Essays on Labor Economics: Parental Preference, Expansion of Education, and Examinations of Causality

Hung-Han Lin

Academic Advisor:

Prof. Yoshiaki Omori

This dissertation has been submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Graduate School of International Social Sciences
Yokohama National University

September 2017

Abstract

This dissertation consists of four self-contained articles that study the causal relationships of labor economics in Taiwanese society. Chapter 1 investigates the effect of childbearing on female labor supply. Chapter 2 empirically examines the existence of a quantity-quality trade-off of children. Chapter 3 focuses on the causal relation between education and health outcomes. Chapter 4 estimates the effect of educational attainment on marriage outcome, and tests the hypothesis of incompatible pool of potential mates.

In society, women generally play the roles of mothers and employees, and those are partly substitute roles to each other; therefore, the contributions of women to the labor market should be considered. Chapter 1 applies the widely observed parental preference for children's sex as an instrumental variable for the number of children in Taiwan and examines the causal effect of childbearing on female labor supply. Ordinary least square estimates reveal that childbearing is associated with reduced female labor supply. However, after using the sex of first and second children as instruments to account for the endogeneity of the number of children, two-stage least square estimates reveal no correlation between the number of children and female labor supply.

Chapter 2 examines the relation between the number of children in Taiwanese families and the educational attainment of those children. To identify the causal relation, our analysis operationalizes the traditional Taiwanese parental preference for male children as an instrumental variable to generate exogenous variations in the

number of siblings. Ordinary least square estimates reveal that a larger number of siblings results in lower educational attainment. However, after addressing for the endogeneity of the number of siblings, our two-stage least square estimates result in doubt of the existence of a trade-off between child quantity and quality within a family.

Studies have consistently reported a positive relationship between education and health status. Chapter 3 refers to Taiwan's educational reforms to account for the well-known endogeneity problem of education. Notably, a substantial growth in the numbers of senior high schools and four-year colleges and universities began in the mid-1990s in Taiwan and has led to an increase in the proportion of college graduates. Multiple measures of health status are used for the estimates herein, namely self-reported health status, out of labor market because of health conditions, and body mass index (BMI) score. The instrumental variable estimates indicate a negative correlation between education and self-reported health status but a positive correlation between education and having no work disability. However, notwithstanding the causal relationship that exist between subjective or objective health outcomes and education, the instrumental variable estimates provide no evidence that education affects BMI scores, overweightness, and obesity.

Using data concerning Taiwan, Chapter 4 empirically investigates the relationship between education and marriage outcome. Taiwan has experienced a great expansion in higher education since the mid-1990s. The rapid increase in the number of four-year colleges and universities has led to a dramatic increase in the number of college graduates. After employing the expansion of education as an instrumental

variable, we find that, for both males and females, more education leads to a higher possibility of not marrying. We test our hypothesis that the educational composition of a married couple corresponds either to educational assortative mating or results in husbands with higher educational levels than their wives' educational levels. We find limited evidence that well-educated women tending to remain unmarried is caused by their facing a smaller pool of potential partners in the marriage market than do less well-educated women.

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS	Two-Stage Least Squares
BMI	Body Mass Index
LFP	Labor Force Participation
OLS	Ordinary Least Squares
PSFD	Panel Study of Chinese Family Dynamics
TSCS	Taiwan Social Change Survey
WH	Work Hour

Chapter 1

Childbearing and Female Labor Supply:

An Empirical Study in Taiwan

1.1 Introduction

The roles of women are influenced by physiological, cultural, and historical factors. In numerous countries, women who bear children also consider their work and career to be crucial. Therefore, the relationship between childbearing and female labor supply has always been a critical topic in the labor market. Generally, the scarcity of time as a result of childbearing causes households to choose between childcare (either stay-at-home parenting or childcare services) and labor force participation (LFP). For example, women who decide to have children may opt to reduce their working hours for childbearing or increase their working hours for the cost of raising children.

Women play two roles in societies, as mothers and employees, those are partly substitute roles to each other. The negative relationships of fertilities and female labor supply are observed in many countries. Several reasons may change the importance of roles of women in developing countries because of that the national progress accompanies social changes and new policies. For instance, the opportunity cost of childbearing is potentially influenced by government policies; an educational policy may change the opportunity cost of that female are not in the labor market. Taiwan has experienced an expansion of education in the past two decades, which increased the average educational attainment of female and their opportunity cost of not being in the labor market. Female tend to participate more in the labor market because of the increase of opportunity cost of childbearing. The birthrate in Taiwan decreased from 15.58‰ in 1993 to 8.53‰ in 2013, whereas the female LFP rate increased from 44.89% in 1993 to 50.4% in 2013.[1] This relationship is not exclusive to Taiwan. Mishra and Smyth (2010) revealed that in Organisation for Economic Co-operation and Development countries, a negative relationship is observed between the female LFP rate and fertility rates as a result of the dual roles of women as workers and mothers.

The time and costs involved in raising children are critical factors when individuals make decisions about participation in the labor market. However, the lack of information limits the feasibility of conducting relevant research. Studies have used the number of children as a proxy for certain variables representing the time and costs involved in raising children. Although married women have increasingly participated in the labor market in recent years, they have continued to retain their role of primary caregiver in many instances, which conflicts with their labor market participation. Studies on the effect of childbearing on female labor supply have determined that the association between childbearing and participation in the labor market can be attributed to various reasons such as the allocation of time, connection with childcare industries, public policy concerns, and other factors of opportunity costs of childbearing. Gronau (1973) estimated the price of time of mothers by using the framework of the shadow price of time. In that study, couples' personal characteristics

[1] These data were sourced from the Directorate General of Budget, Accounting and Statistics (DGBAS) of Executive Yuan, Taiwan. The birthrate is the total number of live births per 1,000 in a population in a year.

and the number and age composition of children were used as covariates. The number of children was used as a proxy for childcare services. The number of children was used to indicate a relationship between childcare services and the price of time. Using a different approach, Blau and Robins (1988) directly examined the effect of childcare costs on family labor supply. They determined that the childcare cost affects household labor supply and childcare arrangement because the childcare costs are directly attributable to the number of children. Brewster and Rindfuss (2000) studied fertility rates and women's employment within industrialized countries. They interpreted the trends of fertility and female LFP and reviewed the childcare policies and arrangements in previous studies.

Ideally, economic agents should jointly make decisions on consumption, labor supply, and fertility. However, this ideal situation would lead to an identification issue, namely the problem of endogeneity, in research. When individuals make decisions on fertility and labor supply according to exogenous factors, such as personal preferences and labor market conditions, it is no longer possible to examine the effect of fertility on labor supply because both factors are determined endogenously.

Thus, when analyzing the impact of childbearing on labor supply, the problem of endogeneity must be taken into account. An appropriate approach entails using instrumental variables to generate exogenous variations in the number of children. The number of children conceived by a particular woman may be related to unobserved characteristics, such as attitude towards fertility, health status, personal financial circumstances, and location or residence. Other unobservable factors are related to cultural dynamics, such as pressure from peers and family members. These aspects also affect labor supply and pose an identification problem, thereby making it difficult to identify causal effects. Causal relationships can be identified by using instrumental variables. Angrist and Krueger (2001) described that the instrumental variable approach could overcome omitted variable problems.

Studies have considered the occurrence of twin births and the sex of children as natural causes of exogenous variations in numbers of children. Pregnancy can be both planned and unplanned, and this fact must be addressed in the first step of research on the relationship between fertility and labor supply. Rosenzweig and Wolpin (1980) addressed this concern, and they also investigated whether fertility is exogenous or endogenous by using a theoretical model and examined the empirical relationship between female labor supply and fertility in the context of a life-cycle decision-making process. Using the occurrence of twin births in a woman's pregnancy as a natural event, they examined the impact on subsequent fertility and labor supply behavior. Their results support that women's labor supply decisions are made considering the intention to have additional children.

Angrist and Evans (1998) used the sex preferences of parents in the United States as instruments to address the problem of endogeneity. The sex preference was used on the basis of the recognized view that Americans consider it ideal to have both male and female children. For instance, assume that an individual has two children; there are three possible sex compositions: one boy and one girl, two boys, or two girls. Having one boy and one girl is often preferred over having two boys or two girls in most instances. Therefore, parents who have only boys or girls intend to have an additional child whose sex is different from that of the previous children. They pointed out that the sex composition of children generates exogenous variations in family size. Hence, they used the widely-observed phenomenon of parental preferences for their research. They stated that sex compositions are randomly assigned; therefore, the sex of the previous children could be used as plausible instruments. The parental sex preference approach introduced by Angrist and Evans (1998) has been applied to other countries, such as South Korea (Chun & Oh, 2002) and Greece (Daouli, Demoussis, & Giannakopoulos, 2009), as well as to Latin American countries (Cruces & Galiani, 2007).

In Taiwan, it is widely observed that parents prefer male children. Traditionally, in Taiwanese culture, male children are more highly regarded because they can extend the family lineage. In addition, there are a number of preconceived perceptions and stereotypes related to males exhibiting higher performance academically and having more favorable career prospects (i.e., earnings) than women. Statistical information on human sex ratios has also shown that the ratios of the third child are much higher than those of the first and second children in Taiwan for the past few decades. For example, in 2013, the ratios of the first and second newborn children were 107.6 and 106.8, respectively, but the ratio of the third child was 114.4.[2] Hence, if the previous two children did not include a male child, son preference induces women to have a third child; this was also verified by Tsay and Chu (2005). Therefore, parents who only have daughters may aim to have additional children for the purpose of bearing a son. This finding implies that the sex of previous children may influence the number

[2] Data from the Ministry of the Interior, Executive Yuan, Taiwan. A sex ratio of 100 implies an equal number of boys and girls; a ratio greater than 100 implies that the number of boys is higher than that of girls; and a ratio below 100 implies that the number of boys is lower than that of girls.

of children that a couple decides to have.

The objective of this study is to apply the instrumental variables proposed by Angrist and Evans (1998) to evaluate the reported preference for a male child in Taiwan, to generate exogenous variations in the number of children, and to examine the effect of childbearing on female labor supply. The difference between American and Taiwanese parental preference of child gender is that American parents prefer mixed-sex composition of children and Taiwanese parents prefer male children. Therefore, the son preference is employed as an instrumental variable in the examinations of this chapter. Because the sex of previous children born within a particular family can impact on a couple's decision to have a subsequent child, and sex composition is almost randomly assigned, the preference for a male child therefore seems to be a plausible instrument to estimate the number of children. This chapter uses the Panel Study of Chinese Family Dynamics (PSFD) of Taiwan to investigate the causal effect of the number of children on female labor supply.

Our key hypothesis is that exogenous changes in number of children do not distinctly lead to the decreases or increases in maternal labor supply. An unexpected child birth may influence female labor supply through two possible sources of childcare. The first is the expenditure on childcare, a reduction in disposal income that may result in less consumption and less leisure hours. The second source is the time spent on childcare: when a woman decides not to engage in the labor market, she misses out on earning wages; accordingly, the cost of time spent on childcare can be determined considering the market wage. In addition, the time spent on childcare reduces the time for market activity and leisure. Nevertheless, because market wages and personal spending preferences other than for childcare are crucial factors, each source of childcare does not have a distinct impact on the labor market hours, thus the number of children does not necessarily decrease or increase maternal labor supply.

The empirical results indicate that the first two children are female have a positive effect on the number of children, and the instrumental variable approach identify no correlation between childbearing and female labor supply in Taiwan. The ordinary least squares (OLS) estimates show that childbearing is associated with reduced female labor supply. However, after accounting for the endogeneity of childbearing by applying the preference for male children as an instrument, two-stage least squares (2SLS) estimates suggest that childbearing does not significantly impact female labor supply. The OLS outcomes may overstate the impact of childbearing on female labor supply.

The remainder of this chapter is structured as follows: Section 2 presents a model of fertility; Section 3 describes the data used in the research and summary statistics; Section 4 presents the empirical methodology; Section 5 describes the empirical results; and Section 6 presents the conclusion.

1.2 Theoretical Framework

In this section, the model of fertility is conducted with an exogenous number of children. The model is based on that of Montgomery and Trussell (1986). Assume that a mother's utility depends on her desired number of children, n, leisure hours, l, and other consumption, y, a utility function that can be expressed as $u(n, l, y)$. A mother's time, T, can be divided into three parts, leisure time, childcare, and being in the labor market, so that $T = l + b + m$, where b is the time spent on childcare and m is market work. Further factors include exogenous husband's income, *h*, and market wage, *w*. In addition to the original model, we include an unexpected number of children as an exogenous variable, $\bar{n}$; this may be induced by twin birth or undesired child gender. The mother's maximization problem is thus expressed as:

$$\begin{aligned} &max \;\; u(n,l,y) \\ &s.t. \;\; y + c(w,n,\bar{n}) + wl - h - wT \end{aligned} \tag{1.1}$$

where $c(w,n,\bar{n})$ is a cost function of childcare and includes time spent on childcare $wb(w,n,\bar{n})$ and expenditure on childcare $e(w,n,\bar{n})$.[3]

The first-order conditions are as follows:

$$\begin{aligned} &u_n - \lambda c_n = 0 \\ &u_l - \lambda w = 0 \\ &u_y - \lambda = 0 \end{aligned} \tag{1.2}$$

We are interested in the effect of unexpected birth on female labor supply, and therefore recognize the potential of $dl/d\bar{n}$ for signifying this. For comparative statics, let the market wage and husband's income remain constant, $dw = 0$ and $dh = 0$, and let the unexpected number of children be potentially changeable, $d\bar{n} \neq 0$. The matrix form is expressed as:

$$\begin{bmatrix} u_{nn} & u_{nl} & u_{ny} & -c_n \\ u_{nl} & u_{ll} & u_{ly} & -w \\ u_{ny} & u_{ly} & u_{yy} & -1 \\ -c_n & -w & -1 & 0 \end{bmatrix} \begin{bmatrix} \dfrac{dn}{d\bar{n}} \\ \dfrac{dl}{d\bar{n}} \\ \dfrac{dy}{d\bar{n}} \\ \dfrac{d\lambda}{d\bar{n}} \end{bmatrix} = \begin{bmatrix} \lambda c_{n\bar{n}} \\ 0 \\ 0 \\ c_{\bar{n}} \end{bmatrix}. \tag{1.3}$$

The effect of an unexpected number of children on leisure time is given by equation

[3] See Montgomery and Trussell (1986) for the cost minimization problem.

(1.4), where H is a 4 × 4 matrix that according to second-order conditions $|H|$ is negative; however, there is still an ambiguous sign in Equation (1.4).

$$\begin{aligned}\frac{dl}{d\bar{n}} = \frac{1}{|H|}[&-\lambda c_{n\bar{n}}(wu_{ny} + c_n u_{ly} - wc_n u_{yy} - u_{nl}) \\ &+c_{\bar{n}}(c_n u_{ny} u_{ly} + wu_{nn} u_{yy} + u_{nl} u_{ny} - c_n u_{nl} u_{yy} - wu_{ny}^2 - u_{nn} u_{ly})]\end{aligned} \tag{1.4}$$

The time spent out of the labor market is comprised of leisure and childcare. An increase in the number of children could cause an increase in time spent on childcare, $db/d\bar{n} > 0$, yet the impact on number of leisure hours is unknown, $dl/d\bar{n} \geq 0$ or $dl/d\bar{n} \leq 0$. Thus, the impact of an exogenous number of children on the time spent out of market work also remains ambiguous.

1.3 Data and Descriptive Statistics

The PSFD of Taiwan is used in this study. The PSFD provides data on individuals' fertility records and other personal information over multiple years; this project was led by a Taiwanese research institute, Academic Sinica. The types, structures, and interactions of families in the Taiwanese society are considerably different from those in other societies. The main respondents of the PSFD are adults from Taiwanese families, and the PSFD uses the following four birth cohorts as the main samples: 1953–64, 1935–54, 1964–76, and 1977–83.

The respondents of the PSFD are Taiwanese adults; thus, these data can clearly indicate parental preferences for male children in the society. In this study, the instrument of the parental preference in Taiwan is applied to examine the effect of childbearing on female labor supply. The examinations require high-quality micro data including observed subjects' complete fertility records and other personal information. Observed subjects' complete fertility records, which are the most crucial, included the number of children and the sex of each child. Because of the availability of information, the effect of previous children's sex on the total number of children could be identified.

For applying the instrumental variables, namely the sex of first and second children, the female subjects with at least two children are chosen from the data set. The human sex ratio of infants in Taiwan encouraged this choice of the sample. The human sex ratio is the ratio of males to females in a population. A sex ratio of 100 implies an equal number of boys and girls; a ratio greater than 100 implies that the number of boys is higher than that of girls; and a ratio below 100 implies that the number of boys is lower than that of girls. In 2013, in Taiwan, the sex ratio of the first and second children was 107.6 and 106.8, respectively, and the ratio of the third child was higher at 114.4. Clearly, son preference more significantly induces women to have a third child. This study uses data from 2009, 2010, and 2011; after pooling and clearing all samples, the number of observed subjects is 3,509.[4]

Table 1.1 reports the descriptive statistics. In the data, the LFP rate for females is approximately 49%, which is close to the actual situation in Taiwan.[5] Children's sex conforms closely to natural events as women have a 50% chance of giving birth to a male or female child. The education level is also a critical factor for labor supply. Thus, we include the education levels of subject and present them as dummy variables. In the data set, 57%, 24%, and 19% of women attended junior high school or below,

[4] The original numbers observed subjects are 3,208 for 2009; 5,073 for 2010; and 4,885 for 2011. After excluding subjects who did not respond, were male, and had only one child or no children, the observed subjects totaled 1,070 for 2009; 1,224 for 2010, and 1,215 for 2011.

[5] The DGBAS provides the official LFP rates in Taiwan. For females, the rates are 49.62%, 49.89%, and 49.97% for 2009, 2010, and 2011, respectively.

senior high school, and higher education, respectively.

Our empirical analysis estimates female labor supply behavior and include other personal factors as controls, such as marital status, educational level, and spousal characteristics. Couples may share their household income and make individual or household decisions (Chiappori, 1988). Married couples are considered in this study, because married people make decisions differently compared with those who are single. Therefore, the analysis includes marital status data and related variables, such as the spouses' employment status and education level.

Table 1.2 illustrates the preference for male children in Taiwan. This study focuses on women with at least two children and determines the number of subjects with exclusively female children who gave birth to an additional child. Therefore, the proportion of subjects with more than two children could indicate the specific sex preference in Taiwan. Approximately 52%–55% of subjects with two children intended to have an additional child when they had at least one boy. However, subjects with only two girls had higher incentive to have a third child; 67.73% had another child when their previous children were all girls. Furthermore, the Pearson's chi-square test is used to determine the relationship between child sex composition and additional children. The *p*-value (0.00) suggests that sex composition and the decision to have additional children are dependent.

1.4 Empirical Model

Previous studies of the effect of childbearing have used the number of children as the proxy variable for the cost childbearing in terms of money and time. Consistent with those approaches, this study uses the number of children as the main explanatory variable. Some unobserved family characteristics and factors related to female labor force may impact childbearing and labor supply simultaneously, such as attitudes toward fertility, the traditional Taiwanese norms, time distribution between child rearing and work, and increasing costs of raising children. Consequently, although it was nearly impossible to obtain complete information from surveys, the lack of variables may have disturbed the results of OLSs. Therefore, the analysis in this study uses instrumental variables to overcome the omitted variable problem. The instrumental variable approach can identify the causal effect of the number of children when the instrument is uncorrelated with the unobserved variables.

The main purpose of the empirical methodology is to determine the causal effect of children on female labor supply. The empirical model comprises two stages, which are shown in Equations (1.5), (1.6), and (1.7). In the first stage, Equation (1.7), the possible endogenous variable, the number of children, is estimated using the instrumental variables; this stage is used to generate exogenous variations to overcome the problem of endogeneity. In the second stage, Equations (1.5) and (1.6), a labor supply equation is formulated to determine the effect of children on female labor supply. In this study, the results of OLSs and logistic regressions are compared with those of 2SLSs.

$$LFP_i = \alpha Child_i + \Gamma X_i + u_i \tag{1.5}$$

$$WH_i = \beta Child_i + \Phi X_i + v_i \tag{1.6}$$

Equations (1.5) and (1.6) represent the labor supply; two dependent variables are used in this study, labor force participation (LFP) and work hours (WH) per week.

First, we employ a linear probability model, where LFP_i is a dummy variable denoting the individual i's state of LFP. When LFP_i is equal to 1, this implies that i is in the labor force; otherwise LFP_i is 0. Second, we use individuals' work hours per week to represent personal labor supply, for which WH_i denotes i's work hours per week.

Regarding the main explanatory variables, $Child_i$ is individual i's number of children, and α and β are its coefficient; those are also the main effect this study intends to examine. X_i is a vector of other independent variables, including a constant, and Γ and Φ are corresponding vectors of coefficients.

In Equation (1.5) and (1.6), u_i and v_i are the error terms. The number of children is considered as the possible endogenous variable in our estimations. Therefore, these error terms are possibly related to the number of children by unobserved characteristics, such as intention to bear children, health status, the traditional Taiwanese norms and family pressure reinforced by the Taiwanese society.

In the first stage, the model estimates the possible endogenous variable by using instrumental variables. The model regresses the number of children by applying the estimation equation shown in Equation (1.7),[6] where d_{1i} and d_{2i} are dummy variables for the sex of individual i's first and second child; these dummy variables are equal to 1 if the respective child is a girl; otherwise, it is 0. As shown in the sex statistics in Section 3 of this chapter, the preference for male children was primarily observed after the second child was born, particularly for people who only had daughters. Thus,

[6] In the original study by Angrist and Evans, s_1 and s_2 were the indicators for first- and second-born children; their instrument was given as $s_1s_2+(1-s_1)(1-s_2)$. The present study focuses on son preference and therefore simply used $d_{1i}d_{2i}$ to denote the birth of two daughters.

this study selected samples with at least two children. Interaction of first and second children's sex is also included in the equation; if $d_{1i}d_{2i}$ is equal to 1, both first and second children are girls. υ_i represents the error term of Equation (1.7). This error term is uncorrelated with the independent variables shown in Equation (1.7).

$$Child_i = \pi_1 d_{1i} + \pi_2 d_{2i} + \pi_3 d_{1i} d_{2i} + \Theta X_i + \upsilon_i \tag{1.7}$$

If the sex of the first two children satisfy the required conditions of instrumental variables, then the sex compositions conform to random assignment. Therefore, the number of children can only influence female labor supply through children's sex. The instruments can generate exogenous variations in the number of children and be used to examine the impact of childbearing on female labor supply behavior. The coefficients α and β shown in the second stage (Equation [1.5] and [1.6]) are the instrumental variable estimates; in other words, these estimates identify the causal effect of childbearing on female labor supply.

This study focuses on the effect of childbearing and included other variables as controls for enhanced precision. These control variables consist of personal characteristics such as individual's age, age squared, education level, and marital status. In addition, we also use the age of the second child as a control because this variable may be a factor contributing to the decision of having a third child or returning to the labor force. Considering the traditional Taiwanese family structure, it is not uncommon for adult children to live with their parents or in-laws after marriage. This is often done for both financial and practical reasons while raising a family. Our empirical analysis control for this phenomenon by using a dummy variable to denote whether a woman lives with her parents or in-laws. The education level is categorized into three levels, each representing dummy variables: junior high school or below, senior high school, and higher education. In the analysis, the highest degrees attained after the completion of junior high school or below is compared with the other two education levels. Furthermore, the analysis includes the husband's employment status and education level. Therefore, the marital status generated additional interactions. The husband's education level is divided into three different levels, similar to individual education, and employment status represented by a dummy variable takes a value of 1 if the spouse is employed and 0 otherwise. The empirical results are shown in Tables 1.4, 1.5, and 1.6. Table 1.4 shows the results of first stage, Table 1.5 reports the results of the LFP model, and Table 1.6 presents the results of the WH model. There are 3,509 observed subjects in the LFP model and 3,493 in the WH model.[7]

1.5 Empirical Results

Valid instrumental variables could affect the dependent variable only through endogenous variables. In this chapter, the genders of children affect maternal labor supply only through the number of children in a family. Before conducting our main estimations, we estimate Equation (1.8) and (1.9), which is the effect of child gender on labor supply. Table 1.3 reports the estimated coefficients. There is no evidence to suggest that previous female children have an impact on female labor supply.

$$LFP_i = \kappa_1 d_{1i} + \kappa_2 d_{2i} + \kappa_3 d_{1i} d_{2i} + \xi_i \tag{1.8}$$

$$WH_i = \rho_1 d_{1i} + \rho_2 d_{2i} + \rho_3 d_{1i} d_{2i} + \psi_i \tag{1.9}$$

[7] The different numbers of observed subjects between the LFP and WH models occurred because some subjects did not report their WHs in the surveys.

The first stage estimates show that if previous children were all girls, there were positive effects on the number of children. According to the estimates in Table 1.4, the coefficients of second children are significant at the 5% level, which implies that more positive effects are exerted on the number of children when the second child was a girl than when the second child was a boy. In both the LFP and WH models, a second-born girl increases the total number of children by approximately 0.08. In addition, the coefficient of interaction of the gender of the first two children also reveals a similar pattern to the aforementioned results, but its effect was much higher than that of the second child's gender. This finding indicates that women with two girls have a higher intention to give birth to an additional child than those who have a different number of girls. The effect of having only two girls on the number of children was nearly 0.52 in both models. Furthermore, the F-test of excluded instruments was significant at the 1% level (in the LFP and WH models, the F-statistic values were 100.30 and 98.93, respectively), which strongly supports the conclusion that if the first child, second child, or both the first and second child are female, the number of children is likely to increase.

The OLS and logistic regression results reveal a weak negative effect of children on female labor supply, and the 2SLS results show no effect on female labor supply in both models. The estimated coefficients of OLSs and the 2SLS for the LFP and WH models are reported in Tables 1.5 and 1.6, respectively. For the binary variable (LFP), logistic regression is also employed; the results are reported in Table 1.5. The first three columns of Tables 1.5 and 1.6 represent OLS estimates for different levels of controls. As shown in the first column, the simple regression model indicates that childbearing decreases a woman's probability of LFP by 14%, and reduces her WHs per week by 6 hours. However, after controlling for personal and spousal characteristics, the estimates in the second and third columns all indicate weak effects, with values of −0.02 for the probability of LFP and −1 for WHs per week. In logistic regression, a significant coefficient is also estimated. The marginal effect is –0.03, similar to the OLS outcome. However, our identification strategy, 2SLS, employs an instrument that causes exogenous variations and accounts for the potential relationship between the number of children and unobserved variables. Considering the estimates in the first stage, the effects of having two daughters (π_3 in Equation [1.7]) on the number of children (Table 1.4) are strong and overwhelm the effects of education level, age, and spousal characteristics. Therefore, these results illustrate that previous children's gender is an appropriate instrument in this study. The results of the 2SLS indicate that fertility has no impact on labor supply among Taiwanese childbearing women. This outcome partly contradicts previous studies that have applied similar instrumental variables. Chun and Oh (2002) used son preference for first-born children in South Korea as the instrument and identified a negative effect. Cruces and Galiani (2007) and Daouli, Demoussis, and Giannakopoulos (2009) applied the same parental gender preference approach introduced by Angrist and Evans (1998), although in these studies, gender composition appeared to be a weak instrument for low-fertility and low-employment countries, (e.g., Greece). These studies have produced similar results, indicating a negative relationship between fertility and female labor supply.

In addition to the effect of the main explanatory variable, the effects of other control variables show the expected effects on maternal labor supply. The coefficients of age and age squared are positive and negative, respectively; this illustrates the common phenomenon that in the early stage of life, labor supply increases with age by a diminishing rate. Regarding the effects of the education level of an individual, well-educated women tend to participate more in the labor market; consequently, they find superior job opportunities and have higher opportunity costs when they exit the labor market. Women who have completed higher education have a higher propensity to engage in the market compared with those who are senior high school graduates and others. On controlling for the current marital status and husbands' education, we find that these factors can be described as an exogenous effect of household income. Women with spouses participate less in the labor market, and families wherein the husbands have completed higher education tend to have higher income, which reduces the necessity of the wife to spend time on market activities. As our estimates suggest, in addition to childbearing, all the aforementioned factors determine female labor supply.

To determine the effect size, we estimate the standardized effects by employing the standard scores of continuous variables. Instead of using actual numbers, we replace the work hours per week, number of children, age, and age squared with the standard scores of these variables from the estimations. In particular, under the WH model, the effect of number of children can be expressed as follows: a change in the number of children by one standard deviation results in a change in the standardized work hours per week. The estimated coefficients are shown in Table 1.7 and are consistent with the results listed in Tables 1.5 and 1.6. In OLS and logistic regression, the number of children causes a reduction in female labor supply. One additional standardized number of children reduces the probability of labor force participation by 2%–3% (OLS coefficient = −0.018; marginal effect of logistic regression = −0.032) and changes the standardized work hours per week by approximately 0.04. Furthermore, the other coefficients of covariates had the expected sign. In brief, under both the WH and LFP models, the estimates present a weak effect of childbearing on maternal labor supply, and the results of 2SLS indicate that the number of children does not affect the decision to participate in the labor market.

Comparing the results of OLSs and 2SLS, the OLS results may overstate the effect of fertility on female labor supply. Based on the OLS results, there was a considerably small and negative effect of fertility on women's LFP and WHs. The birth of an additional child decreases the possibility of women's LFP by approximately 1.92%, and decreased women's WHs by 1 hour per week. These are slight effects but are consistent with the findings of Angrist and Evans (1998), which revealed a negative relationship between female labor supply and the number of children. Moreover, they revealed a smaller effect by using an instrumental variable approach, unlike the outcome of this study that no relationship is observed between fertility and female labor supply. Nevertheless, after applying the instruments, this study found that the OLS results may have exaggerated the impact of fertility on female labor supply in Taiwan.

Most previous empirical studies in this area that used an instrumental variable approach found smaller effects of children on female labor supply than the effects identified by OLSs. Such different results in OLSs and 2SLSs support the conclusion that the correlations of unobserved variables, number of children, and female labor supply generated the exaggerated estimates in OLSs. However, this study shows a slight effect in OLS but no statistical effect of childbearing on female labor supply in 2SLS, with results only partly consistent with those of previous research. This finding may be attributed to the Taiwanese family structure and relationships among family members in Taiwanese society. In the 2010 Taiwan Census, the proportion of stem families was 16.4%, whereas this proportion was 26.5% in the United States, 23.7% in Eastern and Southeastern Europe, and 50.2% in East and South Asia (Ruggles, 2010). In stem families, younger couples' children may be raised by grandparents, and living with parents may influence younger couples' childbearing decisions. If living arrangements or connections with family members are not revealed in surveys, which generates the correlation between the number of children and the error term in OLS estimations. Although this study controlled for the living situations of women (living with parents or in-laws), the relationships among family members were difficult to determine. In Taiwanese society, relationships among family members are crucial; couples support their parents, children, and siblings, although they do not necessarily all live together. For instance, when a couple is not available to take care of their children, they may ask their parents, siblings, in-laws, or other relatives for help. Therefore, a wide range of relatives could provide childbearing support to a couple, enabling women to reenter the labor market after a shorter length of time than they may have otherwise. It seems mothers with such unobserved childcare support have a comparatively higher incentive to stay in the labor force. Use of the instrumental variables approach can account for these unobserved connections among family and other omitted variables. These unobserved family characteristics may have induced the aforementioned significant effects in the OLS.

Finally, this article uses the number of children as a proxy for the cost of childbearing, yet there are various factors that could directly affect fertility decisions and labor supply, such as childcare availability, parental leave provisions, and household structure. Recently, some Asian studies have investigated these factors. Lee and Lee (2014) examined the relationships between childcare availability, female labor force participation, and fertility in Japan. Asai, Kambayashi, and Yamaguchi (2015) analyzed the relationships between childcare availability, household structure, and maternal employment, also in Japan. Chu, Kim, and Tsay (2014) suggested that living with in-laws and working was correlated with a delay in childbearing among married Taiwanese women. Whether childcare availability can encourage fertility or affect female labor supply is one key concern for future government policies, and the effect of informal childcare resulting from family structure on maternal employment is also a major topic in family economics. These dimensions are potential areas of interest for future research.

1.6 Conclusion

This study applies the instrumental variables proposed by Angrist and Evans to determine the effect of childbearing on female labor supply in Taiwan. American parents prefer to have both female and male children. Thus, couples who have only sons or daughters intend to give birth to an additional child in the hopes of having a child of a different gender, meaning that having same-gender children positively affects the number of children per family in the United States. Additionally, the gender composition of previous children could be a plausible instrument because it is randomly assigned. However, Taiwanese families prefer to have at least one son; hence, parents who only have daughters often intend to have an additional child. This chapter therefore employs the gender of first and second children as instrumental variables for the overall number of children per woman.

The results of the first stage of the 2SLS model show that the gender of the first two children significantly affect the overall number of children. If previous two children are all girls, the overall number of children increases by approximately 0.50. This result also supports the existence of a preference for sons and the use of instrumental variables.

The main part of this study examines the causal effect of childbearing on female labor supply. The OLS and logistic regression estimates indicate that childbearing leads to a slight decrease in female labor supply, in terms of both the possibilities of LFP and WHs per week. An additional child decreases the possibility of labor force participation and working hours by 1.7% and 1 hour, separately. However, after employing the instrument of the first two children's gender, the 2SLS estimates exhibit a different outcome from that of the OLSs: there is no statistical effect of childbearing on female labor supply. Finally, comparing the results of these two empirical models, the OLS outcomes appear to exaggerate the impact of childbearing on female labor supply in Taiwan.

Table 1.1 Descriptive Statistics

	Mean	Std. Dev.
Labor force participation	0.489	0.500
Work hours per week	20.166	24.754
Number of children	2.932	1.061
First child's gender (female=1)	0.493	0.500
Second child's gender (female=1)	0.479	0.500
Age	53.533	13.662
Junior high school and below	0.573	0.495
Senior high school	0.238	0.426
Higher education	0.190	0.392
Age of the second child	27.660	15.088
Live with parents or in laws	0.095	0.294
Marital status	0.782	0.413
Spouse’s employment status (employed=1)	0.496	0.500
Spouse’s education, junior high school and below	0.571	0.495
Spouse’s education, senior high school	0.220	0.415
Spouse’s education, higher education	0.209	0.406
Number of observations		3,509

Notes: The data are sourced from the Panel Study of Family Dynamics (PSFD) of Taiwan; this study uses data from 2009, 2010, and 2011. Female subjects having two or more children are chosen as observed subjects.

Table 1.2 Sex Composition of Previous Children

Sex composition	Number of observations	Fraction of sample	Number of having an additional child	Proportion with an additional child	Number of not having additional children	Proportion that not having a third child	*p*-value of Pearson's chi-squared test
Two boys	913	26.02%	481	52.68%	432	47.32%	0.000
First is boy, second is girl	866	24.68%	481	55.54%	385	44.46%	
First is girl, second is boy	914	26.05%	494	54.05%	420	45.95%	
Two girls	816	23.25%	553	67.77%	263	32.23%	
Total	3,509	100.00%	2,009	57.25%	1,500	42.75%	

Notes: "Number of having an additional child" represents the number of observed subjects with two children who at least had a third child; "Proportion with an additional child" represents the proportion of all observed subjects who had additional children; and "Number of not having additional children" and "Proportion that not having a third child" represent the number of observed subjects who had no additional children and the corresponding proportion among all observed subjects, respectively.

Table 1.3 Effect of Gender of Children on Labor Supply

	LFP model	WH model
First child is female	−0.030	−1.564
	(0.023)	(1.173)
Second child is female	0.017	0.190
	(0.024)	(1.175)
Both first and second child are female	0.000	0.775
	(0.034)	(1.676)
Constant	0.500 ***	20.666 ***
	(0.017)	(0.838)
Number of observations	3,509	3,493
Adj. R squared	0.000	0.000

Notes: *, **, and *** denote statistically significant values at the 10%, 5%, and 1% levels. Robust standard errors are reported in parentheses.

Table 1.4 Estimated Coefficients of the First Stage

	LFP model		WH model	
First child is female	0.041	(0.034)	0.041	(0.035)
Second child is female	0.081 **	(0.036)	0.079 **	(0.036)
Both first and second child are female	0.519 ***	(0.054)	0.517 ***	(0.054)
Age	−0.082 ***	(0.009)	−0.083 ***	(0.009)
Age squared	0.000 ***	(0.000)	0.000 ***	(0.000)
Senior high school	−0.236 ***	(0.038)	−0.233 ***	(0.038)
Higher education	−0.160 ***	(0.049)	−0.157 **	(0.049)
Age of second child	0.073 ***	(0.004)	0.074 ***	(0.004)
Live with parents or in laws	0.016	(0.041)	−0.016	(0.041)
Married	0.110 **	(0.046)	0.110 ***	(0.046)
Spouse is employed	0.004	(0.040)	0.007	(0.040)
Spouse's education, senior high school	−0.173 ***	(0.040)	−0.178 ***	(0.040)
Spouse attended higher education	−0.172 ***	(0.047)	−0.175 ***	(0.047)
Constant	4.062 ***	(0.213)	4.086 ***	(0.214)
Number of observations	3,509		3,493	
F-test of excluded instruments	100.30 ***		98.93 ***	
Adj. R squared	0.447		0.448	

Notes: *, **, and *** denote statistically significant values at the 10%, 5%, and 1% levels. Robust standard errors are reported in parentheses.

Table 1.5 Results of Labor Force Participation Model (LFP model)

	OLS	OLS	OLS	Logit	2SLS
Number of children	−0.138 ***	−0.015 *	−0.017 **	−0.030 **	0.019
	(0.007)	(0.008)	(0.008)	(0.013)	(0.028)
Age		0.038 ***	0.038 ***	0.063 ***	0.041 ***
		(0.005)	(0.005)	(0.007)	(0.005)
Age squared		−0.001 ***	−0.001 ***	−0.001 ***	−0.001 ***
		(0.000)	(0.000)	(0.000)	(0.000)
Senior high school		0.074 ***	0.110 ***	0.124 ***	0.119 ***
		(0.023)	(0.025)	(0.031)	(0.026)
Higher education		0.144 ***	0.217 ***	0.269 ***	0.223 ***
		(0.027)	(0.032)	(0.039)	(0.032)
Age of second child		0.004 *	0.004 *	0.007 ***	0.002
		(0.002)	(0.002)	(0.003)	(0.003)
Live with parents or in laws		−0.002	0.011	0.023	0.010
		(0.030)	(0.030)	(0.039)	(0.030)
Married			−0.094 ***	−0.123 ***	−0.098 ***
			(0.020)	(0.031)	(0.021)
Spouse is employed			0.167 ***	0.194 ***	0.167 ***
			(0.022)	(0.026)	(0.022)
Spouse's education, senior high			−0.084 ***	−0.109 ***	−0.076 ***
school			(0.022)	(0.020)	(0.023)
Spouse attended higher education			−0.133 ***	−0.177 ***	−0.128 ***
			(0.027)	(0.036)	(0.028)
Constant	0.894 ***	0.023	−0.038		0.199
	(0.023)	(0.132)	(0.134)		(0.178)
Number of observations	3,509	3,509	3,509	3,509	3,509
Adj. R squared	0.086	0.259	0.277		0.274

Notes: *, **, and *** denote statistically significant values at the 10%, 5%, and 1% levels. Robust standard errors are reported in parentheses. In column 4, the marginal effects of logit regression are reported, log likelihood and chi-squared are -1872.52 and 1117.86.

Table 1.6 Results of Work Hour Model (WH model)

	OLS	OLS	OLS	2SLS
Number of children	−6.277 ***	−0.838 **	−0.938 **	1.089
	(0.330)	(0.404)	(0.394)	(1.458)
Age		1.371 ***	1.376 ***	1.563 ***
		(0.250)	(0.252)	(0.278)
Age squared		−0.021 ***	−0.021 ***	−0.022 ***
		(0.002)	(0.002)	(0.002)
Senior high school		3.867 ***	5.414 ***	5.858 ***
		(1.274)	(1.358)	(1.401)
Higher education		5.977 ***	9.285 ***	9.639 ***
		(1.383)	(1.637)	(1.650)
Age of second child		0.221 **	0.266 **	0.118
		(0.111)	(0.110)	(0.149)
Live with parents or in laws		1.733	2.363	2.346
		(1.608)	(1.616)	(1.628)
Married			−4.203 ***	−4.434 **
			(0.995)	(1.025)
Spouse is employed			8.881 ***	8.834 ***
			(1.166)	(1.171)
Spouse's education, senior high school			−3.641	−3.217 ***
			(1.214)	(1.252)
Spouse attended higher education			−6.225 **	−5.886 ***
			(1.363)	(1.390)
Constant	38.574 ***	6.844	3.149	-5.924
	(1.105)	(6.854)	(6.952)	(9.151)
Number of observations	3,493	3,493	3,493	3,493
Adj. R squared	0.072	0.212	0.230	0.225

Notes: *, **, and *** denote statistically significant values at the 10%, 5%, and 1% levels. Robust standard errors are reported in parentheses.

Table 1.7 Results of Standardized Effects

	LFP model			WH model	
	OLS	Logit	2SLS	OLS	2SLS
Number of children	−0.018 **	−0.032 **	0.020	−0.040 **	0.043
	(0.008)	(0.014)	(0.030)	(0.017)	(0.062)
Age	0.520 ***	0.861 ***	0.565 ***	0.749 ***	0.852 ***
	(0.067)	(0.099)	(0.074)	(0.141)	(0.156)
Age squared	−0.745 ***	−1.187 ***	−0.768 ***	−1.179 ***	−1.237 ***
	(0.061)	(0.097)	(0.063)	(0.126)	(0.131)
Senior high school	0.110 ***	0.124 ***	0.119 ***	0.220 ***	0.237 ***
	(0.025)	(0.032)	(0.026)	(0.055)	(0.057)
Higher education	0.217 ***	0.269 ***	0.223 ***	0.376 ***	0.389 ***
	(0.032)	(0.039)	(0.032)	(0.067)	(0.067)
Age of second child	0.063 **	0.110 **	0.023	0.166 ***	0.079
	(0.033)	(0.044)	(0.043)	(0.068)	(0.091)
Live with parents or in laws	0.010	0.023	0.010	0.098	0.098
	(0.030)	(0.039)	(0.030)	(0.066)	(0.067)
Married	−0.094 ***	−0.123 ***	−0.098 ***	−0.164 ***	−0.173 ***
	(0.020)	(0.031)	(0.021)	(0.040)	(0.041)
Spouse is employed	0.167 ***	0.194 ***	0.167 ***	0.368 ***	0.367 ***
	(0.022)	(0.026)	(0.022)	(0.048)	(0.048)
Spouse's education, senior	−0.084 ***	−0.109 ***	−0.076 ***	−0.156 ***	−0.139 ***
high school	(0.023)	(0.030)	(0.023)	(0.049)	(0.051)
Spouse attended higher	−0.133 ***	−0.177 ***	−0.128 ***	−0.264 ***	−0.250 ***
education	(0.027)	(0.036)	(0.028)	(0.055)	(0.057)
Constant	0.458 ***		0.455 ***	−0.095 ***	−0.100 ***
	(0.019)		(0.019)	(0.041)	(0.041)
Number of observations	3,509	3,509	3,509	3,493	3,493
Adj. R squared	0.277		0.274	0.230	0.226

Notes: The work hours per week, number of children, age, and age squared are replaced by the standard scores in the estimations. *, **, and *** denote statistically significant values at the 10%, 5%, and 1% levels. Robust standard errors are reported in parentheses. In column 2, the marginal effects of logit regression are reported, log likelihood and chi-squared are -1872.52 and 1117.86.

Chapter 2

Number of Siblings and Educational Attainment: Application of Son Preference

2.1 Introduction

Family resources are distributed to and shared among members of a family. Therefore, when holding resources constant as the number of children in a family increases, the share of resources available for each member decreases. This is referred to as the "quantity-quality trade-off of children." Parents must consider this trade-off in light of the increased cost of having additional children and the resources they can devote to their children's education. For the past three decades, the expansion of education and the low fertility rate in Taiwan have resulted in a negative relationship between the quantity and quality of children, as partly evidenced by official data. In 1998, the fertility rate and average family size in Taiwan was 43‰ and 3.44, respectively. By 2013, these figures decreased significantly to 32‰ and 2.80, respectively. In the late 1990s, 21.17% of the population had higher education degrees, but this proportion increased to 40% by 2013.[8]

Studies have theoretically and empirically investigated the quantity-quality trade-off of children. The theoretical frameworks and most empirical evidence indicate an inverse relationship between family size and educational attainment. First introduced by Becker (1960), the theory of quality and quantity of children is recognized as the economic framework that denotes children as durable goods; additional theoretical studies (Becker & Lewis, 1973; Becker & Tomes, 1976) have

---

[8] Data from the Directorate General of Budget, Accounting and Statistics (DGBAS), Executive Yuan, Taiwan.

adopted this framework. In addition, numerous empirical studies have tested the relationship between the quality and quantity of children (e.g. Blake, 1981; Booth & Kee, 2009; Downey, 1995).

In this research, the likelihood of the omitted variables bias must be accounted for. This bias occurs frequently when estimating the effects of sibling size on educational attainment. If parents can predetermine the number of children they wish to raise and their children’s academic attainment, both the quantity of children and the quality of their education can be jointly decided by the parents. Which means if parents make their fertility decisions and investments in children by considering exogenous factors, such as their parental preferences and financial constraints, examining the causal relation between the quantity and quality of children becomes impossible.

To address the identification problem of omitted variables, Angrist and Krueger (2001) proposed the instrumental variable approach. A typical example of instrumental variables is the application of the occurrences of twins to predict the number of children. Bearing twins is denoted as a surprise event that exogenously influences the number of children. The occurrence of twins has been employed for generating exogenous variations in family size (Black, Devereux, & Salvanes 2005; Li, Zhang, & Zhu 2008; Rosenzweig & Wolpin, 1980).

The sex composition of the siblings is another appropriate instrument for predicting the number of children wherein the sex of the children are randomly assigned and correlated with family size. Angrist and Evans (1998) use sex compositions of siblings as instruments to generate exogenous variations in the number of children. Americans prefer to have children of different sexes; families with only sons (or only daughters) prefer a daughter (or a son). Therefore, a positive correlation between previous same sex children and family size can be reasonably assumed.

The objective of this article is to investigate the causal relation between sibling size and academic attainment. The Taiwanese parental preference for a male child is employed as the instrumental variable to generate exogenous variation in sibling size. Tsay and Chu (2005) found that if a woman bears a female child in her first two births, the family is highly likely to go for a third child. In addition, Taiwanese government data partly evidence the aforementioned parental preference for males. In 2014, the male:female ratios of first- and second-born children in Taiwan were 107:100 and 106.6:100, respectively, meaning that male infants outnumber female infants by nearly 7%; nevertheless, the corresponding ratio for third-born children was 112.2:100.[9]

In this chapter, the results of a two-stage least squares (2SLS) analysis indicate that in Taiwan, no correlation exists between the number of siblings and educational attainment; this finding is consistent with previous works that use similar instrumental variables (Angrist, Lavy, & Schlosser 2010; Haan 2010). The ordinary least squares (OLS) estimates reveal a negative relationship between family size and scholastic performance that a larger sibling size reduces the number of years in school, the possibility of senior high school attendance, and the possibility of four-year college attendance. However, after generating exogenous variations in family size, a positive but nonsignificant effect of family size is identified, suggesting that unobserved family characteristics may be simultaneously correlated with the educational attainment and sibling size.

The remainder of the chapter is structured as follows: Section 2 describes the data used in the research and the summary statistics; Section 3 presents the empirical methodology and the results; and Section 4 reports the conclusion.

[9] Data from the Ministry of the Interior, Executive Yuan, Taiwan.

2.2 Data and Descriptive Statistics

This study uses "Social Stratification" data from the Taiwan Social Change Survey (TSCS), conducted by the Academia Sinica (Taiwanese Academic Institution) of Taiwan, to investigate the effect of sibling size on educational attainment. This series of surveys was held in 1997 (sample size 2,596 Taiwanese adults), 2002 (with sample size 1,992), and 2007 (with sample size 2,040) and focused on various topics to track long-term social changes in Taiwan.[10] The data set contains variables that describe the personal characteristics, educational attainment, ethnicity, and complete record of siblings and therefore provides adequate information for applying the instruments of this research.

All survey participants are Taiwanese adults, meaning that Taiwan's societal preference for a male child may be represented in the survey data, thus enabling this study to apply this parental preference as an instrumental variable. The data are partly representative of the actual structure of Taiwanese society: in the data set, 43%, 28%, and 28%, of the participants complete junior high school or less, senior high school, and higher education, respectively, which are close to the actual average percentages of the country (40.16%, 33.16%, and 26.68%, respectively) for 1998, 2002, and 2007.[11] The first two children's sexes and their interactions are applied as the instrumental variables in this study, thus we eliminate individuals without siblings from the analysis. Finally, data of 2,386 respondents from the 1997 survey, 1,761 from the 2002 survey, and 1,862 from the 2007 survey (total 6,009 respondents) were

[10] The "Social Stratification" survey was held in 1992, 1997, 2002, 2007, and 2012; however, this study does not use the 1992 and 2012 data because data on the number of siblings is unavailable for these years.

[11] Data from the Directorate General of Budget, Accounting and Statistics (DGBAS), Executive Yuan, Taiwan. Because data for 1997 was unavailable, data from 1998 was used to compute the averages. In Taiwan, 45.85%, 32.98%, and 21.17%, of the citizens complete junior high school or less, senior high school, and higher education in 1998; 40.19%, 33.82%, and 25.99%, of the citizens complete junior high school or less, senior high school, and higher education in 2002; 34.44%, 32.69%, and 32.87%, of the citizens complete junior high school or less, senior high school, and higher education in 2007.

included in the estimations.

The variables used in this empirical study are listed in Table 2.1, namely educational attainment (years of schooling, senior high school attendance, and four-year college attendance), sibling size, and family characteristics. Years of schooling is defined as 6 years for primary school, 9 for junior high school, 12 for senior high school and vocational school, 14 for junior college, 16 for university and college, 18 for masters, and 22 years for Ph.D. degrees. Sibling size is the total number of children in a family. Because family characteristics strongly influence educational attainment, this study also considers control for parental characteristics, such as the level of education and work status, when the parents were young. Parents' education attainment is categorized as junior high school and below, senior high school, and higher education. The majority of parents were found to have only attended an education level of junior high school or lower (83% for fathers and 93% for mothers). In addition, access to education has increased dramatically in the last three decades in Taiwan. Therefore, educational opportunities differ by generation. The year of birth of the individuals is recorded in the data set, thus enabling age-based identification and classification of individuals (e.g., born in the 1940s or earlier, 1950s, 1960s, 1970s, and 1980s). The descriptive statistics are presented in Table 2.1.

Table 2.2 presents the educational attainment and birth cohort by sibling size. Table 2.1 has shown the average years of education is 10.37, at least 57% of observations attended senior high school, and 15 % of them attended four-year college, but Table 2.2 can express detailed patterns of sibling size and education. When individuals face a larger number of siblings, they also tend to complete lower education. Individuals who have two siblings are with 13 average years of schooling and around 55% of them attended higher education. By contrast, who have nine or more than nine siblings are with 7 average years of schooling and only 10% of them went to higher education. However, the different educational attainments among individuals with different sibling sizes may be induced by generations, people born in older generations have lager family size and less opportunities of education, thus here the negative relation between education and sibling size dose not present any causal evidence. We leave the causal estimations in the next section.

2.3 Empirical Methodology and Results

The identification strategy is employing 2SLS to investigate the effect of number of siblings on academic attainment. The main explanatory variable is the number of siblings, which may also be a possible endogenous variable. Thus, we use previous siblings' sexes as instruments to generate exogenous variations in the number of siblings. The first- and second-stage equations are shown as Equations (2.1) and (2.2), respectively.

$$Sib_i = \pi_1 d_{1i} + \pi_2 d_{2i} + \pi_3 d_{1i} d_{2i} + \theta X_i + \upsilon_i \tag{2.1}$$

$$Edu_i = \beta Sib_i + \gamma X_i + u_i \tag{2.2}$$

Where Sib_i is the number of siblings in individual i's family, and d_{1i} and d_{2i} are dummy variables to denote the sexes of the first and second sibling members (1 for female and 0 for male). In the first stage, the sibling size is regressed using the instrumental variables. Individuals with at least one sibling are selected as our sample for applying the approach, and the first two children in a family and their sexes are applied as the instruments. Furthermore, Equation (2.1) includes the interaction between the first and second sibling's sexes. If both the first and second child are female, $d_{1i}d_{2i}$ equals 1. X_i is the vector of other independent variables, including a constant, and personal and family characteristics, and θ is the corresponding vector of coefficients. υ_i is the error term; this error term is assumed to be uncorrelated with the variables shown in Equation (2.1).

Equation (2.2) estimates the effect of sibling size on educational attainment. Here, Edu_i is the educational outcome of individual i, and β is its coefficient (this coefficient is the main effect this study intends to estimate). We use three indicators for educational attainment, years of schooling, senior high school attendance, and four-year college attendance. Senior high school attendance and four-year college attendance are expressed by dummy variables, Edu_i equals 1 if individuals are in the corresponding status, equals 0 otherwise. Sibling size is considered as a possible endogenous variable, which means unobserved family characteristics can cause the error term, u_i, to be possibly correlated to the number of siblings.

First, we present the OLS estimates in Table 2.3. If the estimation does not account for the endogeneity, the coefficient in the first column of table 2.3 presents a negative and significant effect of sibling size on educational attainment. The results show an additional child in the family reduces the years of schooling by 0.86 years, lowers the possibilities of senior high school attendance and four-year college attendance by 9% and 4% respectively. However, these results are generated by simple regressions, several unobserved factors may bias the estimates.

Therefore, we include more control variables in the estimations, such as personal and parental characteristics, geographic regions, and ethnic backgrounds. Regarding personal characteristics, we generate dummy variables that denote sex and whether the individual is the eldest son. The preference for a male child in Taiwan eventually results in some parents allocating more resources to their sons than their daughters; this discrimination leads to a difference in educational outcomes between male and female children. In addition, the birth order affects education, unfortunately a high correlation between the birth order and the number of children may disturb that using birth order as a control variable. Therefore, rather than directly use the birth order, we use a dummy variable pertaining to whether the first-born child is a male child for the analysis. Furthermore, parental characteristics, such as parental educational attainment and work status during their youth, substantially affect children's education. Parents who have attained higher education levels have a higher likelihood of earning more and thus invest more in their children's education. Thus, the controls include four dummy variables that categorize parents' academic attainments as senior high school and higher education, and parents attended junior high school and lower education as the reference groups. A dummy variable for the father's work status is also used; the survey recorded whether the father was employed when individual was 15; this variable is used for controlling for childhood resources. In addition, access to higher education has substantially increased in Taiwan over the past three decades; therefore, the analysis is also controlled for the generation (i.e., decade) that the individuals were born into.

The OLS estimates partly point out the evidence that the regular regressions may overstate the effect of sibling size on educational attainments because of ignoring personal and family backgrounds. As we adding more controls in the estimations, the effects of sibling size become weaker, which implies that unobserved factors are correlated with sibling size and may disturb the causal estimations. Nevertheless, after controlling for additional variables in the OLS estimation, the effects of sibling size on educational outcomes remain negative and significant. Column 2 of Table 2.3 lists the results after controlling for the gender, whether the individual is the eldest son, parental educational attainment, and the birth cohort; in addition to column 2, column 3 lists the coefficients after controlling for ethnic dummies; and in addition to column 3, column 4 lists the estimates after controlling for regional dummies. For years of schooling, all results show nearly the same effect of sibling size on years of schooling that an additional child in the family reduces the years of schooling by a quarter of year, by 0.26 without controlling for the ethnic and regional variables, by 0.24 without controlling for the regional variable, and by 0.23 when controlling for all variables. Similarly, for the possibilities of attending senior high school and four-year college, the estimates suggest that controlling for more personal and family characteristics the effects of sibling size become weaker.

The estimates in the first stage support the preference for a male child in Taiwan. The results of the first stage in the 2SLS analysis are shown in the last column of Table 2.4; the first stage reveals that the preference for a male child does increase sibling size; furthermore, the *F*-test of excluded instrument*s* is significant at the 1% level, which strongly supports that if the first child, the second child, or both the first and second child are female, the sibling size is likely to increase. The first child being female increases the number of children in the family by 0.26, the second one increases it by 0.17, and when both the first and second child are females, the number of children in the family increases by 0.23.

After addressing the endogeneity of sibling size, this study finds that sibling size does not significantly affect educational attainment. The first, second and third column of Table 2.4 list the estimates of second stage under full control.[12] These estimates indicate that the effect of sibling size does not significantly affect any outcome of educational attainment, none of them can express statistical effect on years of schooling, senior high school attendance, or four-year college attendance.[13] These results are consistent with previous studies that have adopted a similar approach.[14]

[12] Full control includes gender dummy (female=1), eldest son dummy, parents' educational attainment, fathers' employment status when individual was 15, birth cohorts, ethnic dummies and regional dummies.

[13] The 2SLS estimate is estimated for individuals whose number of siblings has been affected by the first two siblings' sexes. Imbens and Angrist (1994) reported that the estimate can be interpreted as a local average treatment effect.

[14] Angrist, Lavy, and Schlosser (2010) applied sex composition of children and twin birth as instrumental variables, and illustrated that no evidence of a quantity-quality trade-off of children. Haan (2010) uses same sex of children and twin birth as instrumental variables, and found that no significant effect of family size on educational attainment.

The results of the 2SLS approach provide no evidence for a relationship between sibling size and educational attainment; this is inconsistent with the results in OLS; this discrepancy may be attributable to unobserved family characteristics. For instance, parents' socioeconomic status may be correlated with number of children and children's academic performance. Parents who have well socioeconomic status may be with lower numbers of children, but are available to provide more resources for children's education, which leads downward biases in OLS. Furthermore, the linkages among Taiwanese family members, for example, interaction among children, are difficult to observe: older siblings may be responsible for caring for their younger siblings, which may negatively affect older siblings' time on study and positively or negatively affect the academic performance of the younger siblings. The other possible unobserved variable is the parents' judgment of their child's academic potential. Parents may provide educational support to the child they believe to be better at studies and thus has more potential to succeed. However, these explanations are hard to evidence through conventional surveys; consequently, the regular least squares approach may entail the unobserved variable problem.

2.4 Conclusion

This chapter examines the effect of sibling size on educational attainment in Taiwan. The preference for a male child in Taiwan justified the use of the sexes of the first two siblings as instrumental variables in this study. Most Taiwanese parents prefer a male child and desire to have at least one male child in the family; those with daughters desire to have a male child in the family. In the first stage of the 2SLS analysis, the estimates show that the first two children being female positively affects the number of siblings. In addition, the sex of siblings is almost randomly assigned; thus, the use of instruments for the number of siblings is plausible. The OLS estimates indicate a significant and negative effect of sibling size on multiple outcomes of educational attainment. However, the results of 2SLS demonstrate no significant effect of family size on educational attainment. These empirical outcomes are similar to those of previous studies that have used similar instruments.

Table 2.1 Descriptive Statistics

Variables	Mean	Std. Dev.	Variables	Mean	Std. Dev.
First sibling is female (female = 1)	0.503	0.500	Father’s education, senior high school	0.100	0.300
Second sibling is female (female = 1)	0.506	0.500	Father’s education, higher education	0.066	0.248
Both first and second siblings are female	0.259	0.438	Mother’s education, junior high school or below	0.926	0.261
Years of schooling	10.367	4.478	Mother’s education, senior high school	0.054	0.228
Junior high school and below	0.431	0.495	Mother’s education, higher education	0.019	0.136
Senior high school	0.284	0.451	Father was employed when individual was 15	0.897	0.304
Higher education	0.285	0.451	Born in the 1940s or earlier	0.248	0.432
Four-year college	0.150	0.357	Born in the 1950s	0.245	0.430
Sibling size	5.058	2.086	Born in the 1960s	0.252	0.434
Female (female = 1)	0.506	0.500	Born in the 1970s	0.174	0.379
Eldest son (eldest son = 1)	0.217	0.412	Born in the 1980s	0.081	0.258
Father’s education, junior high school or below	0.834	0.372			
Number of observations	6,009				

Notes: Data are from the group of Social Stratification of Taiwan Social Change Survey (TSCS) from 1997, 2002, and 2007. Individuals with one or more siblings are involved in the analysis.

Table 2.2 Educational Attainment and Birth Cohort by Sibling Size

	Two siblings	Three siblings	Four siblings	Five siblings
Years of schooling	13.063	12.686	11.374	10.034
	(3.839)	(3.575)	(3.801)	(4.175)
	Share	Share	Share	Share
Junior high school and below	0.166	0.180	0.327	0.470
Senior high school	0.281	0.332	0.344	0.314
Higher education	0.553	0.487	0.329	0.216
Born in the 1940s or earlier	0.131	0.095	0.128	0.219
Born in the 1950s	0.053	0.115	0.213	0.326
Born in the 1960s	0.106	0.252	0.352	0.310
Born in the 1970s	0.275	0.340	0.248	0.129
Born in the 1980s	0.434	0.198	0.060	0.015
Observations	320	1,231	1,260	1,044

	Six siblings	Seven siblings	Eight siblings	Nine or more siblings
Years of schooling	9.021	8.136	7.565	7.265
	(4.598)	(4.290)	(4.261)	(4.440)
	Share	Share	Share	Share
Junior high school and below	0.559	0.679	0.726	0.754
Senior high school	0.259	0.194	0.183	0.147
Higher education	0.182	0.126	0.091	0.098
Born in the 1940s or earlier	0.368	0.408	0.516	0.516
Born in the 1950s	0.317	0.355	0.323	0.323
Born in the 1960s	0.236	0.185	0.148	0.148
Born in the 1970s	0.065	0.049	0.008	0.008
Born in the 1980s	0.013	0.002	0.005	0.005
Observations	829	546	372	407

Notes: Standard deviations are reported in parentheses. The number of siblings includes individual and his/her sibling members. Share is the fraction of observations in the given group.

Table 2.3 Estimated Coefficients of OLS

Variable	Years of Schooling			
Number of siblings	−0.864 ***	−0.258 ***	−0.242 ***	−0.234 ***
	(0.029)	(0.027)	(0.026)	(0.025)
Controls	No	Yes	Yes	Yes
Cohort dummies	No	Yes	Yes	Yes
Ethnic dummies	No	No	Yes	Yes
Regional dummies	No	No	No	Yes
Adj. R squared	0.162	0.444	0.458	0.487
Variable	Attended senior high school			
Number of siblings	−0.091 ***	−0.072 ***	−0.030 ***	−0.028 ***
	(0.003)	(0.003)	(0.003)	(0.003)
Controls	No	Yes	Yes	Yes
Cohort dummies	No	Yes	Yes	Yes
Ethnic dummies	No	No	Yes	Yes
Regional dummies	No	No	No	Yes
Adj. R squared	0.146	0.212	0.353	0.367
Variable	Attended four-year college			
Number of siblings	−0.040 ***	−0.021 ***	−0.012 ***	−0.011 ***
	(0.002)	(0.002)	(0.002)	(0.002)
Controls	No	Yes	Yes	Yes
Cohort dummies	No	Yes	Yes	Yes
Ethnic dummies	No	No	Yes	Yes
Regional dummies	No	No	No	Yes
Adj. R squared	0.055	0.186	0.202	0.208
Observations	6,009	6,009	6,009	6,009

Notes: Controls include the gender dummy (female=1), eldest son dummy, parents' educational attainment, and fathers' employment status when individual was 15. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. Robust standard errors are reported in parentheses.

Table 2.4 Results of 2SLS

Variable	Second stage Years of schooling	Second stage Attended senior high school	Second stage Attended four-year college	First stage Number of siblings
Number of Siblings	0.192 (0.185)	0.027 (0.022)	0.016 (0.017)	
First sibling is female				0.261 *** (0.065)
Second sibling is female				0.170 *** (0.065)
First two siblings are female				0.230 *** (0.090)
Female (female = 1)	−1.109 *** (0.105)	−0.059 *** (0.013)	−0.028 *** (0.010)	−0.284 *** (0.061)
Eldest son (eldest son = 1)	0.591 *** (0.214)	0.090 *** (0.026)	0.033 (0.021)	−1.023 *** (0.061)
Father's education, senior high school	2.129 *** (0.184)	0.219 *** (0.021)	0.178 *** (0.025)	−0.609 *** (0.066)
Father attended higher education	2.761 *** (0.214)	0.248 *** (0.023)	0.266 *** (0.032)	−0.620 *** (0.091)
Mother's education, senior high school	0.515 *** (0.186)	0.007 (0.020)	0.137 *** (0.032)	−0.233 *** (0.091)
Mother attended higher education	0.981 *** (0.357)	-0.002 (0.032)	0.263 *** (0.052)	−0.451 *** (0.144)
Father was employed when individual was 15	0.988 *** (0.156)	0.073 *** (0.018)	0.038 *** (0.012)	0.210 *** (0.086)
Constant	2.511 (1.754)	-0.138 (0.211)	-0.139 (0.171)	6.125 *** (0.571)
Cohort dummies	Yes	Yes	Yes	Yes
Ethnic dummies	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes
Observations	6,009	6,009	6,009	6,009
F-test of excluded instruments				35.20 ***
Adj. R squared	0.459	0.352	0.207	0.306

Notes: *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. Robust standard errors are reported in parentheses.

Chapter 3

Education and Health Outcomes:

Using Taiwanese Education Reform for Identification

3.1 Introduction

Previous studies have thoroughly investigated the relationship between education and health, with substantial evidence suggesting that individuals' educational attainment is correlated with their health status and health-related behaviors.[15] Specifically, most studies have indicated that well-educated people tend to be healthy and engage in more healthy behaviors.[16] The positive effect of education on health outcomes is meaningful not only at an individual level but also at the local and national levels for policy makers to implement long-term policies. A large proportion of the national budget has already been allocated for health care; the positive relationship between education and health suggests that investing similarly in education can improve citizens' educational attainment and reduce healthcare costs in the future. Furthermore, a nation with healthy citizens and a lower mortality rate with higher human capital, which is crucial for national productivity (Arora, 2001; Bloom, Canning & Sevilla, 2004).

Few patterns through which education leads to better health have been suggested, including the receipt of adequate knowledge and collection of information regarding healthy practices, improved social status, and parental and familial characteristics. However, as Michael (1973) notes, education improves not only labor market

---

[15] These studies have used several indicators for health. "Health status" has been alternately defined as self-reported health status, long-term illness, mortality, and work disability, whereas "health-related behaviors" have included body mass index (BMI), amount of exercise, and smoking of cigarettes or consumption of wine.

[16] For an overview of the studies on health and education, see Eide and Showalter (2001).

outcomes but also nonmarket activities; for example, the managerial capacity developed through schooling improves time management and source allocation skills. This is a potential pattern of causality that education leads to better health. A direct explanation is provided by the theoretical framework of Grossman (1972) who predicts that well-educated individuals can benefit from improved efficiency of allocating health inputs, which enhances the marginal productivity of health inputs and leads to higher health productivity, i.e., better health.

The potential identification problem is that unobserved factors may affect education, health status and health-related behaviors simultaneously, which skew the results in regular regression analyses. Because factors such as childhood health status, opportunity costs, time preference, and internal abilities are not commonly evaluated in surveys, their impact on education and health remains unknown. For example, well-educated individuals may have been endowed with superior childhood health, which enables them to complete more years of education and be with better health outcomes in adulthood. Well-educated individuals also encounter high opportunity costs owing to a high future income, therefore they tend to be on healthy behaviors. Moreover, individuals who value future days more than present days, they may prefer to invest more in human capital through pursuing higher education and maintaining better health. These unobserved variables may be positively correlated with both educational attainment and superior health, suggesting that the positive effect of education is overstated in a regular regression analysis. Thus, to accurately determine the causal relationship between education and health, the endogeneity of education cannot be ignored.

One approach to address the identification problem is the use of instrumental variables, which can generate exogenous variations in education. Studies have proposed few instrumental variables that facilitate the investigation of the causal relationship between education and health. For example, the common estimation strategy considers changes in compulsory education laws, which are decided upon by governments and usually increase the number of years that people must be in school. If compulsory education laws do not correlate with citizens' health outcomes, governments exert an exogenous shock on citizens' educational attainment. Changes in compulsory schooling laws have been widely applied in Western countries, including the United States (Lleras-Muney, 2005; Mazumder, 2008), the United Kingdom (Silles, 2009; Braakmann, 2011), France (Albouy & Lequien, 2008), Denmark (Arendt, 2005, 2008), and Germany (Kemptner, Jurges & Reinhold, 2011). Another strategy is applying the changes in education policies, such as the abolition of secondary school fees in Germany (Reinhold, 2009) and the marked expansion of high school education in South Korea in the mid-1970s (Park & Kang, 2008). These changes can form the instruments for causality estimations. Other examples of instrumental variables, such as the quarter of birth and the Vietnam war draft, have been used in studies conducted in the United States.[17]

In Taiwan, education rapidly expanded between the mid-1990s and mid-2000s. The government implemented several education reforms during this period, which facilitated the considerable growth of senior high schools and four-year colleges and universities. During this period, there was a 50% increase in the number of senior high schools and a nearly 140% increase in the number of four-year colleges and universities.[18] This policy encourages more junior high school graduates to attend senior high school and subsequently attend to a four-year college and has transformed Taiwan's general education system into a more academic-centered system. Because of

[17] Adams (2002) applies the quarter of birth as an instrumental variable for education, and Grimard and Parent (2007) and de Walque (2007) use the Vietnam war draft as an instrumental variable for education and college attendance.

[18] More detailed information is presented in Section 2 of this chapter.

its exogenous shock on the education system, we use the rapid increase in secondary and post-secondary education institutions as an identification strategy.

This chapter investigates the causal relationship between education and health status and health-related behaviors (i.e., body mass index [BMI] score, overweightness, and obesity). To address the endogeneity problem of education, we conduct a conventional two-stage least squares (2SLS) regression analysis by using Taiwan's education reform as the instrumental variable. In addition, we use data from the Panel Study of Family Dynamic (PSFD) for empirical estimations. After controlling for the endogeneity of education, we determine that education negatively affects self-reported health status but positively affects objective health outcomes (i.e., no work disability). Furthermore, no causal relationship is noted between education and BMI, and outcome variables defined by BMI values.

The remainder of this chapter is structured as follows. Section 2 presents the educational background and details of education reforms that were implemented in the mid-1990s in Taiwan. Section 3 describes the data used in the estimations and descriptive statistics. Section 4 outlines the methodology underpinning this research. Section 5 discusses the empirical results and research limitations. Section 6 concludes the chapter.

3.2 Education Background and Reforms in Taiwan

The start date of the academic year in Taiwan is the first of September, and every year, children who turn 6 years old before that day begin attending school. Every citizen is required to complete 9 years of education including 6 years in primary school and 3 years in junior high school.[19] After the nine-year compulsory schooling, students

[19] As of September 2014, the Taiwanese government has replaced the nine-year compulsory education system with a twelve-year compulsory education system.

have three options to achieve higher education level: (1) attend a senior high school, (2) attend a vocational school, or (3) attend a five-year junior college. Most students who plan to enter a four-year college enroll in a senior high school because senior high schools are in the rather academic-focused system. Taiwanese government provides few opportunities for students to attend college, such as recommendations from students' schools, and completing the college entrance examination. However, all students who have completed 12 years of school, regardless of whether they attended the academic or vocational school, are eligible to enter college.

The Taiwanese government launched a series of educational reforms in the 1990s including the extensive construction of more senior high schools and four-year colleges and universities. Between the mid-1990s and the early 2000s, the number of these types of schools and the students attending them increased considerably. The overall changes from 1990 to 2012 are depicted in Figure 3.1.[20] Notably, the number of senior high schools and four-year colleges and universities in Taiwan in 1990 was only 170 and 46, respectively; this number increased to 277 and 127, respectively, by 2000 (an increase of over 100%). A similar pattern can also be observed in the number of students attending these schools; there were nearly 210,000 students in senior high school and 240,000 students in undergraduate programs in 1990. However, these numbers increased to approximately 430,000 and 440,000 by 2000. The educational reforms have effectively encouraged citizens' education attainment over the past two decades, and the proportion of people with higher education degrees has increased from 21% in 1998 to 42% in 2015.[21]

[20] Source: Ministry of Education, Taiwan.
[21] Source: Directorate General of Budget, Accounting and Statistics (DGBAS) of Executive Yuan, Taiwan. The available data extends from 1998 to 2015; however, we are unable to show all of the changes during this period (1990– 2010) in Figure 3.1.

3.3 Data and Descriptive Statistics

This study uses multiple data sources including the PSFD for education and personal characteristics, the Ministry of Education of Taiwan for data on the number of schools by regions, and the Ministry of the Interior of Taiwan for data on the population by age. We collect 5 years of data from the PSFD (2009, 2010, 2011, 2012, and 2014). Academic Sinica, a Taiwanese research institute, conducted the survey, and all the participants are Taiwanese adults. A questionnaire that collects participants' birth date and residence information is provided to the participants, which allows us to connect the number of schools with the individuals in a specific school year by the region. Specifically, we use the number of senior high schools and four-year colleges and universities in each region to determine the extent of the educational expansion in the 1990s. Rather than using the actual numbers of schools, this study determines the adjusted number of schools according to the specific age of the population in a region. Therefore, because Taiwanese students usually start their senior high school education at the age of 15 years and their college education at the age of 18 years, the number of people who are aged 15 and 18 years in a region is used to determine the adjusted number of senior high schools and four-year colleges and universities, respectively. Data about the regional number of schools are available from 1990, and the data about single-age populations are available from 1991; thus, we eliminate individuals who were born before September 1976 from this study. In total, 11,323 participants are examined in this study, with age ranging from 26 to 38 years.

Taiwan's education expansion began in the mid-1990s; therefore, individuals who were born before the academic year of 1980 received a substantially different education than that those born after 1980. Table 3.1 provides a summary of the statistics according to the academic year. First, we divide our sample into two groups: individuals who turned 15 before September 1995, and individuals who turned 15 after September 1995. The number of senior high schools is determined by the actual number of schools per 1,000 people aged 15 years in a region, and the number of four-year colleges and universities is determined by the number of colleges and universities per 1,000 people aged 18 years in a region. Overall, we observe that both types of schools increased after 1995. Specifically, 0.49 senior high schools per 1,000 people aged 15 years increased to 0.63, whereas 0.20 four-year colleges and universities per 1,000 people aged 18 years increased to 0.35. The mean years of schooling of the participants who attended senior high school before the expansion and after the expansion was 14.34 and 14.86, respectively. Moreover, a high proportion of students who attended four-year colleges or universities after 1995 (32% vs. 47%) indicates that the reform increased the level of education received by individuals.

We use two indicators to express health status (self-reported health status and work disability) and three indicators to express BMI-based health behaviors (actual BMI value, overweightness, and obesity). Two questions from the PSFD are used to collect information about the participants' assessments of their health status: "*How is your health status in general?*" (participants select one of *very poor*, *poor*, *fair*, *good*, or *great)* and "*Are you not working because of your health condition?*" (participants select *yes* or *no*). The PSFD also contains information about body height and weight; thus, BMI is used as an indicator of health-related behaviors. An individual's BMI is determined by dividing a person's body weight (kg) by the square of their body height (m); BMI is also employed as the standard of obesity. In Taiwan, the Ministry of Health and Welfare has divided BMI scores into four categories: underweight (BMI $<$ 18.5), healthy weight ($18.5 \leq$ BMI < 24), overweight ($24 \leq$ BMI < 27), and obese ($27 \leq$ BMI).

Table 3.2 lists the participants' health statuses and BMI scores according to whether they attended a four-year college or university. Self-reported health status is numerically defined on a scale ranging from one (very poor) to five (great). Overall, our results show that four-year college or university graduates reported a lower health status (3.604) than did noncollege graduates (3.615), which may be because well-educated people are more concerned about their health conditions and therefore tend to refrain from reporting strong subjective health statuses. For example, 11% of four-year college graduates reported "great" health, whereas 15% of non- four-year college graduates reported "great" health. By contrast, when we focus on objective health outcomes, college graduates are less likely to be in work disability compared with their noncollege graduate counterparts. In addition, compared with noncollege graduates who are more likely to be obese, college graduates are more likely to maintain a healthy BMI score.

3.4 Empirical Methodology and Results

A conventional 2SLS regression analysis is used for the estimations to investigate the causal relationship between education and health. The empirical model is set as follows:

$$H_{ij} = \alpha_1 Edu_{ij} + \alpha_3 X_{ij} + \gamma_j + \varepsilon_{ij} \tag{3.1}$$

$$Edu_{ij} = \beta_1 SHS_{ij} + \beta_2 College_{ij} + \beta_3 X_{ij} + \rho_j + \upsilon_{ij} \tag{3.2}$$

where H_{ij} is the health outcome of an individual i in a region j; Edu_{ij} is individual i's education level; and X_{ij} is the vector of other covariates, such as age, age squared, gender, and marital status. Living environment is also a determined factor of health (e.g., urban vs. rural residence and the effect of different local policies); therefore, we use γ_i to determine the regional fixed effect and ε_{ij} as the error term (which may be correlated with education through unobserved factors). Notably, if only ordinary least squares (OLS) is used to estimate Equation (3.1), the approach may yield biased estimates. Hence, we conduct Equation (3.2) using two instrumental variables, namely SHS_{ij} and $College_{ij}$, to address the identification problem. SHS_{ij} is the number of senior high schools, and $College_{ij}$ is the number of four-year colleges and universities, per individual i in a relevant academic year in a region j. Both SHS_{ij} and $College_{ij}$ are adjusted according to the population of 15- and 18-year-olds, respectively. We expect that a higher number of both types of schools positively affects education.

Three indicators are used to determine health outcomes, namely self-reported health status, work disability, and BMI. Notably, the values of self-reported health status do not increase or decrease in equal increments; for instance, the difference between “bad health” and “fair health” or “fair health” and “good health” may be determined by one numeral on our scale; however, the meaning of these differences undoubtedly varies. Thus, rather than directly using the value of self-reported health status, we create two dummy variables for this indicator: “reported good and great health” and “reported great health.” The second indicator of health outcomes, work disability, is assigned a value of 1 if the individual is not working because of health conditions; otherwise, it is 0. The final indicator is the BMI score; in addition to the direct use of BMI values, we create three dummy variables: “healthy BMI score” ($18.5 \leq \text{BMI} < 24$), “overweight and obese BMI score” ($24 \leq \text{BMI}$), and “obese BMI score” ($27 \leq \text{BMI}$).

The first stage estimates support our hypothesis that the education reform had a positive effect on education in Taiwan in terms of both years of education and the possibility of attendance at a four-year college or university. We employ two variables for education. The first is years of education, which can represent the health returns to education directly, the estimated coefficients can describe that an additional year of schooling cause changes in health outcomes. Because people accumulate more knowledge and abilities during college and university than during high school, attendance at a four-year college or university may affect health status and health behaviors. Furthermore, compared with noncollege graduates, college graduates have more job opportunities and a higher social status; therefore, the second variable is a dummy variable of four-year college attendance. An outline of the first stage estimates is listed in Table 3.3. Because the 2009 version of the PSFD does not provide information about body height and weight, the number of observations is different than a full sample would be; specifically, columns (3) and (4) in Table 3.1 remove the observations with no BMI value from our estimations. Moreover, the numbers of both types of schools have a positive effect on the number of years of schooling, with one additional senior high school increasing the years of schooling of students in a region from approximately 0.7 to 0.8, and one additional four-year college or university increasing the years of education of students in a region from 2.3 to 2.4. Although the adjusted number of senior high schools does not significantly affect the probability of four-year college attendance (Column 4, Table 3.3), other estimates still stand on expected sign, and the education reform increases the possibility of people going to a four-year college. Furthermore, the *F*-statistics of the excluded instrumental variables support the use of instrumental variables, and all significantly reject the null hypothesis.

Our estimates from the 2SLS analysis indicate that education has a negative effect on self-reported health status but a positive effect on no work disability. Tables 3.4 and 3.5 list the estimated coefficients of health regressions according to various main explanatory variables. As discussed earlier, the main variables are years of education (Table 3.4) and four-year college attendance (Table 3.5). Notably, when we change the definition of the dummy variables of self-reported health status, the OLS estimates reveal different outcomes. Overall, the effect of education on the “good and great” health status is positive but is negative on the “great” health status. However, the 2SLS estimates provide consistent results for all the variables, including that education has a negative effect on self-reported health status. We specifically observe that an additional year of schooling reduces the possibility of the “good or great” status from 0.03 to 0.04, and four-year college attendance reduces both “good or great” and “great” statuses from 0.13 to 0.18. By contrast, when we use the work disability as a dummy variable, both empirical approaches demonstrate the same result; however, the size of the effects is smaller in the OLS analysis than in the 2SLS analysis. The 2SLS estimates suggest that an additional year of education reduces the possibility of work disability by 0.007, whereas four-year college attendance lowers the possibility by 0.028.

Finally, we investigate the effect of education on the BMI score and related outcomes. The results of the 2SLS analysis do not indicate that higher education is associated with healthy BMI scores or with a reduced likelihood of overweightness or obesity. The same two main explanatory variables used in our other empirical tests are used here, and the results are presented in Tables 3.6 and 3.7. Most of the OLS estimates indicate that higher education is associated with healthy BMI scores and a reduced likelihood of being overweight or obese, except when the four-year college attendance dummy variable is used as the main explanatory variable. Nevertheless, all the estimates in the 2SLS analysis produced the same results, none of which suggest that education significantly affects BMI scores.

3.5 Discussion

Through our empirical estimates, we determine that a negative relationship exists between self-reported health status and education, which is inconsistent with the positive effect of education noted in the literature. However, when we use an objective health indicator (i.e., work disability), the evidence suggests that higher education reduces the possibility of work disability. The negative effect of education on self-reported health status can be explained by the inclusion of both current physical health and personal health concerns. Moreover, people with different levels of education but the same objective health status may subjectively provide different reports; for example, well-educated people tend to have more concerns for their health; thus, they usually report worse health statuses. These diverse self-assessment behaviors between subgroups is a well-identified problem of reporting heterogeneity (Shmueli, 2003; Lindeboom & van Doorslaer, 2004; Etile & Milcent, 2006; Bago d'Uva *et al.*, 2008). Although this problem may have skewed our estimates of the effect of education on health status, we argue that the statistical effect could reflect the correlation between self-health concerns and education.

On the basis of both the OLS and 2SLS estimates, education significantly affects the possibility of work disability, with the OLS estimates indicating a larger effect than the 2SLS estimates. We note few unobserved factors that is correlated to education and work disability simultaneously. The first possible explanation is that the size of the effect is determined by work experience. Given the same age, individuals who have more years of education naturally have less work experience, similarly who are in the status of work disability also have less work experience because of inferior health condition. The second explanation is that the education generates inferior health status before individuals participating in labor market. Taiwan's educational system is a test-centered one, after completing compulsory education students need to participate entrance examination for attending senior high school, vocational school, or fiv-year junior high school, after completing secondary education students still need to participate another examination for achieving higher education. Thus, people who had aimed at higher education degrees might have less time on exercise at young age, which causes a negative impact on health status before being in labor market and present employment status. In short, above two unobserved factors, potential work experience and health status before being in labor market, may be negatively correlated with education and work disability. Both factors are reflected in the upward-trending biased OLS estimates.

We also show the effects of education on BMI scores. The OLS estimates consistently reveal that more years of education is associated with healthy BMI scores and reduced likelihood of being overweight or obese; however, the 2SLS estimates indicate that education does not have a significant effect on BMI scores. These differences may be attributable to some unobserved factors. First, individuals who have higher unobserved abilities tend to effectively manage their health status and perform well in their academic outcomes; thus, unobserved abilities are positively correlated with education. Second, individuals who tend to plan for their future also tend to invest more in human capital (including education) during the present. This can also affect health outcomes because if people are concerned about their future health status, they strive to keep themselves in good health during the present. Third, the opportunity cost for individuals with higher education is also higher than that among individuals with less education. Poor health can increase the risk of losing income in the future, which also encourages well-educated people to engage in healthier behaviors. These unobserved factors possibly generated the significant effects of education on BMI scores in the OLS analysis.

The main limitation of this article is the lack of information. Considering that information on health has been extensively explored, BMI may not be the most appropriate candidate for determining health behaviors. BMI scores are a simple numeric measure that provide a reference for obesity, and medical professionals can use them to understand weight problems and diseases caused by obesity. However, BMI scores are calculated solely according to body height and weight; other crucial factors such as body fat, offal fat, and bone mass cannot be measured using a BMI score. Therefore, a person who exercises frequently may have a higher proportion of muscle that is translated into a high BMI score indicating obesity. Unfortunately, our data set does not provide further details on the participants' health conditions. The second limitation results from the source of the instrumental variables and the availability of data. Taiwan's education reform began in the 1990s, which can generate exogenous variations in education but also limits the age range of our sample. Specifically, the instruments used for such estimates are regional numbers of schools adjusted for the regional populations of regions. However, such information is only available from the academic year of 1991 in Taiwan, which forces us to eliminate individuals who were born before September 1976 and lowers the maximum value of age. Health capital is an accumulative process, and people who are more concerned about their health at a younger age tend to have healthier outcomes later. However, the range of age narrows the process of accumulative health capital and possibly also narrows variations in health outcomes. Nevertheless, according to the instrumental variables, our estimates still indicate the existence of causal relationship between education and few health outcomes.

3.6 Conclusion

This chapter examined the causal relation between education and health outcomes in Taiwan. Because both education and health status may be affected by unobserved factors simultaneously, we had to address the identification problem associated with the education reform that began in the 1990s is used to instrument education. The number of senior high schools and four-year colleges and universities has considerably increased since 1995; thus, the numbers of schools by regions are employed as instrumental variables. We determined that the number of senior high schools and four-year colleges and universities positively affects the number of years that people attend school and particularly increases their attendance at four-year colleges and universities. Three indicators of health status are used for the examinations, namely self-reported health status, work disability, and BMI score. The 2SLS analysis indicates that education has a negative effect on self-reported health outcomes, with just one additional year of education lowering the possibility of a healthy outcome by 3% to 4%, depending on how the outcome variables are defined. However, a positive effect on work attendance is identified, with an additional year of education reducing the possibility of work disability by 0.7% and attendance at a four-year college reducing the possibility by 2.8%. Overall, these estimates from objective or subjective health regression analyses are larger than those obtained from OLS analysis. Finally, the 2SLS results do not provide any evidence that more years of education is associated with a healthier BMI score.

Table 3.1 Summary Statistics

	Full sample		Aged 15 before Sep. 1st, 1995		Aged 15 after Sep. 1st, 1995	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Years of education	14.621	2.518	14.336	2.635	14.860	2.389
Junior high school or below	0.046	0.210	0.061	0.239	0.034	0.182
Senior high school	0.267	0.443	0.290	0.454	0.248	0.432
Some college	0.151	0.358	0.194	0.396	0.114	0.318
Four-year college	0.400	0.490	0.322	0.467	0.466	0.499
Graduate school	0.135	0.342	0.133	0.339	0.138	0.345
No. of senior high school	0.564	0.226	0.488	0.179	0.627	0.242
No. of college and university	0.281	0.159	0.203	0.125	0.346	0.156
Age	31.130	2.552	32.991	1.907	29.568	1.892
Male	0.554	0.497	0.561	0.496	0.548	0.498
Married	0.407	0.491	0.531	0.499	0.303	0.460
Observations	11,323		5,168		6,155	

Note: “Senior high school” includes all individuals who completed 12-year education, regardless of whether the education occurred at a traditional senior high school or vocational school. The number of senior high schools and the number of colleges and universities were calculated by the actual number of schools in a region per 1,000 people aged 15 and 18 years, respectively.

Table 3.2 Health Status and BMI Outcomes by Four-Year College Attendance

	Full sample		Not attended four-year college		Attended four-year college	
Self-reported health status	3.609	(0.771)	3.615	(0.795)	3.604	(0.795)
	Share		Share		Share	
Great	0.131		0.150		0.114	
Good	0.392		0.359		0.420	
Fair	0.436		0.450		0.425	
Poor	0.038		0.038		0.038	
Very poor	0.003		0.003		0.003	
Work-preventing	0.006	(0.074)	0.010	(0.102)	0.001	(0.036)
Observations	11,323		5,260		6,063	
BMI	23.452	(4.041)	23.794	(4.230)	23.169	(3.792)
	Share		Share		Share	
Healthy BMI (18.5≦BMI<24)	0.532		0.509		0.550	
Underweight (BMI<18.5)	0.078		0.073		0.082	
Overweight (24≦BMI<27)	0.217		0.213		0.221	
Obesity (27≦BMI)	0.173		0.205		0.147	
Observations	7,683		3,473		4,210	

Note: Standard deviations are reported in parentheses. Self-reported health status ranged from 1 ("very poor health") to 5 ("great health"). BMI (body mass index) is calculated by weight (kg) divided by height squared (m). "Share" is the fraction of observations in a given group.

Table 3.3 First Stage Estimates

Variable	(1) Years of education	(2) Attended four-year college	(3) Years of education	(4) Attended four-year college
No. of senior high schools	0.832 ***	0.111 *	0.692 *	0.061
	(0.297)	(0.061)	(0.367)	(0.074)
No. of colleges and universities	2.399 ***	0.656 ***	2.283 ***	0.615 ***
	(0.410)	(0.081)	(0.520)	(0.101)
Age	0.220	0.046	0.261	0.046
	(0.198)	(0.039)	(0.262)	(0.052)
Age squared	−0.003	−0.001	−0.004	−0.001
	(0.003)	(0.001)	(0.004)	(0.001)
Male	−0.244 ***	−0.066 ***	−0.189 ***	−0.061 ***
	(0.045)	(0.009)	(0.055)	(0.011)
Married	−0.563 ***	−0.122 ***	−0.423 ***	−0.092 ***
	(0.050)	(0.010)	(0.060)	(0.012)
Constant	10.125 ***	−0.405	9.886 ***	−0.265
	(3.197)	(0.637)	(4.356)	(0.871)
Regional dummies	Yes	Yes	Yes	Yes
Adj. R-squared	0.093	0.091	0.087	0.087
F-statistic of instruments	43.32 ***	58.50 ***	19.57 ***	25.83 ***
Observations	11,323	11,323	7,683	7,683

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. Columns (1) and (2) are used for estimating the effect on self-reported health and absence from work, and columns (3) and (4) are used for estimating the effect on BMI score. The number of senior high schools and the number of colleges and universities were calculated by the actual number of schools per 1,000 people aged 15 and 18 years, respectively.

Table 3.4 Effects of Years of Education on Health Status

Variable	Self-reported good and great		Self-reported great		Work disability	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Years of education	0.008 ***	−0.044 *	−0.008 ***	−0.033 **	−0.002 ***	−0.007 **
	(0.002)	(0.024)	(0.001)	(0.016)	(0.000)	(0.003)
Age	0.035	0.021	0.013	0.006	−0.005	−0.007
	(0.039)	(0.040)	(0.025)	(0.026)	(0.006)	(0.006)
Age squared	−0.001	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.041 ***	0.028 **	0.028 ***	0.022 ***	−0.002 **	−0.004 ***
	(0.010)	(0.012)	(0.006)	(0.008)	(0.001)	(0.001)
Married	0.039 ***	0.009	0.021 ***	0.007	0.001	−0.002
	(0.010)	(0.017)	(0.007)	(0.011)	(0.002)	(0.002)
Constant	−0.108	0.994	0.119	0.651	0.122	0.233 **
	(0.605)	(0.798)	(0.394)	(0.516)	(0.089)	(0.113)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,323	11,323	11,323	11,323	11,323	11,323

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. “Self-reported good and great,” “Self-reported great,” and “Work disability” are binary variables in the corresponding cases.

Table 3.5 Effects of Four-Year College Attendance on Health Status

Variable	Self-reported good and great		Self-reported great		Work disability	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Attended four-year college	0.035 ***	−0.179 *	−0.033 ***	−0.130 **	−0.009 ***	−0.028 **
	(0.010)	(0.098)	(0.007)	(0.066)	(0.002)	(0.013)
Age	0.035	0.022	0.013	0.007	-0.005	-0.007
	(0.040)	(0.040)	(0.025)	(0.026)	(0.006)	(0.006)
Age squared	−0.001	-0.001	0.000	0.000	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.041 ***	0.027 **	0.028 ***	0.021 ***	−0.003 **	−0.004 ***
	(0.010)	(0.012)	(0.006)	(0.008)	(0.001)	(0.002)
Married	0.039 ***	0.012	0.021 ***	0.009	0.001	-0.002
	(0.010)	(0.016)	(0.007)	(0.011)	(0.002)	(0.002)
Constant	−0.006	0.413	0.018	0.208	0.100	0.138
	(0.604)	(0.645)	(0.394)	(0.417)	(0.089)	(0.093)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,323	11,323	11,323	11,323	11,323	11,323

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels. "Self-reported good and great," "Self-reported great," and "Work disability" are binary variables in the corresponding cases.

Table 3.6 Effects of Years of Education on BMI Outcomes

Variable	BMI		Healthy BMI		Overweight and obesity		Obesity	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Years of education	−0.073 ***	0.120	0.006 **	0.002	−0.004 *	0.000	−0.010 ***	0.006
	(0.018)	(0.252)	(0.002)	(0.034)	(0.002)	(0.032)	(0.002)	(0.025)
Age	−0.384	-0.332	0.078 *	0.077	−0.050	−0.049	−0.090 **	−0.086 **
	(0.384)	(0.392)	(0.048)	(0.049)	(0.045)	(0.046)	(0.037)	(0.038)
Age squared	0.008	0.008	−0.001 *	−0.001 *	0.001	0.001	0.002 ***	0.001 ***
	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	2.818 ***	2.855 ***	−0.170 ***	−0.171 ***	0.283 ***	0.284 ***	0.141 ***	0.144 ***
	(0.087)	(0.099)	(0.011)	(0.013)	(0.011)	(0.012)	(0.008)	(0.010)
Married	0.362 ***	0.444 ***	−0.040 ***	−0.041 **	0.058 ***	0.059 ***	0.005	0.012
	(0.096)	(0.141)	(0.012)	(0.019)	(0.012)	(0.018)	(0.009)	(0.014)
Constant	26.480 ***	22.345 ***	−0.558	−0.475	0.775	0.703	1.554 ***	1.195
	(6.120)	(8.136)	(0.769)	(1.056)	(0.722)	(0.997)	(0.590)	(0.794)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,683	7,683	7,683	7,683	7,683	7,683	7,683	7,683

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. "Healthy BMI," "Overweightness and obesity," and "Obesity" are binary variables in the corresponding cases.

Table 3.7 Effects of Four-Year College Attendance on BMI Outcomes

Variable	BMI		Healthy BMI		Overweight and obesity		Obesity	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Attended four-year	−0.330 ***	1.065	0.019	−0.041	−.016	0.051	−0.043 ***	0.076
college	(0.092)	(1.098)	(0.012)	(0.145)	(0.011)	(0.138)	(0.009)	(0.109)
Age	−0.382	−0.309	0.078	0.075	−0.050	−0.046	−0.090 **	−0.083 **
	(0.384)	(0.394)	(0.048)	(0.049)	(0.045)	(0.046)	(0.037)	(0.038)
Age squared	0.008	0.007	−0.001 *	−0.001 *	0.001	0.001	0.002 ***	0.001 **
	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	2.812 ***	2.896 ***	−0.170 ***	−0.174 ***	0.283 ***	0.287 ***	0.141 ***	0.148 ***
	(0.087)	(0.109)	(0.011)	(0.014)	(0.011)	(0.013)	(0.008)	(0.011)
Married	0.362 ***	0.491 ***	−0.040 ***	−0.046 ***	0.058 ***	0.064 ***	0.005	0.016
	(0.096)	(0.138)	(0.012)	(0.018)	(0.012)	(0.017)	(0.009)	(0.014)
Constant	25.569 ***	22.832 ***	−0.475	−0.356	0.726	0.595	1.415 **	1.181 *
	(6.113)	(6.536)	(0.768)	(0.822)	(0.720)	(0.775)	(0.590)	(0.533)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,683	7,683	7,683	7,683	7,683	7,683	7,683	7,683

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. “Healthy BMI,” “Overweightness and obesity,” and “Obesity” are binary variables in the corresponding cases.

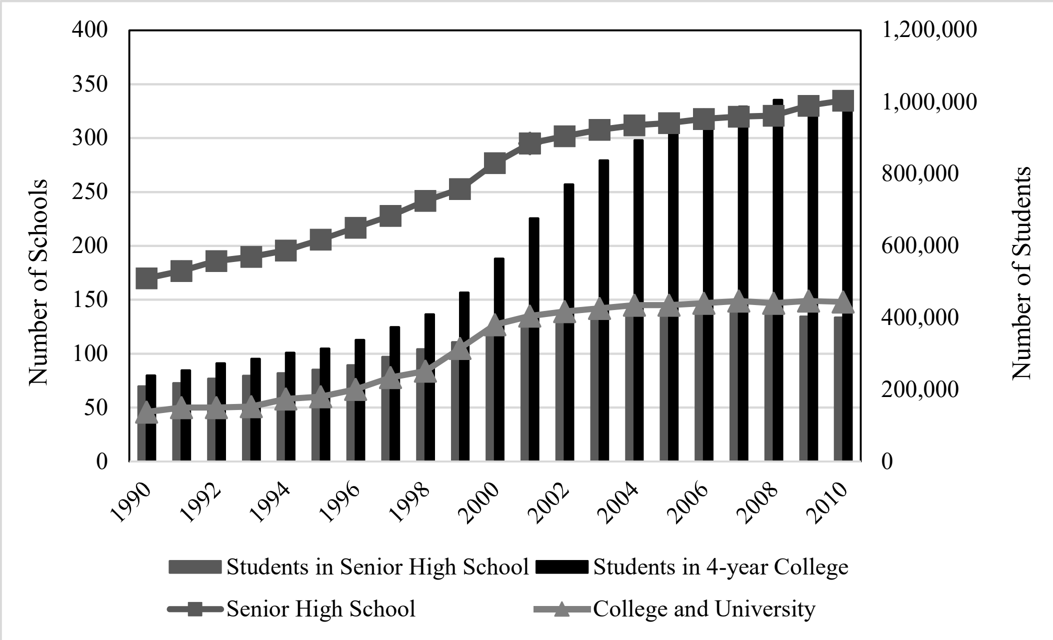

Figure 3.1 Numbers of Schools and Students

Chapter 4

More Education, Fewer Marriages:

Evidence from Spousal Education

4.1 Introduction

Forming a family is a crucial decision in a person's life. Decisions concerning marriage and whom to marry have been an economic question for decades. One well-known argument in the literature regarding marriage is positive assortative mating in the marriage market. Previous studies have shown that people prefer a marriage mate who possesses similar characteristics, such as ethnicity, religion, and economic status (Hout, 1982; Pagnini & Morgan, 1990; Kalmijn, 1991a, Kalmijn, 1991b; Kalmijn, 1994; Watson *et al.*, 2004). Not surprisingly, education is also a substantial determinant of marriage (Kalmijn, 1991a; Kalmijn, 1991b; Mare, 1991): people prefer to marry someone with a similar educational attainment level. Previous evidence on educational assortative mating enhances our interest in exploring the causal relationship between education and marriage outcome. Furthermore, when the causal relationship does exist, is this caused by educational assortative mating?

Lefgen and McIntyre (2006) provide the causal channel between women's education and their marriage outcomes. Educational assortative mating can lead to women tending to marry men with a similar educational attainment through two patterns: First, some women may meet their future spouses at school (Stevens, 1991); therefore, for instance, women's college attendance could increase the possibility of their marrying college graduates. Second, education may result in well-educated women having more prestigious occupations and social status, which enhances their bargaining power in the marriage market and allows them to meet high-quality men after graduation.

Even though education is a critical factor in marriage outcome, surprisingly, the literature barely provides evidence that education has a causal effect on marital status in Western countries (Lefgen & McIntyre, 2006; Anderburg & Zhu, 2014).[22] However, we argue that the role of education in the marriage market in Eastern countries should be different than that in Western countries because the cultures differ. In addition to educational assortative mating, we consider that traditional male gender role expectations in Taiwan induce men to be the main sources of income for their families, which may be reflected in the difference in educational level in a couple. Therefore, we build a hypothesis regarding the pool of potential mates in the marriage market. We assume that the educational composition of a couple should conform with either assortative mating or the man's education being superior to the woman's. Tsay and Wu (2006) outline the popular stereotype of Taiwanese society: their results suggest that well-educated men have a higher propensity to marry women who have a lower level of education. This implies that men with a higher education level may have a larger pool of potential mates; by contrast, women with a higher education level face a smaller pool of potential mates; thus, fewer well-educated men tend to be single, whereas well-educated women tend to be single.

Our empirical strategy is to use the instrumental variable approach to account for the well-known endogeneity of education. The correlation between education, marriage outcomes, and omitted variables may result in ordinary least squares (OLS): such correlation generates substantially biased estimates. A change in educational

[22] Lefgen and McIntyre (2006) use U.S. data and apply a quarter of the birth instruments proposed by Angrist and Krueger (1991). They find that women's education has no effect on the probability of marriage. Similarly, Anderburg and Zhu (2014) use U.K. data and apply the compulsory schooling law instruments to find that education has no effect on the probability of women aged 25 or above being married.

policy provides a source for handling the identification problem. The government launched a series of education reforms in the 1990s in Taiwan. One of these reforms was the expansion of higher education. In the period between the late 1990s and the mid-2000s, this reform dramatically increased citizens' education levels among the younger cohorts: in 1998, 21% of citizens had completed higher education, but this proportion increased to 30% by 2005.[23] Recently, we have gathered enough samples to use the educational expansion as the exogenous factor for educational attainment.

The object of this chapter is two-fold: First, we examine the causal relation between education and the possibility of never marrying by using individual-level data in Taiwan; we introduce the expansion of higher education as an instrumental variable for education. After accounting for the endogeneity of education, we find that more education leads to a higher possibility of never marrying for both men and women. Second, we test our hypothesis from the marriage market by using data concerning spouses' educational attainment. The empirical results imply that well-educated women face a smaller pool of potential partners, which increases their likelihood of remaining single.

The rest of this article is structured as follows: Section 2 describes the data used in the analysis and summary statistics. Section 3 presents the empirical methodology and results, and Section 4 concludes the article.

4.2 Data and Descriptive Statistics

Our estimations rely on the Panel Study of Family Dynamic (PSFD), which is a project conducted by Academia Sinica (a Taiwanese research institute). All participants are Taiwanese adults. The data set includes sufficient information for our

[23] Data source: The Directorate General of Budget, Accounting and Statistics (DGBAS) of the Executive Yuan, Taiwan.

investigation, such as age, marital status, spousal education, and residence. Furthermore, the survey contains month of birth, which allows us to identify the possible school years of individuals attending college and connect those years with the expansion of education. We collect our sample from four waves of the PSFD: 2009, 2010, 2011, and 2012.

The empirical strategy of this chapter is to employ the expansion of higher education in the 1990s as an instrumental variable to address the endogeneity of education. Figure 4.1 provides the number of four-year colleges and universities and the number of students in undergraduate programs in each school year. The expansion started during the mid-1990s. After 1998, both the number of universities and number of college students greatly increased. In the early 1990s, only 46 four-year colleges and universities operated in Taiwan, with approximately 240,000 students in undergraduate programs; these numbers increased to 127 and 564,000, respectively, by 2000. The education reforms dramatically changed citizens' education opportunities; the number of institutions approximately doubled, and the number of college students increased by approximately one and half times in less than a decade.

To describe the expansion of higher education, we use the numbers of four-year colleges and universities in the individuals' relevant academic years. Taiwanese students usually graduate from senior high school at 18, and the population also varies over time; therefore, rather than using the actual number of institutions, we adjust the number by the population aged 18. We collect data on the number of four-year colleges and universities from the Ministry of Education of Taiwan and data on the population aged 18 from the Ministry of the Interior of Taiwan. In the rest of this article, the number of four-year colleges and universities is calculated by the number of institutions per 1,000 people aged 18 years.

Our estimations focus on a rather younger cohort for three main reasons: First, by narrowing the sample to those who just joined the marriage market—for example, those leaving school—an analysis can minimize the estimation bias from age and age-related factors. Second, the expansion of education occurred in the 1990s, and consequently we limit our sample to those who attended college in that decade and whose education was affected by this educational policy. Third, data limitations mean that data for those aged 18 are available only from 1991; fortunately, this is several years before the education policy was implemented, which means we can reasonably use the policy to instrument education. Therefore, we eliminate individuals who were born before September 1976 from this study.[24] In total, 4,853 male and 3,880 female participants are examined in this study, with ages ranging from 26 to 36 years. Table 4.1 provides a summary of statistics.

In Table 4.1, in addition to the full sample, we divide the sample into two subgroups: those aged 18 before September 1, 1998 and those aged 18 after September 1, 1998. The academic year in Taiwan starts September 1, and the 1998 academic year is the year that the expansion of education was launched.[25] Both males and females who turned 18 after September 1, 1998 have more years of education and a higher rate of four-year college attendance. Men who turned 18 before September 1, 1998 had 14.24 years of education; 30% of them had undergone 4 years of college attendance on average. Men who turned 18 after September 1, 1998 had 14.73 years of education; 41% of them had undergone 4 years of college attendance. Women who turned 18 before September 1, 1998 had 14.39 years of education; 36% of them had undergone 4 years of college attendance on average. Women who turned 18 after

[24] People who were born after the end of August 1976 possibly attended college after the 1995 academic year.

[25] The start date of each academic year in Taiwan is September 1, and every year, children who turn 6 years old before that day begin elementary school.

September 1, 1998 had 14.95 years of education; 52% of them had undergone 4 years of college attendance. Furthermore, the proportions of those who had never been married also greatly differ in both subgroups. Both males and females who turned 18 after September 1, 1998 report much higher proportions of having never married: 81% of men are single, and 65% of women are single; among those who turned 18 before September 1, 1998, 56% of men are single, and 37% of women are single. However, the positive relationship between never being married and education does not reveal the causal relation. This relationship may be caused by age and other factors. We leave the examination of causality until the next section.

4.3 Empirical Methodology and Results

4.3.1 The relationship between education and marriage outcome

To examine the causal relationship between education and marriage outcome, we conduct a conventional two-stage least square (2SLS) model. The empirical model is set as follows:

$$NM_{ij} = \alpha_1 Edu_{ij} + \alpha_2 X_{ij} + \eta_j + \varepsilon_{ij} \tag{4.1}$$

$$Edu_{ij} = \beta_1 College_i + \beta_2 X_{ij} + \gamma_j + \upsilon_{ij} \tag{4.2}$$

Equation (4.1) presents the second stage estimation. The suffixes i and j denote that individual i lives in region j. NM_{ij} presents an individual's marital status: when individual i has never been married, NM_{ij} equals 1; NM_{ij} is 0 otherwise. Edu_{ij} is the main explanatory variable. We use the years of education and a dummy variable of four-year college attendance to express it. α_1 is the target effect we attempt to estimate. X_{ij} is a vector of other covariates and the constant covariates contain, such as age and age squared, and α_2 is the corresponding vector of coefficients. We also use η_j to determine regional effects because varying living costs, labor market conditions, and other factors between regions may affect marriage decisions. ε_{ij} is the error term. This may be correlated with an individual's educational attainment through unobserved variables; therefore, if we use a regular regression to estimate equation (4.1), the estimated α_1 tends to be biased. To address the identification problem, we introduce an instrumental variable approach in equation (4.2). $College_i$ is the number of four-year colleges and universities in the academic year that individual i turned 18 by the end of August in the relevant year, which is the instrumental variable used to generate exogenous variations in education. X_{ij} is the same vector of covariates in equation (4.1) with β_2, a corresponding vector of coefficients. Furthermore, in equation (4.2), we use γ_j to control for regional effects.[26]

The expansion of higher education causes higher education levels and a higher possibility of four-year college attendance for both men and women. Table 4.2 reports the first-stage estimates. For men, an additional number of schools increases the years of schooling by 5 years and increases the possibility of four-year college attendance by 109%; for women, an additional number of schools increases the years of schooling by 4 years and increases the possibility of four-year college attendance by 116%.[27] Furthermore, all F-statistics of excluded instruments support the use of instrumental variables. None report an insignificant result.

In Table 4.3, we present the estimated impact of years of education. The OLS estimates are also shown for comparison with the 2SLS estimates. The OLS estimates suggest that an additional year of education increases the possibility of never being

[26] On the basis of the administrative divisions of Taiwan, we define 18 regions in Taiwan: Taipei City, New Taipei City, Taoyuan City, Taichung City, Tainan City, Kaohsiung City, Keelung City, Yilan County, Miaoli County, Changhua County, Nantou County, Yunlin County, Pingtung County, Penghu County, Hualien County, Taitung County, Hsinchu Region, and Chiayi Region. The Kinmen–Matsu region is not used because of the limitations of the sample size. For the male sample, we include 17 regional dummies in the estimations; for the female sample, no sample for Penghu County was used, thus we include 16 regional dummies in the estimations.

[27] In our estimation, the numbers of schools are adjusted by the aged-18 population. These numbers are less than one, thus finding an effect larger than one is reasonable. The mean and standard deviation of the number of four-year colleges and universities are shown in Table 4.1.

married by 1.5% for males and 6.1% for females. The 2SLS estimates also suggest the positive relationship between education and the possibility of never being married; however, the size of effects is larger than that in the OLS. For the male sample, an additional year of education increases the possibility by 7.4%, and for the female sample, an additional year of education increases the possibility by 8.9%. Table 4.4 shows the estimated effect of four-year college attendance. When we use the dummy of four-year college attendance as the main explanatory variable, we also determine similar results. The 2SLS estimates report more substantial effects that men who attend college for 4 years increase their possibility of never being married by 36%, and women who attend college for 4 years increase that possibility by 30%.

Unobserved ability may explain the different sizes of the estimated effect between the results in the OLS and those in the 2SLS. Previous studies have thoroughly accounted for such unobserved ability when studying the effect of education on labor market outcomes.[28] Higher potential ability is positively correlated with education, wage, and employment status, which causes the bias in OLS estimators. Assume that people with a higher unobserved ability, superior academic performance, and intelligence also improve their outcomes in other areas, for instance, in the marriage market. This means that education is more likely positively correlated with the omitted factor. If the unobserved ability is negatively correlated with the possibility of never being married, this omitted factor is reflected in the downward-trending biased OLS estimates.

[28] Economists have argued for a correlation between unobserved ability and education since the 1990s. Most have used the instrumental variable approach or sampled identical twins to account for the ability biases. Studies that use the instrumental variable approach: Angrist and Krueger (1991), Leigh and Ryan (2008). Studies that use the sample of twins for identification: Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998), Behrman and Rosenzweig (1999), and Bonjour et al. (2003).

4.3.2 Education, spousal education, and marriage outcome

Thus far, we have illustrated the positive relationship between education and the possibility of never marrying. But why does more education result in a smaller possibility of marrying? In this subsection, we test a hypothesis regarding the marriage market in Taiwan. We assume that the educational composition of a couple conforms to at least one of the following two patterns. First, educational assortative mating: the husband and wife have a similar level of education. Second, the husband's educational level is superior to his wife's.

Table 4.5 shows the educational attainment of married men and women and of their spouses. We categorize the subjects' education into four categories: years of education equals 12 years or is less than 12 years (senior high school and less), years of education of between 13 and 15 years (some college), years of education equals 16 years (four-year college), and years of education exceeds 16 years (graduate school). Generally, first, the statistics support educational assortative mating for both males and females: the largest proportions of spouses' educational attainment are similar to individuals' own educational level. Second, the statistics also partly support men's education levels being superior to women's within a couple, especially in the well-educated sample. Using individuals who have ever attended graduate school as an example, for men, 46% of their wives attended four-year college and 38% of their wives attended graduate school; for women, 46% of their husbands attended four-year college and 52% of their husbands attended graduate school.

To examine our hypothesis empirically, we apply a multinomial logit model. Assume that individuals have five choices in the marriage market: staying single, marrying a partner with senior high school or less, marrying a partner who attended some college, marrying a partner who attended four-year college, and marrying a partner with a graduate school degree. The model is set as follows:

$$
\begin{aligned}
&\mathrm{Prob}(\mathrm{Y}_i = k) = \frac{\exp(Z_i\delta_k)}{1+\sum_{k=1}^{4}\exp(Z_i\delta_k)} \qquad k\text{=1,2,3,4} \\
&\mathrm{Prob}(\mathrm{Y}_i = k) = \frac{1}{1+\sum_{k=1}^{4}\exp(Z_i\delta_k)} \qquad k = 0
\end{aligned}
\tag{4.3}
$$

Where $\mathrm{Y_i}$ is individual *i*'s choice on marriage, *k* denotes five marriage choices, *k* equals 0 if not married, 1 if the spouse attended senior high school or less, 2 if the spouse attended some college, 3 if the spouse attended four-year college, and 4 if the spouse attended graduate school. Z_i is a vector of explanatory variables, which includes education, age, and age squared, and δ_k is the corresponding vector of coefficients. The education is categorized into senior high school and less, some college, four-year college, and graduate school. Senior high school and less is the omitted group in our estimations.

The estimated coefficients of equation (4.3) are reported in Table 4.6. As equation (4.3) shows, four groups of estimated coefficients are evident. Here, the not married group is the base group. Most education-related coefficients are significant and can be the answer of our hypothesis. Well-educated men have less of a possibility to choose women who attended senior high school and some college. This pattern is also reflected in the female results. The estimated coefficients in Table 4.6 present how these covariates affect the likelihood of falling into each group of marriage choices; however, the coefficients do not present the marginal effects of these covariates. Table 4.7 reports the marginal effects of all explanatory variables at their means. In Table 4.7, we lose some significant signs, especially in the female results. However, we still find some evidence from women who are extremely well educated: women's graduate school attendance reduces the possibility of marrying less-educated men and significantly increases the possibility of marrying men who attended graduate school. In general, the results in Table 4.7 support our hypothesis from the marriage market:

men tend to marry a woman who has a similar or lesser education than they have. This implies that women who have a higher level of education shrink the pool of potential partners and therefore lower their possibility of marrying. The negative relationship between education and being married was determined in Section 4.3.1 of this article.

4.4 Conclusion

This study empirically investigates the causal relationship between education and marriage outcome in Taiwan. To address the endogeneity of education, we employ the expansion of higher education in the 1990s as instrumental variables. The numbers of four-year colleges and universities per 1,000 people aged 18 years in the relevant academic year of individuals are used to determine the expansion of education. The first-stage estimates suggest that, for both males and females, the expansion of education significantly increases the years of education and the possibility of four-year college attendance. 2SLS estimates provide the causal relationship that education leads to a higher possibility of never marrying. For males, an additional year of education increases the possibility of never marrying by 7.4%; for females, an additional year of education increases the possibility of never marrying by 8.9%.

To expose the connection of education and marriage outcome, we provide a hypothesis that the educational composition of a couple follows either educational assortative mating or the husband having a higher education level than his wife does. On the basis of the results of multinomial logistic regressions, we find evidence that the choice of marriage partly conforms to our hypothesis, which implies well-educated women face a smaller pool of potential partners in the marriage market than do less-educated women and so therefore, they have a higher possibility of being single. However, our hypothesis does not explain men's education and their marriage outcomes. Perhaps the causal channel for men is distinct from that for women, thus possibly the explanations that are adequate for females do not apply to men.[29]

[29] Jones and Gubhaju (2009) argued that men's education plays only a minor role in marriage in the low-fertility countries of East and Southeast Asia. One reason may be that men living in urban societies face pressures from work and because of housing affordability and a reluctance to live with or care for their parents-in-law.

Table 4.1 Summary Statistics

	Full sample		Aged 18 before Sept. 1, 1998		Aged 18 after Sept. 1, 1998	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Male						
Never married	0.692	0.462	0.557	0.497	0.811	0.392
Years of education	14.497	2.647	14.235	2.744	14.730	2.536
Junior high school and below	0.054	0.226	0.068	0.252	0.042	0.201
Senior high school	0.302	0.459	0.329	0.470	0.279	0.448
Some college	0.138	0.345	0.166	0.372	0.113	0.317
Four-year college	0.360	0.480	0.299	0.458	0.414	0.493
Graduate school	0.146	0.353	0.139	0.346	0.152	0.359
No. of colleges and universities	0.261	0.082	0.185	0.019	0.328	0.054
Age	30.513	2.226	32.320	1.434	28.911	1.437
Observations	4,853		2,280		2,573	
Female						
Never married	0.525	0.499	0.373	0.484	0.653	0.476
Years of education	14.693	2.311	14.393	2.429	14.945	2.176
Junior high school and below	0.039	0.193	0.053	0.224	0.027	0.161
Senior high school	0.231	0.422	0.249	0.433	0.216	0.412
Some college	0.173	0.378	0.231	0.422	0.123	0.329
Four-year college	0.447	0.497	0.356	0.479	0.523	0.500
Graduate school	0.111	0.314	0.111	0.314	0.111	0.314
No. of colleges and universities	0.264	0.082	0.187	0.019	0.328	0.055
Age	30.432	2.232	32.272	1.474	28.884	1.450
Observations	3,880		1,773		2,107	

Note: The number of colleges and universities were calculated by the actual number of schools in the relevant academic year per 1,000 people aged 18 years.

Table 4.2 Estimated Coefficients of First Stage

	Years of schooling		Four-year college attendance	
Male				
No. of colleges and universities	5.272 ***	(0.886)	1.086 ***	(0.171)
Age	0.814 *	(0.428)	0.077	(0.083)
Age squared	−0.012 *	(0.007)	−0.001	(0.001)
Constant	1.164	(6.727)	−0.866	(1.308)
Regional dummies	Yes		Yes	
Adj. R-squared	0.087		0.083	
F-test of excluded instruments	35.43 ***		40.52 ***	
Observations	4,853		4,853	
Female				
No. of colleges and universities	3.969 ***	(0.861)	1.158 ***	(0.190)
Age	−0.343	(0.407)	0.061	(0.089)
Age squared	0.006	(0.007)	−0.001	(0.001)
Constant	19.504 ***	(6.401)	−0.551	(1.415)
Regional dummies	Yes		Yes	
Adj. R-squared	0.081		0.075	
F-test of excluded instruments	21.27 ***		37.33 ***	
Observations	3,880		3,880	

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. The number of colleges and universities were calculated by the actual number of schools per 1,000 people aged 18 years.

Table 4.3 Effect of Years of Schooling on Never Being Married

	OLS		2SLS	
Male				
Years of schooling	0.015 ***	(0.002)	0.074 ***	(0.030)
Age	0.056	(0.068)	0.055	(0.072)
Age squared	−0.002 *	(0.001)	−0.002	(0.001)
Constant	0.659	(1.017)	−0.372	(1.168)
Regional dummies	Yes		Yes	
Adj. R-squared	0.128		0.022	
Observations	4,853		4,853	
Female				
Years of schooling	0.061 ***	(0.003)	0.089 **	(0.045)
Age	−0.346 ***	(0.072)	−0.319 ***	(0.085)
Age squared	0.005 ***	(0.001)	0.004 ***	(0.001)
Constant	6.033 ***	(1.090)	5.134 ***	(1.808)
Regional dummies	Yes		Yes	
Adj. R-squared	0.220		0.205	
Observations	3,880		3,880	

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively.

Table 4.4 Effect of Four-Year College Attendance on Never Being Married

	OLS		2SLS	
Male				
Four-year college attendance	0.080 ***	(0.013)	0.360 ***	(0.145)
Age	0.063	(0.067)	0.088	(0.072)
Age squared	−0.002 *	(0.001)	−0.002 **	(0.001)
Constant	0.725	(1.017)	0.026	(1.117)
Regional dummies	Yes		Yes	
Adj. R-squared	0.128		0.042	
Observations	4,853		4,853	
Female				
Four-year college attendance	0.259 ***	(0.016)	0.304 **	(0.153)
Age	−0.374 ***	(0.072)	−0.368 ***	(0.075)
Age squared	0.005 ***	(0.001)	0.005 ***	(0.001)
Constant	7.172 ***	(1.090)	7.031 ***	(1.189)
Regional dummies	Yes		Yes	
Adj. R-squared	0.209		0.207	
Observations	3,880		3,880	

Note: Robust standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively.

Table 4.5 Spouse's Educational Attainment by Individual's Education

	Senior high school or less	Some college	College	Graduate school
Male: Wife's education				
Years of schooling	12.835	14.133	15.209	16.497
	(1.950)	(1.844)	(1.729)	(1.965)
	Share	Share	Share	Share
Senior high school or less	0.636	0.343	0.149	0.023
Some college	0.197	0.275	0.156	0.131
College	0.165	0.348	0.623	0.463
Graduate school	0.002	0.034	0.072	0.383
Observations	557	233	416	175
Female: Husband's education				
Years of schooling	12.027	13.879	15.589	17.127
	(1.813)	(1.825)	(1.262)	(1.447)
	Share	Share	Share	Share
Senior high school or less	0.824	0.362	0.161	0.000
Some college	0.105	0.357	0.124	0.027
College	0.064	0.241	0.482	0.455
Graduate school	0.008	0.040	0.234	0.518
Observations	658	373	598	110

Notes: Observations are married individuals. Standard deviations are reported in parentheses. "Share" is the fraction of observations in each group.

Table 4.6 Estimated Coefficients of Multinomial Logit Model

	Male				Female			
	j=1	*j*=2	*j*=3	*j*=4	*j*=1	*j*=2	*j*=3	*j*=4
Some college	−0.588 ***	0.351 **	0.759 ***	2.962 ***	−1.411 ***	0.640 ***	0.675 ***	0.996 *
	(0.140)	(0.170)	(0.166)	(1.063)	(0.132)	(0.173)	(0.209)	(0.526)
College	−1.824 ***	−0.611 ***	0.974 ***	3.417 ***	−2.941 ***	−1.135 ***	0.744 ***	2.165 ***
	(0.146)	(0.164)	(0.131)	(1.018)	(0.131)	(0.182)	(0.182)	(0.462)
Graduate school	−3.702 ***	−0.782 ***	0.673 ***	5.062 ***	−18.737	−3.238 ***	0.033	2.283 ***
	(0.507)	(0.237)	(0.165)	(1.011)	(478.085)	(0.567)	(0.597)	(0.477)
Age	−0.490	1.191	2.061 ***	7.816 ***	2.286 ***	3.477 ***	2.032 ***	3.213 ***
	(0.611)	(0.842)	(0.668)	(2.262)	(0.629)	(0.897)	(0.897)	(1.014)
Age squared	0.013	−0.014	−0.027 ***	−0.113 ***	−0.033 ***	−0.052 ***	−0.027 ***	−0.045 ***
	(0.010)	(0.014)	(0.011)	(0.035)	(0.010)	(0.014)	(0.014)	(0.016)
Constant	1.839	−25.073 *	−39.911 ***	−140.695 ***	−38.272 ***	−58.862 ***	−38.868 ***	−60.190 ***
	(9.443)	(13.066)	(10.402)	(36.322)	(9.630)	(13.850)	(13.850)	(15.843)
Chi-squared	1208.18 ***				1879.26 ***			
Observations	4,736				3,774			

Note: Standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. *j*=0,1,2,3, and 4 mean that the individual is not married, spouse attended senior high school or less, spouse attended some college, spouse attended college, and spouse attended graduate school, respectively. In the estimations, *j*=0 is the base outcome.

Table 4.7 Marginal Effects of Multinomial Logit Model

	Male				
	j=0	*j*=1	*j*=2	*j*=3	*j*=4
Some college	−0.086 ***	−0.035 ***	0.014	0.077 ***	0.030
	(0.027)	(0.006)	(0.010)	(0.021)	(0.021)
College	−0.000	−0.098 ***	−0.029 ***	0.105 ***	0.023 **
	(0.018)	(0.009)	(0.007)	(0.014)	(0.010)
Graduate school	−0.094	−0.102 ***	−0.034 ***	0.059 ***	0.172 **
	(0.069)	(0.006)	(0.007)	(0.023)	(0.083)
Age	−0.196 ***	−0.046	0.050	0.171 ***	0.021 ***
	(0.075)	(0.036)	(0.041)	(0.055)	(0.008)
Age squared	0.002 *	0.001 *	−0.001	−0.002 ***	−0.000 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	4,736				
	Female				
	j=0	*j*=1	*j*=2	*j*=3	*j*=4
Some college	−0.127	−0.031	0.037	0.079	0.042
	(1.316)	(1.634)	(0.096)	(0.186)	(0.051)
College	−0.021	−0.102	−0.076	0.096	0.103 **
	(1.316)	(4.944)	(0.642)	(0.492)	(0.051)
Graduate school	0.039	−0.205 ***	−0.079 ***	0.012	0.233 ***
	(1.316)	(0.010)	(0.006)	(0.026)	(0.066)
Age	−0.523	0.047	0.195	0.182	0.098
	(1.316)	(2.410)	(0.505)	(0.664)	(0.267)
Age squared	0.077	−0.001	−0.003	−0.002	−0.001
	(1.316)	(0.035)	(0.008)	(0.009)	(0.004)
Observations	3,774				

Note: Standard errors are reported in parentheses. *, **, and *** are statistically significant at the 10%, 5%, and 1% levels, respectively. *j*=0,1,2,3 and 4 mean that the individual is not married, spouse attended senior high school or less, spouse attended some college, spouse attended college, and spouse attended graduate school, respectively.

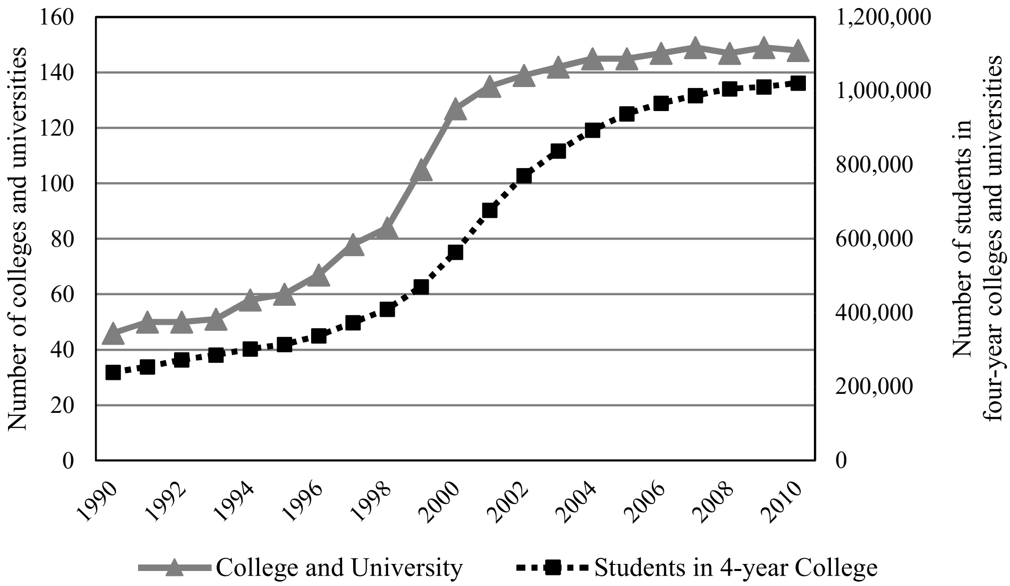

Figure 4.1 Numbers of Four-Year Colleges and Universities, and Students

Bibliography

Adams, S. J. 2002. "Educational Attainment and Health: Evidence from a Sample of Older Adults." *Education Economics*, 10:1, 97–109.

Albouy, V. and L. Lequien. 2009. "Does Compulsory Education Lower Mortality?" *Journal of Health Economics*, 28:1, 155–168.

Anderberg, D. and Y. Zhu. 2014. "What a Difference a Term Makes: The Effect of Educational Attainment on Marital Outcomes in the UK." *Journal of Population Economics*, 27:2, 387–419.

Angrist, J. D. and W. N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review*, 88:3, 450–477.

Angrist, J. D. and A. B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106:4, 979–1014.

Angrist, J. D. and A. B. Krueger. 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., V. Lavy and A. Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics*, 28:4, 773–824.

Arendt, J. N. 2005. "Does Education Cause Better Health? A Panel Data Analysis Using School Reform for Identification." *Economics of Education Review*, 24:2, 149–160.

Arendt, J. N. 2008. "In Sickness and in Health – Till Education Do Us Part: Education Effects on Hospitalization." *Economics of Education Review*, 27:2, 161–172.

Arora, S. 2001. "Health, Human Productivity, and Long-Term Economic Growth." *Journal Economic History*, 61:3, 699–749.

Asai, Y., R. Kambayashi, and S. Yamaguchi. 2015. "Childcare Availability, Household Structure, and Maternal Employment." *Journal of the Japanese and International Economies*, 38, 172–192.

Ashenfelter, O. and A. Krueger. 1994. “Estimates of the Economic Return to Schooling from a New Sample of Twins.” *American Economic Review*, 84:5, 1157–1173.

Ashenfelter, O. and C. Rouse. 1998. “Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins,” *Quarterly Journal of Economics*, 113:1, 253–284.

Bago d'Uva, T., E. V. Doorslaer, M. Lindeboom, and O. O'Donnell. 2008. “Does Reporting Heterogeneity Bias the Measurement of Health Disparities?” *Health Economics*, 17:3, 351–375.

Becker, G. S. 1960. “An Economic Analysis of Fertility.” In *Demographic and Economic Change in Developed Countries*, Princeton, NJ: Princeton University Press, 209–240.

Becker, G. S., and H. G. Lewis. 1973. “On the Interaction between the Quantity and Quality of Children.” *Journal of Political Economy*, 81:2 S279–S288.

Becker, G. S., and N. Tomes. 1976. “Child Endowments and the Quantity and Quality of Children.” *Journal of Political Economy*, 84:4, S143–S162.

Behrman, J. R. and M. R. Rosenzweig. 1999. “‘Ability’ Biases in Schooling Returns and Twins: A Test and New Estimates.” *Economics of Education Review*, 18:2, 159–167.

Black, S., P. J. Devereux, and K. G. Salvanes. 2005. “The More the Merrier? The Effect of Family Composition on Children’s Education.” *Quarterly Journal of Economics*, 120:2, 669–700.

Blake, J. 1981. “Family Size and the Quality of Children.” *Demography*, 18:4, 421–442.

Blau, D. M. and P. K. Robins. 1988. “Child-Care Costs and Family Labor Supply.” *Review of Economics and Statistics*, 70:3, 374–381.

Bloom, D. E., D. Canning, and J. Sevilla. 2004. “The Effect of Health on Economic Growth: A Production Function Approach.” *World Development*, 32:1, 1–13.

Bonjour, D., L. F. Cherkas, J. E. Haskel, D. D. Hawkes, and T. D. Spector. 2003. “Returns to Education: Evidence from U.K. Twins.” *American Economic Review*, 93:5, 1799–1812.

Booth, A. L. and H. J. Kee. 2009. "Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment." *Journal of Population Economics*, 22:2, 367–397.

Braakmann, N. 2011. "The Causal Relationship Between Education, Health and Health Related Behaviour: Evidence from a Natural Experiment in England." *Journal of Health Economics*, 30:4, 753–763.

Brewster, K. L. and R. R. Rindfuss, 2000, "Fertility and Women's Employment in Industrialized Nations." *Annual Review of Sociology*, 26, 271–296.

Chiappori, P.-A. 1988. "Rational Household Labor Supply." *Econometrica*, 56:1, 63–90.

Chu, C. Y. C., S. Kim, and W.-J. Tsay, 2014, "Coresidence with Husband's Parents, Labor Supply, and Duration to First Birth." *Demography*, 51:1, 185-204.

Chun, H., J. Oh. 2002. "An Instrumental Variable Estimate of the Effect of Fertility on the Labor Force Participation of Married Women." *Applied Economics Letters*, 9:10, 631–634.

Cruces, G. and S. Galiani. 2007. "Fertility and Female Labor Supply in Latin America: New Causal Evidence." *Labour Economics*, 14:3, 565–573.

Daouli, J., M. Demoussis and N. Giannakopoulos. 2009. "Sibling-Sex Composition and Its Effects on Fertility and Labor Supply of Greek Mothers." *Economics Letters*, 102:3, 189–191.

De Walque, D. 2007. "Does Education Affect Smoking Behaviors? Evidence Using the Vietnam Draft as an Instrument for College Education." *Journal of Health Economics*, 26:5, 877–895.

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.

Eide, E. R. and M. H. Showalter. 2011. "Estimating the Relation Between Health and Education: What Do We Know and What Do We Need to Know?" *Economics of Education Review*, 30:5, 778–791.

Etile, F. and C. Milcent. 2006. "Income-Related Reporting Heterogeneity in Self-Assessed Health: Evidence from France." *Health Economics*, 15:9, 965–981.

Grimard, F. and D. Parent. 2007. “Education and Smoking: Were Vietnam War Draft Avoiders Also More Likely to Avoid Smoking?” *Journal of Health Economics*, 26:5, 896–926.

Gronau, R. 1973. “The Effect of Children on the Housewife’s Value of Time.” *Journal of Political Economy*, 81:2, 168–199.

Grossman, M. 1972. “On the Concept of Health Capital and the Demand for Health.” *Journal of Political Economy*, 80:2, 223–255.

Haan, D. M. 2010. “Birth Order, Family Size and Educational Attainment.” *Economics of Education*, 29:4, 576–588.

Hout, M. 1982. “The Association between Husbands' and Wives' Occupations in Two-Earner Families.” *American Journal of Sociology*, 88:2, 397–409.

Imbens, G. W., and J. D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica*, 62:2, 467–475.

Jones, G. W. and B. Gubhaju. 2009. “Factors Influencing Changes in Mean Age at First Marriage and Proportions Never Marrying in the Low-Fertility Countries of East and Southeast Asia.” *Asian Population Studies*, 5:3, 237–265.

Kalmijn, M. 1991a. “Status Homogamy in the United States.” *American Journal of Sociology*, 97:2, 496–523.

Kalmijn, M. 1991b. “Shifting Boundaries: Trends in Religious and Educational Homogamy.” *American Sociological Review*, 56:6, 786–800.

Kalmijn, M. 1994. “Assortative Mating by Cultural and Economic Occupational Status.” *American Journal of Sociology*, 100:2, 422–452.

Kemptner, D., H. Jürges, and S. Reinhold. 2011. “Changes in Compulsory Schooling and the Causal Effect of Education on Health: Evidence from Germany.” *Journal of Health Economics*, 30:2, 340–354.

Lee, G. H. Y., S. P. Lee. 2014. “Childcare Availability, Fertility and Female Labor Force Participation in Japan.” *Journal of the Japanese and International Economies*, 32, 71–85.

Lefgren, L. and F. McIntyre. 2006. “The Relationship between Women's Education and Marriage Outcomes.” *Journal of Labor Economics*, 24:4, 787–830.

Leigh, A. and C. Ryan. 2008. "Estimating Returns to Education Using Different Natural Experiment Techniques." *Economics of Education Review*, 27:2, 149–160.

Li, H., J. Zhang and Y. Zhu. 2008. "The Quantity-Quality Trade-Off of Children in a Developing Country: Identification Using Chinese Twins." *Demography*, 45:1, 223–243.

Lindeboom, M. and E. V. Doorslaer. 2004. "Cut Point Shifts and Index Shift in Self-Reported Health." *Journal of Health Economics*, 23, 1083–1099.

Lleras-Muney, A. 2005. "The Relationship Between Education and Adult Mortality in the U.S." *Review of Economic Studies*, 72:1, 189–221.

Mare, R. D. 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review*, 56:1, 15–32.

Mazumder, B. 2008. "Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws." *Economic Perspectives*, 33:2, 2–16.

Michael, R. T. 1973. "Education in Nonmarket Production." *Journal of Political Economy*, 81:2, 306–327.

Mishra, V., and R. Smyth. 2010. "Female Labor Force Participation and Total Fertility Rates in the OECD: New Evidence from Panel Cointegration and Granger Causality Testing." *Journal of Economics and Business*, 62:1, 48–64.

Montgomery, M. and J. Trussell. 1986. "Models of Marital Status and Childbearing." Chapter 3 in *Handbook of labor economics* (Vol. I, pp. 205–271). North Holland, Amsterdam: Elsevier Science.

Pagnini, D. L. and S. P. Morgan. 1990. "Intermarriage and Social Distance Among U.S. Immigrants at the Turn of the Century." *American Journal of Sociology*, 96:2, 405–432.

Park, C. and C. Kang. 2008. "Does Education Induce Healthy Lifestyle?" *Journal of Health Economics*, 27:6, 1516–1531.

Reinhold, S. and H. Jürges. 2010. "Secondary School Fees and the Causal Effect of Schooling on Health Behavior." *Health Economics*, 19:8, 921–938.

Rosenzweig, M. R. and K. I. Wolpin. 1980. "Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models." *Journal of Political Economy*, 88:2, 328–348.

Rosenzweig, M. R. and K. I. Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica*, 48:1, 227–40.

Ruggles, S. 2010. "Stem Families and Joint Families in Comparative Historical Perspective." *Population and Development Review*, 36:3, 563–577.

Shmueli, A. 2003. "Socio-Economic and Demographic Variation in Health and in Its Measures: The Issue of Reporting Heterogeneity." *Social Science & Medicine*, 57:1, 125–134.

Silles, M. A. 2009. "The Causal Effect of Education on Health: Evidence from the United Kingdom." *Economics of Education Review*, 28:1, 122–128.

Stevens, G. 1991. "Propinquity and Educational Homogamy." *Sociological Forum*, 6:4, 715–726.

Tsay, W.-J. and C. Y. C. Chu. 2005. "The pattern of birth spacing during Taiwan's demographic transition." *Journal of Population Economics*, 18:2, 323–336.

Tsay, R.-M. and L.-H. Wu. 2006. "Marrying Someone from an Outside Group: An Analysis of Boundary-Crossing Marriages." *Current Sociology*, 54:2, 165–186.

Watson, D., E. C. Klohnen, A. Casillas, E. N. Simms, J. Haig, and D. Berry. 2004. "Match Makers and Deal Breakers: Analyses of Assortative Mating in Newlywed Couples." *Journal of Personality*, 72, 1029–1068.