

**Empirical Studies of Human Capital Formation:  
The Role of Family, Sibling, and Neighborhood**

Huei-Ling Chen

Dissertation submitted to the Graduate Faculty of  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for  
the degree of

DOCTOR OF PHILOSOPHY  
in  
Economics

Djavad Salehi-Isfahani, Chairman  
Charles Michalopoulos  
Stuart Rosenthal  
Gautam Hazarika  
Russell Murphy

May 1, 1998  
Blacksburg, Virginia

Keywords: sibling effect, neighborhood effect, human capital, household  
model, rate of return

Copyright 1998, Huei-Ling Chen

# **Empirical Studies of Human Capital Formation: The Role of Family, Sibling, and Neighborhood**

Huei-Ling Chen

(Abstract)

The formation of human capital is the main issue in this dissertation. More specifically, this dissertation discusses two alternative types of transferring human capital, in contrast to the transfer of human capital from parents to their children's education. These two types of transfer are sibling effect and neighborhood effect on children's education. Chapter 1 discusses the sibling effect on children's education, "Household Models and Formations of Human Capital with Sibling Effect in Iran." The neighborhood effect on children's education will be discussed in Chapter 2, "Intergenerational Transfer of Human Capital from Parents to Children: Does Neighborhood Matter?" Chapter 3 measures and describes the rate of return on human capital in Taiwan, "Rate of Return on Education by Using Sibling Data from Taiwan."

My empirical results show the following findings. First, the presence of older sisters increases younger sisters' and brothers' education. After controlling the resource contribution factor, the empirical result suggests that a role model effect exists between daughters -- the education of older sisters benefits younger sisters in Iran. Second, assuming that the unobserved parent's preference on children's education is not correlated to the unobserved parent's preference on neighborhood, our results suggest that choosing a "good" neighborhood is important for children's education in Iran. Third, the results suggest that in 1990 data from Taiwan, upward bias in the rate of return on schooling due to the omission of family background factors is significant.

## Acknowledgments

The author wishes to express her appreciation and gratitude to her advisor, Dr. Djavad Salehi-Isfahani, for his guidance, encouragement, and support throughout the dissertation process.

The author is also grateful to

Dr. Charles Michalopoulos, for his kind cooperation and for critically reviewing this dissertation.

Dr. Stuart Rosenthal, for his invaluable advice during the study.

Dr. Gautam Hazarika, for his friendship, helpful comments, and assistance in the proofreading of this dissertation.

Dr. Russell Murphy, for his suggestions and support during my defense.

Dr. Jin-Tan Liu, for his instructing and releasing the precious data from Taiwan.

Dr. Brad Mills, for his helpful comments and suggestions.

My friends, Shu-Wei Yang and Ajay Tandon for their helpful proofreading and discussion.

And my associates, Wai-Hong Ho, Fu-Sheng Hung, Hrachya Kyureghian, Gayaneh Kyureghian, Alexander Outkin, Fatma Aksal, Robert Bowles, and Sudipta Sarangi, for their support and friendship.

All of the members in the weekly meeting of applied econometrics “coffee hour” in the Department of Economics, Virginia Tech.

Finally, I would like to dedicate this dissertation to my dear mother, Yueh-Shya Chen.

H.L.C.

## Table of Contents

Abstract.....	ii
Acknowledgment.....	iii
Table of Contents.....	iv
List of Tables.....	vi
Preface.....	1

### **Chapter 1: Household Models and Formations of Human Capital with Sibling Effect in Iran**

I. Introduction.....	4
II. Data.....	6
III. Model.....	8
IV. Econometric Methodology.....	12
IV.1 Method 1: Dummy for Presence of Siblings.....	13
IV.2 Method 2: Effect of Sibling's Education.....	15
V. Results.....	16
V.1 Do Elder Siblings Increase Younger Sibling's Education?.....	16
V.2 How Do Elder Siblings Increase Younger Sibling's Education? By Role Model .....	20
VI. Conclusion.....	21
Appendix.....	22
References.....	25

### **Chapter 2: Intergenerational Transfer of Human Capital from Parents to Children: Does Neighborhood Matter?**

I. Introduction .....	37
II. Econometric Methodology .....	39
III. Data .....	42
IV. Results .....	45

IV.1	The estimated <i>total</i> influence from parents' education .....	47
IV.2	The estimated <i>direct</i> influence from parents' education .....	48
IV.3	The influence from neighborhood characteristics .....	49
IV.4	Simultaneity problem due to a too narrowed scope of neighborhood .....	50
IV.5	Selection Effect or Neighborhood Effect .....	51
IV.6	Parents' Moving for Their Children? .....	52
V.	Conclusion .....	54
	References .....	55

### **Chapter 3: Rate of Return to Education by Using Sibling Data in Taiwan**

I.	Introduction .....	65
II.	Theory Background .....	66
III.	Econometric Methodology .....	67
IV.	Data .....	71
V.	Results .....	73
	V.1 OLS Regression .....	73
	V.2 Unobserved Ability (Family Factors) .....	74
	V.3 Measurement Error .....	76
	V.4 Summary .....	78
VI.	Conclusion .....	78
	References .....	79

	Appendix A: Documentation for Editing the Iranian Panel Data Set.....	90
--	---	----

	Vita .....	95
--	------------	----

## **List of Tables**

### **Chapter 1: Household Models and Formations of Human**

#### **Capital with Sibling Effect in Iran**

Table 1:	Summary Statistics: Iranian Household Survey Data.....	26
Table 2:	Mean Year of Complete Education for Girl and Boy By Number and Sex of Siblings.....	28
Table 3:	Fixed Effect Result - The Effect from the Birth Order on Education Year .....	29
Table 4:	OLS/Tobit/Fixed Effect Results - The Effect from the Presence of (Elder/Younger) Brothers and Sisters on Respondents' Education Year .....	30
Table 5:	OLS/Tobit Results - The Effect from the Education Year of (Elder/Younger) Brothers and Sisters on Respondents' Education Year .....	31
Table 6:	Nonlinear Regression Result - Sibling's Education Effect on Education Year .....	32
Table 7:	OLS Result - The Effect from the Presence of (Elder/Younger) Brothers and Sisters and Age Difference on Respondents' Education Year .....	33
Table 8:	Tobit Result - The Effect from the Presence of (Elder/Younger) Brothers and Sisters and Age Difference on Respondents' Education Year .....	34
Table 9:	Fixed Effect Result - The Effect from the Presence of (Elder/Younger) Brothers and Sisters and Age Difference on Respondents' Education Year .....	35

### **Chapter 2: Intergenerational Transfer of Human Capital from**

#### **Parents to Children: Does Neighborhood Matter?**

Table 1:	Description of Variables .....	57
Table 2:	Average Socioeconomic Characteristics of Household and	

	Neighborhood with Respect to Rural/Urban and Moving Status .....	58
Table 3:	ANOVA Analysis of Household Socioeconomic Characteristics of Neighborhood: Cluster .....	59
Table 4:	Neighborhood Impact on Girl's Completed Years of Schooling.....	60
Table 5:	Neighborhood Impact on Boy's Completed Years of Schooling.....	61
Table 6:	Neighborhood Impact on Girl's and Boy's Completed Years of Schooling - The Effect from Duration Time in Current Address .....	62
Table 7:	Moving Decision (Probit Analysis) .....	63
Table 8:	Probit Analysis of Moving to a Better Area on the Presence of Children in Specific Age Groups (6-9, 10-14, and 15-18) while Moving .....	64

### **Chapter 3: Rate of Return to Education by Using Sibling Data in Taiwan**

Table 1:	Summary Result of the Samples of Brother Pairs and Sister Pairs .....	81
Table 2:	Mean of the Log Wage Rate of the Samples of Brother Pairs and Sister Pairs .....	82
Table 3:	The OLS Estimation for Rate of Return on Education for the Samples of Brother Pairs and Sister Pairs .....	83
Table 4:	The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (Without Measurement Error) for Brother-Pair Sample .....	84
Table 5:	The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (Without Measurement Error) for Sister-Pair Sample .....	85
Table 6:	The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.05) for Brother-Pair Sample .....	86
Table 7:	The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.05) for Sister-Pair Sample .....	87

Table 8: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.20) for Brother-Pair Sample .....88

Table 9: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.20) for Sister-Pair Sample .....89

## Preface

The formation of human capital is the main issue in this dissertation. More specifically, this dissertation discusses two alternative types of transferring human capital, in contrast to the transfer of human capital from parents to their children's education. These two types of transfer are sibling effect and neighborhood effect on children's education. Chapter 1 discusses the sibling effect on children's education, "Household Models and Formations of Human Capital with Sibling Effect in Iran." The neighborhood effect on children's education will be discussed in Chapter 2, "Intergenerational Transfer of Human Capital from Parents to Children: Does Neighborhood Matter?" Chapter 3 measures and describes the rate of return on human capital in Taiwan, "Rate of Return on Education by Using Sibling Data from Taiwan."

The introduction of each chapter is below.

Chapter 1: The purpose of this chapter is to empirically discern the *role model* effect among siblings. If the role model effect exists from elder siblings to their younger siblings, then the younger sibling's schooling achievement would be superior to his/her peers who do not have elder siblings. Furthermore, issues pertaining to the gender of the children in the family might influence the role model effect. This study extends Behrman, Pollak, and Taubman (1995)'s framework by formulating a theoretical model in which younger siblings benefit from older siblings education. This study then derives predictions regarding how parents allocate family resources on each child's education in the presence of this additional factor. The analysis demonstrates that the positive externality from the first child onto the second child will reduce the reinforcing policy of parents though might not turn over to the compensating policy. In order address the above-mentioned questions, a complete set of household data is needed for the empirical analysis. This study uses data derived from household surveys conducted in Iran in 1987. This paper uses several empirical methods, OLS, Tobit, Fixed Effect, and Nonlinear Least Squares method. My empirical results show that the presence of older sisters increases younger sisters' and brothers' education. After controlling the resource contribution factor, the

empirical result suggests that a role model effect exists between daughters -- the education of older sisters benefits younger sisters.

Chapter 2: The fundamental question raised in Chapter 2 is whether the impact of parental decisions with regard to the choice of neighborhoods on children's educational attainment is more important than the impact from parents spending time with children. The methodology used to test this hypothesis is as follows: if neighborhood choice is an important factor impacting children's educational attainment, then controlling for neighborhood factors will decrease the impact of parental education on children's education. On the other hand, if parents spending time with children is more important than choosing a good neighborhood for children's education, then the coefficients of parental education should not decrease much or become less significant after controlling for neighborhood factors. Assuming that the unobserved parent's preference on children's education is not correlated to the unobserved parent's preference on neighborhood, our results suggest that choosing a "good" neighborhood is important for children's education. Our results also indicate that the neighborhood effect is larger in a rural area than in an urban area.

Chapter 3: The major concern of Chapter 3 is the determination of whether the correlation between earnings and education is due to the correlation between family background and education. In addition, this study poses two other questions. First, which gender has the higher rate of return on education, and how is investment on human capital efficiently allocated between the genders? Second, how sensitive are the estimates of rate of return on education to the presence of measurement error? As suggested by Ashenfelter and Zimmerman (1997) and recent literature in household economics, sibling data maybe treated as a kind of "panel data" because siblings share a common unobservable family factor. Based on this framework, this study adopts a new estimation method: minimum distance estimation (MD). The results suggest that in 1990 data from Taiwan, upward bias in the rate of return on schooling due to the omission of family background factors is significant. The OLS estimation of return on schooling is

downward biased due to conjectured 0.05 level of measurement error and omitted family factors for males. For females, OLS estimation of return on schooling is upward biased due to conjectured 0.05 level of measurement error and omitted family factors.

# **Chapter 1: Household Models and Formations of Human Capital with Sibling Effect in Iran**

## **I. Introduction**

The importance of siblings on a child's education is well chronicled in the literature in labor and development economics. Major factors considered in several studies include the impact of siblings with regard to birth order, gender, and spillover effects<sup>1</sup>. However, explicit considerations regarding the role model effect of siblings on children's educational attainment remains largely unexplored. According to Becker, the total number of children in a family represents an opportunity cost for children's education. This theory concludes that having more siblings would decrease a given child's educational attainment. In contrast to this “quantity-quality” trade-off, the role model effect addresses the quality aspects of siblings on children's education. In reality, siblings may not always be a negative factor on each other's education. Children with better attainment could be other children's role models -- i.e., having higher “quality” siblings may enhance the educational attainment of the other children in a given family. Hence, the role model framework underscores the positive impact among siblings. In contrast to Becker's formulation, this study refers to the positive impact as the “quality-quality” complementary.

Butcher and Case (1994) have considered “spillover effects” from elder brothers to younger sisters. Basically the meaning of spillover effect is that parents treat their daughters in a similar fashion as her elder brothers. So girls could receive more education if she has older brothers when educational attainment is a “masculine” trait. On the other hand, a boy with older sisters receives a more “feminine” trait and less education on average than a boy with older brothers. In other words, the spillover effect decreases a boy's education when he has older sisters and increases a girl's education when she has older brothers. While spillover effect is dependent upon parental decisions, the role model

---

<sup>1</sup> The empirical papers include Garg and Morduch (1996), Sassler (1995), Butcher and Case (1994), Parish and Willis (1993), Pitt, Rosenzweig, and Hassan (1990), etc. The theoretical papers include Becker (1991), Behrman, Pollack, and Taubman (1995).

effect emphasizes children's decisions: younger siblings could learn from their elder siblings<sup>2</sup>.

The purpose of this study is to empirically discern the role model effect among siblings. If the role model effect exists from elder siblings to their younger siblings, then younger sibling's schooling achievement would be superior to his/her peers who do not have elder siblings. In addition, parents could decide a more efficient way to invest in their children's education. For example, investing resources in the elder children's education might be efficient if this investment increases not only the elder children's but also the younger children's schooling achievement, and hence their human capital.

Furthermore, this role model effect might be influenced by issues pertaining to the gender of children in the family. For instance, one important issue is to determine whether differentials exist in the effects of the role model effect based upon gender: are boys more likely to be a role model to siblings than girls are, or siblings benefit more from brothers or sisters?

Besides the role model effect, younger siblings could benefit from their older siblings in other ways. First, earlier born children might contribute their income to family resources, so the latter born children could get more education under a more flexible budget constraint. Second, the education of elder siblings may reduce the cost of schooling for younger siblings.

In order to address the above-mentioned questions, a complete set of household data is needed. I use data derived from household surveys conducted in Iran in 1987. This data set reports on socio-economic variables for every individual living in the same household. I examine how children's education responds to differences in sibling situations by estimating the impact on children's educational attainment from the presence or education of his/her elder or younger and same sex or cross sex sibling.

Behrman, Pollak, and Taubman (1995) elaborate on how household resource distribution could depend upon the ability/talents of children. They conclude that under certain assumptions parents adopt the reinforcement strategy to invest more in the

---

<sup>2</sup> It is possible that a younger sibling who has a better achievement record is the elder sibling's role model. This study does not remove this possibility but it might be rare in reality.

education of the children who are better endowed. I extend their framework by formulating a theoretical model in which younger siblings benefit from the older siblings' education. I then derive predictions regarding how parents allocate family resources on each child's education in the presence of this additional factor. I propose a positive externality from older to younger siblings. Suppose the rate of return on the education of younger siblings increases with an increase in the educational investment for older siblings. The analysis in this paper demonstrates that the positive externality from the first child onto the second child will reduce the reinforcing policy of parents though might not turn over to the compensating policy.

My empirical results show that the presence of older sisters increases younger sisters' and brothers' education. The coefficient of younger same-sex siblings education is less than that of older same-sex siblings education. After controlling the resource contribution factor, the empirical result suggests that there exists a role model effect between daughters -- the education of older sisters benefits younger sisters.

## **II. Data**

This study uses a panel data set derived from the Social and Economic Characteristics of Households in Iran (SECHI) survey. These survey data collected in 1987, contain detailed questions from households regarding individual-level data such as age, education year, gender, working hours, wage rate, and enrollment status for every individual in the household. In addition, wives report their age at first marriage, number of children ever born, number of children who live inside or outside household, respectively. This study generates several household-level variables from individual-level data: total income, total expenditure, parents education, number of children who left household. The original data set also includes household-level background information such as assets and expenditures. Besides the above household data, we generate sibling data from the children who lived in the household from individual-level data.

There are two drawbacks in the data: first, we don't have the individual information of the children who left the household, and hence, the sibling data come from only siblings who physically lived in the household. The second drawback of using the

entire data set is that some children are still in school so their completed education is unknown.

This study uses two subsamples to examine the effect of siblings on the educational attainment of children. One subsample is limited to families in which all the children have completed their education. The other subsample adds respondents whose siblings are students. The results of the two subsamples are very similar, though the size of the second subsample is larger than that of the first subsample.

Summary statistics of the whole sample and these two subsamples are presented in Table 1 by gender. Columns (1) and (2) are male and female data when respondents and their siblings have all completed their education. Columns (3) and (4) are male and female data when respondents have completed their education, but their siblings may or may not have completed theirs. Columns (5) and (6) include all respondents. The mean age of respondents in the two subsamples is higher than those in the entire sample. For example, for male respondents [in column (1) and (3)], the mean ages are 19.5 and 18.2, respectively. In column (5), the mean age is only 10.2. The lower mean age in the entire sample shows that the respondents who have not completed their education are younger than those who have. For this reason, I use the second subsample (respondents have completed their education but their siblings may or may not have completed theirs) for regressions to estimate the effect of siblings on children's education. There is a drawback in using the second subsample: it is likely that there are older siblings who may have left the household upon marriage, for instance. This shortage appears in the subsample: the average number of children living away from the family is 0.738 for male data and 0.623 for female data in the whole sample (column (5) and column (6)), and the same variable is 1.492 for male data and 1.216 for female data in subsample (column (3) and column (4)).

According to the literature [Butcher and Case (1994), Garg and Morduch (1996)], the sex composition of siblings has a significant effect on the educational attainment of children. To read the effect from the sex composition of siblings on Iranian children, Table 2 presents the mean completed education of respondents by number and sex of siblings. In most families, boys with only one sister receive significantly more education than boys without any sisters. An increase in the number of sisters beyond one decreases

the boys' education. The same tendency exists in female data: an increase in the number of sisters decreases girls' mean education. A possible reason for this finding is that more sons means an increase in the family labor force which increases family revenue and therefore benefits all siblings. This reason may reasonably explain why having more sisters (fewer brothers) decreases mean education for both boys and girls. There is an explanation for the finding that boys with one sister get more education than boys without any sisters. The most popular reason is that parents expect their sons to receive a higher wage rate than their daughter so they invest in more education for their sons rather than their daughters<sup>3</sup>. This effect outweighs the decrease in resources per child. With more than one daughter, however, scarce resources dominate.

### **III. Model**

Behrman, Pollak, and Taubman (1995) consider a one-period model to analyze the effect of allocation of the family resources on the distribution of children education ( $S_i, i = 1, \dots, n$ ) among their different talents. They assume that the parental utility function is separable into their own consumption and children's expected earnings ( $C_i, i = 1, \dots, n$ ), so the parental preference for children's earnings can be represented by a separable wealth function,  $U(C_1, \dots, C_n)$ .

Parents face a money budget constraint in which their income equals the value of their own consumption and the cost of their children's education. To focus on the problem of education distribution, not only parental consumption but financial bequest is ignored or assumed to enter the parents' utility function separable from children earnings.

Behrman, Pollak, and Taubman then assume that children's expected earnings depend on their own endowment ( $E_i, i = 1, \dots, n$ ) and education. The earnings function is defined in the Cobb-Douglas formulation:  $C_i = S_i^\alpha E_i^\beta$ , where  $0 < \alpha < 1$  because of the decreasing marginal rate of return, and  $\beta > 0$  because better endowment augments the rate of return.

---

<sup>3</sup> Many studies consider this reason, for example: Parish and Willis (1993), Butcher and Case (1994).

The parental maximization problem in their setting is to decide on the education of each of their  $n$  children to maximize the parental subutility function (1), subject to the partial budget constraint (2) and the earnings budget frontier (3). Assume the subutility function is a CES function and parents are neutral among their children.

$$(1) \quad U = U(C_1, \dots, C_n) = \sum_{i=1}^n C_i^a$$

$$(2) \quad C_i = S_i^\alpha E_i^\beta \quad i = 1, \dots, n$$

$$(3) \quad \sum_{i=1}^n P_i^S S_i < T$$

where  $P_i$  is the price of schooling and  $T$  is the total expenditure for the partial budget constraint. Schooling cost is assumed equal among different children,  $P_i = P_j = P$ . Further, assume that marginal utility decreases as children's earnings increase,  $0 < a < 1$ .

This paper augments the framework from Behrman, Pollak, and Taubman (1995) by incorporating a “role model” issue. For the purpose of analyzing the impact of siblings (or the “sibling effect”), I assume a representative household with two children. Note that this assumption fixes family size, yet in reality the family size is endogenous. This assumption simplifies our model to help focus on the discussion on the role model issue. The expected earnings of the first child is  $C_1$  which is then defined in equation (4). To consider the positive externality from the elder sibling’s education on the younger sibling's rate of return, I assume that the expected earnings of the second child,  $C_2$ , depends not only on his/her own schooling investments,  $s_2$ , but also on that of the first child,  $S_1$  [See equation (5)].

$$(4) \quad C_1 = S_1^\alpha E_1^\beta \quad , \quad 0 < \alpha < 1 \quad , \quad \beta > 0$$

$$(5) \quad C_2 = S_2^\alpha E_2^\beta S_1^\gamma \quad , \quad 0 < \gamma < 1$$

Where  $0 < \gamma < 1$  because the first child's education could increase the second child's marginal rate of return. Parents maximize utility function (1), subject to (4), (5), and the partial budget constraint (3). The endogenous variables are  $S_1, S_2, \lambda$ , where  $\lambda$  is the Lagrangian multiplier. The exogenous variables are  $E_1, E_2, \alpha, \beta$  and  $\gamma$ .

The first order conditions are

$$(6) \quad \alpha\alpha S_1^{a\alpha-1} E_1^{a\beta} + a\gamma S_2^{a\alpha} E_2^{a\beta} S_1^{a\gamma-1} - \lambda P = 0$$

$$(7) \quad \alpha\alpha S_2^{a\alpha-1} E_2^{a\beta} S_1^{a\gamma} - \lambda P = 0$$

$$(8) \quad T - PS_1 - PS_2 = 0$$

Solving for the equilibrium ratios of schooling and earnings from the first-order conditions,

$$(9) \quad \frac{S_1}{S_2} = \left( \frac{E_1}{E_2} \right)^{\frac{\beta}{1-\alpha}} \left( \frac{1}{S_1} \right)^{\frac{\gamma}{1-\alpha}} + \left( \frac{\gamma}{\alpha} \right)^{\frac{1}{1-\alpha}} \left( \frac{S_1}{S_2} \right)^{\frac{-\alpha}{1-\alpha}}$$

The clear relationship among the endowment, wage rate, and education couldn't be achieved when the role model exists (See equation (9)). In Behrman, Pollak, and Taubman (1995), there is no interaction between siblings, i.e.  $\gamma = 0$ , and the equilibrium ratio is

$$(10) \quad \frac{S_1}{S_2} = \left( \frac{E_1}{E_2} \right)^{\frac{\beta}{1-\alpha}}$$

Equation (10) indicates that parents adopt a reinforcing strategy to invest in their children because greater genetic endowments are associated with higher marginal returns to education ( $\beta > 0$ ). So a better endowed child could get a higher education.

This constrained maximization problem satisfies the second order sufficient condition, so there exists a unique equilibrium.

To obtain the effect of a change in the endowment,  $E_1$  or  $E_2$ , on the education of children,  $S_1$  or  $S_2$ , totally differentiate equation (6)-(8). The effect from the first child's wage rate on his/her own and sibling's education can be readily shown that

$$(11) \quad \frac{dS_1}{dE_1} = \frac{1}{|H|} P^2 a^2 \alpha \beta S_1^{a\alpha-1} E_1^{a\beta-1} > 0$$

$$(12) \quad \frac{dS_2}{dE_1} = \frac{-1}{|H|} P^2 a^2 \alpha \beta S_1^{a\alpha-1} E_1^{a\beta-1} < 0$$

Where  $|H|$  is the determinant of the bordered Hessian matrix which is the second order condition. Clearly the increase of the first child's endowment increases his/her own schooling year and decreases the second child's schooling year.

On the other hand, the effect from the change of the second child's endowment is not as clear as that of the first case. The comparative static of the second child's endowment on his/her own and the sibling's education is shown that

$$(13) \quad \frac{dS_1}{dE_2} = \frac{-1}{|H|} P^2 a^2 \beta S_2^{\alpha-1} E_2^{a\beta-1} S_1^{a\gamma-1} [\alpha S_1 - \gamma S_2]$$

$$(14) \quad \frac{dS_2}{dE_2} = \frac{1}{|H|} P^2 a^2 \beta S_2^{\alpha-1} E_2^{a\beta-1} S_1^{a\gamma-1} [\alpha S_1 - \gamma S_2]$$

Because of the role model effect, the sign of equation (13) and (14) is ambiguous. If we assume  $\gamma = 0$ , then the role model effect does not exist and the result is symmetric with the equation (11) and (12). Even when  $\gamma > 0$  but is comparatively small such that  $\gamma S_2 < \alpha S_1$ , the sign of  $\frac{dS_1}{dE_2}$  still could be negative and the sign  $\frac{dS_2}{dE_2}$  could be positive. In

a word, the positive externality from the first child onto the second child will alleviate the reinforcing policy of parents though might not turn over to the compensating policy.

Another comparative static result is the impact from the role model effect ( $\gamma$ ) to children's year of schooling. We expect that an elder sibling has more chance to receive a higher education when his/her education could benefit his/her younger sibling's rate of return on education, compared to the child who could not benefit his/her younger sibling.

$$(15) \quad \frac{dS_1}{d\gamma} = \frac{-1}{|H|} P^2 a S_2^{\alpha-1} E_2^{a\beta} S_1^{a\gamma-1} [a(\alpha S_1 - \gamma S_2) \ln S_1 - S_2]$$

$$(16) \quad \frac{dS_2}{d\gamma} = \frac{1}{|H|} P^2 a S_2^{\alpha-1} E_2^{a\beta} S_1^{a\gamma-1} [a(\alpha S_1 - \gamma S_2) \ln S_1 - S_2]$$

The sign of the equation (15) and (16) could not be decided, so we could not conclude that the first child could get more education as he/she has the positive influence on his/her younger sibling.

To distinguish the “role model” and the “spillover effect,” this study amends its model to analyze “spillover effect.” Consider a case which the first child is a boy and the second child is a girl and assume that parents have different preference among boys and girls, presumably higher preference for boys. The utility function is:

$$(17) \quad U = U(C_1, C_2) = C_1^a + C_2^b$$

where  $a > b$  (because parents prefer boys) without considering spillover effect. Note that the comparative statics of  $a$  on  $S_1$  and  $b$  on  $S_2$  are positive. If spillover effect exists, parents treat their daughter with “masculine trait” and hence  $b$  increases and reaches  $a$ . A girl will receive a higher level of education if she has an older brother.

However, spillover effect fails to explain the other possible positive impacts: from the presence of elder sisters to younger brothers and sisters and from the presence of elder brothers to younger brothers.

#### **IV. Econometric Methodology**

To build a bridge between the theoretical model and the empirical work, this study simplifies the children's education function from the literature [Behrman, Pollak, and Taubman (1995), Becker (1991)] and the above analysis.

Children's education could be represented as a production function of individual characteristics and common family factors,

$$(18) \quad y_{ij} = f(x_{ij}, F_j)$$

where  $ij$  refers to the  $i$ th child in the  $j$ th household,  $y_{ij}$  is education (in years),  $x_{ij}$  are individual characteristics, (e.g., ability, preference for education), and  $F_j$  includes household factors (e.g., household income, parent's preference for children's education, own consumption, number of children, and neighborhood environment). If we consider the role model effect in the production function, and assume there are two children in the family. The production of the first and second child's education could be as below:

$$(19) \quad y_{1j} = A_1 + \beta x_{1j} + \delta F_j$$

$$(20) \quad y_{2j} = A_2 + \beta x_{2j} + \delta F_j$$

where  $A_1$  and  $A_2$  are the first and the second child's education due to the birth order effect, after controlling for other factors. If role model effect exists,  $A_1$  is smaller than  $A_2$  because the second child could benefit from his/her elder sibling's better education in addition to parents planning education investment.

Though this method is direct and clear, it needs a strong assumption: there is no change in family characteristic among siblings when they were of school age. For

instance, we need to assume the following factors are constant: household income, budget constraint, neighborhood environment, social cultural, and parent's preference. Based on these assumptions, we can use fixed effect model to absorb the family characteristics.

Another way to describe siblings' impacts on the children's education attainment is as follows:

$$(21) \quad y_{ij} = X_{ij}\beta + S_{ij}\gamma + \delta F_j + \varepsilon_{ij}$$

where  $y_{ij}$  is education (in years) of the  $i$ th child in the  $j$ th household,  $X_{ij}$  is a vector of individual characteristics (including age, age squared, and birth order),  $S_{ij}$  is a vector of sibling characteristics, and  $F_j$  includes household factors (e.g., log of household income, mother's education, father's education, number of children, number of children who live outside the household, and living area). The unobserved error term,  $\varepsilon_{ij}$ , is distributed normally with mean 0 and variance  $\sigma^2$ .

To estimate the existence of the role model effect, this study applies two different approaches

#### IV.1 Method 1: Dummy for Presence of Siblings

The first approach is to use the presence of the siblings for  $S_{ij}$ , the sibling characteristics<sup>4</sup>. Sibling variables include a dummy variable for having any elder brothers ( $S_{ij}^{ob}$ ), a dummy variable for having any elder sisters ( $S_{ij}^{os}$ ), a dummy variable for having any younger brothers ( $S_{ij}^{yb}$ ), and a dummy variable for having any younger sisters ( $S_{ij}^{ys}$ ).

Equation (21) is rewritten:

$$(22) \quad y_{ij} = \beta x_{ij} + \gamma_1 S_{ij}^{ob} + \gamma_2 S_{ij}^{os} + \gamma_3 S_{ij}^{yb} + \gamma_4 S_{ij}^{ys} + \delta F_j + \varepsilon_{ij}$$

As mentioned in the [I. Introduction](#), there are two modes through which older siblings could impact a younger sibling's educational attainment: as role models or as resource providers. Both of these modes result in a larger impact from the presence of older siblings to the younger sibling's educational attainment than the other direction. That is,

---

<sup>4</sup> This method has been adopted in the literature of sibling effect. Butcher and Case (1994) estimate the influence of the presence of brothers and the sisters, as well. Garg and Morduch (1996) estimate the influence of at least one brother.

$\gamma_1$  and  $\gamma_2$  would be greater than  $\gamma_3$  and  $\gamma_4$ . To absorb the influence from the resource providers, we add dummies for older brothers' income, older sisters' income, younger brother's income, and younger sister's income in the explanatory variables.

Note that two drawbacks exist in the data. First, detailed information regarding the children who left the household is not reported, and hence sibling data refers only to siblings who physically live in the household. Second, some children are still students so their completed education is unknown.

To correct the first sample selection bias, researchers may use two methods. One method is adding the number of children who live outside the household as an independent variable in order to consider the influence from unknown characteristics of siblings who are not living in the household. Another method is to use a fixed effect model to absorb family characteristics. Assume these children are the oldest kids in the household, and every respondent from the same family receives the same impact from these unknown elder siblings. Hence the influence from the unknown common impact from the eldest siblings away from the family could be removed by differencing out their effects.

For the second potential bias, we run a Tobit estimation model to allow for censored student status<sup>5</sup>.

---

<sup>5</sup> Assuming that decisions regarding student status and years of completed schooling have the same explanatory variables. The model is:

$$y^* = z'\alpha$$

where  $z = \{x, S, F\}$  and  $\alpha = \{\beta, \gamma, \delta\}$

$$d = 1 \text{ (nonstudents) } \quad (i \in N)$$

$$d = 0 \text{ (students) } \quad (i \in S)$$

observe

$$y = y^* \quad \text{if } d = 1$$

$$y \geq y^* \quad \text{if } d = 0$$

( $y$  = years of schooling).

The log-likelihood function is

$$\log L = -\frac{1}{2} \sum_{i \in N} \left[ \left( \frac{y_{ij} - z'\alpha}{\sigma} \right)^2 + \log 2\pi\sigma^2 \right] + \sum_{i \in S} \log \left[ 1 - \Phi \left( \frac{y_{ij} - z'\alpha}{\sigma} \right) \right]$$

## IV.2 Method 2: Effect of Sibling's Education

The second approach is to replace the presence of a sibling as an explanatory variable with average sibling's education under the same four categories: older brother, older sister, younger brother, and younger sister<sup>6</sup>.

In poorer families, older siblings might leave school in order to work, and their younger siblings might receive a higher education because the extra earnings of the older sibling pay for the younger sibling's education. Likewise, if an older sibling stays in school, that may take away many resources that the younger sibling has to leave school earlier -- again a negative relationship between the older sibling's education and the younger sibling's. However, if there is a large enough age difference, then an older sibling has enough time to get more education *and* to find a job in order to contribute resources back to the family.

That suggests the following pattern, if resources are important:

### **Poor families —**

if siblings are close in age then

higher education for older → less education for younger

if siblings are far apart in age then

higher education for older → higher education for younger

### **Wealthy families —**

no difference in effect by age difference

The above pattern implies that we can use the relationship between older and younger siblings' education levels to determine the effect of role models, and use the difference between that and the relationship in poorer families to infer the role of resources.

Directly entering sibling education into the explanatory variables in the above OLS regression to measure education outcomes will create some bias. There are at least two children in the same household who simultaneously enter their education in the dependent

---

<sup>6</sup> This methodology has been used in the neighborhood effect paper such as Case and Katz (1991).

variable and enter each other's education in the explanatory variables. It implies that children's education influences their sibling's education while at the same time being influenced by these siblings' education so the explanatory variable is correlated with the random error. To correct this bias, I use a nonlinear regression to estimate equation (2)<sup>7</sup>.

## **V. Results**

This section is organized as follows: First, this study examines the existence of a positive effect between elder siblings and younger siblings. This extra positive effect (if extant) could be either a role model effect or a resource contribution effect. A positive effect from elder siblings to younger siblings is expected. The second stage is to distinguish between the role model effect and the resource effect.

### **V.1 Do Elder Siblings Increase Younger Sibling's Education?**

There are three ways to test if elder siblings help increase the younger sibling's education or vice versa. The first method is to apply *birth order effect* to the respondent's education (reported in Table 3). The second method is to compare the effect of *the presence of siblings* to children's education (reported in Table 4). The third method is to compare the effect of the *sibling's education* to children's education (reported in Table 5 and Table 6).

Table 3 reports the birth order effect. The sample consists of respondents who are not in school. We estimate the impact from both birth order and gender. Three dummy variables are generated for respondents: the second oldest child, the third oldest child, and the child who was born after three elder siblings. The constant term in Table 3 represents the oldest child. The other independent variables include the respondent's age and the squared term of age.

Table 3 shows that earlier born boys receive more education than later born boys. In contrast, later born girls receive more education than the earlier born girls. For example, a boy who is the second oldest child receives 0.514 less years of schooling than a boy who is the oldest child in the family. A boy who is the third oldest child receives

---

<sup>7</sup> The details are displayed in the Appendix.

1.089 less years of schooling than a boy who is the oldest child. On the other hand, a girl who is the second oldest child receives 0.489 more years of schooling than a girl who is the oldest child; a girl who is the third oldest child receives 0.785 more years of schooling than a girl who is the oldest child.

These results show that younger sisters benefit from birth order effect, which is associated with the role model effect and maybe the resource contribution effect as well. On the other hand, brothers benefit from birth order effect in a different direction: older brothers receive more education than younger brothers.

We need to make a strong assumption if we want to draw the conclusion that the above results come from role model effect. The assumption should be that at least the following factors are fixed over time: household income, budget constraint, neighborhood environment, social cultural, and parent's preference.

In Table 4, we compare the impact from the presence of elder siblings and from the presence of younger siblings in order to examine the positive effect. Literature [Butcher and Case (1994), Parish and Willis (1993)] suggests several theories to explain the impact of siblings. These theories imply that the impact from the presence of elder and younger siblings should be equal. One example is the reference group theory that suggests the existence of other sisters would make the parents treat the daughters as a group and give them a different goal which is often accompanied with lower education. Both elder sister and younger sister should have the same negative impact on the respondent's education year. Unlike these theories, a role model effect from elder siblings to younger siblings generates a larger impact from the presence of elder siblings.

Table 4 reports the result. The dummy variable, “the presence of any elder brother”, “the presence of any elder sister”, “the presence of any younger brother”, and “the presence of any younger sister” are the main independent variables. The other independent variables are dummy variable for any elder brother's income, dummy variable for any elder sister's income, dummy variable for any younger brother's income, dummy variable for any younger sister's income, mother's education and squared, father's education and squared, log of household income, the share of mother's income in

household income, living area (rural or urban), age, the squared term of age, number of children living in the household, and number of children living away from the household.

Due to the data selection problem which is described in Section II, Table 4 provides three regression methods, OLS, Tobit, and Fixed Effect model. This study restricts the sample applied in OLS and Fixed Effect model to the respondents who have completed their schooling. The sample restriction might lose some information of the student-status respondents and then raise another sample selection problem. That is, the respondents are more likely to have lower schooling achievement (earlier to be out of school). To justify the sample selection bias, the Tobit model is used to include the information of the student-status respondents.

Another issue pertaining to the data is that the data does not provide detailed information, like age, gender, and education about the siblings who are not living in the household at the time of the survey. The only known information is the number of children living away from the household. In order to record the impact from these siblings, this variable is included in the independent variables list in OLS and Tobit regression. A better way to deal with this problem is to assume these children are the oldest children in the household, and hence every respondent from the same family receives the same impact from these unknown elder siblings. The fixed effect model could absorb the household factors and surely include the unknown common impact from the eldest siblings away from the family.

The result in Table 4 shows a significant impact from the presence of elder sisters and elder brothers to the respondents' education year among three different regression methods for both the boys' and the girls' education. The presence of any younger sisters is not significant and younger brothers is either insignificant in some cases (girl's education in OLS and both genders' education in the Fixed Effect Model) or has smaller impact than the elder siblings (though the impact is significant) in the other cases. The result proves an extra positive impact from the presence of elder siblings. Also, the presence of elder sisters has larger impact than the coefficient from the presence of elder brothers. Noted that the coefficient from the presence of elder sisters to girls' education is prominently significant (this variable is significantly positive in OLS, Tobit, and Fixed Effect).

The positive influence from elder sisters might come from the following reasons: (1) older sisters leave schooling materials at home to save younger siblings' schooling cost or contribute their income to the household, (2) as a role model, they instruct their younger siblings to pursue a higher achievement in education. Since the dummies for older brother's and older sister's income absorb the influence from the resource contribution, the positive significance of older sibling's presence is more likely due to the role model effect. Besides, the presence of an elder sister is not significantly positive for both the boy's and the girl's education (except in the Tobit regression). It might imply that the elder sister is a role model for younger sister.

Another way to measure the sibling's effect is to use a sibling's education as a sibling character instead of using the presence of siblings. Table 5 is the OLS regression and Tobit regression based on the same methodology as Table 4. The sample size is far smaller here than in Table 4 because it only includes respondents who have both older siblings and younger siblings.

The result shows a common tendency for both males and females. The coefficient of younger same-sex siblings' education is less than that of older same-sex siblings' education. In contrast, the coefficient of older cross sex siblings' education is less than that of younger cross sex siblings' education. For example, one elder sister's schooling year correlates with 0.385 increase in female respondent's schooling year; and one more schooling year for the younger sister correlates with 0.263 more for the schooling year of female respondents.

Directly entering sibling education into the explanatory variables in the above OLS regression to measure education outcomes will create some bias. There are at least two children in the same household who simultaneously enter their education in the dependent variable and enter each other's education in the explanatory variables. It implies that children's education influences their sibling's education while at the same time being influenced by these siblings' education, so the explanatory variable is correlated with the random error. To correct this bias, this study uses a nonlinear regression to estimate Table 6. (See the details in Appendix.)

## **V.2 How Do Elder Siblings Increase Younger Sibling's Education? By Role Model Effect or Resource Contribution Effect**

People might suspect that the dummy for sibling's income cannot catch all of the resource contribution effect. The positive impact from the presence of an elder sibling or the elder sibling's education comes from the resource contribution effect. The next step is to distinguish between the role model effect and the resource contribution effect using the method introduced in Section IV. This study will compare the coefficients for different household income levels and age differences among siblings. If the resource contribution effect is an important factor in explaining the benefit of having an elder brother or sister, then larger age differences among siblings could give poor family children an extra positive impact from elder siblings than a smaller age differences among siblings. The reason is because a larger age difference could alleviate the competition among siblings for household resources and also elder siblings could work and contribute their income. Two subsamples based on income levels are used to do the comparative analysis. The high income families are those for which income is at the top quartile of the survey data. In contrast, low income families are those for which income is at the bottom quartile of the survey data. In other words, this study drops the middle-level households. This clean separation helps us observe the income effect more clearly.

Two variables are added to consider the effect of age differences between elder siblings and respondents. This study sets a dummy variable of high age difference between the elder sibling and respondents as one if the average elder sibling's age is at least 3 years older than the respondents' age; otherwise it is 0. The added independent variables are the dummy variable of high age difference between elder brothers (sisters) times the dummy variable of the presence of elder brothers (sisters). The coefficients of these two variables represent the difference of the impact from the presence of elder siblings between high age differences and small age differences.

Table 7 reports the results using OLS. Table 8 reports Tobit results. Table 9 reports Fixed Effect results. Most of the results do not show a significantly different impact from elder siblings due to the age difference in low income family .

The conclusions from the above results imply that there is little evidence of a positive impact from elder sibling's resource contribution. From this study, researchers could conclude that the positive influence from elder siblings, especially from elder sisters, is due to the role model effect. The possible mechanism behind the second reason, which creates a larger impact from the presence of elder sisters than elder brothers is due to societal gender roles in Iranian society. In Iran, most females do not work outside the home and they tend to stay at home until they are married. So, if both the older sister and older brother are educated, the older sister is more potentially able to help younger siblings in their homework. This result shows that there is a human capital mechanism that increases younger siblings endowment and motivates younger siblings to increase their education.

## **VI. Conclusion**

My empirical results indicate some evidence for the presence of a positive sibling effect which depends on gender. The presence of elder sisters has a significant impact on younger sisters. The results imply that there is an additional positive effect from elder sisters to younger sisters. This study refers to this effect as a role model effect because the resource constraint is not a significant factor here. This study also runs a regression of children's education on their siblings mean education. The result shows a common tendency for both males and females: the coefficient of younger same-sex siblings education is less than that of older same-sex siblings education. Since Iranian women have lower labor market participation rate than Iranian men, females have larger likelihood of instructing younger siblings and transferring their human capital.

My summary result shows that an increase in the number of sisters decreases boys education as well as girls mean education. The fact that boys with a sister receive significantly more education than boys without any sisters suggests that parents invest more in boys education if they have daughters. The reason is because having more daughters might tighten family resources and parents would invest less in the education of both sons and daughters.

Though the expected earning in the labor market for women might be lower than that for men, the accumulation of human capital might be more important than the expected earning in the labor market for a developing country. A suggestion based on these results is that parents should invest more in a daughter's education when they are more likely to help their younger siblings and transfer the human capital.

The contribution for the discussion about sibling effects in this paper is that it considers an important factor in the family: the role model effect among siblings. The empirical results suggest the existence of this effect.

### **Appendix**

Table 6 presents the OLS estimators of sibling education on children's education. Basically the model is of the form:

$$(A1) \quad Y = \phi WY + X\beta + u$$

where  $Y$  is an  $(N \times 1)$  vector of years of education,  $u$  is an  $(N \times 1)$  vector of disturbance terms distributed normal  $N(0, \sigma^2)$  but not independent within the same household (Yet it could be assumed to be independent if the household factor is absorbed using a fixed effect model),  $W$  is an  $(N \times N)$  matrix that assigns to each dependent variable the mean of siblings' education, so  $W$  should be expressed as below:

$$W = \begin{pmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & W_m \end{pmatrix}$$

where

$$W_i = \begin{pmatrix} 0 & \frac{1}{n_i - 1} & \dots & \frac{1}{n_i - 1} \\ \frac{1}{n_i - 1} & 0 & & \vdots \\ \vdots & & \ddots & \frac{1}{n_i - 1} \\ \frac{1}{n_i - 1} & \dots & \frac{1}{n_i - 1} & 0 \end{pmatrix}$$

and  $n_i$  is the number of reporters who are from  $i$ th family ( $i = 1 \dots m$ ).  $N = \sum_{i=1}^m n_i$  is the total number of the reporters in the data. So  $\phi$  is the parameter measuring the extent of sibling education influences.

Two major problems exist in this formulation. First, the household effect needs to be absorbed using the fixed effect model. To deal with this problem, this study transform the model to the deviation form. Second, the explanatory variables are not independent of the disturbance term:

$$\begin{aligned} \text{Cov}(WY, u) &= \text{Cov}(W[\phi WY + X\beta + u], u) \\ &= \phi W \text{Cov}(WY, u) + W \cdot \text{Var}(u) \end{aligned}$$

so,  $\text{Cov}(WY, u) = (I - \phi W)^{-1} W \sigma^2 \neq 0$

Researchers could solve the above problem using the IV approach if they could find an instrument  $Z$  that satisfies the following conditions:

- 1)  $\text{Cov}(Z, u) = 0$
- 2)  $\text{Cov}(Z, WY)$  is significant.

Researchers should pick up only “one” instrumental variable to serve this objective and prevent an over identification problem. The drawback is that the choice of a “bad” instrumental variable might cause more bias.

A paper by Case and Katz (1991) offers an alternative method to solve a similar problem (neighborhood effect) without relying on IV method. First, equation (A1) could be written in the following form

$$(A2) \quad Y = (I - \phi W)^{-1} X\beta + (I - \phi W)^{-1} u$$

Here no correlation exists between the disturbance term and the explanatory variables because of the assumption that  $X$  are nonstochastic variables. But the variance of  $Y$  does not satisfy the homoskedasticity assumption for the OLS regression

$$\text{Var}(Y) = (I - \phi W)^{-1} \sigma^2 I (I - \phi W)^{-1}$$

furthermore, equation (A2) is nonlinear in explanatory variables.

Note that the matrix associated with  $X\beta$  can be expanded to

$$(I - \phi W)^{-1} = I + \phi W + \phi^2 W^2 + \dots \approx I + \phi W$$

as long as  $\phi$  is less than one in absolute value. The equation (A2) could be expressed as the following simpler nonlinear form:

$$Y = X\beta + WX\beta\phi + (I - \phi W)^{-1}u$$

Newton-Raphson method could estimate the parameters  $\beta, \phi$  and diminish the heteroskedasticity problem by minimizing the following generalized sum of squares:

$$\begin{aligned}\hat{\theta} &= \arg \min \hat{Q}_N(\theta) \\ &= \arg \min (Y - X\beta - WX\beta\phi)' \cdot VC^{-1} \cdot (Y - X\beta - WX\beta\phi)\end{aligned}$$

where  $\theta = (\beta, \phi)$  and  $VC = (I - \hat{\phi}W)^{-1}\sigma^2(I - \hat{\phi}W)^{-1}$ .

The details of the estimation procedure are as follows:

Set  $\hat{\theta}_1$  is an initial estimator and

$$g_1 = \left. \frac{\partial Q}{\partial \theta} \right|_{\hat{\theta}_1} \quad \text{and} \quad H_1 = \left. \frac{\partial^2 Q}{\partial \theta \partial \theta'} \right|_{\hat{\theta}_1}$$

The second round estimator  $\hat{\theta}_2$  of the Newton-Raphson iteration is obtained by the following

$$\hat{\theta}_2 = \hat{\theta}_1 - H_1^{-1}g_1$$

The iteration is to be repeated until the sequence  $\{\hat{\theta}_n\}$  thus obtained converges. [For a detailed exposition see Amemiya (1985)].

Note the  $g_1$  and  $H_1$  in terms of  $W, Y, X, \beta, \phi$  are:

$$\frac{\partial \hat{Q}(\theta)}{\partial \beta} = \frac{-2}{N} (Y - X\beta - WX\beta\phi) \cdot VC^{-1} \cdot (X + WX\phi)$$

$$\frac{\partial \hat{Q}(\theta)}{\partial \phi} = \frac{-2}{N} (Y - X\beta - WX\beta\phi) \cdot VC^{-1} \cdot WX\beta$$

$$\frac{\partial^2 \hat{Q}(\theta)}{\partial \beta \partial \beta'} = \frac{2}{N} (X + WX\phi)' \cdot VC^{-1} \cdot (X + WX\phi)$$

$$\frac{\partial^2 \hat{Q}(\theta)}{\partial \beta \partial \phi} = \frac{2}{N} [(WX\beta)' \cdot VC^{-1} \cdot (X + WX\phi)]$$

$$\frac{\partial^2 \hat{Q}(\theta)}{\partial \phi \partial \phi'} = \frac{2}{N} (\text{WX}\beta)' \cdot \text{VC}^{-1} \cdot (\text{WX}\beta)$$

The original form of the cross derivative on  $\beta, \phi$  is

$$\frac{\partial^2 \hat{Q}(\theta)}{\partial \beta \partial \phi} = \frac{2}{N} \left[ (\text{WX}\beta)' \cdot \text{VC}^{-1} \cdot (\text{X} + \text{WX}\phi) - (\text{Y} - \text{X}\beta - \text{WX}\beta\phi)' \cdot \text{VC}^{-1} \cdot \text{WX} \right]$$

In order to ensure that  $H_i$  has full rank, the literature generally assumes that the second term is zero [for example see Amemiya (1985)]. This paper also makes the same adjustment.

Using asymptotic theory, the NLLS estimators converge to the asymptotic normal distribution:

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}(0, \Omega^{-1} \Sigma \Omega^{-1})$$

where  $\Omega = E\left(\frac{\partial^2 \hat{Q}(\theta)}{\partial \theta \partial \theta''}\right)$  and  $\Sigma = E\left(\frac{\partial \hat{Q}(\theta)}{\partial \theta} \frac{\partial \hat{Q}(\theta)}{\partial \theta}\right)$ , the consistent estimator for the

variance is  $\frac{1}{N} \sum \left(\frac{\partial^2 \hat{Q}(\theta)}{\partial \theta \partial \theta''}\right)^{-1}$ .

Another method to estimate  $\phi$  is to rewrite the following objective function:

$$\begin{aligned} \hat{\theta} &= \arg \min (\text{Y} - \text{X}\beta - \text{WX}\beta\phi)' \cdot \text{VC}^{-1} \cdot (\text{Y} - \text{X}\beta - \text{WX}\beta\phi) \\ &= \arg \min \frac{1}{\sigma^2} (\text{Y} - \text{X}\beta - \text{WX}\beta\phi)' \cdot (\text{I} - \phi\text{W}) \cdot (\text{I} - \phi\text{W}) \cdot (\text{Y} - \text{X}\beta - \text{WX}\beta\phi) \\ &= \arg \min \frac{1}{\sigma^2} (\text{Y} - \text{X}\beta - \phi\text{WY} + \text{W}^2\text{X}\beta\phi^2)' \cdot (\text{Y} - \text{X}\beta - \phi\text{WY} + \text{W}^2\text{X}\beta\phi^2) \end{aligned}$$

with homoskedasticity variance-covariance matrix. The NLLS method could be used to estimate  $\beta$  and  $\phi$ .

## **References**

1. Becker, Gary S.(1991) , *A Treatise on the Family*. 2nd ed. Cambridge, Mass.: Harvard Univ. Press.
2. Behrman, Jere R., Robert Pollack, and Paul Taubman (1995), *From Parents to Child: Intrahousehold Allocations and Intergenerational Relations in the United States*, University of Chicago Press.

3. Butcher, Kristin F. and Anne Case, (1994) "The effect of Sibling Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, vol. 104, no. 3, August, 531 - 563.
4. Case, Anne C. and Lawrence F. Katz (1991), "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," *NBER working paper*, no. 3705.
5. Garg, Ashish and Jonathan Morduch (1996), "Sibling Rivalry, Resource Constraints, and the Health of Children," Harvard University (draft).
6. Parish, William L., and Robert J. Willis (1993), "Daughters, Education and Family Budgets: Taiwan Experiences," *Journal of Human Resources*, vol. 28 , no. 4, 863-898.
7. Pitt, Mark M, Mark Rosenzweig and Nazmul Hassan (1990), "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries," *American Economic Review*, vol. 80, no. 5, Dec., 1139-1156.
8. Rosenzweig, Mark R. and Paul Schultz (1982), "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India," *American Economic Review*, vol. 72, September, 803-815.
9. Sessler, Sharon (1995), "Trade-Offs in the Family: Sibling Effects on Daughters Activities in 1910," *Demography*, vol. 32, no. 4, Nov., 557 - 575
10. Salehi-Isfahani, Djavad, (1996), "Economics of Fertility in Iran," Virginia Tech. (draft).
11. Salehi-Isfahani, Djavad, (1995), "Intergeneration Transfer of Human Capital: Evidence for Gender Role from Iran," Virginia Tech. (draft).

**Table 1: Summary Statistics; Iranian Household Survey Data  
(SECHI 1987)  
(standard deviation in parentheses)**

Samples	<u>Respondents' and Their Siblings' Complete Education</u>		<u>Respondents Complete Education but Their Siblings' Might Not be</u>		<u>The Whole Sample</u>	
	<u>Boys</u>	<u>Girls</u>	<u>Boys</u>	<u>Girls</u>	<u>Boys</u>	<u>Girls</u>
<b><u>Variables</u></b>						
Education Years	5.285 (4.487)	3.575 (4.638)	5.506 (4.263)	3.498 (4.266)	5.083 (3.595)	4.131 (3.695)
Any Brothers	0.599 (0.490)	0.636 (0.482)	0.836 (0.371)	0.858 (0.349)	0.839 (0.368)	0.835 (0.371)
Any Sisters	0.487 (0.500)	0.619 (0.486)	0.728 (0.445)	0.810 (0.392)	0.784 (0.412)	0.797 (0.402)
Any Elder Brothers	0.309 (0.462)	0.341 (0.475)	0.379 (0.485)	0.447 (0.497)	0.549 (0.498)	0.569 (0.495)
Any Elder Sisters	0.182 (0.386)	0.306 (0.461)	0.237 (0.425)	0.360 (0.480)	0.471 (0.499)	0.511 (0.500)
Any Younger Brothers	0.410 (0.492)	0.432 (0.496)	0.708 (0.455)	0.715 (0.451)	0.549 (0.498)	0.536 (0.499)
Any Younger Sisters	0.362 (0.481)	0.465 (0.499)	0.637 (0.481)	0.678 (0.468)	0.514 (0.500)	0.511 (0.500)
Percent Sisters	0.432 (0.165)	0.754 (0.193)	0.410 (0.158)	0.627 (0.196)	0.419 (0.163)	0.650 (0.200)
Mother's Education	1.032 (2.941)	0.986 (2.765)	0.849 (2.481)	0.783 (2.409)	1.674 (3.352)	1.898 (3.567)
Father's Education	1.879 (3.784)	1.849 (3.748)	1.774 (3.370)	1.820 (3.508)	3.018 (4.219)	3.270 (4.326)
Number of Children Living with the Family	2.864 (1.756)	3.388 (2.072)	4.558 (2.136)	5.038 (2.226)	4.584 (2.015)	4.655 (2.078)
Number of Children Living away the family	2.086 (2.001)	1.564 (1.855)	1.492 (1.754)	1.216 (1.547)	0.738 (1.374)	0.623 (1.229)

Age	19.522 (7.627)	17.206 (9.229)	18.221 (6.631)	15.596 (7.084)	10.181 (7.362)	8.953 (6.926)
Living area (Urban=1)	0.504 (0.500)	0.439 (0.497)	0.534 (0.499)	0.419 (0.494)	0.571 (0.495)	0.571 (0.495)
Birth Order	1.647 (0.956)	1.914 (1.207)	1.881 (1.189)	2.241 (1.421)	2.745 (1.739)	2.882 (1.775)
Observations	544	428	1801	1616	7311	6298

**Table 2: Mean Year of Complete Education for Girl and Boy By Number and Sex of Siblings**

Number	Sex composition	Men	Obs	Women	Obs
0 siblings		6.031	129	4.597	72
1 siblings	0 sisters, 1 brother	6.060	117	5.886	79
	1 sister, 0 brothers	6.845	71	4.949	59
2 siblings	0 sisters, 2 brothers	5.759	112	4.55	60
	1 sister, 1 brother	5.774	133	4.516	93
	2 sisters, 0 brothers	5.127	55	4.273	44
3 siblings	0 sisters, 3 brothers	4.930	71	4.265	49
	1 sister, 2 brothers	6.170	141	3.606	104
	2 sisters, 1 brother	5.963	82	3.5	92
	3 sisters, 0 brothers	4.692	26	2.781	32
4 siblings	0 sisters, 4 brothers	5.405	37	3.318	22
	1 sister, 3 brothers	6.172	76	3.377	69
	2 sisters, 2 brothers	5.033	92	3.193	109
	3 sisters, 1 brother	4.512	41	3.185	65
	4 sisters, 0 brothers	3.556	9	8.556	9

**Table 3: Fixed Effect Result<sup>1</sup>**  
**The Effect from the Birth Order on Education Year**  
 (standard error is in the parenthesis)

<b>Dependent Variables</b>	<b>Boys Education</b>	<b>Girls Education</b>
<b><u>Independent Variables</u></b>		
Constant (the first child in the family)	0.536 (1.413)	-3.767** (0.935)
Dummy (the second child in the family)	-0.514 (0.278)	0.489 (0.253)
Dummy (the third child in the family)	-1.089* (0.459)	0.785* (0.350)
Dummy (born after three children in the family)	-1.829** (0.671)	1.197* (0.490)
Number of observations	1531	1438

\* The coefficient is significant at the 5% level

\*\* The coefficient is significant at the 1% level

1. The data set for fixed effect model include only respondents who already complete their education. The household factor is absorbed in the fixed effect model. And hence the other independent variables include only individual characters: age and the squared term of age.

**Table 4: OLS / Tobit / Fixed Effect Results**  
**The Effect from the Presence of (Elder / Younger) Brothers and Sisters**  
**on Respondents Education Year**  
 (standard error is in the parenthesis)

Models	OLS <sup>1</sup>		Tobit <sup>2</sup>		Fixed Effect <sup>3</sup>	
	Boy's Education	Girl's Education	(censored: students) Boy's Education	(censored: students) Girl's Education	(fixed: household) Boy's Education	(fixed: household) Girl's Education
<b>Independent Variables</b>						
Any Elder Brothers	0.445* (0.225)	0.528* (0.206)	0.691** (0.217)	1.112** (0.215)	-0.099 (0.315)	0.063 (0.378)
Any Elder Sisters	0.302 (0.214)	0.641** (0.195)	0.960** (0.206)	0.899** (0.203)	-0.480 (0.525)	0.693** (0.231)
Any Younger Brothers	0.575** (0.217)	0.017 (0.209)	0.588** (0.219)	0.492* (0.221)	0.360 (0.346)	0.834 (0.444)
Any Younger Sisters	0.205 (0.200)	-0.399* (0.200)	0.083 (0.206)	-0.083 (0.213)	0.347 (0.699)	0.088 (0.259)
Observations	1514	1408	4306/1514	3444/1408	1531	1438

\* The coefficient is significant at the 5% level

\*\* The coefficient is significant at the 1% level

1. The data set for OLS regression includes respondents who already complete their education. The other independent variables include a dummy variable for elder brother's income, a dummy variable for elder sister's income, a dummy variable for younger brother's income, a dummy variable for younger sister's income, mother's education, the squared term of mother's education, father's education, the squared term of father's education, log household income, the share from mother's income among household income, living area (urban or rural), age, squared term of age, number of children living in the household, and number of children living away from the household.

2. The data set for Tobit regression include all of the respondents. The student status is censored. The number on the left side of slash "/" is the number of all of the respondents and on the right side is uncensored number. The other independent variables are same as in the OLS regression.

3. The data set for fixed effect model include only respondents who already complete their education. The household factor is absorbed in the fixed effect model. And hence the other independent variables include a dummy variable for elder brother's income, a dummy variable for elder sister's income, a dummy variable for younger brother's income, a dummy variable for younger sister's income, and individual characters: age and the squared term of age.

**Table 5: OLS / Tobit Results**  
**The Effect from the Education Year of (Elder / Younger) Brothers and Sisters**  
**on Respondents Education Year**  
(standard error is in the parenthesis)

Models	OLS <sup>1</sup>		Tobit <sup>2</sup> (censored: students)					
	Boy's Education	Girl's Education	Boy's Education	Girl's Education				
<u>Independent Variables</u>								
Elder Brother's Education	0.276** (0.048)	0.161** (0.055)	0.418** (0.052)	0.455** (0.068)				
Elder Sister's Education	0.094 (0.073)	0.385** (0.057)	0.314** (0.082)	0.626** (0.062)				
Younger Brother's Education	0.213** (0.066)	0.355** (0.081)	0.240** (0.074)	0.055 (0.102)				
Younger Sister's Education	0.355** (0.119)	0.263** (0.083)	0.347* (0.135)	0.383** (0.101)				
Observations	305	144	283	195	855/305	575/144	659/283	528/195

\* The coefficient is significant at the 5% level

\*\* The coefficient is significant at the 1% level

1. The data set for OLS regression includes respondents who already complete their education. The other independent variables include a dummy variable for elder brother's (or elder sister's) income, a dummy variable for younger brother's (or younger sister's) income, mother's education, the squared term of mother's education, father's education, the squared term of father's education, log household income, the share from mother's income among household income, living area (urban or rural), age, squared term of age, number of children living in the household, and number of children living away from the household.

2. The data set for Tobit regression includes all of the respondents. The student status is censored. The number on the left side of slash "/" is the number of all of the respondents and on the right side is uncensored number. The other independent variables are same as in the OLS regression.

**Table 6: Nonlinear<sup>+</sup> Regression Result**  
**Siblings Education Effect on Education Year**  
 (standard error is in the parentheses)

	<u>Dependent Variables</u>			
	Boys' Education	Obs.	Girls' Education	Obs.
<u>Independent Variables</u>				
Elder sister's mean education	0.664** (0.053)	136	-0.314** (0.069)	228
Younger sister's mean education	0.845** (0.048)	493	-0.481** (0.057)	360
Elder brother's mean education	-0.563** (0.043)	272	0.389** (0.097)	280
Younger brother's mean education	-0.625** (0.034)	584	-0.171 (0.126)	375

The data set includes respondents who have sibling data.

+ Least Squared Method is used here to estimate the nonlinear form and correct the heteroskedascity problem.

\* The coefficient is significant at the 5% level.

\*\* The coefficient is significant at the 1% level.

**Table 7: OLS Result<sup>1</sup>**  
**The Effect from the Presence of (Elder / Younger) Brothers and Sisters**  
**and Age Difference**  
**on Respondents' Education Year (Subsamples: High / Low Household income)**  
(standard error is in the parenthesis)

Sample	High Income Family <sup>2</sup>		Low Income Family <sup>2</sup>	
	Boy's Education	Girl's Education	Boy's Education	Girl's Education
<u>Independent Variables</u>				
Dummy (Age Difference > 3) <sup>3</sup> * Any Elder Brother	-0.047 (0.482)	1.221* (0.524)	0.382 (0.691)	0.710 (0.468)
Dummy (Age Difference > 3) <sup>3</sup> * Any Elder Sister	-0.736 (0.628)	1.144 (0.585)	0.475 (0.670)	0.373 (0.414)
Any Elder Brother	0.676 (0.423)	0.646 (0.527)	-0.439 (0.505)	-0.257 (0.310)
Any Elder Sister	0.349 (0.496)	0.415 (0.534)	0.132 (0.564)	0.055 (0.357)
Any Younger Brother	0.339 (0.386)	0.419 (0.472)	0.476 (0.491)	-0.132 (0.323)
Any Younger Sister	0.301 (0.347)	-0.314 (0.451)	-0.046 (0.468)	-0.040 (0.310)
Observations	503	341	344	449

\* The coefficient is significant at the 5% level

\*\* The coefficient is significant at the 1% level

1. The data set for OLS regression includes respondents who already complete their education. The other independent variables include a dummy variable for elder brother's income, a dummy variable for elder sister's income, a dummy variable for younger brother's income, a dummy variable for younger sister's income, mother's education, the squared term of mother's education, father's education, the squared term of father's education, log household income, the share from mother's income among household income, living area (urban or rural), age, squared term of age, number of children living in the household, and number of children living away from the household.

2. The first and the third quartile of log household income in the survey data are 12.8739 and 13.75841, respectively. We define low income family as a family whose log household income is below 12.8739 and high income family as a family whose log household income is above 13.75841.

3. Set dummy of large age difference as 1 if the difference between the average elder brother's (sister's) age and respondent's age is greater than 3 years.

**Table 8: Tobit Result<sup>1</sup>**  
**The Effect from the Presence of (Elder / Younger) Brothers and Sisters**  
**and Age Difference**  
**on Respondents Education Year (Subsamples: High / Low Household income)**  
(standard error is in the parenthesis)

Sample	High Income Family <sup>2</sup>		Low Income Family <sup>2</sup>	
	Boy's Education	Girl's Education	Boy's Education	Girl's Education
<u>Independent Variables</u>				
Dummy (Age Difference > 3) <sup>3</sup> * Any Elder Brother	0.970* (0.447)	1.663** (0.514)	0.380 (0.651)	1.328** (0.500)
Dummy (Age Difference > 3) <sup>3</sup> * Any Elder Sister	-0.230 (0.553)	1.549** (0.554)	0.399 (0.636)	0.320 (0.463)
Any Elder Brother	0.974* (0.397)	1.045* (0.489)	-0.008 (0.495)	0.919** (0.343)
Any Elder Sister	0.791 (0.473)	0.477 (0.505)	0.662 (0.563)	0.445 (0.406)
Any Younger Brother	0.351 (0.376)	1.041* (0.438)	0.498 (0.505)	0.646 (0.376)
Any Younger Sister	-0.115 (0.340)	0.560 (0.434)	-0.340 (0.492)	0.213 (0.355)
Observations	1242/503	914/341	995/344	837/449

\* The coefficient is significant at the 5% level

\*\* The coefficient is significant at the 1% level

1. The data set for Tobit regression include all of the respondents. The student status is censored. The number on the left side of slash "/" is the number of all of the respondents and on the right side is uncensored number. The other independent variables include a dummy variable for elder brother's income, a dummy variable for elder sister's income, a dummy variable for younger brother's income, a dummy variable for younger sister's income, mother's education, the squared term of mother's education, father's education, the squared term of father's education, log household income, the share from mother's income among household income, living area (urban or rural), age, squared term of age, number of children living in the household, and number of children living away from the household.

2. The first and the third quartile of log household income in the survey data are 12.8739 and 13.75841, respectively. We define low income family as a family whose log household income is below 12.8739 and high income family as a family whose log household income is above 13.75841.

3. Set dummy of large age difference as 1 if the difference between the average elder brother's (sister's) age and respondent's age is greater than 3 years.

**Table 9: Fixed Effect Result<sup>1</sup>**  
**The Effect from the Presence of (Elder / Younger) Brothers and Sisters**  
**and Age Difference**  
**on Respondents Education Year (Subsamples: High / Low Household income)**  
(standard error is in the parenthesis)

Sample	High Income Family <sup>2</sup>		Low Income Family <sup>2</sup>	
	Boy's Education	Girl's Education	Boy's Education	Girl's Education
<u>Independent Variables</u>				
Dummy (Age Difference > 3) <sup>3</sup> * Any Elder Brother	-0.132 (0.782)	0.867 (0.754)	1.015 (1.156)	0.112 (0.581)
Dummy (Age Difference > 3) <sup>3</sup> * Any Elder Sister	-0.448 (0.776)	0.769 (0.748)	-2.036 (1.295)	1.135* (0.562)
Any Elder Brother	-0.086 (0.525)	-0.826 (1.029)	0.048 (0.745)	-0.114 (0.498)
Any Elder Sister	-0.688 (0.845)	0.331 (0.504)	-1.295 (1.240)	0.637 (0.365)
Any Younger Brother	0.115 (0.527)	-1.204 (0.968)	0.121 (0.830)	0.207 (0.662)
Any Younger Sister	-1.532 (1.283)	0.496 (0.550)	0.030 (1.490)	-0.221 (0.387)
Observations	517	366	344	450

\* The coefficient is significant at the 5% level

\*\* The coefficient is significant at the 1% level

1. The data set for fixed effect model include only respondents who already complete their education. The household factor is absorbed in the fixed effect model. And hence the other independent variables include a dummy variable for elder brother's income, a dummy variable for elder sister's income, a dummy variable for younger brother's income, a dummy variable for younger sister's income, individual characters which are age and the squared term of age.

2. The first and the third quartile of log household income in the survey data are 12.8739 and 13.75841, respectively. We define low income family as a family whose log household income is below 12.8739 and high income family as a family whose log household income is above 13.75841.

3. Set dummy of large age difference as 1 if the difference between the average elder brother's (sister's) age and respondent's age is greater than 3 years.

## Chapter 2: Intergenerational Transfer of Human Capital from Parents to Children: Does Neighborhood Matter?

### I. Introduction

Empirical analysis of data from several countries indicates that parental education has a positive impact on the educational attainment of children. Intergenerational transfers of human capital are evident in studies by Behrman and Deolalikar (1988) (using cross-country survey data), Thomas, Schoeni, and Strauss (1996) (using Brazilian data), and Salehi (1997) (using Iranian data). These papers suggest two primary factors as an explanation of this phenomenon: one focuses on the *family budget constraint*, and the other on the *production function for human capital* within households. In general, parents with higher education have higher incomes, and hence their children could receive more education due to the existence of a more flexible family budget constraint. A second argument regarding the influence of parental education suggests that the time parents spend with their children at home (*input*) is an important factor in determining the educational attainment of children (*output*). Relatively well-educated parents, cognizant of the importance of human capital in determining socioeconomic well-being, are arguably more willing and able to personally help in the education of their progeny.

Apart from the above-mentioned *direct* means of intergenerational human capital transfers, well-educated (and/or rich) parents may also influence children's educational attainment by choosing to live in *good areas* – neighborhoods with relatively-better schools, for instance. Well-educated neighbors can be positive role models, and their children be desirable peers for one's own children. Hence, there can exist a significant positive impact from living in "good" neighborhoods for children's educational attainment. Hence, a positive influence from parents on their children's education might result from this neighborhood effect, since parents can select their neighborhood. In other terms, researchers can deconstruct parental influence on children's education into two types of effects: *direct* effects and *indirect* effects. *Direct* effects refer to the impact of parental input (either monetary or time) into educating their children. *Indirect* effects refer to the

fact that parents may choose *good* neighborhoods to live in – in order for their children to benefit from being in an environment more conducive to educational attainment.

The above discussion leads us to believe that estimates of the direct effect of parents on children's education will be biased if one ignores the neighborhood effect. This paper estimates the relationship between parental education, neighborhood choice, and children's education using the Social and Economic Characteristics in Iran (SECHI) panel survey of households. More specifically, this paper focuses explicitly on the role of neighborhoods in influencing educational attainment.

Some studies already show that the influence from neighborhood is an important influence on the educational attainment of children. For example, Aaronson (1996) finds that the poverty rate and the dropout rate in neighborhoods have a significant impact on children's high school graduation. Datcher (1982) suggests that the achievement in education and earnings are affected by the average neighbor's income and the percentage of neighbors who are white.

One way to account for neighborhoods is to include a dummy variable for household clusters in order to capture both observed and unobserved attributes of the neighborhood. After neighborhood effects are controlled for, unbiased *direct* influence of parental education on children's education can be estimated. Another methodology would entail an explicit account of key neighborhood characteristics in the model, and a comparison of the results to a fixed effect model using cluster dummies. If the neighborhood variables could represent key neighborhood characteristics, then the influence from neighborhoods on children's education can be measured.

Before analyzing the data, two assumptions need to be specified for the model. First, this study assumes that the error term regarding parental preference for neighborhoods is not related to the error term regarding their preference for the educational achievement of their children. Second, this study assumes parents do not choose their neighborhood due to the same education level or income level. If the above assumptions are not sustained, the cluster fixed effect model will raise the simultaneous problem. A solution to solve this problem is to use a broad scope of location as a

neighborhood dummy variable instead of using the narrow scope of location, a cluster. The paper presents details later.

This study also focuses the analysis on the comparison between rural versus urban areas, and the education of boys versus girls in Iran. Since mobility is higher in urban areas compared to rural area (suggested by the data), a significantly different result from parent's educational impact between rural and urban areas is expected due to neighborhood effect.

The results of this study suggest that choosing a "good" neighborhood is important for children's education. Furthermore, they suggest that neighborhood effect is more significant in rural area than that in urban area.

The remainder of the paper is organized as follows. The econometric framework is elaborated in Section II. Data from the Iranian survey are described in Section III. Results are presented in Section IV, and conclusions are in Section V.

## **II. Econometric Methodology**

Assuming that parents are decision makers for children's education, factors affecting children's education include parent's education, household income, and the unobserved preference of parents for children's education. Well-educated parents might be more efficient teachers for teaching their children at home. The teaching process generates a positive influence from parents' education to children's education. As mentioned earlier, this process is defined as a *direct* effect in this paper. In addition, well-educated parents are more capable of moving to an ideal neighborhood which could positively impact children's education. For example, if the parents in a neighborhood are concerned for their children's education, they might make an effort to increase the quality of the school that their child attends. A school composed of students with high quality will also upgrade the quality of the school. Hence, a good neighborhood could increase the quality of the school and distribute a positive influence on children's education. Although parents do not spend much time on their children's education, their selection of neighborhood could have a positive influence on children's education. This effect is defined as an *indirect* effect from parent's education to children's education in this paper.

To put these factors in the model, the following equation is established:

$$(1) \quad Y_{ij} = \gamma_1 X_{ij} + \gamma_2 N_i + e_{ij}$$

where  $ij$  refers to  $j$ th child in the  $i$ th neighborhood.  $Y_{ij}$  represents the completed children's education,  $X_{ij}$  represents parent's characteristics,  $N_i$  represents neighborhood characteristics, and  $e_{ij}$  represents the unobserved error term such as parent's preference on children's education. Completed children's education is influenced by parent's characteristics, neighborhood characteristics, and the error term. The parent's characteristics include parent's education and household income. The unobserved error term,  $e_{ij}$ , is distributed normally with mean zero and variance  $\sigma_e^2$ .  $\gamma_1$  represents the *direct* influence from parent's characteristics on children's education, and  $\gamma_2$  the influence from neighborhood on children's education.

Since well-educated or rich parents are more capable of moving their families to an ideal residence, neighborhood effect ( $N_i$ ) is mainly determined by parents and therefore influenced by both parents' characteristics ( $X_{ij}$ ) and unobserved preference of parents on neighborhood effect ( $w_{ij}$ ). The unobserved error term,  $w_{ij}$ , is also distributed normally with mean zero and variance  $\sigma_w^2$ . The above description is indicated as equation (2).

$$(2) \quad N_i = \beta X_{ij} + w_{ij}$$

where  $\beta$  represents the influence from parent's characteristics on the selection of neighborhood.

Assuming that the unobserved parent's preference on children's education and neighborhood is not correlated, i.e.  $Cor(w_{ij}, e_{ij}) = 0$ , an OLS estimation of  $\gamma_1$  in equation (1) will be unbiased and consistent. If neighborhood characteristics  $N_i$  are omitted in the OLS regression of the equation (1),  $\tilde{\gamma}_1$  could be estimated from the following regression

$$(3) \quad Y_{ij} = \tilde{\gamma}_1 X_{ij},$$

where  $Y_{ij}$  represents completed children's education and  $X_{ij}$  represents parent's characteristics. Equation (3) is named the *OLS model without neighborhood variables* in

later analyses. The OLS estimator  $\tilde{\gamma}_1$  will be a biased estimator of  $\gamma_1$ , since  $E(\tilde{\gamma}_1)$  is equal to  $\gamma_1$  plus  $\gamma_2 \cdot \beta$ , i.e.,

$$E(\tilde{\gamma}_1) = \gamma_1 + \gamma_2 \cdot \beta.$$

Hence, the estimated coefficient ( $\tilde{\gamma}_1$ ) in the above equation represents parents' *direct* impact ( $\gamma_1$ ) plus an *indirect* impact ( $\gamma_2 \cdot \beta$ ) on children's education. The *indirect* impact is defined as the influence from parent's characteristics on neighborhood ( $\beta$ ) times the impact from neighborhood on children's education ( $\gamma_2$ ). The meaning of  $\tilde{\gamma}_1$  is that one additional year of parent's schooling could increase children's schooling by  $\tilde{\gamma}_1$  years when parents tutor their children in the household and choose a good neighborhood for their children.

In order to measure the *direct* influence ( $\gamma_1$ ) from the parent's education, researchers should add neighborhood characteristics ( $N_i$ ) in the regression equation. Under the same assumption,  $Cor(e_{ij}, w_{ij}) = 0$ , the *direct* influence from parent's education is measurable without bias and inconsistency. To absorb all the observed and unobserved characteristics of the neighborhood, a fixed effect model is applied by using cluster-level dummy variables to account for neighborhood characteristics (so called *cluster-fixed-effect* model in the later analysis). Besides the *cluster-fixed-effect* model, we use certain cluster-specific variables to represent some specific neighborhood characteristics. For instance, we use the influence from the average of mothers' education in the cluster on children's education. This model is named *OLS model with neighborhood variable* in later analyses. These neighborhood variables help to identify which neighborhood characteristics are more important on children's education.

The above analysis assumes that unobserved parental preference for children's education and the neighborhood effect are not correlated. If this assumption is sustained, the estimation of equation (1) will be unbiased and consistent.

Note that there is another possibility to generate a biased estimator in equation (1). If parents choose their neighborhood according to the level of education or income, the *cluster-fixed-effect* model might suffer from the simultaneous problem. For example, three kinds of neighborhoods are possible in this model: first, all well-educated parents;

second, all low-educated parents; third, a mixed neighborhood with some well-educated parents and some low-educated parents. Well-educated parents might choose to live in a mixed neighborhood because their lower ambition (preference) for children's education; while the low-educated parents might choose to live in a mixed neighborhood with the high-educated neighbors because of their higher ambition (preference) for children's education. Be aware that the ambition is the unobserved preference of parents on children's education ( $e_{ij}$ ). Hence, a negative correlation exists between parent's education subtracting out the average of parents' education and the unobserved error term. The unobserved error term is the parent's ambition subtracting the average parents' ambition in the *cluster-fixed-effect* model. Therefore, the estimated coefficient of parent's education will be downward biased in the *cluster-fixed-effect* model.

As suggested by Gabriel and Rosenthal (1997), a fixed effect model using provinces as dummy variables is less likely to suffer from the simultaneous problem. On average, provinces contain ten clusters consisting of a range of different levels of education and income. Therefore, it is less likely that a correlation exists between parent's education and the error term of the within province difference. The fixed effect model using provinces as dummy variables is called *province-fixed-effect* model in later analyses.

On the other hand, households might not choose neighborhoods by the same level of education or income. Instead, the selection of neighborhoods depends on the same religion, ethnic, or "social culture." Therefore, the negative correlation between independent variables and error term in *cluster-fixed-effect* model might not exist. Researchers could ignore this case.

### **III. Data**

The data set requires two features in the data set in order to analyze the neighborhood effect. One is the socioeconomic character of neighborhood, including the average parental education of neighbors, average household income of neighbors, and average children's education of neighbors. Another feature is the moving history of households. This information could assist in relating the selection of neighborhoods with parental characteristics.

A panel data set derived from the Social and Economic Characteristics of Households in Iran (SECHI) survey is adopted in this paper. These survey data were collected from 1987 until 1989 for continuous twelve rounds, with three months per round. The data set contains answers to a detailed questionnaire for households – and includes information such as age, years of schooling, sex, working hours, wage rate, and status of student or non-student for every individual in the household. In addition, each individual reported his/her birth place (rural, urban, or overseas), present living area (urban or rural), previous living area (urban or rural), and the period of time that they have been living at the present address.

Neighborhood variables derived from other households in the same cluster include average mothers' education of neighbors, average fathers' education of neighbors, log of average household income of neighbors, and average children's education of neighbors. Table 1 shows the abbreviations and the descriptions of the variables which are used in the text and the rest of the tables.

Table 2 reports household variables and neighborhood variables related to living areas (rural/urban) and moving history. The socio-economic characteristics of the household include mother's years of schooling (weduc), father's years of schooling (heduc), log of household income (lhinc), children's years of schooling (educyr), and the percentage of siblings who have attended high school (shigh). The results show significant differences in all household characteristics (weduc, heduc, lhinc, educyr, and shigh in Table 2) between households living in rural and urban areas. For instance, the average father's years of schooling (1.7 years) for the households that never moved is lower in rural areas than those (5.9 years) for the households that never moved in urban areas.

Other household characteristics are also significantly different based upon different moving histories. There are three types of moving histories: 1) households that never moved; 2) households that moved from rural to rural or urban area; 3) households that moved from urban to rural or urban area. It is discovered that households that moved from urban area are more likely to have higher average parents' education than other households (never moved or moved from rural area). For example, the average father's education (heduc) of the urban households that moved from another urban area is 7.1

years, and that of the urban households that moved from a rural area is 3.5 years. This pattern is also found in rural households: the average father's education of the rural households that moved from urban area is 4.9 years, and that of the rural households that moved from another rural area is 1.7 years. Another fact is that households that moved within the same area also have higher parents' education than households that never moved. Household income has the same pattern regarding to moving history and living area (rural/urban). In other words, the results suggest that having been born or living in an urban area is positively related to parents' education and household income. The results also show that moving or not is positively related to parents' education and household income.

The average children's education is higher in urban areas than in rural areas. However, a significant difference does not exist in children's education regarding household's moving history. For the urban households, the average children's education (educyr) is 6.2 and 5.6 years for the households that moved from other urban areas and that haven't moved, respectively. The difference is not significant in these two figures.

In order to see if neighborhood characteristics are related to moving history and living in rural or urban area, this study proceeds as follows. The average father's education of neighbors (bheduc), the average mother's education of neighbors (bweduc), the log of average income of neighbors (blhhinc), the average children's education of neighbors (beducyr), and the percentage of children who go to high school in the neighborhood (bhigh) are derived from other households in the same clusters. The bottom part of Table 2 shows that the figures of these variables (bheduc, bweduc, blhhinc, beducyr, and bhigh) are different for rural and urban, but there is no significant difference due to a different moving history. For instance, the average fathers' education in the neighborhood (bheduc) is 6.8 and 5.3 years for the urban households that moved from other urban area and that haven't moved, respectively. The difference is not significant between these two figures. While parents' characteristics are quite different for different moving history, children's education and neighborhood characteristics are not apparently different for the different moving history.

An appropriate geographic scope of neighborhood is critical to the analysis. If the cluster does not represent the “real” neighborhood, the characteristics of the households in the same cluster are less likely to be homogeneous, and the characteristics obtained across different clusters are comparatively homogeneous to each other. In order to test whether the cluster represents the “real” neighborhood, a hypothesis related to whether the households across different clusters share common characteristics is tested. The tested characteristics include fathers’ years of schooling (heduc), mothers’ years of schooling (weduc), and children’s education (educyr). ANOVA (Analysis of Variance Model) tests this hypothesis and the results are shown in Table 3. The  $F$  statistics<sup>8</sup> from ANOVA method are 3.3 and 9.3 for mother’s education in rural and urban area, respectively. The  $F$  statistics are 4.1 and 9.3 for father’s education in rural and urban area, respectively. The  $F$  statistics are 9.7 and 8.5 for log of household income in rural and urban area, respectively. The figures are higher than the critical values, therefore, the hypothesis that there is no significant difference in the average household characteristics among different clusters is rejected. Hence, the cluster could represent the *real* neighborhood.

#### **IV. Results**

The fundamental question raised in this paper is whether the impact of parental decisions with regard to the choice of neighborhoods on children’s educational attainment is more important than the impact from parents’ spending time with children. The

---

<sup>8</sup> The statistic of ANOVA test is

$$F = \frac{SSR / df_1}{SSE / df_2} \sim F(df_1, df_2)$$

where  $df_1$  is the number of clusters minus 1 and  $df_2$  is total observations minus number of clusters.  $df$  represents degree of freedom.

SSR, called “cluster” sum of squares, is a sum of squares composed of deviations of the estimated cluster means  $\bar{Y}_j$  around the overall mean  $\bar{\bar{Y}}$ .

$$SSR = \sum_j n_j (\bar{Y}_j - \bar{\bar{Y}})^2$$

SSE, error sum of squares, measures the variability of the observations  $Y_{ij}$  around the estimated cluster means  $\bar{Y}_i$ .

$$SSE = \sum_j \sum_i (Y_{ij} - \bar{Y}_i)^2$$

methodology utilized to test this hypothesis is as follows. If neighborhood choice is an important factor, then controlling for neighborhood factors shall decrease the impact of parental education on children's education. On the other hand, if parents' spending time with children is more important than choosing a good neighborhood, then the coefficients of parental education should not decrease much or become less significant after controlling for neighborhood factors. Another question is whether father's spending time with children is more important than mother's spending time with children for children's education. The question could be answered by comparing the coefficients of paternal education and maternal education on children's education, after controlling for neighborhood effects.

From previous studies, parental education is known to be important for children's education. In addition, maternal education has a larger impact than paternal education on a daughter's education. Paternal education has a larger impact than maternal education on a son's education. The results of this study are consistent with these findings, without controlling for neighborhood effects. After controlling for neighborhood effects, the impact from maternal and paternal education decreases for both boys' and girls' education. This result suggests that neighborhood factors are an important determinant of children's education.

After the neighborhood factors are controlled by adding the neighborhood variables, the estimated coefficients of neighborhood variables are significant for boys' education in rural area and insignificant in urban area. The results indicate that the neighborhood effect is more significant in a rural area than in a urban area.

This paper estimated the impact of household characteristics on children's education<sup>9</sup> for four cases: 1) *OLS model without neighborhood variables*: ignoring neighborhood effects as in most previous studies; 2) *OLS model with neighborhood variables*: controlling neighborhood effects with neighborhood variables; 3) *province-fixed-effect* model: controlling neighborhood effects at the province level; 4) *cluster-fixed-*

---

<sup>9</sup> Note that the data set for regressions in tables 4, 5, and 6 include respondents who already completed their education. Dropping the respondents who haven't completed their schooling might cause a selection

*effect* model: controlling neighborhood effects at the cluster level. These four cases allow for a distinction between the impact on children’s education from household characteristics and from neighborhood effects. Tables 4 and 5 present the results for completed education for girls and boys, respectively. In addition, the regressions are run separately for rural and urban households. The *OLS model without neighborhood variables* (columns 1 and 5 for rural and urban areas, respectively) includes the following independent variables: household characteristics such as mother’s education, mother’s education squared, father’s education, father’s education squared, log of household income, the share of mother’s income in household income, age of child, and the age of child squared. In the *OLS model with neighborhood variables* (columns 2 and 6 for rural and urban, respectively), four neighborhood variables are added to column 1: mothers’ average education, fathers’ average education, log of average household income, and percentage of children who have received a high school education. These variables are derived from other households in the same cluster (same neighborhood). In *province-fixed-effect* model (columns 3 and 7 for rural and urban, respectively), dummy variables for provinces are added in order to capture all the observed and unobserved influences of the local province. In *cluster-fixed-effect* model (columns 4 and 8 for rural and urban, respectively), dummy variables for clusters are added in order to capture all the observed and unobserved influences of the neighborhood cluster.

#### IV.1 The estimated *total* influence from parents’ education

Assuming that the unobserved parent’s preference on children’s education and neighborhood are not correlated, i.e.  $\text{Cor}(e_{ij}, w_{ij}) = 0$ , the bias and inconsistent problem may be ignored. The estimates of the impact from parents’ education on children’s education from the *OLS model without neighborhood variables* indicates the *total* effect. The total effect,  $E(\tilde{\gamma}_1) = \gamma_1 + \gamma_2 \cdot \beta$ , is shown in columns 1 and 5 of both Tables 4 and 5. The coefficients are comparable to previous estimates in the literature. The result shows that the completed schooling of daughters is more strongly influenced by maternal

---

problem because the respondents who already completed their education tend to be older than the students.

education than by paternal education. On the other hand, education of sons is more influenced by paternal education than by maternal education. These results are consistent with the literature [Thomas, Schoeni, and Strauss (1996) and Salehi (1997)]. These results may reflect gender-specific differences in the inter-generational transfer of human capital [as suggested in Thomas, Schoeni, and Strauss (1996)].

An additional year of mother's schooling (*weduc*) increases daughter's schooling by 0.6 years, and an additional year of father's schooling (*heduc*) increases daughter's schooling for a rural girl's education (see column 1 in Table 4) by 0.4 years. This positive impact from parents' education on children's education might be due to parents' direct assistance in the household, or a positive influence from "good" neighborhood selection on the part of parents.

The estimated coefficients from parents' education to girls' education are larger and more statistically significant than that to boys' education. This suggests that a girls' education depends more on parents' characteristics than boys' education does. In other terms, in comparison with girls' education, boys' education appears to be more *inelastic* with regard to parental education.

#### IV.2 The estimated *direct* influence from parents' education

In columns 1 (rural) and 5 (urban) of Table 4, the estimated coefficients of parents' education (*weduc* and *heduc*) represent the *total* effect of parent's education on girl's education. In Table 5, the estimated coefficients of parents' education (*weduc* and *heduc*) represent the total effect of parent's education on boy's education. Addition of cluster dummies to measure the influence of children's education from neighborhood is shown in columns 4 and 8 for rural and urban areas, respectively, in both Tables 4 and 5. The estimated coefficients of parents' education represent the *direct* impact ( $\gamma_1$ , described in Section II) of parent's characteristics on children's education (estimated by *cluster-fixed-effect model*). For a rural girl's education, for instance, the estimated coefficient of the *direct* impact associated with mother's education (*weduc*) derived from the *cluster-fixed-effect model* is 0.3 which is lower than the coefficient, 0.6, derived from the *OLS model without neighborhood variables* (Table 4). The coefficients of parents' characteristics are

smaller in the *cluster-fixed-effect* model than those in *OLS (without neighborhood variables)* model. A similar pattern is observed to urban girls' education (column 8 in Table 4), rural boys' education (column 4 in Table 5), and urban boys' education (column 8 in Table 5).

The coefficients of mother's education (*weduc*) on children's education are not significant in rural girls' education (column 4 of Table 4), urban girls' education (column 8 of Table 4), rural boys' education (column 4 of Table 5), and urban boys' education (column 8 of Table 5). The results might be due to the small within-difference of mother's education in a cluster. Hence, the results do not imply that mother does not tutor her children at home or her assistance is not an important factor for children's education.

On the other hand, the coefficients of father's education (*heduc*) on children's education are significant in rural girls' education (column 4 of Table 4), urban girls' education (column 8 of Table 4), and urban boys' education (column 8 of Table 5).

After controlling the neighborhood factor, the estimated coefficients of parent's education decreases less in rural area than in urban area for girls' education. For instance, the difference between columns 1 (the coefficient of *weduc* is 0.6) and 4 (the coefficient is 0.3) in Table 4 is 0.3, while the difference between columns 5 (0.5) and 8 (0.1) is 0.4.

#### IV.3 The influence from neighborhood characteristics

The impact on children's education from neighborhood characteristics is measurable by using neighborhood variables instead of cluster dummy variables, if the neighborhood variables represent the *real* neighborhood characteristics. The estimated results are shown in columns 2 and 6 for rural and urban areas, respectively.

The estimated coefficients of parents' characteristics are very close to the coefficients in the *cluster-fixed-effect* model. This result suggests that the neighborhood variables could represent the *real* neighborhood characteristics.

This study expects that the magnitude of the coefficients of parents' characteristics to be lower in column 2 than that in column 1 in Table 4. The same pattern is observed in Table 5. From Table 4, the result shows that the average mother's education in neighborhood households has a significant positive effect on girl's education in both rural

and urban areas. From Table 5, the results indicate that the positive influence from average father's education in neighborhood households is only significant for a rural boy but not for an urban boy. Average mothers' education in the neighborhood seems to have larger influence on children's education than average fathers' education does. However, notice that average mothers' education could be a proxy variable of neighborhood attribute such as low crime rate. Hence, the results show that the average mothers' education in the neighborhood is positively correlated with children's education.

For boy's education (in Table 5), neighborhood variables are significant in rural areas (column 2) and insignificant in urban areas (column 6). Neighborhood effect is more significant in rural areas. A possible explanation is that urban areas have higher mobility and the clusters are close to each other, and hence, the impact from neighborhood becomes non-distinguishable.

#### IV.4 Simultaneity problem due to a too narrowed scope of neighborhood

It is possible that the unobserved parent's preference on children's education is correlated with parent's characteristics. As suggested by Gabriel and Rosenthal (1997), the geographic scope of the neighborhood should not be too narrow, otherwise it will cause a simultaneity problem. For example, parents with low education but high ambition to educate their children may choose to live in a neighborhood with *good* socio-economic characteristics such as a high level of education. If parents choose their neighborhoods on the basis of education, average parents' characteristics will be negatively correlated with average unobserved ambition. Thus, these findings from the *cluster-fixed-effect* model are subject to simultaneity bias. The estimated coefficients of parent's education on children's education will be downward biased.

However, the *province-fixed-effect* model (adding province dummy instead of clusters dummy) is likely to avoid both omitted variable and simultaneity bias. The reasons are that a province consists of several clusters with different education levels, and that no simultaneity bias exists due to the calculation of the within difference in a province.

In Table 4, the *province* is used as the geographic scope of the neighborhood, shown in columns 3 and 7 for rural and urban areas, respectively, and fixed effects are

applied to the estimations. For girls, the estimated coefficient of mother's education in column 3 (7) is larger than that in column 4 (8), for rural (urban) area. In the rural area, the coefficient of mother's education is 0.5 (column 3) for neighborhood in province level and 0.3 (column 4) for neighborhood in cluster level. In the urban area, the coefficient of mother's education is 0.4 (column 7) for neighborhood in province level and 0.1 (column 8) for neighborhood in cluster level. The same situation happens in boy's data which is shown in Table 5. These results imply that parents choose their neighborhood based on the mother's education. The estimated coefficient of mother's education on girls' education in cluster-fixed-effect model is slightly downward biased if too narrow neighborhood (cluster) is chosen. These results are also found in boys' data (Table 5). The estimated coefficient of father's education on girls' and boys' education is also downward biased.

The results obtained from *province-fixed-effect* model have similar pattern compared to those obtained from *OLS model without neighborhood variables*, which are shown in columns 3 and 7 in Table 4 for girls and Table 5 for boys. Using the *province-fixed-effect* model to control omitted variable bias without raising the simultaneity problem, maternal education remains a larger influence on daughters' education, while paternal education remains a larger influence on sons' education.

#### IV.5 Selection Effect or Neighborhood Effect

After the recognition of the existence of neighborhood factors, we want to know whether they arise from parental selection or a real influence from the neighborhood. It is possible that parents select a good neighborhood, but there is no influence from good neighborhood to children's education for some reasons. In this case, the coefficient of neighborhood characteristics will only represent a positive correlation with children's education. On the other hand, the neighborhood effect might not be only a parental selection but also actually influences children's education. We could expect that the longer period for a child living in a cluster, the higher the influence from neighbors to the child's education.

Table 6 reports OLS regression in which five additional regressors were added to *OLS model with neighborhood variables* in both Tables 4 and 5. One variable is the duration time in the current residence (*resdur*), and another four are the neighborhood variables (*bheduc*, *bweduc*, *blhhinc*, and *bhigh*) timed with “*resdur*” individually. In Table 6, columns 1 and 2 report girl’s education in rural and urban areas, respectively; and columns 3 and 4 report boy’s education in rural and urban areas, respectively. If neighborhood affects children’s education, the impact would increase by the period of time that children have lived in that cluster. Hence, we expect that the coefficients from the five new regressors are significant. On the other hand, if neighborhood is simply a parent’s selection, the coefficients of the four neighborhood variables are expected to be significant. Table 6 does not show significance of the coefficients of either the four neighborhood characteristics (*bweduc*, *bheduc*, *blhhinc*, and *bhigh*) or the five new variables (*resdur*, *resdur\*bweduc*, *resdur\*bheduc*, *resdur\*blhhinc*, and *resdur\*bhigh*). The reason might come from that the four neighborhood characteristics and the five new variables are highly correlated, and therefore, all variables are insignificant.

#### IV.6 Parents’ Moving for Their Children?

To estimate parent’s motivation of moving, Probit analysis of moving to a good neighborhood is carried out. The question to be clarified is whether parents choose their neighborhood for their children’s education. Probit analysis for the moving decision is reported in Table 7. The first column shows the estimation from the data of all households for rural and urban areas. Dependent variable is a binary variable which could be “1” representing “move” or “0” representing “stay”. Independent variables include their current living area (rural/urban), their born area (rural/urban), father’s education, mother’s education, and the four neighborhood variables. Results show that people who currently live in urban area and people who were born in rural area have higher probability to move than those people living in rural or born in urban area (column 1 in Table 7). By notice that while their own education is not significant, the average mother’s education in the neighborhood households has positive impact on the probability of moving. Where they stay and where they are from are important factors for moving. An interesting

question raised is whether neighborhood characteristics have a different impact on people who move from rural to rural, from rural to urban, or from urban to urban. Columns 2, 3, and 4 in Table 7 report these three cases, respectively. Interestingly we found that the average mother's education in the neighborhood household is an important factor for the decision of moving; while the average father's education in the neighborhood is only important for people who moved from rural to rural area (column 2). Average household income of neighborhood is an important factor for households which moved to urban areas (columns 3 and 4).

Hence, this study could conclude that the average mother's education in the neighborhood and average household income of neighborhood are the two most important factors to people's decisions regarding to where to move.

Estimation of the correlation between parents' moving decision and whether parents have children with schooling age while moving could assist to clarify whether parents choose to move to a good neighborhood for their children. The dependent variable is defined as whether parents move to a cluster in which neighbors' characteristics are better than parent's characteristics. For example, parents might move to a cluster which has a higher average father's education of neighbors than the father's own education. The independent variables are the dummy variables representing whether there is any school-age child when they move. The dummies are divided into three age groups: 6 to 9 years old, 10 to 14 years old, and 15 to 18 years old. Table 8 reports the Probit analysis regarding to three dependent variables: moving to a neighborhood with a higher average father's education of neighbors than the father's own education (columns 1 and 2), moving to a neighborhood with a higher average mother's education of neighbors than the mother's own education (columns 3 and 4), moving to a neighborhood with a higher average household income of neighbors than the household income (columns 5 and 6). In addition, columns 1, 3, and 5 consider the gender of child. Columns 2, 4, and 6 do not consider the gender of child.

The result in column 3 of Table 8 shows that the coefficient of the existence of a son with 6-9 years old (c69b) is significantly positive. Which means parents are more likely to move to a neighborhood with a high average mother's education of neighbors

(mark of a good neighborhood) if parents have a son 6 to 9 years old. Another significant result is the moving destination to a higher income area. The coefficients of c1518g, c69b, and c1014b in column 5 of Table 8 are significantly positive. Parents who have a son 6-9 years old or 10-14 years old are more likely to move to a richer area (column 5, Table 8). The only significant impact from daughter is in column 5 in Table 8. The results show that parents prefer to move because of their sons rather than for their daughters in Iran. Average mother's education of neighbors and average household income in the neighborhood are two important factors considered by the parents while they make decision about moving.

## **V. Conclusion**

Parent's influence on their children's education refers to two types of effects: *direct* effect and *indirect* effect. *Direct* effect means that parents teach their children. *Indirect* effect indicates that parents select a good neighborhood for their children.

Assuming that the unobserved parent's preference on children's education is not correlated to the unobserved parent's preference on neighborhood, the biased and inconsistent problem will not arise. The result indicates that the completed schooling of daughters is more strongly influenced by maternal education than by paternal education. On the other hand, sons' education is more influenced by paternal education than by maternal education. The results are consistent with those in the literature [Thomas, Schoeni, and Strauss (1996) and Salehi (1997)]. These positive impacts from parents' education on children's education might be due to parents' direct assistance in the household or a positive influence from a good neighborhood from parents' selection.

After controlling for neighborhood effects, the impact from maternal and paternal education decreases for both boys' and girls' education. This result suggests that choosing a *good* neighborhood is important for children's education.

Comparing the results of rural area versus urban area, the impact from neighborhood variables is more significant in a rural area than those in an urban area for boys' education after controlling the neighborhood factor. The results indicate that the neighborhood effect is more significant in rural areas than in urban areas.

## References

1. Aaronson, Daniel (1996), "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," *Federal Reserve Bank of Chicago*, working paper no. WP-96-19.
2. Behrman, Jere R. and Anil B. Deolaikar (1988), "Health and Nutrition," in Hollis B. Chenery and T.N. Srinivasan, eds., *Handbook on Economic Development*, Vol. 1, Amsterdam: North Holland Publishing Co., 631-711.
3. Benabou, Roland (1994), "Education, Income Distribution, and Growth: The Local Connection," (draft).
4. Borjas, George. (1995), "Ethnicity, Neighborhoods, and Human Capital Externalities," *American Economic Review*. p. 365-390.
5. Case, Anne and Lawrence Katz. (1991), "The company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth." mimeo, *National Bureau of Economic Research*.
6. Datcher, Linda (1982), "Effects of Community and Family Background on Achievement," *Review of Economics and Statistics*, vol. 64, p.32-41.
7. Datcher-Loury, Linda (1989), "Family Background and School Achievement among Low Income Blacks," *Journal of Human Resources*, vol. 24, no. 3
8. Evans, William and Wallace Oates and Robert Schwab. (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior." *Journal of Political Economy*. p. 966-991.
9. Gabriel, Stuart A. and Stuart S. Rosenthal (1997), "Location and the Effect of Demographic Traits on Earnings," (draft)
10. Manski, Charles. (1993), "Identification of Endogenous Social Effects: The Reflection Problem?" *Review of Economic Studies*. p. 531-542.
11. Thomas, Duncan and Robert F. Schoeni and John Strauss (1996), "Parental Investments in Schooling: The Roles of Gender and Resources in Urban Brazil," *Rand Labor and Population Program Working Paper Series 96-02*.

12. Salehi, Isfahani Djavad (1997), "Catching Up: the Role of the Family in the Closing of the Gender Gap in Education in Iran", mimeo.

**Table 1: Description of Variables**

<u>Variables</u>	<u>Description</u>
<u>Household:</u>	
educyr	Child's Years of Schooling
weduc	Mother's Years of Schooling
wedsq	Mother's Years of Schooling - Squared Term
heduc	Father's Years of Schooling
hedsq	Father's Years of Schooling - Squared Term
lhhinc	Log of Household Income
wincsh	The Share of Mother's Income among Household Income
shigh	Percentage of Siblings Attending High School or above Derived from other children in the same Household
age	Child's Age
agesq	Child's Age - Squared Term
sex	Child's Gender (1 is Boy; 0 is Girl)
ru	Currently living in Urban (1) or Rural (0)
bru	Previous Lived in Urban (1) or Rural (0)
resdur	Duration Years in Current Residence
c69g	Any Daughters Whose Age Were from 6 to 9 While Moving
c1014g	Any Daughters Whose Age Were from 10 to 14 While Moving
c1518g	Any Daughters Whose Age Were from 15 to 18 While Moving
c69b	Any Sons Whose Age Were from 6 to 9 While Moving
c1014b	Any Sons Whose Age Were from 10 to 14 While Moving
c1518b	Any Sons Whose Age Were from 15 to 18 While Moving
c69d	Any Child Whose Age Were from 6 to 9 While Moving
c1014d	Any Child Whose Age Were from 10 to 14 While Moving
c1518d	Any Child Whose Age Were from 15 to 18 While Moving
<u>Neighborhood:</u>	
bweduc	Average Mother's Education Derived from other Households in the same Branch Cluster
bheduc	Average Father's Education Derived from other Households in the same Branch Cluster
blhhinc	Average Log of Household Income Derived from other Households in the same Branch Cluster
beducyr	Average Children's Education Derived from other Households in the same Branch Cluster
bhigh	Percentage of Children Attending High School or above Derived from other Households in the same Branch Cluster

**Table 2: Average Socioeconomic Characteristics of Household and Neighborhood with Respect to Rural / Urban and Moving Status**

<b>Currently Living Area</b>	<b>Rural</b>			<b>Urban</b>		
<b>Moving / Staying</b>	<b>Staying</b>	<b>Moving from other Rural</b>	<b>Moving from Urban</b>	<b>Staying</b>	<b>Moving from Rural</b>	<b>Moving from other Urban</b>
Variables:						
Household Characteristics						
heduc	1.734	1.702	4.967	5.895	3.505	7.065
weduc	0.587	0.596	2.600	4.451	2.027	5.491
lhhinc	12.861	12.854	13.333	13.397	13.368	13.526
educyr	3.570	3.488	3.454	5.573	5.005	6.225
shigh	0.025	0.025	0.017	0.119	0.082	0.173
Neighborhood Characteristics (Neighborhood: Cluster)						
bheduc	1.698	2.032	2.351	5.299	4.268	6.843
bweduc	0.605	0.749	0.968	3.763	2.993	5.289
blhhinc	12.753	12.806	12.856	13.302	13.302	13.528
beducyr	3.565	3.508	3.752	5.557	4.999	6.325
bhigh	0.025	0.024	0.033	0.121	0.086	0.176
Number of Households	1096	275	30	912	839	434

**Table 3: ANOVA Analysis of Household Socioeconomic Characteristics of Neighborhood: Cluster**

	<b>Rural</b>					<b>Urban</b>				
	Source	Partial SS	df	MS	F*	Source	Partial SS	df	MS	F*
weduc	Clusters (SSR)	637.1	61	10.4	3.3	Clusters (SSR)	13605.2	96	141.7	9.3
	<u>Residual (SSE)</u>	<u>4249.7</u>	<u>1341</u>	3.2		<u>Residual (SSE)</u>	<u>32025.9</u>	<u>2095</u>	15.3	
	Total (SST)	4886.8	1402			Total (SST)	45631.1	2191		
heduc	Clusters (SSR)	2180.9	61	35.8	4.1	Clusters (SSR)	15962.1	96	166.3	9.3
	<u>Residual (SSE)</u>	<u>11618.2</u>	<u>1341</u>	8.7		<u>Residual (SSE)</u>	<u>37598.5</u>	<u>2095</u>	17.9	
	Total (SST)	13799.1	1402			Total (SST)	53560.6	2191		
lhinc	Clusters (SSR)	319.5	61	5.2	7.5	Clusters (SSR)	281.8	96	2.9	4.4
	<u>Residual (SSE)</u>	<u>918.8</u>	<u>1316</u>	0.7		<u>Residual (SSE)</u>	<u>1354.9</u>	<u>2050</u>	0.7	
	Total (SST)	1238.7	1377			Total (SST)	1636.7	2146		
educyr	Clusters (SSR)	4831.8	61	79.2	9.7	Clusters (SSR)	10773.1	96	112.2	8.5
	<u>Residual (SSE)</u>	<u>31775.1</u>	<u>3877</u>	8.2		<u>Residual (SSE)</u>	<u>66341.8</u>	<u>5053</u>	13.1	
	Total (SST)	36606.9	3938			Total (SST)	77114.8	5149		

\*Critical values of F distribution:  $F(61,1341,\alpha=0.05)\cong 1.32$ ;  $F(61,3877,\alpha=0.05)\cong 1.32$ ;  $F(96,2095,\alpha=0.05)\cong 1.25$ ;  $F(96,5053,\alpha=0.05)\cong 1.25$ .

**Table 4: Neighborhood Impact on Girl s Completed Years of Schooling<sup>1</sup>**

Sample	Girl s Completed Years of Schooling in Rural Area				Girl s Completed Years of Schooling in Urban Area			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS (without Neighborhood Variables)	OLS (with Neighborhood Variables)	Province-Fixed-Effect Model	Cluster-Fixed-Effect Model	OLS (without Neighborhood Variables)	OLS (with Neighborhood Variables)	Province-Fixed-Effect Model	Cluster-Fixed-Effect Model
<b>Independent Variables</b>								
weduc	0.552 (1.065)	0.278 (0.571)	0.527 (1.062)	0.279 (0.612)	0.484** (3.800)	0.229 (1.897)	0.377** (2.938)	0.143 (1.084)
heduc	0.401** (3.564)	0.350** (3.303)	0.326** (2.993)	0.255* (2.476)	0.397** (4.360)	0.269** (3.138)	0.367** (4.002)	0.190* (1.980)
lhinc	0.332** (3.896)	0.252** (2.749)	0.269** (3.055)	0.280** (3.038)	0.516** (3.451)	0.393** (2.729)	0.354* (2.286)	0.385* (2.452)
bweduc		0.606** (3.173)				0.520** (3.363)		
bheduc		0.086 (0.945)				0.088 (0.642)		
blhinc		0.207 (1.160)				-0.752 (-1.788)		
bhigh		22.308** (6.533)				3.049 (1.489)		
Observations	882	882	882	882	575	575	575	575
Number of Provinces (or Clusters)			23	62			21	96
R <sup>2</sup>	0.233	0.321	0.322	0.481	0.588	0.654	0.615	0.724

t-value is in the parenthesis. \*t-value is significant at the 5% level. \*\* t-value is significant at the 1% level.

1. The data set for regressions includes respondents who already complete their education. The other independent variables include the squared term of mother's education, the squared term of father's education, the share from mother's income among household income, age of the child, and age of the child squared.

**Table 5: Neighborhood Impact on Boy s Completed Years of Schooling<sup>1</sup>**

Sample	Boy s Completed Years of Schooling in Rural Area				Boy s Completed Years of Schooling in Urban Area			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS (without Neighborhood Variables)	OLS (with Neighborhood Variables)	Province-Fixed-Effect Model	Cluster-Fixed-Effect Model	OLS (without Neighborhood Variables)	OLS (with Neighborhood Variables)	Province-Fixed-Effect Model	Cluster-Fixed-Effect Model
<u>Independent Variables</u>								
weduc	0.281 (0.408)	-0.268 (-0.404)	-0.019 (-0.028)	-0.312 (-0.451)	0.327** (2.995)	0.237* (2.169)	0.276* (2.504)	0.200 (1.748)
heduc	0.346* (2.251)	0.147 (0.992)	0.361* (2.290)	0.160 (0.979)	0.346** (4.553)	0.276** (3.589)	0.335** (4.390)	0.312** (3.728)
lhinc	0.363** (3.073)	0.511** (4.119)	0.045** (3.644)	0.549** (4.071)	0.241 (1.903)	0.138 (1.048)	0.221 (1.680)	0.156 (1.079)
bweduc		0.758** (3.120)				0.116 (0.955)		
bheduc		0.454** (3.607)				0.127 (1.177)		
blhinc		-0.798** (-3.193)				0.221 (0.652)		
bhigh		2.396 (0.613)				0.061 (0.039)		
Observations	748	748	748	748	825	825	825	825
Number of Provinces (or Clusters)			23	62			22	97
R <sup>2</sup>	0.365	0.430	0.418	0.487	0.540	0.554	0.565	0.633

t-value is in the parenthesis. \*t-value is significant at the 5% level. \*\* t-value is significant at the 1% level.

1. The data set for regressions includes respondents who already complete their education. The other independent variables include the squared term of mother's education, the squared term of father's education, the share from mother's income among household income, age of the child, and age of the child squared.

**Table 6: Neighborhood Impact on Girls and Boys Completed Years of Schooling<sup>1</sup>**  
**--- The Effect from Duration Time in Current Address**

Sample	Girls in Rural	Girls in Urban	Boys in Rural	Boys in Urban
weduc	0.474 (0.953)	0.239 (1.923)	-0.263 (-0.393)	0.205 (1.847)
heduc	0.420** (3.897)	0.282** (3.221)	0.161 (1.075)	0.282** (3.588)
lhinc	0.232* (2.492)	0.376* (2.562)	0.494** (3.904)	0.157 (1.163)
bweduc	0.225 (0.364)	0.405 (1.240)	0.655 (0.729)	0.206 (0.780)
bheduc	-0.142 (-0.567)	0.285 (0.947)	-0.217 (-0.551)	-0.002 (-0.008)
blhinc	-0.779 (-1.395)	-0.278 (-0.312)	0.074 (0.093)	0.440 (0.568)
bhigh	48.088** (3.926)	-2.644 (-0.601)	-16.199 (-0.990)	-2.368 (-0.635)
resdur	-0.306 (-1.884)	0.153 (0.565)	0.196 (0.871)	0.076 (0.351)
resdur*bweduc	0.011 (0.871)	0.003 (0.329)	0.001 (0.063)	-0.004 (-0.625)
resdur*bheduc	0.006 (0.994)	-0.005 (-0.629)	0.015 (1.864)	0.004 (0.777)
resdur*blhinc	0.024 (1.921)	-0.010 (-0.495)	-0.017 (-1.015)	-0.006 (-0.342)
resdur*bhigh	-0.613* (-2.457)	0.133 (1.169)	0.345 (1.082)	0.064 (0.698)
Observations	852	557	716	799
R <sup>2</sup>	0.339	0.662	0.442	0.558

t-value is in the parenthesis. \* t-value is significant at the 5% level. \*\* t-value is significant at the 1% level.

1. The data set for regressions includes respondents who already complete their education. The other independent variables include the squared term of mother's education, the squared term of father's education, the share from mother's income among household income, age of the child, and age of the child squared.

**Table 7: Moving Decision (Probit Analysis)**

	(1)	(2)	(3)	(4)
Born Place	Rural/Urban	Rural	Rural	Urban
Move/Stay	Move=1, Stay=0	Moving to Rural (Move=1, Stay=0)	Moving to Urban (Move=1, Stay=0)	Moving to Urban (Move=1, Stay=0)
<b>Independent Variables</b>				
bru	-2.162** (-25.428)			
ru	2.024** (23.228)			
heduc	0.001 (0.137)	-0.015 (-0.970)	-0.005 (-0.396)	-0.004 (-0.411)
weduc	-0.009 (-0.950)	0.006 (0.239)	0.010 (0.475)	-0.010 (-0.869)
bheduc	0.043 (1.528)	0.127** (2.625)	-0.028 (-0.556)	0.018 (0.455)
bweduc	0.083* (2.539)	0.214* (2.255)	1.083** (14.752)	0.104* (2.468)
blhhinc	0.253** (3.910)	0.173 (1.913)	1.205** (10.438)	0.273* (2.504)
bhigh	-0.746 (-1.678)	-7.269** (-4.538)	-3.820 (-3.338)	0.540 (1.069)
Observations	3525	1349	1897	1327
Log Likelihood	-1679.786	-655.783	-619.831	-778.147

z-value is in the parenthesis

\*z-value is significant at 5% level

\*\* z-value is significant at 1% level

**Table 8: Probit Analysis of Moving to a Better Area on the Presence of Children in Specific Age Groups (6-9, 10-14, and 15-18) while Moving**

Dependent Variable	move to higher father's education =1, otherwise=0		move to higher mother's education =1, otherwise=0		move to higher household income =1, otherwise=0	
	(1)	(2)	(3)	(4)	(5)	(6)
c69g	0.081 (0.510)		-0.047 (-0.267)		0.084 (0.555)	
c1014g	0.319 (1.646)		0.263 (1.261)		0.163 (0.859)	
c1518g	0.095 (0.272)		-0.589 (-1.170)		0.758* (2.349)	
c69b	0.112 (0.834)		0.280* (1.993)		0.535** (4.369)	
c1014b	0.055 (0.326)		0.099 (0.547)		0.556** (3.562)	
c1518b	-0.170 (-0.592)		-0.109 (-0.353)		0.092 (0.369)	
c69d		0.195 (1.655)		0.252* (2.007)		0.531** (4.805)
c1014d		0.219 (1.481)		0.237 (1.506)		0.489** (3.504)
c1518d		-0.080 (-0.338)		-0.254 (-0.949)		0.393 (1.865)
Observations	3595	3976	3595	3976	3595	3976
Log Likelihood	-1631.811	-1697.884	-1360.069	-1410.171	-1599.865	-1660.963

z-value is in the parenthesis.

\* z-value is significant at the 5% level

\*\* z-value is significant at the 1% level

The other independent variable is bru.

## Chapter 3: Rate of Return to Education by Using Sibling Data in Taiwan

### I. Introduction

The major concern of this study is the determination of whether the correlation between earnings and education is due to the correlation between family background and education. In addition, this study poses two other questions. First, which gender has the higher rate of return on education, and how is the investment of human capital efficiently allocated between the genders? Second, how sensitive are the estimates of rate of return on education to the presence of measurement error?

There has been considerable empirical research on the issue of the rate of return on education<sup>10</sup>. OLS estimates of the rate of return are considered upward biased due to omitted family factors or unobserved ability. One solution to the problem has been the use of sibling's or identical twin's education to capture omitted family factors or unobserved ability. Recent literature in this branch includes two papers by Miller, Mulvey, and Martin (1997), and Ashenfelter and Zimmerman (1997). The estimation methods in these two papers are very similar. Using Australian twin sample data, Miller, Mulvey, and Martin (1997) suggests that family background is a stronger influence on the relation between schooling and wages in the male sample than in the female sample. On the other hand, using the sample of brother-pairs for the United States, Ashenfelter and Zimmerman (1997) find that family background factors do not influence the relationship between schooling and wages.

As suggested by Ashenfelter and Zimmerman as well as recent literature in household economics, sibling data maybe treated as a kind of *panel data* because siblings share a common unobservable family factor. Based on this framework, this study adopts a new estimation method: minimum distance estimation (MD), which has been proved [(Chamberlain (1982), Carey (1997), Islam (1995), Griliches and Mairesse (1995), and Griliches and Hausman (1986)] to be more efficient for the analysis of panel data than the traditional "within" estimation method. In addition, this paper complements research on data from industrialized countries, by analyzing data from a developing Asian country, Taiwan.

Also, while Ashenfelter & Zimmerman (1997) use only years of schooling as the measure of education, this study measures education both in terms of years of schooling and a dummy variable indicating the earning of at least a bachelor's degree. As suggested in some papers<sup>11</sup>, a diploma is a signal in the labor market and this credential raises earnings. Some papers describe this effect as a "sheepskin effect" or "signaling effect".

The results suggest that in 1990 data from Taiwan, upward bias in the rate of return on schooling due to the omission of family background factors is significant. The family background factor is measured as the schooling including earned diplomas of the individual and the same-sex sibling. If earned diplomas are excluded from the measure of family background, family background ceased to play an important role in the sample of brother-pairs, which is consistent with the results of Ashenfelter and Zimmerman (1997). In 1995 data, the significance of family background factors disappears for both the sample of brother-pairs and sister-pairs.

It is also found that there may be efficiency gains from raising the schooling of girls since the rate of return on schooling is higher among them. This result is consistent with the cross-country analysis in Psacharopoulos (1985). A female also has a greater increase in wages from a diploma than a male, so we conclude that households should encourage their daughters to go to a college. This results may also be applicable to other developing countries.

It is well known that there may be a downward bias in OLS estimates of the rate of return on schooling due to measurement error. The OLS estimation of return on schooling is downward biased due to conjectured 0.05 level of measurement error and omitted family factors for males. For females, OLS estimation of return on schooling is upward biased due to conjectured 0.05 level of measurement error and omitted family factors.

## **II. Theory Background**

There are two major theories to explain the relationship between earnings and education. One theory is the human capital model which is developed by Becker (1964). Becker sees

---

<sup>10</sup> Literature includes Ashenfelter and Zimmerman (1997), Miller, Mulvey, and Martin (1997), Blackburn and Neumark (1995), Ashenfelter and Krueger (1994), Card (1993), Angrist and Krueger (1991), and Griliches (1977), etc.

<sup>11</sup> Some papers suggest the education is a signaling of ability rather than an investment in the human capital in the labor market. Theoretical papers include Arrow(1973) and Spence (1973). Empirical papers include Hungerford and Solon (1987), and Kroch and Sjoblom (1994).

schooling as skill enhancing. A positive rate of return on schooling is seen by him to result from the increased human capital from an extra year of schooling.

Unlike the human capital model, another theory proposes that education is not skill enhancing but is a signal for unobserved ability. In a pure signaling model, individuals with higher ability accumulate more schooling because the accumulation of schooling is cheaper for them. Wages increase in schooling because firms recognize this type of sorting behavior. The screening theory is developed by Arrow (1973), and Spence (1973).

In reality, perhaps both theories are applicable. Hence, in the empirical work, both the human capital and signaling attributes of schooling are taken into account. The human capital contribution of schooling is measured by years of schooling whereas the signaling aspect of schooling is measured by a dummy variable indicating the earning of a bachelor's degree. The following earnings function is considered:

$$Y=f(\text{Years of Schooling, Ability, others})$$

$Y$  is earnings which is a function  $f$  of several factors: years of schooling (the human capital contribution of schooling), ability (captured by diploma or credential), other factors including random factors, etc.

Like other empirical work in this area, this paper deals with both the issues of ability bias and measurement error.

### **III. Econometric Methodology**

The log of the wage rate of child  $i$ ,  $i=1, 2$ , in family  $j$  is a function of years of schooling, a dummy variable indicating the earning of a bachelor's degree, age, unobserved family characteristics, and an individual specific error,

$$(1) \quad Y_{ij} = \beta_i S_{ij} + \gamma_i D_{ij} + \alpha_i \text{Age}_{ij} + F_j + v_{ij}$$

where  $Y$  indicates logarithm of wages;  $S$  is years of schooling;  $D$  represents the dummy variables indicating the earning of a bachelor's degree;  $F_j$  is an unobserved family characteristic; and  $v$  terms indicate the individual specific error. Clearly,  $\beta$  represents the rate of return on schooling. Positive  $\gamma$  would imply value to the earning of a bachelor's degree. The  $\alpha$  parameters would perhaps measure the contribution of work experience to earnings.

Assume that family characteristics,  $F_j$ , are represented as

$$(2) \quad F_j = \lambda_1 S_{1j} + \lambda_2 S_{2j} + \delta_1 D_{1j} + \delta_2 D_{2j} + \kappa_1 \text{Age}_{1j} + \kappa_2 \text{Age}_{2j} + \varepsilon_j$$

The error term  $\varepsilon_i$  is assumed to be uncorrelated with the independent variables.  $\lambda_i, i = 1, 2$ , represents the extent of bias in OLS estimates of the rate of return on education of sibling  $i$  due to the omission of family background factors from the wage equation.  $\delta_i, i = 1, 2$ , indicates the extent of bias in OLS estimates of the value to the earning of a bachelor's degree of sibling  $i$  due to the omission of family background factors from the wage equation.

Substituting (2) into (1) yields the reduced forms

$$(3) \quad \begin{aligned} Y_{1j} &= (\beta_1 + \lambda_1)S_{1j} + \lambda_2 S_{2j} + (\gamma_1 + \delta_1)D_{1j} + \delta_2 D_{2j} + (\alpha_1 + \kappa_1)\text{Age}_{1j} + \kappa_2 \text{Age}_{2j} + w_{1j} \\ Y_{2j} &= \lambda_1 S_{1j} + (\beta_2 + \lambda_2)S_{2j} + \delta_1 D_{1j} + (\gamma_2 + \delta_2)D_{2j} + \kappa_1 \text{Age}_{1j} + (\alpha_2 + \kappa_2)\text{Age}_{2j} + w_{2j} \end{aligned}$$

Rewrite (3) as

$$(4) \quad \begin{aligned} Y_{1j} &= \pi_{11}S_{1j} + \pi_{12}S_{2j} + \pi_{13}D_{1j} + \pi_{14}D_{2j} + \pi_{15}\text{Age}_{1j} + \pi_{16}\text{Age}_{2j} + w_{1j} \\ Y_{2j} &= \pi_{21}S_{1j} + \pi_{22}S_{2j} + \pi_{23}D_{1j} + \pi_{24}D_{2j} + \pi_{25}\text{Age}_{1j} + \pi_{26}\text{Age}_{2j} + w_{2j} \end{aligned}$$

In matrix notation, (3) and (4) could be expressed as

$$\begin{aligned} \begin{pmatrix} Y_{1j} \\ Y_{2j} \end{pmatrix} &= \begin{pmatrix} \beta_1 + \lambda_1 & \lambda_2 & \gamma_1 + \delta_1 & \delta_2 & \alpha_1 + \kappa_1 & \kappa_2 \\ \lambda_1 & \beta_2 + \lambda_2 & \delta_1 & \gamma_2 + \delta_2 & \kappa_1 & \alpha_2 + \kappa_2 \end{pmatrix} \begin{pmatrix} S_{1j} \\ S_{2j} \\ D_{1j} \\ D_{2j} \\ \text{Age}_{1j} \\ \text{Age}_{2j} \end{pmatrix} + \begin{pmatrix} w_{1j} \\ w_{2j} \end{pmatrix} \\ &= \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} & \pi_{16} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} & \pi_{25} & \pi_{26} \end{pmatrix} \begin{pmatrix} S_{1j} \\ S_{2j} \\ D_{1j} \\ D_{2j} \\ \text{Age}_{1j} \\ \text{Age}_{2j} \end{pmatrix} + \begin{pmatrix} w_{1j} \\ w_{2j} \end{pmatrix} \end{aligned}$$

Let

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} & \pi_{16} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} & \pi_{25} & \pi_{26} \end{bmatrix}$$

and

$$G(\theta) = \begin{pmatrix} \beta_1 + \lambda_1 & \lambda_2 & \gamma_1 + \delta_1 & \delta_2 & \alpha_1 + \kappa_1 & \kappa_2 \\ \lambda_1 & \beta_2 + \lambda_2 & \delta_1 & \gamma_2 + \delta_2 & \kappa_1 & \alpha_2 + \kappa_2 \end{pmatrix}$$

where  $\theta$  consists of the elements of  $G(\theta)$ .  $\theta$  is estimated by minimizing the distance between the elements in the  $\Pi$  and  $G(\theta)$  matrices. The function to be minimized is

$$\hat{\theta} = \arg \min(\text{vec}\Pi - \text{vec}G(\theta))'W^{-1}(\text{vec}\Pi - \text{vec}G(\theta))$$

where  $\text{vec}\Pi$  are the elements in  $\Pi$ ,  $\text{vec}G(\theta)$  are the elements in  $G(\theta)$ , and  $W$  is a diagonal matrix whose diagonal elements are  $\text{Asy.Var}[\text{vec}\Pi]$ . The details of the procedure have been described by Chamberlain (1982), Islam (1995), and Greene (1995).

Three special restrictions are discussed below:

(1)  $\beta_1 = \beta_2 = \beta$ ,  $\gamma_1 = \gamma_2 = \gamma$ , and  $\alpha_1 = \alpha_2 = \alpha$  in the equation (3). This is the traditional fixed effect model.  $G(\theta)$  takes the form

$$\begin{pmatrix} \beta + \lambda_1 & \lambda_2 & \gamma + \delta_1 & \delta_2 & \alpha + \kappa_1 & \kappa_2 \\ \lambda_1 & \beta + \lambda_2 & \delta_1 & \gamma + \delta_2 & \kappa_1 & \alpha + \kappa_2 \end{pmatrix}$$

in which there are nine parameters required to be estimated.

(2)  $\lambda_1 = \lambda_2 = \lambda$ ,  $\delta_1 = \delta_2 = \delta$ , and  $\kappa_1 = \kappa_2 = \kappa$ .  $G(\theta)$  takes the form

$$\begin{pmatrix} \beta_1 + \lambda & \lambda & \gamma_1 + \delta & \delta & \alpha_1 + \kappa & \kappa \\ \lambda & \beta_2 + \lambda & \delta & \gamma_2 + \delta & \kappa & \alpha_2 + \kappa \end{pmatrix}$$

in which there are nine parameters required to be estimated.

(3)  $\beta_1 = \beta_2 = \beta$ ,  $\gamma_1 = \gamma_2 = \gamma$ ,  $\alpha_1 = \alpha_2 = \alpha$ ,  $\lambda_1 = \lambda_2 = \lambda$ ,  $\delta_1 = \delta_2 = \delta$ , and  $\kappa_1 = \kappa_2 = \kappa$ .

Therefore, the restricted form is

$$G(\theta) = \begin{pmatrix} \beta + \lambda & \lambda & \gamma + \delta & \delta & \alpha + \kappa & \kappa \\ \lambda & \beta + \lambda & \delta & \gamma + \delta & \kappa & \alpha + \kappa \end{pmatrix}$$

in which there are only six parameters left to be estimated.

To consider measurement error in the reporting education in the analysis, this study adopts a method used by Ashenfelter and Zimmerman (1997). It is suggested that equation (4) should be expressed as equations (6) and (7) when the measurement error is considered in the model.

$$\begin{aligned}
(6) \quad Y_{1j} &= \left[ (\beta_1 + \lambda_1) - \frac{(\beta_1 + \lambda_1)\varphi_1 - \frac{\text{cov}(S_1, S_2)}{\text{var}(S_1)}\lambda_2\varphi_2}{1 - \rho^2} \right] S_{1j} \\
&+ \left[ \lambda_2 - \frac{\lambda_2\varphi_2 - \frac{\text{cov}(S_1, S_2)}{\text{var}(S_2)}(\beta_1 + \lambda_1)\varphi_1}{1 - \rho^2} \right] S_{2j} + w_{1j} \\
(7) \quad Y_{2j} &= \left[ \lambda_1 - \frac{\lambda_1\varphi_1 - \frac{\text{cov}(S_1, S_2)}{\text{var}(S_1)}(\beta_2 + \lambda_2)\varphi_2}{1 - \rho^2} \right] S_{1j} \\
&+ \left[ (\beta_2 + \lambda_2) - \frac{(\beta_2 + \lambda_2)\varphi_2 - \frac{\text{cov}(S_1, S_2)}{\text{var}(S_2)}\lambda_1\varphi_1}{1 - \rho^2} \right] S_{2j} + w_{2j}
\end{aligned}$$

where  $\varphi_1$  equals to  $\text{var}(m_1)/\text{var}(S_1)$  and  $\varphi_2$  equals to  $\text{var}(m_2)/\text{var}(S_2)$ .  $m_1$  and  $m_2$  are measurement error terms for the years of schooling,  $S_1$  and  $S_2$ .

Because the measurement error is unobservable, a certain values of  $\varphi_1$  and  $\varphi_2$  should be specified. Following the method developed by Ashenfelter and Zimmerman (1997), an identical measurement error for two siblings is assumed, that is,  $\varphi_1 = \varphi_2$ . High and low levels of measurement error (0.20 and 0.05) are specified. Equations (6) and (7) represent the unrestricted case. However, above restrictions are applied, and the results are presented in Section V.

### Family Hypothesis

The correlation between individual's wage rate and his (her) education might partially come from family factors. If brothers and sisters could affect each other, the coefficient of siblings' education ( $\lambda_1, \lambda_2, \delta_1, \delta_2, \kappa_1$ , and  $\kappa_2$ ) on individual's wage rate should be significantly positive. In addition, the positive impact from elder or younger siblings could be distinguished by comparing the magnitude and significance of  $\lambda_1$  ( $\delta_1$ ) and  $\lambda_2$  ( $\delta_2$ ). If there is a positive impact (so called the role model effect, usually from elder siblings to younger siblings (Chen,

Michalopoulos, and Salehi (1998)),  $\lambda_1$  will be higher than  $\lambda_2$ . If positive impact does not exist,  $\lambda_1$  should be equal to  $\lambda_2$ . Another possible reason is the birth order effect, which is the first born child could receive more attention from parents than the other children. Hence, a higher correlation exists between elder sibling and family background.

The factors of family budget constraint or parents preference factor are not considered in this model. Since the samples of the same-sex sibling pairs are used, the gender issue is not raised in the model. The results of wage comparison between brother and sister samples will be discussed later as well as gender issue. It is more likely that  $\lambda_1$  is greater than  $\lambda_2$  in the female sample than in the male sample because a daughter's education has higher impact on the younger sibling's education in Taiwan [suggested from some literature: Parish and Willis (1993), Sassler (1995)].

#### **IV. Data**

This paper uses two data sets of household survey of human resources in Taiwan collected in 1990 and 1995. This survey focuses on family members with age sixteen and above. After excluding the respondents who are currently not working, the data collected in 1990 contains 4097 households in which there are totally 6702 individuals with age sixteen and above who are currently employed. The data collected in 1995 contain 5143 households in which there are 8074 individuals with age sixteen and above who are currently employed. The sibling data containing brother pairs and sister pairs for both survey in 1990 and 1995 are generated. Four variables from the sibling data sets are summarized in Table 1. The variables presented in Table 1 include a logarithm of hourly wage, completed years of schooling, dummy variable indicating the earning of a bachelor's degree, and age.

This study considers two principles when generating these sibling data sets. The first consideration is to choose the elder siblings instead of the younger siblings because the jobs of the elder respondents are usually more permanent than those of the younger respondents. The young respondents may work for temporary jobs and, therefore, their wage rate may not reflect their education status. The second consideration is to minimize the age difference between two siblings in order to diminish the effect of cohort or life cycle. Based on these two considerations, the

oldest two brothers having paid job in the last week from the reporting date are chosen for further analysis. The same rule is applied to the data sets of sister pairs.

In the 1990 data, there are 741 observations in the sample of brother pairs and 173 observations in the sample of sister pairs. The average ages for elder brother and younger brother are 31 and 27 years old, respectively. The average ages for the elder sister and younger sister are 27 and 24 years old, respectively.

In the 1995 data, there are 949 observations and 221 observations for brother pairs and sister pairs, respectively. The average ages for the elder brother and younger brother are 32 and 28 years old, respectively; for sister are 27 and 25 years old. The mean ages of sister pairs are apparently lower than that of brother pairs in both year surveys. The reason could be that in Taiwan's tradition, daughters usually move out of the households after they get married. In contrast, sons are more likely to stay in the household after they get married. The average years of schooling for women are around 12 years in 1990 and 13 years in 1995 which are two years more than the men's average years of schooling, i.e., 10 years in 1990 and 11 years in 1995. Women receive equal or higher degree of bachelor's degree in a higher percentage, 7.5% in 1990 and 23.5% in 1995 than men do, 5% in 1990 and 8% in 1995. This evidence suggests that the data may suffer from a problem related to censoring. In other words, the females in the sample are more likely to be single and receive a high wage rate associated with more education.

Notice that the average wage rate of the men is higher than that of women even though men have lower average education. Men have 4.6 of log hourly wage rate and women have 4.3 of log hourly wage rate. The log wage rate by education year was decomposed and listed in Table 2. For male, the log wage rate increases steadily with increase of years of schooling before 15 years of schooling, i.e., before receiving bachelor's degree. The log wage rate jumps from 4.7 to 4.9 when males receive a bachelor's degree or higher. For females, this wage jump starts from receiving a degree from community college (from 4.3 of log wage for high school graduates to 4.6 of log wage in community college graduates) and have another jump when receiving a bachelor's degree (from 4.6 of log wage in community college graduates to 4.7 of log wage in bachelor's degree).

## V. Results

### V.1 OLS Regression

Table 3 presents the OLS estimations of the rate of return on schooling. Notice that OLS estimations of return on schooling might be downward biased due to the measurement error of schooling and upward biased due to the omitted family factor. Sections V.2 and V.3 present individual estimations of these two possible biases.

The dependent variable is logarithm of the wage rate. The major independent variables are years of schooling and a dummy variable indicating the earning of a bachelor's degree. The coefficients represent the rate of return on schooling ( $\beta_1$  for elder siblings and  $\beta_2$  for younger siblings) and the value of the earning of a bachelor's degree ( $\gamma_1$  for elder siblings and  $\gamma_2$  for younger siblings). The other independent variables include age and age squared.

In the sample of brother pairs (the top part in Table 3), the OLS estimations show that one additional year of schooling increases the wage rate by 2.5% for the elder brothers (coefficients  $\beta_1$  are 0.025 in 1990 data and 0.023 in 1995 data) and by 1% for the younger brothers (coefficients  $\beta_2$  are 0.010 for both 1990 and 1995 data set). This result indicates that the rate of return on schooling in Taiwan is lower than that in industrialized countries. The relatively low rate of return on schooling in Taiwan may due to the introduction of the earning of a bachelor's degree in the model. Taiwan is a country with a high population and a high average education. Since the over population and high average education can make the job opportunity more competitive, an employer could hire a highly-educated employee at a relatively low wage rate. The estimations of the coefficients in industrialized countries are usually in the range of 0.06 to 0.12 in the literature<sup>12</sup>.

Notice that in the 1990 data, the earning of a bachelor's degree significantly increases the wage rate by 20% for elder brothers (coefficient  $\gamma_1$  is 0.20) and by 27% for younger brothers (coefficient  $\gamma_2$  is 0.27). The status of the earning of a bachelor's degree could be an important signal representing the respondent's ability in Taiwan and the wage rate is highly dependent on this factor for males. The importance of earning a bachelor's degree decreases in the 1995 data,

---

<sup>12</sup> The literature includes Miller, Mulvery, and Martin (1997), Ashenfelter and Zimmerman (1997), Ashenfelter and Krueger (1994), Angrist and Krueger (1991), and Griliches (1977).

in which the coefficients of earning of a bachelor's degree drop to 0.141 for elder brothers ( $\gamma_1$ ) and 0.211 for younger brothers ( $\gamma_2$ ). Since the percentage of males receiving a bachelor's degree is higher in 1995 than in 1990 (see Table 1), there are more high-educated males competing for the limited job offers, which might be the reason that the importance of earning a bachelor's degree drops from 1990 to 1995.

In the sample of sister pairs (the bottom part in Table 3), the OLS estimations show that in 1990 data, the wage rate increases by 4.2% for elder sisters (the coefficient  $\beta_1$  is 0.042) and 4.5% for younger sisters (the coefficient  $\beta_2$  is 0.045) with an additional year of education. In 1995 survey, the estimated coefficients of years of schooling for females increase to 4.9% for elder sisters ( $\beta_1$ ) and 6.7% for the younger sisters ( $\beta_2$ ). The rate of return on schooling for females is higher than that for males. This result is consistent with the major finding for other countries in the literature [for instance, Psacharopoulos (1985)]. The higher rate of return on education for females than for males might be a *relative concept*, as suggested by Psacharopoulos (1985). That is, high-educated females in the female labor forces is relatively less competitive than high-educated males in male labor forces.

Upon the earning of a bachelor's degree, elder sisters receive higher log wage rate ( $\gamma_1=0.219$ ) and younger sisters also receive higher log wage rate ( $\gamma_2=0.158$ ) in 1990 survey than those in 1995 survey. In 1995, the coefficients for the earning of a bachelor's degree drop to 0.11 ( $\gamma_1$ ) for elder sisters and 0.117 ( $\gamma_2$ ) for younger sisters. Receiving a bachelor's degree becomes less important for the wages in 1995. Since the percentage of females receiving a bachelor's degree in 1995 is also higher than in 1990 (see Table 1), more females with bachelor's degree are competing for limited job offers in recent years.

## V.2 Unobserved Ability (Family Factors)

The above analysis in Section V.1 ignores unobserved ability, which is the family factors. Hence, the estimated coefficients in the OLS might be upward biased due to the omitted family factors. The estimated coefficients of family factors include the elder siblings' years of schooling ( $\lambda_1$ ), younger siblings' years of schooling ( $\lambda_2$ ), a dummy variable indicating that the elder sibling has earned a bachelor's degree ( $\delta_1$ ), and a dummy variable indicating that the younger sibling has earned a bachelor's degree ( $\delta_2$ ).

This study uses the Minimum Distance (MD) method in order to consider unobserved ability (or family factor) into the model. Without considering the measurement error in the model, the MD estimations for the brother pairs and sister pairs are listed in Tables 4 and 5, respectively. The family factors are symbolized by  $\lambda_1$ ,  $\lambda_2$ ,  $\delta_1$ , and  $\delta_2$ . Four cases are reported in both Tables 4 and 5: unrestricted case (column 1), fixed effect case (column 2), identical impact from siblings to family (column 3), and both restrictions in columns 2 and 3 (column 4). These four cases have been introduced in Section III. The  $p$ -values of  $\chi^2$  statistics for testing the specification of fixed effect are low, for instance, 0.0470 in 1990 data and 0.0763 in 1995 data in Table 4. However, the  $p$ -values are higher than the critical value 0.01. Hence, the specification of the restrictions cannot be rejected. The  $p$ -values of  $\chi^2$  statistics for testing the other restrictions are high, the hypothesis cannot be rejected, either.

For the 1990 data, both  $\lambda_1$  and  $\lambda_2$  for males appear to be insignificant. This result is consistent with the finding in the sample of brother pairs from the United States [Ashenfelter and Zimmerman (1997)]. In their article, they conclude that the upward bias for the OLS estimation of return on schooling due to the omitted factors of family background is small. However, if another two factors of family background, dummy variables of elder brother's and younger brother's earning of a bachelor's degree ( $D_1$  and  $D_2$ , respectively), are considered as the family factors, family background obviously plays an important role for wage rate ( $\delta$  coefficients are significant in the 1990 survey in Table 4). This result implies that the omission of family factors might not upward bias the OLS estimation of return on schooling, instead, it might upward bias the value of the earning of a bachelor's degree. The same results are obtained in the three restricted cases (columns 2, 3, and 4 in Table 4).

For the sample of sister pairs in the 1990 survey, the bias of omitted family factor is also significant;  $\lambda_1$  is significant (in column 1 in Table 5). For the 1995 data, there is no significant impact from the family factors on wages in the samples of both brother pairs and sister pairs. All the coefficients of family characteristics ( $\lambda_1, \lambda_2, \delta_1, \delta_2$ ) are not significant in the samples (1995 survey in Tables 4 and 5). In a word, the family factor is significant in 1990 but not significant in 1995 for both males and females.

The gap of the estimated rate of return on schooling between the samples of brother pairs and sister pairs is narrower than that of the OLS estimation, after considering the bias of omitted factors of family background. In 1990, females received 2.2% and 3.3% ( $\beta_1$  and  $\beta_2$ , respectively, in column 1 of Table 5) higher wages for elder and younger siblings. Males received 3.1% and 1.5 % ( $\beta_1$  and  $\beta_2$ , respectively, in column 1 of Table 4) higher wages for an additional year of schooling for elder siblings and younger siblings. In Table 3 of the 1990 survey, the OLS estimates of  $\beta_1$  and  $\beta_2$  are 0.042 and 0.045, respectively, for females, and  $\beta_1$  and  $\beta_2$  are 0.025 and 0.010, respectively, for males. In Table 5 of the 1995 survey, females receive 3.2% ( $\beta_1$ ) and 5.7% ( $\beta_2$ ) higher wages for an additional year of schooling and males receive 2.4% ( $\beta_1$ ) and 1.8% ( $\beta_2$ ) higher wages for an additional year of schooling. In Table 3 of the 1995 survey, the OLS estimates of  $\beta_1$  and  $\beta_2$  are 0.049 and 0.067, respectively, for females, and  $\beta_1$  and  $\beta_2$  are 0.023 and 0.010, respectively, for males.

Notice that the estimation of return on schooling in 1995 is higher than in 1990 for females after considering the family factors (Table 5). In contrast, the estimation of return on schooling in 1995 is lower than in 1990 for males, after considering the family factors (Table 4). The finding suggests that the job market is more competitive for males, and less competitive for females, in 1995 than in 1990 in Taiwan.

### V.3 Measurement Error

The bias raised by a conjecture magnitude of measurement error is estimated in this section. Tables 6 and 7 present the MD estimation of return on schooling under a specific conjecture level, 0.05, of measurement error with the family factors, using the samples of brother pairs and sister pairs, respectively. Tables 8 and 9 present the MD estimation of return on schooling under a specific level, 0.20, of measurement error and with the family factors for brother pairs and sister pairs, respectively.

In Table 6, estimated returns on schooling are 3.4% ( $\beta_1$ ) for elder brothers and 1.8% ( $\beta_2$ ) for younger brothers in the unrestricted case (column 1) in the 1990 data. Comparing these figures to Table 4, 0.3% of bias to the estimated returns on schooling for both elder brother and younger brothers is created when there is 0.05 level of measurement error in reporting years of schooling ( $\beta_1=0.031$  and  $\beta_2=0.015$  in the without-measurement-error case in 1990 data). In Table

7, estimated returns are 2.5% ( $\beta_1$ ) for elder sisters and 3.7% ( $\beta_2$ ) for younger sisters in the unrestricted case (column 1) in the 1990 data. Comparing these figures to those in Table 5, 0.3% of bias of the estimated returns on schooling for elder sisters and 0.4% for younger sisters is created when there is 0.05 level of measurement error in reporting years of schooling ( $\beta_1=0.022$  and  $\beta_2=0.033$  in the without-measurement-error case in 1990 data). These results show that a small downward bias, 0.3%, is created by 0.05 level of the measurement error for both sister-pair data and brother-pair data.

The downward bias due to 0.05 level of measurement error is 0.3% for brother-pair data in 1995 (comparing the columns 1 in Tables 4 and 6). For sister-pair data in 1995, the downward biases due to 0.05 level of measurement error are 0.5% and 0.7% for elder sisters and younger sisters, respectively.

If the measurement error is 0.20 for the schooling variance, returns are 5.2% ( $\beta_1$ ) for elder brothers and 3.6% ( $\beta_2$ ) for younger brothers in the unrestricted case (see column 1 in Table 8). This indicates that the 0.20 level of measurement error creates a 2.1% downward bias of the estimated return on schooling for both elder and younger brothers [comparing 0.052 (Table 8) to 0.031 (Table 4) for elder brother and comparing 0.036 (Table 8) to 0.015 (Table 4) for younger brother]. Comparing the unrestricted case for the sample of sister pairs in Tables 5 and 9, a 0.20 level of measurement error creates a 2.4% downward bias for elder sisters and 2.7% downward bias for younger sisters ( $\beta_1$  increases from 0.022 in Table 5 to 0.046 in Table 9 and  $\beta_2$  increases from 0.033 in Table 5 to 0.060 in Table 9).

The estimated return on schooling is 3.4% and 1.8% for elder and younger brothers, respectively, in 1990, after removing the downward bias derived from 5% level of measurement error and upward bias derived from the omitted family factors. These figures are higher than those in OLS estimations. For elder and younger sisters, the estimated returns on schooling are 2.5% and 3.7%, respectively, after removing the downward bias derived from 5% level of measurement error and upward bias derived from the omitted family factors. These figures are lower than those in the OLS estimation.

#### V.4 Summary

The major findings of these results are: 1) The earning of a bachelor's degree for each individual in the family is considered as an important variable of family background, and the upward bias in the rate of return on education derived from the omitted family factors is found in the 1990 data but not in the 1995 data. 2) Females have a higher rate of return on schooling than have males in Taiwan. 3) The OLS estimation of return to schooling is downward biased due to conjectured 0.05 level of measurement error and omitted family factors for males. In contrast for females, OLS estimation of return on schooling is upward biased due to conjectured 0.05 level of measurement error and omitted family factors.

#### **VI. Conclusion**

This study uses Minimum Distance Estimation (MD) to measure the rate of return on education for the sibling data collected in Taiwan in 1990 and 1995.

The bias due to the omitted family factors is significant in 1990 but not in 1995. According to the results in the 1990 data, there is a significantly positive impact on log wage rate from siblings having a bachelor's degree.

While the literature suggests that the range of rate of return on schooling for industrialized countries is between 0.06 and 0.12 [Miller, Mulvey, and Martin (1997), Ashenfelter and Zimmerman (1997), Ashenfelter and Krueger (1994), Angrist and Krueger (1991), Griliches (1977)], our results show that the range of rate of return on schooling is 0.018 to 0.034 for males and 0.025 to 0.064 for females in Taiwan (considering the correlated family factors and conjectured measurement error with level 0.05). These figures are much lower than the values found in industrialized countries. However, the estimated values for another education variable, earning a bachelor's degree (0.021 to 0.107 for male and 0.161 to 0.163 for female, for elder and younger siblings, respectively), are much higher than those of the rate of return on schooling. The results suggest that the high rate of return on years of schooling in the literature might come from the omitted signal variable, usually a bachelor's degree.

The higher rate of return on the education for females relative to males indicates that households get higher return when they invest more in their daughters' education. This suggests that obtaining a bachelor's degree for daughters should be encouraged in Taiwan.

For males, the OLS estimation of return on schooling is downward biased due to conjectured 0.05 level of measurement error and omitted family factors. For females, OLS estimation of return on schooling is upward biased due to conjectured 0.05 level of measurement error and omitted family factors. Ashenfelter and Zimmerman (1997) use the United States data and find that the downward bias due to the measurement error and the upward bias due to the unobserved family background are canceled out when using the OLS method. Hence, they conclude that the OLS method will not cause bias. However, this conclusion does not hold by using Taiwan data.

### **References:**

1. Angrist, Joshua D., and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106:4 (1991), 979-1014.
2. Arrow, Kenneth J., "Higher Education as a Filter," *Journal of Public Economics*, 2:3 (1973), 193-216.
3. Ashenfelter, Orley, and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review* 84:5 (1994), 1157-1173.
4. Ashenfelter, Orley, and David J. Zimmerman, "Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons, and Brothers," *The Review of Economics and Statistics* 79:1 (1997), 1-9.
5. Becker, Gary S., *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, (1964) New York: National Bureau of Economic Research.
6. Behrman, Jere R., and Barbara L. Wolfe, "The Socioeconomic Impact of Schooling in A Developing Country," *The Review of Economics and Statistics* 66:2 (1984), 296-303.
7. Berndt, Ernst R., *The Practice of Econometrics: Classic and Contemporary*, (1991), Addison Wesley Publishing Company.
8. Blackburn, McKinley L., and David Neumark, "Are OLS Estimates of the Return to Schooling Biased Downward? Another Look," *The Review of Economics and Statistics* (1995), 217-29.
9. Bound, John, Zvi Griliches, and Bronwyn H. Hall, "Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ?" *International Economic Review* 27:1 (1986), 77-105.

10. Carey, Kathleen, "A Panel Data Design for Estimation of Hospital Cost Functions," *The Review of Economics and Statistics* 79:3 (1997), 443-53.
11. Chamberlain, Gary, "Multivariate Regression Models for Panel Data," *Journal of Econometrics* 18:5 (1982), 5-46.
12. Chen, Huei-Ling, Djavad Salehi-Isfahani, and Charles Michalopoulos, "Household Models and Formations of Human Capital with Sibling Effect in Iran," mimeo (1998).
13. Griliches, Zvi, "Estimating The Returns to Schooling: Some Econometric Problems," *Econometrica* 45:1 (1977), 1-22.
14. Griliches, Zvi, and Jerry A. Hausman, "Errors in Variables in Panel Data," *Journal of Econometrics* 31 (1986), 93-118.
15. Griliches, Zvi, and Jacques Mairesse, "Production Functions: The Search for Identification," *NBER Working Paper No. 1719* (1995).
16. Hungerford, Thomas, and Gary Solon, "Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics* 69:1 (1987), 175-77.
17. Islam, Nazrul, "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics* (1995), 1127-1170
18. Kroch, Eugene A. and Kriss Sjoblom, "Schooling as Human Capital or a Signal," *Journal of Human Resources* 29:1 (1994), 156-180.
19. Miller, Paul, Charles Mulvey, and Nick Martin, "Family Characteristics and the Returns to Schooling," *Economica* 64 (1997), 119-36.
20. Psacharopoulos, George, "Returns to Education: A Further International Update and Implications," *Journal of Human Resources* 20:4 (1985), 583-604.
21. Sassler, Sharon, "Trade-Offs in the Family: Sibling Effects on Daughters' Activities in 1910," *Demography* 32:4 (1995), 557-575.
22. Spence, M., "Job Market Signaling," *Quarterly Journal of Economics*, 1973.
23. Thomas, Duncan, "Intra-Household Resource Allocation: An Inferential Approach," *The Journal of Human Resources* 25:4 (1990), 635-664.

**Table 1: Summary Result of the Samples of Brother Pairs and Sister Pairs**  
**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

Samples	Survey 1990				Survey 1995			
	Elder Sibling		Younger Sibling		Elder Sibling		Younger Sibling	
Characteristics	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
<u>Brother Sample</u>								
Log Hourly Wage	4.658	0.380	4.581	0.373	5.078	0.529	5.039	0.343
Years of Schooling	10.524	3.067	10.849	2.720	11.168	2.901	11.366	2.526
Age	30.937	5.150	26.923	4.077	31.849	5.572	28.064	4.569
Percentage of Receiving a bachelor's Degree	0.058		0.063		0.08		0.054	
Sample Size	741				949			
<u>Sister Sample</u>								
Log Hourly Wage	4.375	0.350	4.340	0.369	4.854	0.345	4.771	0.331
Years of Schooling	12.179	2.630	12.208	2.469	13.385	2.539	13.158	2.217
Age	27.069	4.525	24.145	3.500	26.986	3.457	24.403	2.754
Percentage of Receiving a bachelor's Degree	0.110		0.075		0.235		0.149	
Sample Size	173				221			

**Table 2: Mean of the Log Wage Rate of the Samples of Brother Pairs and Sister Pairs**  
**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

Samples	Survey 1990				Survey 1995			
	Elder Sibling		Younger Sibling		Elder Sibling		Younger Sibling	
<u>Schooling Year</u>	<u>Mean</u>	<u>Obs.</u>	<u>Mean</u>	<u>Obs.</u>	<u>Mean</u>	<u>Obs.</u>	<u>Mean</u>	<u>Obs.</u>
<b><u>Brother Sample</u></b>								
6 years (elementary school)	4.573	141	4.688	71	4.994	99	5.079	45
9 years (junior high school)	4.606	224	4.504	275	5.032	289	5.011	306
12 years (senior high school)	4.674	251	4.545	280	5.059	370	5.010	423
15 years (community college)	4.722	82	4.686	68	5.168	115	5.081	124
16 years above (bachelor's degree or higher degree)	4.994	43	4.935	47	5.323	76	5.316	51
Total Observations	741				949			
<b><u>Sister Sample</u></b>								
6 years (elementary school)	4.240	12	4.205	7	4.562	6	4.452	2
9 years (junior high school)	4.188	21	4.238	28	4.720	15	4.469	19
12 years (senior high school)	4.295	91	4.247	89	4.722	93	4.660	104
15 years (community college)	4.572	30	4.552	36	4.919	55	4.886	63
16 years above (bachelor's degree or higher degree)	4.740	19	4.691	13	5.093	52	5.092	33
Total Observations	173				221			

**Table 3: The OLS Estimation for Rate of Return on Education for the Samples of Brother Pairs and Sister Pairs**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

Samples	Survey 1990		Survey 1995	
<u>Brother Sample</u>	<u>Elder Brother</u>	<u>Younger Brother</u>	<u>Elder Brother</u>	<u>Younger Brother</u>
$\beta_1$	0.025** (4.814)		0.023** (3.235)	
$\beta_2$		0.010 (1.823)		0.010* (2.113)
$\gamma_1$	0.200** (3.104)		0.141 (1.924)	
$\gamma_2$		0.270** (4.387)		0.211** (3.953)
<u>Sister Sample</u>	<u>Elder Sister</u>	<u>Younger Sister</u>	<u>Elder Sister</u>	<u>Younger Sister</u>
$\beta_1$	0.042** (3.941)		0.049** (4.758)	
$\beta_2$		0.045** (3.742)		0.067** (6.465)
$\gamma_1$	0.219* (2.483)		0.110 (1.796)	
$\gamma_2$		0.158 (1.418)		0.117 (1.803)

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_1$  and  $\beta_2$  parameters represent the estimated rates of return on schooling for elder and younger siblings, respectively. The  $\gamma_1$  and  $\gamma_2$  parameters represent the estimated value to the earning of a bachelor's degree for elder and younger siblings, respectively.

**Table 4: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (Without Measurement Error) for Brother-Pair Sample**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

	Survey 1990				Survey 1995			
	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$
$\beta_1$	0.031** (4.347)		0.030** (4.359)		0.024* (2.037)		0.019* (2.307)	
$\beta_2$	0.015 (1.854)		0.016* (2.115)		0.018 (1.093)		0.010 (1.283)	
$\beta_1=\beta_2=\beta$		0.026** (3.832)		0.025** (3.915)		0.024** (2.794)		0.018** (2.800)
$\gamma_1$	0.021 (0.230)		0.033 (0.425)		0.103 (1.716)		0.104 (1.822)	
$\gamma_2$	0.127 (1.490)		0.084 (1.141)		0.093 (1.093)		0.148** (2.414)	
$\gamma_1=\gamma_2=\gamma$		0.059 (0.957)		0.068 (1.112)		0.099 (1.936)		0.121** (2.528)
$\lambda_1$	-0.005 (-1.068)	-0.004 (-0.707)			0.002 (0.317)	0.003 (0.532)		
$\lambda_2$	-0.004 (-0.537)	-0.010 (-1.611)			-0.009 (-0.587)	-0.014 (-1.444)		
$\lambda_1=\lambda_2=\lambda$			-0.005 (-1.208)	-0.006 (-1.501)			-0.0001 (-0.016)	-0.002 (-0.556)
$\delta_1$	0.156 (1.850)	0.142* (2.168)			0.013 (0.307)	0.013 (0.305)		
$\delta_2$	0.131** (2.397)	0.178** (3.455)			0.112 (1.319)	0.106 (1.888)		
$\delta_1=\delta_2=\delta$			0.139** (3.184)	0.143** (3.440)			0.030 (0.779)	0.040 (1.142)
Prob> $\chi^2$	-	0.0470	0.9635	0.5285	-	0.0763	0.7212	0.5278

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_i$  parameters represent the estimated rates of return on schooling. The  $\gamma_i$  parameters represent the estimated value to the earning of a bachelor's degree. The  $\lambda_i$  parameters represent the extent of bias in OLS estimates of the rate of return on education due to the omission of family background factors from the wage equation. The  $\delta_i$  parameters indicate the extent of bias in OLS estimates of the value to the earning of a bachelor's degree due to the omission of family background factors from the wage equation.  $i, i=1,2$ , refers to elder and younger siblings, respectively.

**Table 5: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (Without Measurement Error) for Sister-Pair Sample**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

	Survey 1990				Survey 1995			
	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$
$\beta_1$	0.022** (2.360)		0.025** (2.763)		0.032* (1.967)		0.030 (1.859)	
$\beta_2$	0.033** (2.520)		0.033** (2.545)		0.057** (3.100)		0.046** (2.698)	
$\beta_1=\beta_2=\beta$		0.024** (2.685)		0.024** (2.773)		0.039** (2.617)		0.041** (2.852)
$\gamma_1$	0.135 (1.287)		0.172 (1.748)		0.150 (1.867)		0.117 (1.516)	
$\gamma_2$	0.163 (1.743)		0.170 (1.838)		0.061 (0.720)		0.092 (1.159)	
$\gamma_1=\gamma_2=\gamma$		0.181* (2.285)		0.181** (2.339)		0.084 (1.383)		0.071 (1.171)
$\lambda_1$	0.026** (2.349)	0.024* (2.170)			0.009 (0.860)	0.006 (0.581)		
$\lambda_2$	-0.008 (-0.674)	-0.004 (-0.343)			0.006 (0.463)	0.014 (1.086)		
$\lambda_1=\lambda_2=\lambda$			0.008 (1.257)	0.011 (1.677)			0.012 (1.456)	0.011 (1.396)
$\delta_1$	0.082 (0.685)	0.056 (0.516)			-0.048 (-0.771)	-0.018 (-0.326)		
$\delta_2$	-0.002 (-0.024)	-0.022 (-0.217)			0.067 (0.939)	0.057 (0.887)		
$\delta_1=\delta_2=\delta$			0.034 (0.469)	0.016 (0.237)			0.001 (0.014)	0.013 (0.325)
Prob> $\chi^2$	-	0.2548	0.2327	0.2269	-	0.1589	0.4421	0.3363

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_i$  parameters represent the estimated rates of return on schooling. The  $\gamma_i$  parameters represent the estimated value to the earning of a bachelor's degree. The  $\lambda_i$  parameters represent the extent of bias in OLS estimates of the rate of return on education due to the omission of family background factors from the wage equation. The  $\delta_i$  parameters indicate the extent of bias in OLS estimates of the value to the earning of a bachelor's degree due to the omission of family background factors from the wage equation.  $i, i=1,2$ , refers to elder and younger siblings, respectively.

**Table 6: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.05) for Brother-Pair Sample**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

	Survey 1990				Survey 1995			
	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$
$\beta_1$	0.034** (4.325)		0.033** (4.333)		0.027* (2.015)		0.021* (2.285)	
$\beta_2$	0.018* (2.003)		0.019* (2.220)		0.020 (1.256)		0.011 (1.297)	
$\beta_1=\beta_2=\beta$		0.029** (3.966)		0.029** (3.937)		0.031** (3.291)		0.020** (2.807)
$\gamma_1$	0.021 (0.230)		0.034 (0.430)		0.103 (1.716)		0.105 (1.843)	
$\gamma_2$	0.107 (1.167)		0.095 (1.135)		0.093 (1.093)		0.149** (2.425)	
$\gamma_1=\gamma_2=\gamma$		0.069 (1.116)		0.068 (1.099)		0.088 (1.731)		0.121** (2.516)
$\lambda_1$	-0.007 (-1.130)	-0.005 (-0.892)			0.001 (0.251)	0.002 (0.366)		
$\lambda_2$	-0.006 (-0.708)	-0.012 (-1.637)			-0.011 (-0.644)	-0.020 (-1.875)		
$\lambda_1=\lambda_2=\lambda$			-0.006 (-1.376)	-0.008 (-1.709)			-0.0005 (-0.102)	-0.003 (-0.670)
$\delta_1$	0.156 (1.850)	0.129* (1.963)			0.013 (0.307)	0.014 (0.324)		
$\delta_2$	0.131** (2.397)	0.157** (3.059)			0.112 (1.319)	0.119* (2.122)		
$\delta_1=\delta_2=\delta$			0.138** (3.157)	0.143** (3.440)			0.029 (0.769)	0.040 (1.152)
Prob> $\chi^2$	-	0.1852	0.9667	0.6992	-	0.5974	0.7203	0.5346

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_i$  parameters represent the estimated rates of return on schooling. The  $\gamma_i$  parameters represent the estimated value to the earning of a bachelor's degree. The  $\lambda_i$  parameters represent the extent of bias in OLS estimates of the rate of return on education due to the omission of family background factors from the wage equation. The  $\delta_i$  parameters indicate the extent of bias in OLS estimates of the value to the earning of a bachelor's degree due to the omission of family background factors from the wage equation.  $i, i=1,2$ , refers to elder and younger siblings, respectively.

**Table 7: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.05) for Sister-Pair Sample**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

	Survey 1990				Survey 1995			
	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$
$\beta_1$	0.025** (2.443)		0.029** (2.811)		0.037* (2.057)		0.034 (1.927)	
$\beta_2$	0.037** (2.561)		0.037** (2.639)		0.064** (3.121)		0.052** (2.709)	
$\beta_1=\beta_2=\beta$		0.026** (2.545)		0.027** (2.761)		0.051** (3.137)		0.045** (2.823)
$\gamma_1$	0.135 (1.287)		0.174 (1.763)		0.150 (1.867)		0.116 (1.506)	
$\gamma_2$	0.163 (1.743)		0.167 (1.807)		0.061 (0.720)		0.093 (1.160)	
$\gamma_1=\gamma_2=\gamma$		0.173* (2.195)		0.182** (2.349)		0.084 (1.381)		0.071 (1.166)
$\lambda_1$	0.027* (2.182)	0.028* (2.220)			0.007 (0.620)	0.002 (0.156)		
$\lambda_2$	-0.012 (-0.854)	-0.007 (-0.516)			0.004 (0.301)	0.011 (0.796)		
$\lambda_1=\lambda_2=\lambda$			0.007 (0.972)	0.010 (1.427)			0.010 (1.167)	0.010 (1.125)
$\delta_1$	0.081 (0.685)	0.050 (0.460)			-0.048 (-0.771)	-0.027 (-0.489)		
$\delta_2$	-0.002 (-0.024)	-0.015 (-0.142)			0.067 (0.939)	0.056 (0.869)		
$\delta_1=\delta_2=\delta$			0.034 (0.474)	0.016 (0.233)			0.001 (0.018)	0.013 (0.322)
Prob> $\chi^2$	-	0.1532	0.2357	0.2209	-	0.4744	0.4413	0.3263

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_i$  parameters represent the estimated rates of return on schooling. The  $\gamma_i$  parameters represent the estimated value to the earning of a bachelor's degree. The  $\lambda_i$  parameters represent the extent of bias in OLS estimates of the rate of return on education due to the omission of family background factors from the wage equation. The  $\delta_i$  parameters indicate the extent of bias in OLS estimates of the value to the earning of a bachelor's degree due to the omission of family background factors from the wage equation.  $i, i=1,2$ , refers to elder and younger siblings, respectively.

**Table 8: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.20) for Brother-Pair Sample**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

	Survey 1990				Survey 1995			
	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$
$\beta_1$	0.052** (4.176)		0.051** (4.163)		0.040 (1.933)		0.028* (2.167)	
$\beta_2$	0.036** (2.536)		0.037* (2.633)		0.034 (1.377)		0.017 (1.346)	
$\beta_1=\beta_2=\beta$		0.050** (4.086)		0.049** (4.052)		0.049** (3.369)		0.031** (2.855)
$\gamma_1$	0.021 (0.230)		0.036 (0.468)		0.103 (1.716)		0.108 (1.923)	
$\gamma_2$	0.107 (1.167)		0.092 (1.100)		0.093 (1.093)		0.151** (2.462)	
$\gamma_1=\gamma_2=\gamma$		0.065 (1.047)		0.064 (1.036)		0.085 (1.653)		0.118** (2.457)
$\lambda_1$	-0.013 (-1.344)	-0.013 (-1.344)			-0.0003 (0.003)	-0.00009 (0.013)		
$\lambda_2$	-0.016 (-1.354)	-0.024* (-2.116)			-0.021 (-0.846)	-0.035* (-2.161)		
$\lambda_1=\lambda_2=\lambda$			-0.014* (-2.040)	-0.017** (-2.542)			-0.003 (-0.435)	-0.008 (-1.265)
$\delta_1$	0.156 (1.850)	0.132* (2.002)			0.013 (0.307)	0.016 (0.366)		
$\delta_2$	0.131** (2.397)	0.155** (3.027)			0.112 (1.319)	0.123* (2.178)		
$\delta_1=\delta_2=\delta$			0.134** (3.061)	0.143** (3.436)			0.028 (0.736)	0.042 (1.199)
Prob> $\chi^2$	-	0.2650	0.9589	0.6480	-	0.6896	0.7116	0.5696

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_i$  parameters represent the estimated rates of return on schooling. The  $\gamma_i$  parameters represent the estimated value to the earning of a bachelor's degree. The  $\lambda_i$  parameters represent the extent of bias in OLS estimates of the rate of return on education due to the omission of family background factors from the wage equation. The  $\delta_i$  parameters indicate the extent of bias in OLS estimates of the value to the earning of a bachelor's degree due to the omission of family background factors from the wage equation.  $i, i=1,2$ , refers to elder and younger siblings, respectively.

**Table 9: The Minimum Distance Estimation for Rate of Return on Education and the Impact from Family Factors (With Measurement Error 0.20) for Sister-Pair Sample**

**Data: Household Survey of Human Resources in Taiwan in 1990 and 1995**

	Survey 1990				Survey 1995			
	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$	Unrestricted	$\beta_1=\beta_2$ $\gamma_1=\gamma_2$	$\lambda_1=\lambda_2$ $\delta_1=\delta_2$	$\beta_1=\beta_2, \lambda_1=\lambda_2$ $\gamma_1=\gamma_2, \delta_1=\delta_2$
$\beta_1$	0.046** (2.682)		0.050** (2.980)		0.065** (2.370)		0.059* (2.170)	
$\beta_2$	0.060** (2.696)		0.064** (2.945)		0.098** (3.154)		0.080** (2.874)	
$\beta_1=\beta_2=\beta$		0.042** (2.472)		0.045** (2.692)		0.079** (2.975)		0.068** (2.647)
$\gamma_1$	0.135 (1.287)		0.178 (1.812)		0.150 (1.867)		0.113 (1.477)	
$\gamma_2$	0.163 (1.743)		0.158 (1.704)		0.061 (0.720)		0.093 (1.163)	
$\gamma_1=\gamma_2=\gamma$		0.177* (2.251)		0.186** (2.408)		0.083 (1.340)		0.071 (1.153)
$\lambda_1$	0.032 (1.537)	0.035 (1.681)			-0.005 (-0.262)	-0.009 (-0.549)		
$\lambda_2$	-0.034 (-1.525)	-0.026 (-1.215)			-0.007 (-0.315)	0.002 (0.115)		
$\lambda_1=\lambda_2=\lambda$			-0.003 (-0.259)	0.003 (0.351)			0.001 (0.053)	0.002 (0.152)
$\delta_1$	0.081 (0.685)	0.047 (0.436)			-0.048 (-0.771)	-0.026 (-0.471)		
$\delta_2$	-0.002 (-0.024)	-0.016 (-0.161)			0.067 (0.939)	0.056 (0.876)		
$\delta_1=\delta_2=\delta$			0.035 (0.492)	0.014 (0.203)			0.001 (0.033)	0.012 (0.309)
Prob> $\chi^2$	-	0.1330	0.2524	0.1921	-	0.3412	0.4389	0.2640

Note: Figures in parentheses are t statistics. Age and age squared are controlled in the models.

\*The coefficient is significant at the 5% level.

\*\*The coefficient is significant at the 1% level.

The  $\beta_i$  parameters represent the estimated rates of return on schooling. The  $\gamma_i$  parameters represent the estimated value to the earning of a bachelor's degree. The  $\lambda_i$  parameters represent the extent of bias in OLS estimates of the rate of return on education due to the omission of family background factors from the wage equation. The  $\delta_i$  parameters indicate the extent of bias in OLS estimates of the value to the earning of a bachelor's degree due to the omission of family background factors from the wage equation.  $i, i=1,2$ , refers to elder and younger siblings, respectively.

## **Appendix A: Documentation for Editing the Iranian Panel Data Set**

Empirical Work in the first two chapters in this dissertation is based on raw data in Iran. Below I described the process by which raw data is transformed into working files.

### **Data set:**

---

A panel data set is derived from the Social and Economic Characteristics of Households in Iran (SECHI) survey. These survey data were collected from 1987 until 1989 for continuous 12 rounds. This data set contains detailed questions from households regarding individual-level data such as age, education year, gender, working hours, wage rate, and enrollment status for every individual in the household.

The following work is working on the first round data files:

ind01.dat: reports every individuals' economic and demographic data.

hhb01.dat: reports every household's expenditure data.

hha01.dat: reports every household's asset data and mortality data in the household.

### **Dictionary Files:**

---

The above three data sets are ASCII files so we use STATA dictionary file to read the data and convert them to STATA data sets (dta file). Four dictionary files are used here:

demic.dct: Read [*round, ru prov branch hhold, record, indiv, relation, sex, residence present, residence war migrant, birth month, birth year, age, education variables, married, polygamy, afm, ceb, cebin, cebout, ceb12m, ceb12f*] from ind01.dat. Save these variables as inddem.dta STATA file.

econdic.dct: Read [*round, ru prov, branch, hhold, record, indiv, relation, sex, average number of hours worked per day, average number of days worked per week, average net income last week, average net income last month, average net income last 12 months, average number of hours worked per day for other jobs, average number of days worked per week for other jobs, average net income last week for other jobs, average net income last month for other jobs, average net income past 12 months for other jobs, unearned income (month), unearned income (3 months), unearned income (12 months), transfer income last month from government, transfer income last month from other, transfer income last 3 months from government, transfer income last 3 months from others, transfer income last 12 months from government, transfer income last 12 months from others*] from ind01.dat. Save these variables as indecon.dta file.

expdic.dct: Read [*round, ru, prov, branch, hhold, record, code1, exp1, code2, exp2, code3, exp3, code4, exp4, code5, exp5, code6, exp6, code7, exp7, code8, exp8*] from hhb01.dat. Save these variables as hhexp.dta file.

astdic.dct: Read [*round, ru prov, branch, hhold, record, car, tv, etc. the assets owned in the household; and deathyr, deaths1, deathdy1, deathmn1, dage1, deaths2, deathdy2, deathmn2, dage2, deaths3, deathdy3, deathmn3, dage3, house (1-8), water (1-2), fuel (c1,c2,h1,h2,ot1,ot2)*] from hha01.dat. Save these variables as hhast.dta file.

## Do Files, First Step:

Since the further 4 dta sets are in the string format, not in the numeric format. The first step of the following do files is to convert the string variables to numerical format so that dashes are read as missing variables. And then the data set is organized by doing mathematical operation.

### inddem.do:

- converting the demographic inddem dataset into a numerical format.
- generating education variables from educ1, educ2, and educ3. and recode educyr so that it equals number of years of schooling.
  - Note: Recode educyr missing as 0 if the respondent's age is greater than 5 years old. And hence, if the respondent's age is below or equal to 5, the educyr will remain missing.
- delete 4 bad observations.
- replace the original inddem.dta file.

Cleandem.do: correct the observations which have inconsistent age or birth year. Notice that the year of survey is in Iranian calendar 1366, and birth year variable is reported as two digits. For example, if a person is born in Iranian calendar 1355 year, he will report birth year as 55. We need to replace birth year by adding 1300 if birth years is smaller than 67 and we add 1200 to birth year if it is larger or equal to 67. This work is under an assumption that no one lives longer than 100 years. We calculate a new age variable from this new birth year by subtracting birth year from 1366. Compare the new age variable and reported age variable. There are several inconsistent observations about age and new age. We recode the age or the birth year to a conjecture right number. The following 155 observations are changed:

- newage - age >20: 27 observations.
- 11 < newage - age <=20 6 observations
- newage - age = 11 23 observations
- newage - age =10 6 observations
- 4 < newage - age <= 9 9 observations
- 2 < newage - age <=4 13 observations
- newage - age <-60 6 observations
- -60 <= newage - age < -40 15 observations
- -40 <= newage - age < -20 8 observations
- -20 <= newage - age < -10 8 observations
- -10 <= newage - age < -6 21 observations
- -6 <= newage - age < -2 13 observations
- replace the original inddem.dta file.

### Indecon.do:

- First, convert the income and employment data so that dashes are read in missing.
- Second, consolidate the income and employment data. (Note: the following variables are individual variables, not household variables.)
  - $unearned\ income = (month)*12 + (3\ months)*4 + (12\ months).$
  - $transfer\ income = from\ government [(last\ month) + (last\ 3\ months) + (last\ 12\ months)] + from\ other [transfer\ income\ (last\ month) + (last\ 3\ months) + (last\ 12\ months)].$
  - $annual\ net\ income = (last\ week)*48 + (last\ month)*11 + (last\ 12\ months)$

- $annual\ net\ income\ for\ other\ jobs = (last\ week)*48 + (last\ month)*11 + (last\ 12\ months)$
- $hours\ worked\ per\ year = average\ number\ of\ hours\ worked\ per\ day*average\ number\ of\ days\ worked\ per\ week$ 
  - $hours\ worked\ per\ year\ for\ other\ jobs = average\ number\ of\ hours\ worked\ per\ day\ for\ other\ jobs*average\ number\ of\ days\ worked\ per\ week\ for\ other\ jobs.$
- $wages = annual\ net\ income / hours\ worked\ per\ year$ 
  - $wages\ for\ other\ jobs = annual\ net\ income\ for\ other\ jobs / hours\ worked\ per\ year\ for\ other\ jobs.$
- $total\ annual\ income = unearned\ income + transfer\ income + annual\ net\ income + annual\ net\ income\ from\ other\ jobs.$
- $annual\ earned\ income = annual\ net\ income + annual\ net\ income\ from\ other\ jobs.$
- replace the original indecon.dta file.

#### hhast.do:

- this do file recode all of the 1\2 data to 1\0 data from the hhast.dta file.
- replace the original hhast.dta file.

#### hhexp.do:

- this do file generates the items expenditure and total expenditure for every household. Since every household might report up to 4 observations of 8 expenditure variables. Each item of the 8 expenditure represents a specific expense, receptively. So we need to add up each expenditure across multi-observations for the same household. We use “*reshape*” STATA command to rearrange the multi-observations of the same household to one observation. And sum the 8 kinds of expenditure, respectively. Most importantly, we generate total household expenditure from these item expenditures.
- replace the original hhexp.dta file.

### **Do files, Second Step:**

This final step consists of 3 major do files: hhold.do, wife.do, and child.do. Three major STATA data files will be generated: hhold.dta, wife.dta, child.dta. Plus, there will be two small do files in this step: one is clean.do which clean the bad-expenditure or may-be-wrong income-expenditure data. Another small do file is move.do which generates moving-status variables and merges them into wife.dta and child.dta.

#### hhold.do:

- this file constructs household variables such as size and income from individual data sets inddem.dta, indecon.dta, hhexp.dta, and hhast.dta.
- from inddem.dta :
  - calculate the number of individuals present in the survey for every household and keep one record for each household containing only an address list [round ru prov branch hhold] and hhsz, which means household size. we save it as hhdem.dta.
  - wives report their children ever born who is currently living in the household and away from the household in the inddem.dta. We add these two variables and label it as “wife’s cebin+cebout”. Two variables “boys even born for the past 12 months” and “girls ever born for the past 12 months” are also reported here. We save these three variables into hhdem.dta.

- Note that relation==2 represents the spouse of the househead. But some women are household heads. So we recognize wives as those who are female and are either household heads or household head's spouse.
- from indecon.dta:
  - we generate household total income from indecon.dta, save as hhinc.dta. The command here we use is still “reshape”. But actually there is a simpler way to sum up observations with same household id, which is “egen newvarname = sum (varname), by (household id)”
  - generate husband's income and earned income variables and merge them into hhinc.dta.
    - Note that we recognize husbands as those who are male and are either household heads or household head's spouse.
  - generate wife's income and earned income variables and merge them into hhinc.dta.
- merge hhdem.dta and hhinc.dta to a new file: hhold.dta.
- from hhast.dta:
  - We only use tv variable and merge it into hhold.dta. If researchers want to add new variables from hhast.dta. They could simply add them here.
- from hhexp.dta:
  - We use total household expenditure from hhexp.dta and merge it into hhold.dta.

clean.do : this do file cleans the bad expenditure observations and unreasonable income-expenditure data from hhold.dta We recode either the expenditure or income variables as missing by the following rules:

- replace household total expenditure as missing
  - if individual expenditure is less than one hundred dollars annually.
  - if household total expenditure is ten millions dollars more than household total income.
  - if household total expenditure is less than one thousand dollars yet the household income is greater than 100,000 dollars.
- replace husband income, wife income, husband earned income, wife earned income, and household total income as missing if household income is reported as zero.
- save a new file as hholdc.dta.

wife.do: this do file creates a data set of wives.

- Use women democratic data from inddem.dta and save it to wife.dta. Variables include wife's education, wife's literacy, ceb, cebin, cebout, children mortality (which is generated by the difference between “children ever born (ceb)” and “children ever born living in the household (cebin) + living away from the household (cebout)”), afm, age, duration of marriage (which is generated by differencing age of the first marriage (afm) and age). Save these variables to a new data set wife.dta.
- Merge husband democratic data from inddem.dta into wife.dta. Variables include husband's education year, husband's literacy, husband's polygamy status.
- Merge wife working hours and wage data from indecon.dta to wife.dta. Three variables are added into wife.dta. The variables are wife's wage (wwage1), wife's wage-other (wwage2), and wife's working hours per year (whr), which is the sum of working hours and working hours-other.
- Merge husband working hours and wage data from indecon.dta to wife.dta. Save as above.

- Merge income (household, husband, wife) and expenditure (household) from hholdc.dta to wife.dta. Here we generate the following variables:
  - $lhinc = \log(hhinc)$ .
  - $lhinc = \log(hinc)$ .
  - $lwinc = \log(winc)$
  - $wincsh = winc/hhinc$
  - $lwhinc = \log(winc+hinc)$
  - $lwearn = \log(wearn)$
  - $lhearn = \log(hearn)$

Note that the  $\log(var)=0$  if  $var=0$ . We keep  $hhinc$ ,  $whinc$ ,  $hinc$ ,  $winc$ ,  $lhinc$ ,  $lwhinc$ ,  $lhinc$ ,  $lwinc$ ,  $lhearn$ ,  $lwearn$ ,  $wincsh$ ,  $totexp$ ,  $hhsz$ ,  $size$ ,  $ceb12m$ ,  $ceb12f$  and merge them into wife.dta
- Recode  $ceb12m$  “2 =1”, Recode  $ceb12f$  “2=1”.
- Recode ru 1/2 as 1/0.
- Generate household id as one variable by combining the household id list :  $id = hhold + branch * 1000 + prov * 100000 + ru * 10000000$
- Drop observations which husband has at least two wives.

child.do: This do file creates data set child.dta from inddem.dta, wife.dta, indecon.dta.

- Generate a child base data by keeping all of the reporters whose relation is “3” to househead. We keep individual code (indiv), sex, age, educyr, literate, student. Save it as child.dta.
- Merge household data from wife.dta into child.dta.
  - Note that the ru code is still 2/1 for child.dta for this moment. So we need to recode ru code in wife.dta back to 2/1. Another point is to rename age as mage (mother’s age) in wife.dta because child.dta has age which is referred to children’s age. This recode and rename is to avoid confusion.
- Generate child income from indecon.dta into child.dta. The variables we generate are as below:
  - $linc = \log(\text{children's income})$
  - $learn = \log(\text{children's earned income})$
  - $lcinc = \log(\text{total children's income in the household})$
  - $llearn = \log(\text{total children's earned income in the household})$
- Generate birth order variable (parity).
- Recode ru 2/1 to 0/1 (so, rural is 0 and urban is 1)
- Recode sex 2/1 to 0/1 (so girl is 0 and boy is 1)
- Recode student 2/1 to 0/1 (so student is 1 and nonstudent is 0)
- Save all of these work in child.dta

move.do: This do file generates data from ind01.dat which contains household head's birth place and previous residence and duration time for the current address. The procedure is as below:

- Write a inddem1.dct and read the above variables. And convert the string variables to numeric variables. Recode ru 2/1 to 0/1 and then keep only household head’s record. Save it as temp.dta.
- The temp file will be merged into wife.dta and child.dta.

**VITAE**  
**Huei-Ling Chen**

**Education**

Ph.D in Economics, Virginia Polytechnic Institute and State University, May 1998  
M.A. in Economics, National Taiwan University, Taipei, Taiwan, June 1993  
B.A. in Agricultural Economics, National Taiwan University, Taipei, Taiwan, June 1991

**Fields of Specialization**

Household Economics, Econometrics, Labor Economics

**Work Experience**

**Teaching Assistant:**

1994-1998; Virginia Polytechnic Institute and State University, Blacksburg, VA; Advanced Econometrics, Qualitative Methods, Principles of Economics.  
1992-1993; Department of Economics, National Taiwan University, Taipei, Taiwan; Qualitative Methods.

**Research Assistant:**

1993; Institute of Economics, Academia Sinica, Taipei, Taiwan; Asymmetric Information Economics.  
1994; Department of Agricultural Economics, National Taiwan University, Taipei, Taiwan;  
Environmental Economics

**Teaching Interests**

**Undergraduate:** Applied Statistics & Econometrics, Applied Microeconomics, Labor Economics;  
**Graduate:** Labor Economics, Econometrics.

**Awards and Honors**

Liu, Ta-Chung Scholarship, Department of Economics, National Taiwan University, 1992.

---

Huei-Ling Chen