found that the decline in urban fertility was not only due to the government's birth control policy but was also related to the wife's education, compact housing, and work participation. Education profoundly impacts women's economic independence, family planning, and conception of childbearing.

Many studies argued that fertility decline is closely related to an increase in woman's education level. Shirahase (2000) found that educational background found to be an essential factor in determining marriage timing. According to Zhang and Zhao (2023), an additional year of schooling reduces the number of children a woman would have by around 0.09 and the probability of having second or more children reduces if the first child was a girl. However, some studies have obtained conflicting findings. Martin (1995) observes that in several of the study's least-developed nations, education favored fertility at the lower levels of the educational range.

The main purpose of this thesis is to investigate the relationship between the increase in women's education and a decline in fertility rates within the context of China—a populous developing country. After accounting for significant control factors such as age, marital status, working hours, and income, we attempt to isolate the unique impact of education on fertility rates and give a deeper comprehension of the relationship by controlling for significant factors known to affect fertility rates and may be associated with educational attainment.

To answer these research questions, we employ panel data for 2018 and 2020, and compared models with and without including control variables. Based on the initial findings obtained from the Ordinary Least Squares (OLS) model, we employed the Fixed Effects (FE) model to further investigate the relationship between women's education and fertility rates. While the OLS model provided valuable insights, the FE model offers distinct advantages in addressing unobserved individual heterogeneity and endogeneity concerns. However, it is important to acknowledge that endogeneity may persist, where education and fertility rates could mutually influence each other. Therefore, we extended our analysis to tackle this issue by employing an Instrumental Variables (IV) model using education expenditure as the instrumental variable.

The result shows that an additional year of schooling lowered the number of children a woman would have by approximately 0.588 when using the IV model with control variables. The negative relationship also shows in the OLS model but with a much lower coefficient (0.0474). However, the result in the FE model shows a positive relationship, which may cause by insufficient within-person variation over the years or the impact of time-invariant individual characteristics. In addition, including the control variable makes education's effect on fertility smaller than the result without control variables.

We also investigate the association between education for women and fertility rates across three age groups: 17-27, 28-38, and 39-49. The result shows that the impact of education on fertility is particularly pronounced among individuals in the youngest age group (17-27). This emphasizes the significance of educational pursuits, career planning, and delayed family formation during this life stage. In contrast to the OLS and FE models that show an age-related increase in fertility rates for older age groups, the IV model reveals a divergent pattern with a negative association between age and fertility. This negative effect with older age may signify the impact of biological constraints and evolving individual and familial priorities on fertility decisions.

The dataset is nationally representative, encompassing a rich array of information on socioeconomic backgrounds, educational achievements, and fertility outcomes. However, the interpretations of our findings should consider the limitations related to the data and model selection. Due to the challenges presented by data limitations and the presence of missing values, it became necessary to forgo the utilization of an extensive longitudinal dataset in this study.

The remainder of this paper is organized as follows. Section 2 gives a basic background of China: female challenges and the decline in fertility rates and literature reviews. The data part is included in section 3, while section 4 presents the methodology for this paper. Section 5 reports the empirical analysis. And section 6 concludes the whole paper.

2. Background

2.1 Female challenge and Declining Fertility rate in China

Despite China's progressive strides toward gender equality, women in the workplace face significant challenges, particularly concerning pregnancy and childbirth. The intersection of the crowded labor market and societal expectations creates a complex environment that often disadvantages working women when they plan to start a family (Song & Dong 2011).

While China has laws protecting pregnant rights and nursing women in the workplace, enforcement is often inconsistent. Discrimination against pregnant employees and woman who planning pregnancy is not uncommon and the enforcement mechanism for Chinese anti-discrimination is far from mature (Hou, 2012). Those challenges range from direct actions such as dismissal or demotion to more subtle forms like exclusion from training and promotional opportunities. The lack of sufficient maternity leave and the paucity of support for childcare services in many areas make the situation more complicate. These factors can lead to women delaying childbirth or even deciding against it, contributing to the declining fertility rate.

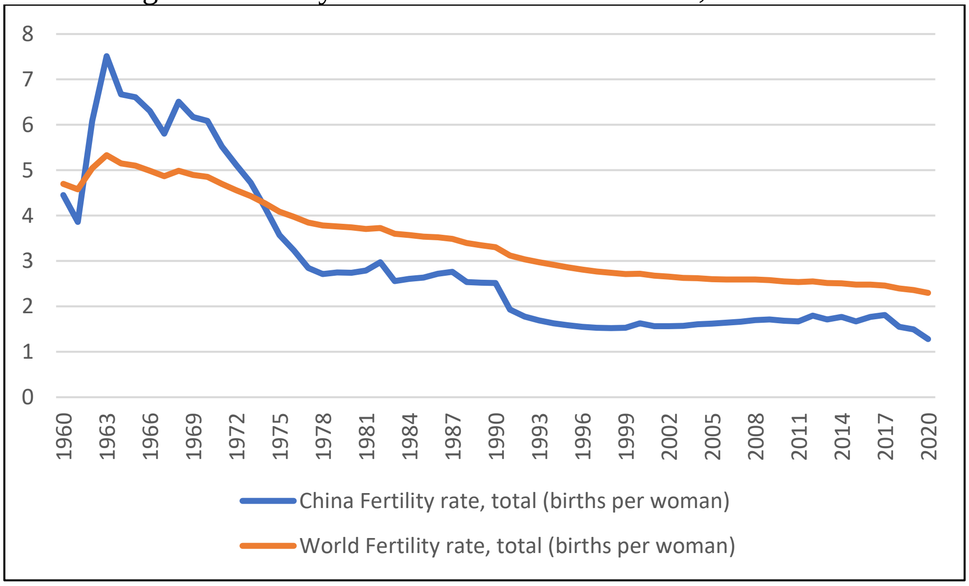
Understanding the relationship between women's education and fertility rates becomes even more important in this situation. As educational attainment rises, women often aspire to professional development and financial independence, which may be inconsistent with the traditional motherhood expectation. This study tries to investigate these complex changes, clarify the mechanisms through which education may influence fertility decisions among Chinese women.

Over half a century, China has seen significant policy changes, shifting from the strict enactment of the one-child policy to the present subtle policy of encouraging birth. Regardless of these changes, the influence on China's population increase may seem modest. Surprisingly, after the one-child policy is fully implemented, many families would prefer to pay a fine to have another child. However, with the beginning of the two-child policy, many people have lost their desire to have children, and China's fertility rate continues to fall year after year.

Before beginning the primary analysis, we can look at the fertility trend in China and the World from a macro perspective (Figure 1). This figure shows that China's fertility rates have changed dramatically over the past half-century. Before 1973, China's fertility rate was relatively high, exceeding the world average. In order to control this rapid population growth, China implemented the "one-child policy" in 1979, which effectively controlled the excessive population growth. Until 2020, China's fertility rate has continually fallen to 1.3, which is still below the global average.

While the world fertility rates are also declining continually but more slowly. According to the World Bank, the global fertility rate fell from around 5.0 in the 1960s to around 2.3 in 2020. Fertility rates vary widely across regions and countries, with fertility rates still relatively high in some developing countries and very low in some developed countries.

Figure 1 Fertility rate in China and the World, 1960-2020



Data Sources: World Development Indicators

Note: The total fertility rate signifies the number of children a woman would have throughout her reproductive years, assuming she experiences the age-specific fertility rates of a given year throughout her childbearing lifespan.

2.2 Previous Research for the Effect of Education on Fertility

Shirahase (2000) explores the relationship between rising higher education levels among women and descending fertility rates in Japan. Her research incorporates two aspects. In the first part, she uses Cox regression analysis to examine marital and reproductive behavior, considering variables such as family background, educational credentials, and first job as indicators of social stratification. The second part explores changes in values resulting from increased education and investigates the relationship between views on the sexual division of labor and men's participation in housework. Women who access higher education tend to postpone marriage and childbirth, decreasing fertility rates. This trend is affected by various variables, including women's increased labor-force involvement, changes in gender roles and expectations, and rising child-rearing expenditures.

Zhang and Zhao (2023) utilize the Compulsory Education Law of China (CELC), implemented in the 1980s, to empirically investigate the causal effect of women's education on fertility in rural China by employing difference-in-differences methods. They concluded that an additional year of schooling reduces the number of children a woman would have by approximately 0.09, delays the age of first childbirth by 0.7 years, and decreases the likelihood of having a second child or more children by 0.18 for mothers whose first child was a girl. In this study, due to the endogeneity of education, there may be a bidirectional causal relationship between education and fertility behavior, which means that directly estimating their relationship may be biased. To address this issue, researchers exploit the CELC as a natural experiment to correct this potential bias, since it was implemented at different times across various provinces, resulting in a quasi-random variation in educational exposure. This variation provides researchers with a valuable tool to more accurately estimate the causal impact of education on fertility behavior, thereby addressing the potential issue of endogeneity bias.

Sohn and Lee (2019) utilize the higher education reform of 1993 as an exogenous variation to examine the causal impact of a college degree on fertility. Their Regression Kink Design (RKD) analysis reveals that having a college degree reduces the likelihood of childbirth by 23 percentage points and the total number of childbirths by 1.3. The study highlights the labor market as a significant mechanism driving this relationship, indicating that women with college degrees are more likely to be wage earners, hold professional occupations, and have lower unemployment rates. The study emphasizes the value of education for society and suggests addressing the opportunity cost of fertility associated with higher education to alleviate declining fertility rates.

Cygan-Rehm and Maeder (2013) used panel data to investigate the causal effect of education on fertility in Germany, focusing on women born between 1937 and 1961. The study exploited a school reform that extended compulsory education by a year, implemented staggered across German states, as a natural experiment and instrumental variable. They found that an additional year of schooling led to a significant reduction in fertility, decreasing the number of children by more than 0.1 and increasing the probability of childlessness by 2-5 percentage points. Increased education was also found to decrease the likelihood of teenage motherhood.

There are some exceptions to this general trend, despite substantial evidence showing a negative relationship between women's education levels and fertility rates. Martin (1995) notes that education was found to favor fertility at the lower end of the educational spectrum in several of the study's least-developed countries and the impact of education on fertility has become less common over time. The mechanism of this research includes examining statistical associations and exploring how education influences reproductive behavior. It emphasizes how different societies' socioeconomic growth, social structures, and cultural contexts have different effects on how education affects fertility. It also discusses policy debates about the most effective approaches to reducing fertility, such as supporting secondary education or boosting elementary education. The analysis follows up on the case for widespread literacy and universal access to primary school. The study also emphasizes the vital link between female education and contraceptive use. Women with Higher Education use contraceptives at higher rates and are more likely to employ effective methods.

2.3 Previous Research for the Models

In our research, the principal issue confronted was identifying the causal influence of education on fertility rates, which is often affected by endogeneity concerns and the potential bias introduced by omitted variables. To address this problem, we employ the Fixed Effect model and Instrumental Variable (IV) estimation, the importance and applicability of which are discussed in our literature review. This methodology is effectively exemplified in the study by Zhang and Zhao, where they utilized a natural experiment as an IV to deduce the effects of education on fertility.

Hedges (1994) discusses fixed effects models used in econometric panel data analysis. These models help minimize bias from unobserved heterogeneity. Time-invariant heterogeneity is managed by subtracting the mean of the observed units. The book compares fixed and random-effects models: fixed effects models focus on changes within units, while random-effects models focus on changes within and outside units. The choice depends on the research topic, data characteristics, and the suspected relationship between unobserved heterogeneity and explanatory variables. The Hausman test can help compare these models statistically.

Brüderl and Ludwig's paper (2015) provides a thorough introduction to fixed-effects models for panel data analysis. It emphasizes controlling for time-invariant unobserved heterogeneity, allowing for more accurate estimates of variables' effects. These models also facilitate the estimation of individual-specific effects. An example is in labor economics, where the fixed-effects model can help estimate the impact of education on wages while controlling for individual factors like ability or motivation, leading to a more nuanced understanding of the education-wage relationship.

The instrumental variables method is a vital technique in econometrics, instrumental in estimating demand and supply curves and tackling issues related to measurement errors in explanatory variables (Angrist & Krueger, 2001). This technique has been used to mitigate the bias towards zero in a bivariate ordinary least squares regression, which occurs when an explanatory variable contains additive random errors (Angrist & Pischke, 2009). Instrumental variables are independent of both the measurement error and the equation error, and they provide a consistent estimate despite the measurement error (Imbens &Angrist, 1994). The instrumental variable method has been employed extensively to address omitted variables problems, enabling more accurate estimation of causal relationships (Bound, Jaeger, and Baker, 1995). As such, understanding the operational mechanisms and institutional determinants of the regressor of interest is crucial in identifying good instruments.

3. Data

3.1 Data Sources

This study mainly utilizes panel data from the China Family Panel Studies (CFPS), with the primary data source being the adult survey questionnaire collected at two-year intervals in 2018 and 2020.

The China Family Panel Studies (CFPS) is a national, large-scale, multidisciplinary social tracking survey project that provides a data foundation for academic research and public policy analysis. Collecting data on individuals, families, and communities reflects changes in China's society, economy, population, education, and health. The CFPS covers various research topics, such as economic activities, educational outcomes, family relationships and dynamics, population migration, and health.

The survey encompasses 25 provinces/municipalities/autonomous regions and targets 16,000 households. It started in 2010 after pilot surveys and follow-ups in 2008 and 2009. All baseline family members and their future blood/adoptive children will be permanently tracked. The CFPS includes four main types of questionnaires: community, family, adult, and child, with additional questionnaire types developed for different family members.

In this study, we also employ education expenditure across various regions in China as the instrumental variable, drawing data from the National Bureau of Statistics of China (NBS). The NBS, a government agency under the direct administration of the State Council of the People's Republic of China, is tasked with collecting, compiling, analyzing, and disseminating a comprehensive range of statistical data on the country. Since its establishment in 1952, the NBS has provided vital information to facilitate China's economic planning, development, and decision-making processes.

3.2 Variable Definitions and Descriptions

In our study, education and fertility are the main variables, quantified by years of schooling and the number of children, respectively. The control variables used to capture potential confounding effects include age, income, working hours, and marital status. Age is specifically considered within the conventional fertility-relevant range of 17 to 49 years. We used annual income figures and weekly working hours to measure economic status and time commitment. Furthermore, we denote marital status as a dummy variable, assigning 1 for married individuals and 0 for those not married.

3.21 Main Variables:

Education attainment is measured by years of schooling. The average years of education being 11.8 indicates that on average (Table 1), the individuals in the dataset have completed approximately 12 years of formal education. This indicates the completion of secondary education. The dataset only includes the highest completed education level, it may omit some unfinished years of actual education and cause potential bias.

Fertility is operationalized as the number of children a woman has given birth to. In the dataset utilized for this study, the average number of children per woman is calculated to be 2.28.

3.22 Control Variables:

Age is a clear control variable because it's universally accepted that fertility decreases as women age, irrespective of education level. However, it is also true that women with higher education levels may choose to delay childbearing due to the pursuit of advanced degrees or careers. Therefore, age can be both an outcome of educational choices and a determinant of fertility. The study from Shirahase (2000) confirms this, which reveals that while educational attainment significantly influences the timing of marital commitments, it does not substantially impact the decision to procreate. Instead, age becomes the main reason influencing fertility decisions. This revelation underscores the rationale behind including age as a control variable in our study.

The rationale for incorporating income and working hours as control variables in our analysis stems from multiple considerations. Primarily, individual income has been shown to influence fertility decisions considerably, with resource constraints as a key determinant. In parallel, the potential income loss women encounter due to childbearing and child-rearing, referred to as substitution costs, is another crucial factor. Furthermore, the delicate equilibrium women need to maintain between professional commitments and familial responsibilities, alongside the influence of the family's financial security perception on fertility intentions, are instrumental in shaping fertility decisions. As such, these variables are integral to a comprehensive analysis of the impact of education on fertility rates.

Similarly, marital status can be both a determinant and outcome of fertility and education. In many societies, marriage is considered a prerequisite for childbearing, and China is no exception. Individuals who are unmarried or divorced may face more significant social pressure to influence their reproductive decisions (social norm). Married couples are generally more likely to plan and raise children. Marital stability may make couples more willing to bear the costs and responsibilities of childbearing. Thus, marital status is a crucial control variable in our relationship analysis between education and fertility rates.

3.23 Instrumental Variable:

Educational expenditure in each province may affect an individual's educational level because higher educational expenditure usually means better educational quality and resources, thus increasing the individual's educational level. However, there may not be a direct causal relationship between education spending itself and fertility. Therefore, by using education expenditure as an instrumental variable, the causal effect of education on fertility can be estimated more accurately, excluding other possible confounding factors.

Table 1

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Number of children	3772	2.281	1.624	0	6
Education (years)	3772	11.822	4.037	0	21
Age	3772	34.328	7.836	17	49
Income	3772	41123	35905	0	500000
Marital status (1 = married)	3772	.837	.369	0	1
Working hours (One weeks)	3772	48.7	16.488	1	144
Education expenditure (10k RMB)	3772	20123998	12219799	2346978	53869558

3.3 Limitation of the Data

Despite the robustness of the data sources employed in this study, which included the China Family Panel Studies (CFPS) and the National Bureau of Statistics of China (NBS), there are several limitations to be aware of.

When merging panel data with matched survey respondents, we observed varying levels of changes in the sample of survey respondents across year in the database. Additionally, the variables of interest also exhibited varying degrees of missingness. Using panel data from multiple years could result in insufficient data due to these limitations. We decided to restrict data within two years to achieve enough of a sample for analysis. Angrist and Krueger (2001) mentioned that researchers using instrumental variables should aim to work with large samples.

As the study does not capture more temporal dynamics of the association between education and reproduction rates, this might lead to potential bias. Furthermore, because data filtering and selection were required, some necessary variables may have been removed, which may have contributed to an inadequate depiction of the factors impacting fertility rates.

Since the dataset only includes the highest completed level of education, it may not account for an individual's whole educational experience. For instance, a person who attended university but did not finish their degree would be recorded as having completed only secondary education, despite already having time spent at the university. This could potentially lead to an underestimation of the actual years of education and cause bias in our results.

While the data utilized in this study is comprehensive and reliable, it is susceptible to the constraints inherent in secondary data analysis, such as potential variations in collection methods and reporting standards across different years and areas. Our data is limited to specific demographics, which may not reflect the entire population. This could lead to selection bias and restrict the generalizability of our findings. Moreover, the self-reported data we rely on will cause measurement errors due to social desirability bias. It may be influenced by respondents' desire to present themselves in a favorable light or conform to societal norms.

4. Methodology

This study using panel data model to investigate the relationship between woman's education and fertility rate. In designing our methodology, we have controlled for multiple factors, such as age, income, marital status, and working hours, to minimize their potential influence on fertility The Ordinary Least Squares (OLS) model used as our baseline model. Based on this, we adopted the Fixed Effects (FE) model. FE models are preferable to OLS model in panel data analysis as they control for unobserved, time-invariant individual characteristics, thereby reducing bias and providing more accurate estimates, especially when these unobserved variables are correlated with the explanatory variables. Furthermore, to address the endogeneity issue and enhance the accuracy of our estimation, we also utilized the Instrumental Variables (IV) model.

The pooled OLS model is a linear regression model used to predict a dependent variable (fertility rate, in your case) based on a set of independent variables (education, age, income, working hours, and marital status, in your case). The model can be presented as follows:

$$Fert_i = \alpha + \beta_1 X_{1,i} + \cdots + \beta_k X_{k,i} + \varepsilon_i \tag{1}$$

Here, $Fert_i$ represents the fertility rate of individual i. $X_{k,i}$ is control variables including the education, age, income, working hours, and marital status of individual i at time t respectively, and ε_i represents the error term.

The Fixed Effects model is a type of linear regression model used for panel data that controls for all time-invariant, unobserved individual characteristics. In this model, all explanatory variables are time-varying. The FE regression model is presented as follows:

$$Fert_{it} = \alpha_i + \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it} + \varepsilon_{it} \tag{2}$$

Here, α_i is an individual-specific intercept term, capturing all time-invariant individual characteristics.

The Instrumental Variables (IV) approach is a method used in econometrics to address issues of endogeneity, where explanatory variables are correlated with the error term, which violates one of the key assumptions of pooled OLS regression. This correlation often arises due to omitted variables, measurement errors, or simultaneous causality. The IV approach aims to provide consistent and unbiased estimates in the presence of such endogeneity.

The IV method consists of two stages. In the first stage, the endogenous variable (the variable that is correlated with the error term) is regressed on the instrumental variable. The instrumental variable is a variable that is correlated with the endogenous variable but uncorrelated with the error term.

$$Edu_{it} = \alpha_0 + \alpha_1 Z_{it} + \gamma^1 X_{1,it} + \cdots + \gamma^k X_{k,it} + \eta_i + \varepsilon_{it} \quad (3)$$

Here, Z_{it} represents as instrument variable (education expenditure) of individual I for time t. η_i represents individual fixed effects.

In the second stage, the dependent variable is regressed on the predicted values of the endogenous variable obtained from the first stage. The predicted values of the endogenous variable are uncorrelated with the error term, thus addressing the endogeneity problem.

$$Fert_{it} = \beta_0 + \beta_1 \widehat{Edu_{it}} + \delta^1 X_{1,it} + \cdots + \delta^k X_{k,it} + \eta_i + u_{it} \quad (4)$$

Here, Edu_i is the endogenous explanatory variable, education. Z_{it} represents as instrument variable (education expenditure). $\widehat{Edu_{it}}$ is the predicted value of Edu_{it} from the first stage regression. ε_{it} and u_{it} are error terms. γ and δ are vectors of parameters associated with the control variables in the first and second stage regressions respectively.

5. Empirical Analysis

In the process of our research, it was crucial to ensure that the instrumental variable utilized was indeed robust and valid. To substantiate this, we implemented a strong instrument test. The results (Appendix table 1) from the test corroborated the strength of our chosen instrumental variable, substantiating its relevance and suitability for our IV estimation model. This confirmation reinforces the credibility of our analysis, as a strong instrumental variable is fundamental to the validity of our IV estimation results. The analysis of our results is structured in a systematic and comprehensive manner, divided into two distinct stages.

In the first part, we have the regression results in Table 2. This table presents the results of three models, Pooled Ordinary Least Squares (OLS), Fixed Effects (FE), and Instrumental Variables (IV), each estimated with and without additional control variables. The dependent variable is the number of children, and the key independent variable is years of schooling.

The Pooled OLS model (columns 1 and 2) suggests a negative and significant relationship between years of schooling and the number of children. This implies that an increase in education leads to a decrease in the number of children, consistent with existing literature suggesting a relationship between education and fertility (Cygan-Rehm & Maeder, 2013; Sohn & Lee, 2019; Zhang & Zhao, 2023). This impact becomes more significant after excluding control variables. When control variables are removed in columns 2, the coefficient for years of schooling increases from -0.0474 to -0.0905. This increase could suggest that the control variables were capturing some variations previously attributed to years of schooling. Moreover, the overall R-squared decreases from 0.155 to 0.0506, indicating that the model without control variables has less explanatory power.

However, the FE model results (columns 3 and 4) differ. The impact of education on fertility turns statistically insignificant both with and without control variables. This suggests that when accounting for individual time-invariant characteristics (the fixed effects), years of schooling do not have an obvious impact on the number of children. However, age has a strong positive relationship with the number of children in these models. This lack of significance for years of schooling in the FE models could suggest a substantial influence from time-invariant individual characteristics or a lack of sufficient within-person variation between the two time periods (2018 and 2020).

In the Instrumental Variables (IV) models (columns 5 and 6), education expenditure is used as an instrumental variable to correct potential educational endogeneity. The result shows an additional year of schooling significantly reduces the number of children by about 0.588 with control variables. The effect appears even more significant when control variables are not included (columns 6), which implies that the control variables were suppressing some of the impact of years of schooling on the number of children.

Regarding the control variables, age positively correlates with the number of children in the OLS and FE models but shows a negative relationship in the IV model. Age positively correlates with fertility in the OLS and FE models because women may have more opportunities to have more children as time passes. However, the negative coefficient in the IV model may indicate that the relationship between age and fertility can be different after addressing potential endogeneity issues. Here, the negative coefficient might suggest that as women age, the rate of having additional children could decrease. This could be due to biological limitations, changing social norms, and career or educational commitments that might lead women to delay childbearing. In the OLS model, income has a positive but weak effect on fertility, suggesting that financial stability might encourage childbearing, albeit modestly. However, this relationship is not apparent in the FE model but re-emerges strongly in the IV model. Marital status consistently correlates positively with fertility across all models, reiterating the traditional expectation that marriage precedes children. The variable 'working hours' shows varied effects, indicating a complex interplay between work-life balance, financial considerations, and childbearing decisions.

Table 2

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Number of children					
	Pooled OLS		FE		IV	
Year of schooling	-0.0474***	-0.0905***	0.0285	0.245	-0.588***	-0.835***
	(-6.55)	(-14.18)	(0.39)	(-1.72)	(-5.11)	(-3.42)
Age	0.0414*** (10.72)		1.165*** (70.29)		-0.046** (-2.36)	
Income	0. 176 * (2.4)		-0. 176 (-1.58)		2.24 *** (4.96)	
Marital status	0.825***		0.332*		0.659***	
	(10.58)		(2.08)		(5.17)	
Working hours	0.00289		-0.00165		-0.028***	
	(1.88)		(-0.93)		(-4.01)	
Constant	0.516** (2.69)		-38.18*** (-38.13)		10.70*** (4.91)	
Control Variables	Yes	No	Yes	No	Yes	No
R-squared	0.155	0.0506	0.7431	0.0016	-	-
N	3772		3772		3772	

Data Source: CFPS
Notes: T statistics in parentheses
* p < 0.1; ** p < 0.05; *** p < 0.01.
Notes: The dependent variable in the model is 'Number of Children,' which represents the total number of children that a woman has. It serves as a proxy for fertility rates. 'Year of Schooling' measures the total years of formal education a woman has completed. It serves as a proxy for educational attainment. 'Age' measures the age of the woman. The typical reproductive age span of 17 to 49 years is considered in the study. 'Income' is the annual income of the individual, rescaled by a factor of 100,000 for easier interpretation. Hence, the coefficients for income should be interpreted as the change in the number of children per 100,000 unit change in income. 'Marital Status' is a dummy variable, with 1 indicating a married woman and 0 indicating an unmarried woman. 'Working hours' represents the number of hours a woman works in a week.

Based on our results, the Ordinary Least Squares (OLS) method tends to underestimate the impact of education on fertility rates, which is consistent with previous studies using compulsory schooling to estimate the relationship between education and the fertility rate. Leon (2004) explains downward bias in OLS by individual heterogeneity and non-linear fertility responses to education. Zhang and Zhao (2023) argue that unobserved factors encouraging schooling (e.g., ability, credit access) may positively impact fertility rates, conditional on the dominance of the income effect over the substitution effect. In the context of son preference in China, educated women were more likely to be part of affluent families that valued larger family sizes. As such, the fertility-reducing effect commonly associated with increased education was often offset by these families' preference for having more children.

Since our instrument variable is education expenditure, the rationale behind this could be that higher education expenditure, which might include tuition fees, textbooks, and other educational resources, often leads to a higher quality or level of education. This is because it can provide better resources, more qualified teachers, and more comprehensive curricula. Then, with a higher level of education, individuals might make different life choices that can affect fertility rates. For instance, they might delay starting a family and focusing on their careers. Alternatively, they might have fewer children because they are more aware of the cost and time commitments of raising children.

Regarding the different results produced by the FE model and IV model, one possibility is that the FE model can control for time-invariant unobserved variables, it cannot account for unobserved variables that change over time. If these variables are correlated with the explanatory variables, they could still bias the estimates. Another explanation is that the FE model cannot deal with endogeneity arising from reverse causality or simultaneous causality, where not only does education affect the fertility rate, but fertility can also impact the level of education attained. This is the problem we try an Instrumental Variable (IV) approach to solve it. Lastly, FE models also cannot handle the omitted variable bias that arises from omitting relevant variables that vary over time. The estimates can still be biased if such variables are omitted from the model.

The second part represents regression results from three age groups (17-27, 28-38, 39-49), which include in the appendix table 2.3.4. In the 17-27 age group, both the OLS and IV models reveal that education has the most significant impact on fertility rates. This can be attributed to the concurrent pursuit of higher education or vocational training, forward-thinking attitudes toward planning for the future, and significant life transitions experienced during this stage of life. Individuals in this age range prioritize educational attainment and personal development, often resulting in delayed marriage and childbearing as they focus on building a solid foundation for their careers and financial stability. The influence of education on fertility rates is particularly pronounced in this age group due to the interplay of these factors, highlighting the importance of educational opportunities in shaping reproductive decisions.

All three models (OLS, FE, IV) show a significant positive effect of age on fertility rates in the 17-27 age group. However, in the other two age groups, the IV model consistently shows a negative effect of age on fertility rates (including in Table 2). In this case, the IV model seems more reasonable. Since individuals progress through a younger age range, their likelihood of having children increases. This finding aligns with the notion that younger adults are more prone to initiating and expanding their families. However, the relationship between age and fertility is negative beyond this age group. This implies that as individuals advance in age, their fertility diminishes. This phenomenon may be attributed to physiological changes associated with aging, shifting personal priorities, or deliberate decisions regarding family planning.

The results show that the R-squared values rise when control variables are included. This suggests that these control variables have helped improve the model's explanatory power over the data. Specifically, this could be because the control variables contain important information related to the dependent variable (fertility rates), so when they are included in the model, the model can better capture the overall trends in the data.

6. Conclusion

This study examined the impact of women's education on fertility rates in China using data from 2018 and 2020. Three distinct econometric approaches were used: Pooled Ordinary Least Squares (OLS), Fixed Effects (FE), and Instrumental Variables (IV), each of which provided distinct insights into this intricate connection. The models were adjusted for relevant control variables including age, income, working hours, and marital status.

When analyzing the results for the entire age group, we found that the OLS and IV models consistently demonstrated a significant negative relationship between women's education and fertility rates. The inclusion of control variables mitigated this relationship, leading to a lower negative correlation between years of schooling and fertility. These results align with our theoretical expectations and existing literature, signifying that women tend to have fewer children as they acquire more education. This trend may be attributed to various reasons, such as higher opportunity costs associated with childbearing and rearing, increased career aspirations, or changes in personal values and preferences. However, the Fixed Effects model presents a positive association. This discrepancy might be attributed to insufficient within-person variation over the years studied (2018 and 2020) or the overwhelming influence of time-invariant individual characteristics. In IV model, which employs education expenditure as the instrumental variable to address the potential endogeneity of education, reinstates the significant role of education in determining fertility rates. With IV estimation producing significantly larger results compared to OLS, it indicates the presence of endogeneity issues in the OLS estimation. IV method, by using instrumental variables, corrects for endogeneity bias and offers more reliable estimates.

When we delve deeper into the age-specific effects of education on fertility, the results suggest that the impact of education is most significant in the youngest age group (17-27), reflecting the life-stage-specific nature of fertility decisions. The prominence of education in this age group may be driven by the synergies between pursuing higher education, career planning, and delayed family formation. For the older age groups, while an age-induced increase in fertility is observed in the OLS and FE models, the IV model presents a contrasting picture with a negative effect of age. With increasing age, this negative effect may indicate biological limitations and shifts in personal and family priorities. These age-related trends unearth the dynamic nature of fertility decisions, showing that as individuals progress through different life stages, the determinants and their effects may shift accordingly.

This study is also with some limitations. While our dataset is robust, it only includes data from 2018 and 2020, restricting our ability to capture long-term trends or cyclical changes. Additionally, the data does not capture all potential confounding variables, such as cultural factors or individual preferences, which could influence fertility decisions. Furthermore, while the IV model has strengths in addressing endogeneity concerns, it assumes a perfect correlation between the instrument and the endogenous variable, which might not hold in every scenario.

Given these results, policymakers should consider broader societal and economic implications. As women's educational attainment rises, fertility rates might decline further, creating difficulties in China's demographic balance and social welfare programs. As a result, it is critical for policymakers to consider these potential effects when devising strategies related to education and family planning. To counter the declining fertility trend, policies could focus on promoting a balance between family-centred lifestyle and personal development, which might encourage a shift towards having more children.

Future research could build upon this work by incorporating data from a broader time frame to capture longer-term trends. It would also be beneficial to explore other potential factors that could influence fertility decisions, such as cultural norms or societal values. Furthermore, refining the instrumental variable or exploring other econometric models could provide additional insights into the causal relationship between women's education and fertility rates.

7. Reference

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics, 106(4), 979-1014.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. Journal of Economic perspectives, 15(4), 69-85.

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Basu, A. M. (2002). Why does education lead to lower fertility? A critical review of some of the possibilities. World Development, 30(10), 1779-1790.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association, 90(430), 443-450.

Brüderl, J., & Ludwig, V. (2015). Fixed-effects panel regression. The Sage handbook of regression analysis and causal inference, 327, 357.

Croll, E. (2000). Endangered Daughters: Discrimination and Development in Asia. Routledge.

Cygan-Rehm, K., & Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. Labour Economics, 25, 35-48.

Dasgupta, P. S. (1995). Population, poverty, and the local environment. Scientific American, 272(2), 40-45.

Hedges, L. V. (1994). Fixed effects models. The handbook of research synthesis, 285, 299.

Hou, Y. (2012). Means of Transformation? The Role of Enforcement Mechanisms in Providing Protection against Pregnancy Discrimination in Employment (Master's thesis). Retrieved from https://www.duo.uio.no/handle/10852/35074

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. Econometrica, 62(2), 467-475.

Institute of Social Science Survey, Peking University. (2023). China Family Panel Studies (CFPS). Retrieved April 27, 2023, from http://www.isss.pku.edu.cn/cfps/index.htm.

Jia, N., & Dong, X.Y. (2013). Economic Transition and the Motherhood Wage Penalty in Urban China: Investigation using Panel Data. Cambridge Journal of Economics, 37(4), 819-843.

Kravdal, Ø. (2001). The high fertility of college educated women in Norway: An artefact of the separate modelling of each parity transition. Demographic research, 5, 187-216.

Lavely, W., & Freedman, R. (1990). The origins of the Chinese fertility decline. Demography, 27(3), 357-367.

Leon, A. (2004). The effect of education on fertility: Evidence from compulsory schooling laws. unpublished paper. University of Pittsburgh.

Martin, T. C. (1995). Women's education and fertility: results from 26 Demographic and Health Surveys. Studies in family planning, 187-202.

Martin, T. C., & Juarez, F. (1995). The impact of women's education on fertility in Latin America: Searching for explanations. International family planning perspectives, 52-80.

Maurer-Fazio, M., Connelly, R., Chen, L., & Tang, L. (2011). Childcare, eldercare, and labor force participation of married women in urban China, 1982–2000. Journal of human resources, 46(2), 261-294.

National Bureau of Statistics of China. (2023). National Bureau of Statistics of China. Retrieved April 27, 2023, from https://data.stats.gov.cn/index.htm.

Shirahase, S. (2000). Women's increased higher education and the declining fertility rate in Japan. Review of population and social policy, 9(2000), 47-63.

Sohn, H., & Lee, S. W. (2019). Causal impact of having a college degree on women’s fertility: Evidence from regression kink designs. Demography, 56(3), 969-990.

Song, L., & Dong, X.Y. (2011). Gender and Occupational Mobility in Urban China during the Economic Transition. Economics of Transition, 19(1), 191-219.

White, T., & Parish, W. L. (1984). The link between urbanization and fertility: An analysis of Chinese data. Sociological Forum, 1-24.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

World Bank. (2023). Fertility rate, total (births per woman). Retrieved from https://data.worldbank.org/indicator/SP.DYN.TFRT.IN.

Zhang, Z., & Zhao, Z. (2023). Women Education and Fertility in China. SSRN. https://doi.org/10.2139/ssrn.4136221.

8. Appendix

Table 1 Weak identification test

Weak identification test (Cragg-Donald Wald F statistic):		39.332
Stock-Yogo weak ID test critical values:	10% maximal IV size	16.38
	15% maximal IV size	8.96
	20% maximal IV size	6.66
	25% maximal IV size	5.53

Table 2 Age group of females between 17 and 27

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Number of children					
	Pooled OLS		FE		IV	
Year of schooling	-0.135***	-0.0955***	0.103	0.473*	-0.715**	-0.895*
	(-6.55)	(-5.25)	(1.00)	(2.16)	(-2.92)	(-2.16)
Age	0.187***		1.235***		0.458***	
	(7.19)		(29.76)		(3.84)	
Income	0.368*		-0.248		1.23**	
	(1.97)		(-0.94)		(2.76)	
Marital status	0.311*		0.466*		-0.841	
	(2.58)		(2.06)		(-1.64)	
Working hours	0.00311		0.00237		-0.0258*	
	(0.91)		(0.61)		(-1.98)	
Constant	-1.837**	2.576***	-30.07***	-4.869	0.748	13.04*
	(-3.03)	(10.56)	(-19.38)	(-1.70)	(0.54)	(2.40)
Control Variables	Yes	No	Yes	No	Yes	No
R-squared	0.1542	0.0326	0.7878	0.0138	.	.
N	820	820	820	820	820	820

Data Source: CFPS

Notes: T statistics in parentheses

* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 3 Age group of females between 28 and 38

	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
	Pooled OLS		FE		IV	
Year of schooling	-0.0621***	-0.0797***	-0.0892	0.0459	-0.592***	-0.683**
	(-5.49)	(-8.03)	(-0.65)	(0.18)	(-3.66)	(-2.98)
Age	0.0622***		1.084***		-0.0196	
	(5.21)		(40.36)		(-0.64)	
Income	0.0389		-0.0986		1.70**	
	(0.40)		(-0.62)		(3.24)	
Marital status	0.589***		0.0874		-0.102	
	(3.93)		(0.25)		(-0.33)	
Working hours	0.00209		-0.00633*		-0.0296**	
	(0.88)		(-2.05)		(-2.89)	
Constant	0.484	3.397***	-31.41***	1.805	11.25***	11.05***
	(1.05)	(25.94)	(-16.19)	(0.57)	(3.37)	(3.80)
Control Variables	Yes	No	Yes	No	Yes	No
R-squared	0.0641	0.0357	0.7037	0	.	.
N	1740	1740	1740	1740	1740	1740

Data Source: CFPS

Notes: T statistics in parentheses

* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 4 Age group of females between 39 and 49

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled OLS		FE		IV	
Year of schooling	-0.0320**	-0.0417***	0.182	0.305	-0.638*	-0.782
	(-2.84)	(-4.18)	(1.05)	(0.86)	(-2.50)	(-1.81)
Age	0.0741***		1.198***		-0.0945	
	(4.68)		(41.95)		(-1.23)	
Income	0.161		-0.0859		3.14*	
	(1.11)		(-0.36)		(2.45)	
Marital status	0.408		-1.246		-0.210	
	(0.75)		(-1.88)		(-0.20)	
Working hours	0.00343		-0.000621		-0.0231	
	(1.38)		(-0.21)		(-1.92)	
Constant	-0.798	3.182***	-50.23***	-0.192	13.37*	10.39*
	(-0.88)	(29.86)	(-23.02)	(-0.06)	(2.16)	(2.46)
Control Variables	Yes	No	Yes	No	Yes	No
R-squared	0.0344	0.0142	0.7667	0.0013	.	.
N	1212	1212	1212	1212	1212	1212

Data Source: CFPS

Notes: T statistics in parentheses
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.