Journal of the Knowledge Economy (2024) 15:4056–4071
https://doi.org/10.1007/s13132-023-01333-6

The Casual Effect of Family Size on the Human Capital in Iranian Society

Nooshin Rashti Mohammad[1] · Ali Taieb Nia[1] · Sajjad Barkhordari[1]

Received: 15 March 2022 / Accepted: 25 February 2023 / Published online: 25 March 2023

Abstract

This study examines the causal effect of family size on the educational level of children in Iranian society using the urban and rural household income and expenditure survey-2019 and 2SLS method. The instrumental variable of the gender related to the first child born in the family was used to address the endogeneity of the family size variable, due to the clear preference of Iranian families for having boys and extension of the family until having a boy. The results indicated the negative and significant effect of family size on the educational level of children regardless of the probability of endogeneity of this variable, while family size had a positive and significant effect on the education of children after considering the probability of endogeneity and using the instrumental variable. In addition, results showed the birth order of children had a negative and significant effect on their educational level. Therefore, younger children were less likely to graduate compared to older ones.

Keywords Family size · Education · Children · Endogeneity · Instrumental variable

JEL J13 · C31 · C26

✉ Sajjad Barkhordari
barkhordari@ut.ac.ir

Nooshin Rashti Mohammad
nooshin.rashti@ut.ac.ir

Ali Taieb Nia
taiebnia@ut.ac.ir

[1] Faculty of economics, University of Tehran, North Kargar Street, Jalal-e-Al Ahmad Street, P. O. Box 141556445, Tehran, Iran

Introduction

Education plays an important role in human communities regarding its impact on economic variables and on the well-being of individuals. Regarding the considerable impact of education on the economic and social structure of communities, many social scientists, such as economists, psychologists, and demographers, have considered the determinants of education. Based on the results of studies, the family background and characteristics of individuals have been widely recognized as essential factors that affect education (Feng, 2020). Accordingly, family size is one of the main factors that affect the education of children in a family along with household income level and parents' education level (Duncan et al. 2017).

Many studies have considered this case due to the quantity and quality of labor force in the factor market, population growth, and human capital are the main factors of economic growth and provide the field for financing the public pension systems (Huang, 2021). Hence, economists and politicians of society should find why, how, and to what extent the number of children in the family affects their education level. Before the beginning of the twenty-first century, empirical studies concentrated on the simple relationship between the number of children and education level. The potential endogeneity of family size is the reason for any specific conflict in results in the studies. Therefore, empirical literature was created after introducing control variables for family characteristics and instrumental variable approaches over the recent 20 years (Becker & Tomes, 1976; Blake, 1981; Angrist et al. 2010; Conley & Glauber, 2006; Kugler & Kumar, 2017; Lee, 2008).

Iranian society is a young one, and education is the main factor that can improve economic development. Therefore, an analysis of the family size on the education level is an important subject for policymakers. No studies in Iran, to the best of our knowledge, have not considered this subject, and the results of this study can help policymakers to implicate policies for improving the quality of education. The contribution of this paper is related to applying an instrumental approach for investigating the effect of family size on the quality of education in the case of Iranian society. Hence, three questions were designed: do children with more siblings have lower education levels (is there a negative correlation between family size and education of children)? Is there a different effect of family size on the rural groups comparing urban groups? Is there a similar effect of family size on the uneducated or low-educated mothers comparing the educated ones?

This study is organized into six sections. After the introduction section, "Literature Review" section reviews the theoretical foundations. "Data" section introduces and describes the applied data. The research model is specified, and variables are defined in the "Model and Variables" section. "Emperical Findings" section explains the empirical findings. "Discussion" section presents a discussion and summary of the study and policy implications, respectively.

Literature Review

A large number of empirical studies have assessed the association between the quantity and quality of children to expand the related literature such as Argys and Averett (2015), Black, Devereux, and Salvanes (2010), Conley and Glauber (2006), Dang and Rogers (2016), Feng (2020), Kugler and Kumar (2017), and Blanden et al. (2022). Different results have been obtained from empirical studies. The related studies can be classified into three categories.

The first group focused on the size and direction of the relationship between quantity and quality of children, and any consensus is not obtained among studies. Although many empirical studies have indicated the negative and significant impact of the number of children on their quality (Black et al., 2010; Downey, 1995; Feng, 2020; Mosli, Kaciroti, Corwyn, Bradley, & Lumeng, 2016), the other ones have reported no significant impact (Angrist & Evans, 1998; Black, Devereux, & Salvanes, 2005). In addition, some other studies have shown the positive and significant impact of family size on the quality of children (Adongo et al., 2022; Qian, 2009; Zhang, 2017). This group of studies argues that an increase in the number of children in a family can improve the quality of experience among children by creating an opportunity to learn from each other. In addition, these studies introduce an economy of scale as another source for the positive relationship between the quantity and quality of children. Based on this argument, the higher the number of children in the family induces lower the marginal cost of care for children.

The second scope of related studies considered the moderating role of contextual factors in individuals, families, governance, and community. Therefore, a heterogeneous relationship between family size and the quality of children can observe based on the time study and contextual factors. A large of studies focused on the heterogeneous effect of family size on the different groups of society. These studies classified the sample based on sex, urban/rural residence, born groups, education and income levels of parents, etc., and then compared the impact and intensity rates of family size on different classes. A large number of studies proved the heterogeneous effect of family size on different groups of communities without proposing similar results owing to the difference in contextual factors existing in societies. Some studies have shown the considerable negative effect of family size on the education level of children among girls, poor people, individuals living in rural residences, and children with less-educated parents (Barra & Boccia, 2022; Kamanda et al., 2016; Kugler & Kumar, 2017; Wang, 2019; Li, Zhang, & Zhu, 2008). In addition, some studies rejected the significant heterogeneity among different social groups, such as Black et al. (2005).

The third category of studies was conducted to find some solutions to address the endogenous variable of the model family size. The QQ trade-off model cannot be tested for society simply, which the problem of the estimation of the causal effect of family size on the quality among children is the endogeneity in the number of children among families. Since this variable is influenced by household income, child expenditure, contraception knowledge, lack of confidence, and relative preferences or tastes of parents, thus ordinary least squares (OLS) estimators of regression may

be biased. Among empirical papers, some of the studies estimated the causal effect of family size on the education level of children through addressing the endogeneity of this variable.

In general, studies have chosen an instrumental variable approach to overcome the endogeneity problem. Three main strategies can consider in this approach. Using the twin births strategy as an instrumental variable is the first common method in empirical studies to address the endogeneity of family size. For example, Angrist, Lavy, & Schlosser (2010); Black et al. (2005); Glick, Marini, & Sahn (2007); Lee (2008); Rosenzweig & Zhang (2009) are some of the outstanding studies that applied this strategy. However, the use of this strategy may be problematic due to some reasons; firstly, twins are born at the same time, and the difference in birth time between them equals zero, thus parents allocate few sources to twin births compared to the singleton. This case can cause a biased estimation of the QQ trade-off model. Secondly, the weight of twins is usually lower than the weight of singletons, which can directly affect the quality of individuals (education level) during adulthood. Thirdly, twin births can only be used in case of unplanned pregnancy and lack of medical interventions in determining the number of children. However, twin birth cannot be considered as an exogenous phenomenon that is out of the control of parents due to the developed assisted reproductive technology. Hence, the results may be doubtful if this instrumental variable is used. However, using this strategy is not possible as an instrumental variable in the present study due to only the birth year (not month and day) of individuals is accessible in the survey database of household income and expenditure published by the statistical center of Iran. Hence, the children born in one specific year are not necessarily twins or multiple twins.

The second strategy used in the related literature is based on public policies as an exogenous phenomenon that is out of control among families. This case has been widely considered recently, such as the one-child (OCP) and two-child policies (TCP) in China can be named as the most popular public policies on the family planning program in the world, which banned families from having more than one or two children for more than 30 years. Therefore, this strategy is used in East Asian countries (i.e., China, Hong Kong, and South Korea) that have experienced population-widening policies. Kwan (2009); Li et al. (2008); Liu (2014) employed these policies as endogenous phenomena in their studies.

As the third solution, the gender of the first children in the family or gender composition among children was used as an instrumental variable to address the endogeneity of family size. Families' preferences for having children with a specific gender (boy or girl) or having both genders of children encourage them to have more children. Angrist et al. (2010), Conley and Glauber (2006), Kugler and Kumar (2017), and Lee (2008) employed this method for the data from South Korea, USA, Israel, and India, respectively. Similar to the disadvantage of using twin births as an instrumental variable, this method is faced with problems due to the spread of medical science and the possibility of using sex-determining methods as well as aborting by parents. These phenomena cast doubt on the sex of the children as a potential exogenous variable. It is worth mentioning, as a general rule, none of these strategies addressed for solving the endogeneity of family size is preferable to the other. However, an appropriate strategy can be selected based on the studied period,

public policies, institutional arrangements, and data-collecting methods in a specific society.

Data

The census process is annually considered in Iran by the statistical center of Iran, in which some households are selected randomly, and their information is recorded. The initial data are open to the public as databases through Access software. In addition, adjusting and organizing the data should be considered according to the need of studies and to the method of data collected (classifying data based on the household ID and relativeness code of household members with the household head), which the present study prepared and integrated data using STATA software.

In the present study, households' income-expenditure survey microdata-2019 were employed, which the total sample included social-economic data of 132,541 for rural and urban individuals. In addition, an age limit of 6–18 was applied to household children; one-child households and incomplete samples were removed, and then the individual and household data of 13,244 children (from 7426 Iranian households) were studied. This database included characteristics of households, such as income, expenditure, age, gender, education level, job status, family size, and marital status of household members. The main observations of this study included children so a unique ID was used to match the information of other family members to the personal information of children. Quantity of siblings or family size considered the total number of 0–18 years old children who lived together at the time of the census. The instrumental variable (gender of the first child born in the family) was determined based on the gender of the oldest child of the family at the time of the census which equaled one when the first child was a daughter and zero otherwise.

The dependent variable of the model is the last education degree of children and considered as a human capital index, which comprised eight intervals from uneducated to Ph.D. degrees. Control variables included personal characteristics of the child (including age, gender, and birth order) and family characteristics (such as income, age, education level of parents, and households' type of dwelling (rural or urban)).

The sample size was limited based on some constraints. First, no-child and one-child families were excluded for using the gender of the first child in the family as an instrumental variable in the model. Second, the school-age range was subjected to 6–18 age due to formal education beginning at age of 6 or 7 and ends at 18 in Iran. Third, data of children whose mothers were older than 40 were removed for minimizing the probability that adult children have left the family before the census.

Table 1 reports a summary of descriptive statistics of the personal and household characteristics among the sample. The average age of subjects equaled 10.91, and the average education level of them was about 1.47 units. Based on Table 1, about 47% of the first children of families were girls. The average age of fathers and mothers equaled 40.13 and 34.66, respectively; accordingly, fathers were older than mothers were. The education level of fathers was also on average higher than mothers of selected children. The average family size was about 2.73, and almost 0.48%

Table 1 Descriptive statistics of the sample

Variables	All	Firstborn girl	Firstborn boy
Children's age	10.91 (3.38)	11.78 (3.39)	11.88 (3.45)
Firstborn girl	0.47 (0.49)	–	–
Educational level	1.47 (0.78)	1.65 (0.89)	1.65 (0.88)
Mother's age	34.66 (3.91)	33.85 (4.07)	33.88 (4.13)
Father's age	40.13 (5.33)	39.02 (5.09)	39.09 (5.21)
Mother's educational level	2.36 (1.70)	2.62 (1.72)	2.59 (1.72)
Father's educational level	3.00 (1.83)	3.20 (1.81)	3.21 (1.82)
Family size	2.73 (0.97)	2.53 (0.79)	2.46 (0.76)
Dwelling (urban = 1)	0.48 (0.49)	0.51 (0.49)	0.50 (0.50)
Family income	288,000,000 (226,000,000)	289,000,000 (226,000,000)	293,000,000 (221,000,000)
Number of observations	13,244	13,244	13,244

standard errors in the parentheses

of selected subjects lived in rural dwellings and the rest in urban. The same statistics were separately extracted for two groups of households whose first children were boys and households whose first children were girls.

Model and Variables

Based on the literature review, the first step is estimating the effect of family size on the education level of children by using the OLS model. The model was specified as follows:

$$E_i = \beta_0 + \beta_1 \text{sibsize}_i + X'_i\beta_2 + X'_j\beta_3 + u_i \quad (1)$$

where E_i represents the education level of child I, sibsize indicates the total number of children younger than 18 years who live together with the studied child in a dwelling at the census time, β_1 reflects the effect of family size on the education status of children, X_i is a vector of child-level covariates including age, age square, gender, and birth order, X_j represents a vector of parent/household-level covariates including income, age, age square, education level of parents, and a dwelling dummy variable, and μ_i is the error term.

β_1 reflects the causal effect of family size on the educational status of children, whether sibsize is determined exogenously. OLS estimates may have upward or downward biases regarding the endogeneity source. Therefore, the addressing of endogeneity of family size is the most challenging point in the examined relationship that is influenced by family income, child expenditures, contraception knowledge, lack of confidence, and relative preferences or "tastes" of parents. The tastes might be determined based on the religion, race, age, and other characteristics of a family (Becker, 1960). Kugler and Kumar (2017) and Lee (2008) used the gender of

the first-born child, and Angrist and Evans (1998) used the gender of the two first children as an instrumental variable for the family size. Therefore, the gender of the firstborn child was introduced as an instrumental variable for family size for overcoming the endogeneity issue. Finally, the model was specified as follows:

$$E_i = \pi_0 + \pi_1 \text{sibsize}_i + X'_i \pi_2 + X'_j \pi_3 + \delta_i \tag{2}$$

$$\text{sibsize}_i = \alpha_0 + \alpha_1 FBG_i + X'_i \alpha_2 + X'_j \alpha_3 + \varepsilon_i \tag{3}$$

Equation (3) indicates the first regression step, in which *FBG* is a dummy variable. *FBG* equals one when the first child of the family is a girl and zero otherwise. Other variables of Eq. (3) are similar to the variables specified in Eq. (1). Equation (2) shows the second step in which the education level of children is determined based on the predicted sibsize variable in Eq. (3) and other control variables.

In the based econometrics literature, using the 2SLS approach is dependent on the validity of the instrumental variable in regression, which means sibsize is highly correlated to the gender of the first child of the family. This behavior associates with the preference for a specific gender in society. The male gender has been always preferred in Iranian society due to different cultural factors, customs, and beliefs (to keep the family name alive, to achieve a better job position, and to support parents when they become old). However, this preference has become less; but still, parents prefer to have more children, when their first child is a girl. Nevertheless, this case occurs in absence of medical interference to determine the gender of children and no aborting female gender (necessity of exogeneity of the gender in the first child). One-sample t-test was used for examining this presumption. As shown in the next section, the results of this test indicated that the null hypothesis is rejected which means the average household size in the target sample (families whose first child is a boy) is not equal to the predictive value of the family size of all sample families. Therefore, Iranian families prefer having boys, and the gender of the first child of the family affects the family size. Hence, FBG could be considered as an appropriate and reliable instrumental variable for predicting sibsize.

Empirical Findings

As seen in "Model and Variables" section, the effect of family size on the education level was first estimated by using the OLS model, and Table 2 reports the results in five types of models. The family size was the only variable that entered into model I. In model II, personal variables of children (age, squared age, gender, and birth order of child) were considered as control variables. In models III and IV, variables related to parents' characteristics were employed separately along with the personal variables of children. Finally, individual and household variables of children were entered into model V. Comparing the results of different regressions allows us to better understand the importance of the presence of family characteristics of children.

As reported in Table 2, an increase in the number of births in a family leads to a decrease in the probability of graduating from an education level. In addition, a

Table 2 The effect of family size on education resulting from the OLS approach

Variable	Model I	Model II	Model III	Model IV	Model V
Sibsize	0030 (0.007).	043 (0.004).−0	024 (0.005).−0	026 (0.005).−0	−0.021* (0.005)
Child's age	–	357 (0.007).−0	354 (0.009).−0	358 (0.009).−0	−0.355* (0.009)
Child's age squared	–	0023 (0.0003).	0023 (0.000).	0023 (0.000).	0.023* (0.0004)
Gender (female=1)	–	0016 (0.007).	0015 (0.006).	0015 (0.006).	0.015* (0.006)
Birth order	–	0010 (0.005).	0012 (0.006).	0005 (0.006).	0.007* (0.006)
Mother is illiterate	–	–	169 (0.017).−0	–	−0.086* (0.022)
Mother is primary schooled	–	–	073 (0.007).−0	–	−0.025* (0.009)
Father is illiterate	–	–	–	195 (0.031).−0	−0.111* (0.037)
Father is primary schooled	–	–	–	096 (0.009).−0	−0.061* (0.012)
Mother's age	–	–	032 (0.010).−0	–	−0.037* (0.010)
Mother's age squared	–	–	0000 (0.000).	–	0000 (0.000).
Father's age	–	–	–	010 (0.005).	0.009* (0.005)
Father's age squared	–	–	–	000 (0.000).−0	000 (0.000).−0
Family income	–	–	–	–	0.041* (0.004)
Dwelling (urban = 1)	–	–	–	–	0.037* (0.007)
Cons	1391 (0.020).	2377 (0.042).	2826 (0.174).	2092 (0.112).	2.693* (0.188)
R^2	0.0014	0.73	0.74	0.74	0.74
Number of observations	13,244	13,244	13,244	13,244	13,244

standard errors in the parentheses

*$p < 0.05$

negative correlation between family size and the education level of children is obtained. Furthermore, the results indicate that the parents' characteristics have a significant effect on education, but the signs of the effects are unexpected. For example, we expect the positive effect of parents' education on the educational level of children, but the results indicate negative effects. Therefore, the results are not valid and need to improve the estimation methods.

Regarding the specified constraint (probability of endogeneity in family size) of the OLS model in "Empirical Findings" section, this equation was re-estimated by using the 2SLS model. The validity of the instrumental variable (gender of the first child of the family) must be examined before presenting the results of this estimation. In other words, the hypothesis of Iranian families' preference to have boys must be validated. To do so, a one-sample T-test was used for the total selected households and families whose first children were boys. Table 3 reports the result of the mentioned t-test.

Table 3 Sample means and test of mean differences

Variable	Whole sample mean (μ)	Mean difference ($\mu_m - \mu$)
Number of children	2.498	−0.028* (0.012)
Sample size	7433	3899

mean differences are defined as the sample means of households having a son as the first child (μ_m) minus the sample means of all households (μ). Numbers in parentheses are standard errors

*$p < 0.05$

In this table, the mean difference is defined as the mean difference between total sample families and sample families whose first children were boys. According to the obtained results, the number of children in an Iranian family highly depends on the gender of the first child born in the family. The calculated mean difference of −0.028 was significant at the level of 1%. Therefore, if the first child of an Iranian family is a boy, then they are less likely to prefer to have a second child. Hence, the gender of the first child affects the family size significantly. Table 4 reports the

Table 4 First step of 2SLS estimates of the effect of family size on education

Variable	Coefficient	Standard error	t-statistic	$p > \|t\|$
FBG (female = 1)	0.094*	0.016	5.68	0.000
Child's age	0.102*	0.014	7.01	0.000
Child's age squared	−0.000	0.000	−0.60	0.548
Gender (female = 1)	0.008	0.016	0.50	0.614
Birth order	0.649*	0.010	60.15	0.000
Mother is illiterate	0.655*	0.037	17.59	0.000
Mother is primary schooled	0.191*	0.020	9.27	0.000
Father is illiterate	0.316*	0.052	6.06	0.000
Father is primary schooled	0.069*	0.022	3.05	0.002
Mother's age	0.137*	0.025	5.31	0.000
Mother's age squared	−0.002*	0.000	−6.36	0.000
Father's age	−0.008	0.010	−0.88	0.377
Father's age squared	−0.000	0.000	−1.30	0.193
Family income	−0.129*	−0.031	−4.09	0.000
Dwelling (urban = 1)	0.013	0.014	−0.95	0.341
Cons	0.720	0.439	−1.64	0.101
R-squared	0.34			
Adj *R*-squared	0.34			
F-statistic	456.28			
Prob(*F*-statistic)	0.000			
Number of observations	13244			

*$p < 0.05$

results obtained from the first step of 2SLS estimation in which this hypothesis is explained in another way.

According to the results obtained from the first step of 2SLS regression, the instrumental variable (FBG) applied in this study was at a high significance level of 99% and had a positive correlation with the family size. Therefore, parents whose first child is a girl prefer to have more children until they have a boy, so the family size will increase with a higher probability in such families.

Results of the second step of 2SLS regression have been reported in Table 5. These results indicate the weak significance of the level of family size, as well as its positive effect on the education level of children contrary to OLS estimates. In addition, the comparing 2SLS and OLS estimation indicated that the removal of the endogeneity aspect of family size changes the results completely. Such change in previous empirical studies appeared in form of the smaller coefficients or changed direction and or changed the significance level of variables. Furthermore, the results indicate that mother's education, mother's age, and family income have a positive and significant effect on the educational level of children. Furthermore, the effect of gender on the educational level of children is not significant. Finally, the results show the positive effect of dwelling on the educational level of children.

Table 6 reports comparing OLS and 2SLS estimation. The results of 2SLS estimation indicate that birth order and mother age have a negative and positive effect

Table 5 Second step of 2SLS estimates of the effect of family size on education

Variable	Coefficient	Standard error	t-statistic	$p > \|t\|$
Sibsize	0.157	0.093	1.69	0.091
Child's age	−0.373	0.012	−30.23	0.000
Child's age squared	0.023	0.0003	70.35	0.000
Gender (female = 1)	0.004	0.009	0.50	0.619
Birth order	−0.109	0.061	−1.79	0.073
Mother is illiterate	−0.203	0.064	−3.17	0.002
Mother is primary schooled	0.059	0.020	−2.83	0.005
Father is illiterate	−0.168	0.040	−4.16	0.000
Father is primary schooled	−0.074	0.013	−5.40	0.000
Mother's age	0.062	0.018	−3.31	0.001
Mother's age squared	0.001	0.0003	3.29	0.001
Father's age	0.010	0.005	1.97	0.049
Father's age squared	−0.000	0.000	−0.74	0.458
Family income	0.064	0.020	3.12	0.002
Dwelling (urban = 1)	0.025	0.007	3.21	0.001
Cons	2.831	0.244	11.60	0.000
R-squared	0.71			
Adj *R*-squared	0.70			
F-statistic	2260.83			
Prob(*F*-statistic)	0.000			
Number of observations	13,244			

Table 6 Comparison of OLS regression results and 2SLS regression results

Variable	OLS results	2SLS results
Sibsize	−0.021*** (0.005)	0.157* (0.093)
Child's age	−0.355*** (0.009)	−0.373*** (0.012)
Child's age squared	0.023*** (0.0004)	0.023*** (0.000)
Gender (female)	0.015** (0.006)	0.004 (0.009)
Birth order	0.007 (0.006)	−0.109* (0.061)
Mother is illiterate	−0.086*** (0.022)	−0.203*** (0.064)
Mother is primary schooled	−0.025** (0.009)	−0.059*** (0.020)
Father is illiterate	−0.111*** (0.037)	−0.168*** (0.040)
Father is primary schooled	−0.061*** (0.012)	−0.074*** (0.013)
Mother's age	−0.037*** (0.010)	0.062*** (0.018)
Mother's age squared	0.000*** (0.000)	0.001*** (0.000)
Father's age	0.009* (0.005)	0.010** (0.005)
Father's age squared	−0.000 (0.000)	−0.000 (0.000)
Family income	0.041*** (0.004)	0.064*** (0.20)
Dwelling (urban = 1)	0.027*** (0.007)	0.025*** (0.007)
Cons	2.693*** (0.188)	2.831*** (0.244)
R-squared	0.74	0.71
F-statistic	2554.20	2260.83
Prob (*F*-statistic)	0.000	0.000
Number of observations	13,244	13,244

standard errors in the parentheses

***$p < .01$

**$p < .05$

*$p < .1$

Table 7 Results of the Durbin–Wu–Hausman test

Exogeneity test	Test-statistic	Probability
FAQ	Chi2(1) = 4.17	0.041
The Durbin–Wu–Hausman	$F(1,13227) = 4.17$	0.041

Table 8 The Breusch–Pagan test results

The Breusch–Pagan test	Test-statstic	Probability
Variance homogeneity versus variance heterogeneity	11,658.69	0.0000

on the educational level of children, respectively, while the results of OLS estimation are contrary to the results of 2SLS estimation. In addition, employing the 2SLS method improves the accuracy of model coefficients.

Table 7 indicates the results of the exogeneity of the Durbin–Wu–Hausman test. H_0 tests the exogeneity of the variable. According to the results reported in this

Table, H_0 (exogeneity of family size) was rejected at the confidence level of 95%; hence, this variable is endogenous.

Variance homogeneity is a presumption of regression equation indicating fixed variance of residuals. If the variance of residuals is not fixed, variance heterogeneity exists. Breusch–Pagan's test was used to examine variance homogeneity. As reported in Table 8, H_0 (variance homogeneity) was rejected; therefore, the 2SLS-based estimate had variance heterogeneity. The white test was used to correct this issue. To this end, the standard deviation of coefficients was calculated by consideration of variance heterogeneity. The results obtained from the 2SLS model after correcting the variance heterogeneity were compared to the OLS results in Table 9. As reported in Table 9, family size had a positive and significant effect on the education level of children at the significance level of 95%. Also, estimated results indicate a negative and significant effect of birth order on the education level of children; it means younger children of the family will be less likely to be graduated compared to older ones. Furthermore, the effect of other variables is the same as in Table 6.

Table 9 Comparing OLS with 2SLS regression results after addressing variance heterogeneity

Variable	OLS regression	2SLS regression
Sibsize	−0.021*** (0.005)	0.157** (0.078)
Child's age	−0.355*** (0.009)	−0.373*** (0.012)
Child's age squared	0.023*** (0.0004)	0.023*** (0.000)
Gender (female = 1)	0.015** (0.006)	0.004 (0.007)
Birth order	0.007 (0.006)	−0.109** (0.051)
Mother is illiterate	−0.086*** (0.022)	***−0.203 (0.056)
Mother is primary schooled	−0.025** (0.009)	−0.059*** (0.017)
Father is illiterate	−0.111*** (0.037)	−0.168*** (0.045)
Father is primary schooled	−0.061*** (0.012)	−0.074*** (0.013)
Mother's age	−0.037*** (0.010)	0.062*** (0.015)
Mother's age squared	0.000*** (0.000)	0.001*** (0.000)
Father's age	0.009* (0.005)	0.010** (0.005)
Father's age squared	−0.000 (0.000)	−0.000 (0.000)
Family income	0.041*** (0.004)	0.064*** (0.018)
Dwelling (urban = 1)	0.027*** (0.007)	0.025*** (0.007)
Cons	2.693*** (0.188)	2.831*** (0.196)
R-squared	0.74	0.74
F-statistic	2554.20	1559.59
Prob (*F*-statistic)	0.0000	0.0000
Number of observations	13,244	13,244

standard errors in the parentheses

***$p < .01$

**$p < .05$

*$p < .1$

Table 10 2SLS estimation by residence

Variable	Educational level (rural households)	Educational level (urban households)
Sibsize	0.108 (0.083)	0.240 (0.156)
Children's controls	Yes	Yes
Parent's controls	Yes	Yes

standard errors in the parentheses

Table 11 2SLS estimation by mother's education

Variable	Educational level (primary and less)	Educational level (more than primary)
sibsize	0.108 (0.097)	0.121 (0.105)
Children's controls	Yes	Yes
Parent's controls	yes	Yes

standard errors in the parentheses

For answering the 2nd and 3rd questions, we applied the 2SLS estimation among urban/rural groups and uneducated/educated mothers. The results obtained after the removal of heteroscedasticity indicated the positive and insignificant effect of the family size on the education level of children in all four groups studied in this research. In addition, the effect of family size in urban households and educated mothers is greater than in rural households and uneducated mothers, respectively. Tables 10 and 11 report 2SLS estimation in four groups of rural and urban families (Table 10), as well as uneducated/less educated and educated mothers (Table 11).

Discussion

Although the quantity-quality trade-off model benefits from an acceptable theoretical and empirical background in many countries, few studies have examined this association based on the causal effects. Family size is one of the main factors that affect the education of children in a family along with household income level and parents' education level (Duncan et al. 2017).

In this paper, we focused on three questions: do children with more siblings have lower education levels (is there a negative correlation between family size and the education of children)? Is there a different effect of family size on the rural groups comparing urban groups? Is there a similar effect of family size on the uneducated or low-educated mothers comparing the educated ones?

According to the obtained results, the OLS estimate indicated a negative and significant effect of family size on the education level of children, while the 2SLS method showed a positive and significant effect of family size on education. The

negative association between the aforementioned two variables in this study and other empirical studies may stem from the unobserved household characteristics or bias in removed variables. To ensure the reliability of these results, this estimation was done two other times by consideration of the total number of children (child or adult) living in a dwelling and by excluding children older than 15. However, the findings remained unchanged. Another significant finding was Iranian families' preference for having boys that affected family size.

These results verify the empirical results in Duncan et al. (2017).

The potential endogeneity of family size is the reason for any specific conflict in results in the studies. So, for answering the 2nd and 3rd questions, we applied the 2SLS estimation among urban/rural groups and uneducated/educated mothers. The empirical literature was created after introducing control variables for family characteristics and instrumental variable approaches over the recent 20 years (Becker & Tomes, 1976; Blake, 1981; Angrist et al. 2010; Conley & Glauber, 2006; Kugler & Kumar, 2017; Lee, 2008).

The results obtained after the removal of variance heterogeneity indicated the positive and insignificant effect of the family size on the education level of children in all four groups studied in this research. In addition, the effect of family size in urban households and educated mothers is greater than in rural households and uneducated mothers, respectively. These results verify the results of Adongo et al. (2022); Qian (2009); Zhang (2017).

Conclusion and Policy Implications

Iranian society is a young one, and education is the main factor that can improve the economic development. Therefore, an analysis of the family size on education level is an important subject for policymakers and the future of Iran. The results of this paper indicate that the effect of family size on Iranian society is affected by households' characteristics. In addition, the social and economic status of households is the important for the relationship between family size and human capital. Therefore, it needs to consider the household's characteristics and the social and economic status of households for analyzing the relationship between family size and children's education in developing countries such as Iran.

The results of this paper indicate that to improve the way for development, especially in developing countries such as Iranian society, it needs to correct the population age structure, solve pension system crises, reduce potential recession and economic crises, alleviate poverty, and reduce inequality. In addition, on the base of these results, we suggest policymakers to improve human capital in developing countries such as Iran and should consider mother's education, especially in rural households. Furthermore, policymakers should regard improving the social and economic status of households to improve human capital that can achieve the development goals.

Applying the information about the number of births given by the mother, detailed information about the educational status of individuals and blood and non-blood relationship between parents and children can be considered in future studies. In addition, it is essential to find the effect of public demographic and educational policies on the enhanced investment in human capital for studies in developing countries.

Funding Faculty of Economics, the University of Tehran funded this article.

Declarations

Ethical Approval This article does not contain any studies with human participants or animal performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

References

Adongo, A. A., Dapaah, J. M., & Wireko, D. (2022). The influence of family size on academic performance of high school students in Ghana. *SN Social Sciences, 2*(9), 1–26.

Angrist, J., & Evans, W. N. (1996). *Children and their parents' labor supply: Evidence from exogenous variation in family size*. Cambridge: National Bureau of Economic Research.

Angrist, J., Lavy, V., & Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics, 28*(4), 773–824.

Argys, L. M., & Averett, S. L. (2015). The effect of family size on education: New evidence from China's one child policy. Discussion Paper No. 9196, IZA

Barra, C., & Boccia, M. (2022). What matters in educational performance? Evidence from OECD and non-OECD countries. *Quality & Quantity, 56*(6), 4335–4394.

Blanden, J., Del Bono, E., Hansen, K., & Rabe, B. (2022). Quantity and quality of childcare and children's educational outcomes. *Journal of Population Economics, 35*(2), 785–828.

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries* (pp. 209–240). Columbia University Press.

Becker, G. S., & Tomes, N. (1976). Child endowments and the quantity and quality of children. *Journal of political Economy, 84*(4, Part 2), S143–S162.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (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.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2010). Small family, smart family? Family size and the IQ scores of young men. *Journal of Human Resources, 45*(1), 33–58.

Blake, J. (1981). Family size and the quality of children. *Demography, 18*(4), 421–442.

Conley, D., & Glauber, R. (2006). Parental educational investment and children's academic risk estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of Human Resources, 41*(4), 722–737.

Dang, H.-A. H., & Rogers, F. H. (2016). The decision to invest in child quality over quantity: Household size and household investment in education in Vietnam. *The World Bank Economic Review, 30*(1), 104–142.

Downey, D. B. (1995). When bigger is not better: Family size, parental resources, and children's educational performance. *American Sociological Review, 60*(5), 746–761.

Duncan, G. J., Kalil, A., & Ziol-Guest, K. M. (2017). Increasing inequality in parent incomes and children's schooling. *Demography, 54*(5), 1603–1626.

Feng, N. (2020). The effect of sibling size on children's educational attainment: Evidence from Indonesia. *ECNU Review of Education, 4*(4), 830–856.

Glick, P. J., Marini, A., & Sahn, D. E. (2007). Estimating the consequences of unintended fertility for child health and education in Romania: An analysis using twins data. *Oxford Bulletin of Economics and Statistics, 69*(5), 667–691.

Huang, Y. (2022). Family size and children's education: Evidence from the one-child policy in China. *Population Research and Policy Review, 41*(1), 1–26.

Kamanda, M., Madise, N., & Schnepf, S. (2016). Does living in a community with more educated mothers enhance children's school attendance? Evidence from Sierra Leone. *International Journal of Educational Development, 46*, 114–124.

Kugler, A. D., & Kumar, S. (2017). Preference for boys, family size, and educational attainment in India. *Demography, 54*(3), 835–859.

Kwan, S. (2009). Framing the fat body: Contested meanings between government, activists, and industry. *Sociological Inquiry, 79*(1), 25–50.

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

Li, H., Zhang, J., & Zhu, Y. (2008). The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography, 45*(1), 223–243.

Liu, H. (2014). The quality–quantity trade-off: Evidence from the relaxation of China's one-child policy. *Journal of Population Economics, 27*(2), 565–602.

Mosli, R. H., Kaciroti, N., Corwyn, R. F., Bradley, R. H., & Lumeng, J. C. (2016). Effect of sibling birth on BMI trajectory in the first 6 years of life. *Pediatrics, 137*(4), e20152456.

Qian, N. (2009). Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China. NBER Working Paper No. 14973, National Bureau of Economic Research.

Rosenzweig, M. R., & Zhang, J. (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's "one-child" policy. *The Review of Economic Studies, 76*(3), 1149–1174.

Wang, D. (2019). Reduction but not elimination: Health inequalities among urban, migrant, and rural children in China—the moderating effect of the fathers' education level. *BMC Public Health, 19*(1), 1–12.

Zhang, J. (2017). The evolution of China's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives, 31*(1), 141–160.