Unobservable Family Effects

and the Apparent External Benefits of Education

John Gibson

Department of Economics

University of Waikato

Private Bag 3105

Hamilton

NEW ZEALAND

Fax: 64-7 838-4031

Email: jkgibson@waikato.ac.nz

Running head: “Family effects and external benefits”

Abstract

Estimates of the returns to education usually ignore private and external non-marketed benefits, which, if counted, may double the social rate of return (Haveman and Wolfe, 1984). There are formidable difficulties in measuring these non-market benefits, so before trying it is worth checking whether the claimed benefits are in fact caused by education. This paper examines the effect of education on volunteer work, which is listed by Wolfe (1994) as a non-marketed output benefiting both private individuals and the general public. The positive correlation between education and the probability of volunteering seems to support this claim, but previous studies have ignored the role of family unobservables. This study uses data from a sample of twins to test the relationship between education and volunteering, holding unobservable family effects constant. The results show that education significantly reduces the probability of volunteering and the supply of volunteer hours, so volunteering may not be an external benefit of education.

JEL: D64, I20

Keywords: human capital, economic impact

I. Introduction

Estimates of the returns to education usually ignore private and external non-marketed benefits, which, if counted, may double the social rate of return (Haveman and Wolfe, 1984). There are formidable difficulties in measuring these non-market benefits, so before trying it is worth checking whether the claimed benefits are in fact caused by education. This paper examines the effect of education on volunteer work, which is listed by Wolfe (1994) as a non-marketed output benefiting both private individuals and the general public. The positive correlation between education and the probability of doing volunteer work (de Combray, 1987; Hayghe, 1991; van Dijk and Boin, 1993; Vaillancourt, 1994; Freeman, 1997) and between education and the amount of volunteer time given (van Dijk and Boin, 1993; Freeman, 1997) seems to support the notion that more education will mean more volunteering. But an alternative explanation for this correlation is that family unobservables determine both volunteering (Janoski, 1995) and schooling (Morgan, Alwin and Griffin, 1979). Thus, providing more resources for education may have no effect, and may even reduce volunteering due to the higher opportunity cost of time for highly educated people.

This paper uses data from a sample of twins to test the relationship between education and volunteering. Data on twins allow unobservable family effects to be held constant so that the structural effect of education on the probability of volunteering and on the supply of volunteer hours can be identified. For this purpose, there is no requirement that twins be representative of the general population – a requirement which is unlikely to hold. Instead, the point in using data on twins is to control for unobservable family effects in a way that is not possible with usual samples and the relevant comparison is not between the sample of twins and the population but between the results for twins with family effects controlled for versus the results for twins without the controls. While twins data have been used to answer questions about the role of unobservable family effects in wage equations (Ashenfelter and Rouse, 1998), they have not previously been used in the study of voluntary work.

Although voluntary work is only one of many claimed non-market benefits of education, evidence that it is not caused by education may reduce confidence in some of the other claimed benefits, such as reduced criminal activity and improved health. Moreover, voluntary work is an important outcome in its own right, and is receiving increased attention from economists. For example, Freeman (1997) reports that volunteering in the U.S. augments paid work hours by between 3-7%. Similarly, volunteer labor is equivalent to about 7% of the number of full-time equivalent paid jobs in the Dutch economy (van Dijk and Boin, 1993), and about 4% in the Canadian economy (Vaillancourt, 1994). Estimates of the value of volunteer time are just as impressive: in the U.S., volunteer time given in 1985 was worth $110 billion (de Combray, 1987) and time given in 1991 was worth roughly $116 billion (Freeman, 1997).

II. The Model

A. Conceptual Framework

A frequently used model in the twins literatue is the "within-twins" or fixed effects estimator (Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998). In this model, the difference in the outcome of interest (e.g., wage rates, volunteer hours) between the two twins in a pair is related to differences in the educational attainment and other observable characteristics of each twin. This fixed effects model provides an estimate of the impact of education on the outcome variable which is not biased by the omission of family background variables (Miller, Mulvey and Martin, 1995).

For example, let h_{1i} and h_{2i} represent the hours of volunteer work by the first and second twins in the ith pair. These hours are assumed to depend on an unobservable component that varies by family, m_i, observable components, such as age and sex, that vary by family but not across twins, $\mathbf{X}_i$, observable components, such as schooling and marital status, that may vary across the twins in a pair, $\mathbf{Z}_{1i}$ and $\mathbf{Z}_{2i}$, and unobservable individual components, e_{1i} and e_{2i}. This implies:

$$h_{1i} = \alpha\mathbf{X}_i + \beta\mathbf{Z}_{1i} + m_i + e_{1i} \tag{1}$$

and

$$h_{2i} = \alpha\mathbf{X}_i + \beta\mathbf{Z}_{2i} + m_i + e_{2i} \tag{2}$$

where we assume that the equations are identical for the two twins, who have been randomly numbered as either #1 or #2. In other words, applying an equation to one-half of a sample should give the same estimates as applying it to the other half, if the splitting has been random.

The difficulty in estimating equations (1) and (2), or the 'stacked' version if the two half-samples are combined, is that there are no data available on the family unobservables, m_i. If these family unobservables are correlated with schooling levels, and also affect the decision to volunteer, biased estimates of the component of $\boldsymbol{\beta}$ related to schooling will result, due to the omission of a relevant variable. One way to solve this problem is to look at the *difference* between the two twins in a pair because the family unobservables, being common to the two twins, then drop out:

$$h_{1i} - h_{2i} = b\left(\mathbf{Z}_{1i} - \mathbf{Z}_{2i}\right) + e_{1i} - e_{2i} \ . \qquad (3)$$

Thus, the within-pair difference in hours of volunteer work (or the probability of volunteering) is related to differences in observable components that vary across the twins in a pair. For example, if education increases volunteering, the more highly schooled individual in a pair of twins should supply more hours of volunteer work (or be more likely to volunteer) than their sibling, and the component of $\boldsymbol{\beta}$ that relates to schooling in equation (3) should be positive.

Although the fixed effects estimator provides a simple and intuitive framework for estimation, it does not allow us to directly estimate the correlation between the unobservable family effects and schooling. It is this correlation which may cause usual estimates of volunteer work equations to suffer omitted variables bias. The “selection effects” model, introduced by Ashenfelter and Krueger (1994), does allow explicit consideration of the correlation with unobservable family effects in the volunteering equation. If the correlation between the unobservable family effects m_i and the observables is:

$$m_i = \boldsymbol{\gamma}\mathbf{Z}_{1i} + \boldsymbol{\gamma}\mathbf{Z}_{2i} + \boldsymbol{\delta}\mathbf{X}_i + w_i \tag{4}$$

where the common $\boldsymbol{\gamma}$ shows that the correlation is the same with each twin's observables, and ω_i is an uncorrelated, random error, then substituting (4) into (1) and (2) gives reduced form volunteer hours equations in terms of just observables and individual random errors:

$$h_{1i} = [\boldsymbol{\alpha}+\boldsymbol{\delta}]\mathbf{X}_i + [\boldsymbol{\beta}+\boldsymbol{\gamma}]\mathbf{Z}_{1i} + \boldsymbol{\gamma}\mathbf{Z}_{2i} + e'_{1i} \tag{5}$$

and

$$h_{2i} = [\boldsymbol{\alpha}+\boldsymbol{\delta}]\mathbf{X}_i + \boldsymbol{\gamma}\mathbf{Z}_{1i} + [\boldsymbol{\beta}+\boldsymbol{\gamma}]\mathbf{Z}_{2i} + e'_{2i} \tag{6}$$

where $e'_{1i} = w_i + e_{1i}$ and $e'_{2i} = w_i + e_{2i}$.

In this framework, the coefficient on the co-twin’s observable characteristics, $\boldsymbol{\gamma}$ provides an estimate of the correlation between schooling and unobservable family effects (Miller, Mulvey and Martin, 1995). Intuitively, the volunteering of one twin, say #1, depends on the observable characteristics of the other twin, $\mathbf{Z}_{2i}$ because those co-twin characteristics act as an indicator of the unobserved family (and genetic) background that is common to the two twins. Equations (5) and (6) also show that the coefficients on own-observable effects (e.g., the coefficient on $\mathbf{Z}_{1i}$ for twin #1) are comprised of two parts [β+γ], where β captures the structural effect and γ captures the effect due to family unobservables. Ordinarily, these two effects cannot be untangled, so the estimated coefficient will be a biased estimate of β unless γ=0. For example, if families with an otherwise high probability of volunteering are also more likely to educate their children, the component of γ for schooling will be positive and the estimated coefficient on own-education, which equals [β+γ], will overstate the structural effect of education. But with the estimate of γ from the coefficient on the co-twin's observables, the structural effect can be retrieved as [β+γ]-γ. In other words, the coefficient on the co-twin's educational level provides an estimate of the impact of family effects, which can be subtracted from the coefficient on the own-education variable to derive an estimate of the pure impact of schooling on volunteering (Miller, Mulvey and Martin, 1995). Note that this derived pure effect should be similar to the fixed effects estimate of the impact of education, obtained from equation (3).

B. Specification of Variables

A recent model of volunteering included the following explanatory variables: schooling, age, gender, marital status, employment status, family income, urban location, and number of adults and children in the household (Freeman, 1997).[1] This list of demographic, schooling and income variables is similar to other models of volunteering, so these variables, where available in the data, are used for the specification of the current model. Of these variables, the observable characteristics that vary across families but not across siblings are age, gender, and race.[2] The characteristics that may vary across the twins in a pair are years of schooling, employment status, marital status, family income, and the number of adults and children in the household.[3] In the equation (5) and (6) framework it would be possible to include each of these sibling characteristics in the participation equation of the co-twin. But following the example of Ashenfelter and Krueger (1994), only the sibling's education level is included, where the coefficient on this variable is a measure of γ, the correlation between family unobservables and education.

III. Data

The data are from a survey of adult twins, carried out in New Zealand in 1994. The survey covered 253 individuals, but this analysis concentrates on the 85 sets of identical and same-sex fraternal twins where both siblings had completed their schooling. Compared with the population, the sample is younger, is disproportionately female, has a higher employment rate, and appears to be more highly educated (Table 1).[4] However, the sample has the same relationship between education and volunteering as in the population: The participation rate in volunteer work in the population of people with tertiary qualification is 45.2%, while it is only 36.4% for those without tertiary qualifications. Similarly, the participation rate in volunteer work for those sampled twins with tertiary qualifications is 43.7%, while for the less-qualified twins it is only 26.9%.[5] Hence, there is no reason to believe that the sample is unfavourable to the hypothesis that education raises the probability of volunteering.

(Table 1 about here)

There are also some characteristics of the sample that cannot be compared with the population. The first of these is that self-reported years of education averaged 13.4 years. Second, only 36% of the sample had the same number of years of schooling as their sibling, and the correlation between years of schooling of siblings was 0.66. This imperfect correlation is a useful feature of the data because without this within-pair variation, the fixed effects estimator would not work.

Twins reported their own and their sibling's schooling, and the correlation between the report made by one person on their own school years and the report on their school years made by their twin is 0.91. The fact that the correlation between the two reports on the same variable is not 1.0 indicates that self-reported years of schooling may have slight measurement error. Apparently, individuals who report their own schooling with error are also more likely to report their sibling's schooling with some error.[6] This correlation between a potential instrument (the sibling report) and the explanatory variable may make IV estimation inconsistent (Ashenfelter and Krueger, 1994), and also seems to rule out the method suggested by Iwata (1992) for correcting attenuation bias in probit models. However, as a check on the robustness of the findings, the model is re-estimated using averages of the two reports of schooling for each twin as an explanatory variable because averaging ameliorates the effect of measurement error.

IV. Estimation Methods and Results

The first models estimated are of the decision to volunteer. The probit estimator is used for the selection effects model (equations (5) and (6)) because the dependent variable is the 0-1 measure of whether the person volunteered in the week preceding the survey or not. Probit analysis has previously been used in the study of volunteering by Vaillancourt (1994), but an improvement here is to use a heteroscedastically-robust covariance estimator that takes account of the intra-pair correlation in the disturbances.[7] Marginal effects are also reported to show the effect of independent variables on the probability of volunteering. The fixed effects model (equation (3)) is based on the within-twins difference in the 0-1 variable indicating volunteering. There are three possible outcomes: the dependent variable equals one if the twin randomly numbered as #1 within a set volunteered in the week of the survey while their sibling did not; it equals zero if the volunteer status of both twins was the same; and it equals minus one if twin #1 did not volunteer while their sibling did. Given this ordering, the model is estimated with an ordered probit, and also with OLS, which corresponding to the linear probability model and is comparable to the marginal effects calculated from the probit estimation of the selection effects model.

(Table 2 about here)

The first two columns of Table 2 present estimates of the fixed effects model, which use self-reported schooling and so ignore the slight measurement errors. Both OLS and ordered probit coefficients suggest that the more highly educated twin within a pair is less likely to volunteer than is their less educated sibling. According to the OLS estimates, an extra year of education for one twin over their sibling reduces their probability of volunteering by 4.8 percentage points (a reduction of about one-eighth at the mean probability). None of the other variables are statistically significant, although there is weak evidence that if one twin is employed while the other is not, the employed twin is less likely to volunteer than is the twin who is not working. When averages of the two reports on each person's schooling are used, to alleviate measurement error (the last two columns of Table 2), there is a slight rise in the size of the schooling coefficients and a slight fall in their statistical significance but there is nothing to suggest that measurement error is a cause of the negative effect of education on volunteering.

The fixed effects estimates give a result that is at variance with most of the literature, which suggests that education raises the probability of volunteering. It is easy to produce the 'standard' positive effect with the current sample, just by estimating a model that does not control for family unobservables. This model is reported in the first column of Table 3, where equations (1) and (2) are 'stacked' (so the cross-equation restrictions hold). Each additional year of schooling appears to raise the probability of volunteering by 4.4 percentage points (significant at $p<0.01$) when family unobservables are ignored in the estimates in the first column of Table 3.[8] The only other statistically significant variables in the first column suggest that the probability of volunteering is higher for males and for whites.

(Table 3 about here)

The second model in Table 3 stacks equations (5) and (6) and shows that unobserved family effects are a source of the correlation between schooling and volunteering, because the coefficient on sibling's schooling, which gives an estimate of γ, is positive and statistically significant ($p<0.01$). In other words, families whose unobservable characteristics cause them to have a high likelihood of volunteering are also more likely to educate their children, so the relationship between schooling and volunteering is just a correlation caused by an excluded common cause. Moreover, once unobservable family effects are controlled for, the coefficient on own years of education becomes quantitatively small and statistically insignificant ($p<0.57$). The structural effect of schooling on volunteering, which is given by the coefficient on own years of education minus the coefficient on sibling's years of education, suggests that each additional year of education reduces the probability of volunteering by 4.8 percentage points (significant at the $p<0.06$ level). This probability impact is unchanged if averages of the two reports on each person's school years are used in the model.[9] The implied probability impact of –0.048 is the same as the linear probability estimated by OLS for the fixed effects model, while in terms of the probit coefficients, the implied structural effect of schooling in the selection effects model is –0.13 while in the fixed effects model it is –0.12.

The estimates that include family effects in Table 3 may suffer from multicollinearity problems, because of the correlation between the schooling levels of twins.[10] To see if this was the case, variance inflation factors (VIF) were calculated for each of the independent variables, where:

$$\mathrm{VIF}\left(x_j\right) = \frac{1}{1 - R_j^2}$$

where R_j^2 is the square of the multiple correlation coefficient that results when x_j is regressed against all the other explanatory variables. Chatterjee and Price (1991) suggest that there is evidence for multicollinearity if the largest VIF is greater than 10 but the highest VIF amongst the explanatory variables in Table 3 was only 1.95, for own school years (1.81 for sibling's school years). The results of a more comprehensive detection approach, suggested by Belsey et al. (1980) also indicated no grounds for concern about multicollinearity problems (see Appendix 1).

When the selection effects model is re-estimated on the sub-sample of identical twins the results are largely the same (Table 4). When family unobservables are ignored, each additional year of schooling appears to raise the probability of volunteering by 5.7 percentage points. But once sibling's education is added to the model, the coefficient on own schooling becomes small and statistically insignificant ($p<0.27$). Subtracting the coefficient on sibling's school years from that on own school years suggests that the pure effect of each additional year of education is to lower the probability of volunteering by 4.6 percentage points, which is almost the same as the result when fraternal twins were included in the sample.

(Table 4 about here)

Table 5 reports estimates of both selection effects and fixed effects volunteer hours equations. The estimates in column (i) are of a 'standard' model with no controls for family unobservables and they suggest that hours volunteered are greater for whiles and less for males, while there is a positive but statistically insignificant effect of own-education. Once the sibling's schooling level is included, as a control for unobservable family effects, the implied structural effect of education is negative and statistically significant – each additional year of schooling appears to reduce volunteering by one hour per week. The estimates from the fixed effects model, in column (iii), are very similar – the within-pair difference in school years is negatively related to the difference in hours of volunteering, while none of the other differences in observable characteristics between the two twins matter.

(Table 5 about here)

V. Conclusions

The positive correlation between education and volunteering in a model that ignores unobservable family effects is reversed once these unobservable family effects are controlled for. Usual estimates of volunteer participation and hours equations apparently suffer omitted variables bias because they do not include controls for family unobservables, which are positively correlated with both schooling and volunteering. The current study controls for family unobservables by using data on identical twins. This method raises the question of why the twins in a pair vary – perhaps they are not so identical after all? However, almost one-half of the variation in education levels comes after the completion of high school, when each individual may be subject to idiosyncratic shocks because they do not live in the same household, and so it need not invalidate the approach used here.

The results reported here suggest that increased volunteer work does not belong on the list of external benefits of education. In fact, additional schooling appears to lower the probability of volunteering and the average hours of volunteer work supplied. This finding also raises doubts about inferences from other correlations – such as between education and reduced crime – which may also just reflect the omitted effects of family unobservables.

References

Ashenfelter, O., Krueger, A., 1994. Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* 84, 1157-1173.

Ashenfelter, O., Rouse, C. 1998. Income, schooling, and ability: evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113(1): 253-284.

Belsey, D., Kuh, E., Welsch, R., 1980. *Regression Diagnostics*. John Wiley and Sons, New York.

Chatterjee, S., Price, B., 1991. *Regression Analysis by Example*. John Wiley and Sons, New York.

de Combray, N., 1987. Volunteering in America. *American Demograhics* 9, 50-52.

Fomby, T., Hill, C., Johnson, S., 1984. *Advanced Econometric Methods*. Springer-Verlag, Needham, MA.

Freeman, R.B., 1997. Working for nothing: the supply of volunteer labor. *Journal of Labor Economics*, S140-S166.

Gibson, J., 1998. *Ethnicity and Schooling in New Zealand: An Economic Analysis Using a Survey of Twins*. Institute of Policy Studies, Wellington.

Haveman, R., Wolfe, B., 1984. Schooling and economic well-being: the role of nonmarket effects. *Journal of Human Resources* 19, 377-407.

Hayghe, H.V., 1991. Volunteers in the United States: who donates the time? *Monthly Labor Review* 114, 17-23.

Iwata, S., 1992. Errors-in-variables regression using estimated latent variables. *Econometric Reviews* 11, 195-200.

Janoski, T., 1995. Pathways to voluntarism: family socialization and status transmission models. *Social Forces* 74, 271-292.

Judge, G., Hill, C., Griffiths, W., Lutkepohl, H., Lee, T., 1982. *Introduction to the Theory and Practice of Econometrics*. John Wiley and Sons, New York.

Miller, P., Mulvey, C., Martin, N. 1995. What do twins studies reveal about the economic returns to education? A comparison of Australian and U.S. findings. *American Economic Review* 85(3): 586-589.

Morgan, W., Alwin, D., Griffin, L., 1979. Social origins, parental values, and the transmission of inequality.” *American Journal of Sociology* 85, 156-166.

Vaillancourt, F., 1994. To volunteer or not: Canada, 1987. *Canadian Journal of Economics* 27, 813-826.

Van Dijk, J., Boin, R., 1993. Volunteer labor supply in the Netherlands. *De Economist* 141, 402-418.

Wolfe, B., 1994. External benefits of education. In Husen, T., Postlethwaite, N. (Ed.) *The International Encyclopaedia of Education*, Pergamon, Oxford, pp. 2208-2212.

Table 1: Descriptive Statistics for the Sample

Variable	Means (standard deviations in parentheses) Twins[a]	Population[b]
Participation rate in volunteer work	0.37 (0.48)	0.41[c]
Self-reported years of education	13.39 (2.51)	…
Own-school years = sibling's-years	0.36 (0.48)	…
Without school qualifications (=1)	0.23 (0.42)	0.37
Male (=1)	0.24 (0.43)	0.48
White (=1)	0.81 (0.39)	0.84
Age	38.87 (14.23)	44
Married (=1)	0.55 (0.50)	0.51
Employment rate	0.74 (0.44)	0.66
ln (annual household income)	10.75 (0.60)	10.68
No. of adults in household	2.33 (1.04)	2.15
No. of children in household	0.89 (1.26)	0.74
Sample size	170	…

[a] Source: Postal survey of the schooling and labour market experience of twins, Nov 1994 – Feb 1995.

[b] Source: 1996 Census of Population. Means are based on bracketed data for age 18 and over.

[c] Based on a question about volunteering in the previous four weeks, whereas the survey of twins used a question about volunteering in the previous week. In the 1991 Census, where the one week reference period was used, the participation rate in volunteer work was 0.19.

Table 2: Fixed Effects Estimates of the Effect of Education on the Probability of Volunteering

	Ignoring Errors in Self-Reported Schooling		Using Averages of Schooling Reports	
	OLS coefficients	Ordered Probits[a]	OLS coefficients	Ordered Probits[a]
ΔSchool years[b] $S_1^1 - S_2^2$	-0.048 (2.01)	-0.120 (2.23)	...	...
ΔSchool years[b] $[(S_1^1 + S_1^2)/2] - [(S_2^2 + S_2^1)/2]$	...	...	-0.050 (1.78)	-0.125 (1.97)
ΔMarried	-0.010 (0.09)	-0.016 (0.07)	-0.009 (0.08)	-0.014 (0.06)
ΔEmployed	-0.237 (1.59)	-0.475 (1.45)	-0.247 (1.67)	-0.499 (1.53)
Δln (annual household income)	0.021 (0.22)	0.076 (0.37)	0.021 (0.22)	0.077 (0.37)
ΔNo. of adults in household	0.013 (0.24)	0.034 (0.27)	0.018 (0.32)	0.046 (0.37)
ΔNo. of children in household	-0.008 (0.17)	-0.006 (0.05)	-0.008 (0.16)	-0.005 (0.04)
Cut-points		-1.225 0.914		-1.229 0.908
R^2	0.087	0.056	0.086	0.055

Notes: Numbers in ( ) are *t*-statistics, calculated from heteroscedastically-robust standard errors. Dependent variable = 1 if the twin randomly numbered as # 1 within a set volunteered in the week of the survey while their sibling, numbered #2, did not, = 0 if the volunteer status of both twins was the same, and =-1 if twin #1 did not volunteer while their sibling did. The explanatory variables are the intra-pair differences in the variables listed in the first column. *N*=85.

[a] The ordered probit coefficients show the effect of the independent variable on a linear index, while the cut-points give the value of the linear index (plus random error) needed to allocate a person into one of the three volunteer status groups.

[b] The education level of the *n*th twin as reported by the *m*th twin is denoted S_n^m, *m*,*n*=1,2.

Table 3: Probit Selection Effects Estimates of the Effect of Education on the Probability of Volunteering

	Omitting Family Effects		Including Family Effects	
	Probit coefficients	Probability impact [a]	Probit coefficients	Probability impact [a]
Own school Years	0.118 (2.71)	0.044	0.025 (0.57)	0.009
Sibling's school Years	...	...	0.156 (3.76)	0.058
Male (=1)	-0.634 (2.21)	-0.220	-0.698 (2.39)	-0.234
White (=1)	0.873 (2.29)	0.276	0.843 (2.10)	0.269
Age	0.014 (1.40)	0.005	0.016 (1.55)	0.006
Married (=1)	-0.200 (0.70)	-0.074	-0.181 (0.69)	-0.067
Employed (=1)	-0.402 (1.24)	-0.153	-0.378 (1.18)	-0.144
ln (annual household income)	0.010 (0.05)	0.004	-0.044 (0.20)	-0.016
No. of adults in household	0.044 (0.34)	0.016	0.070 (0.51)	0.026
No. of children in household	0.154 (1.59)	0.057	0.146 (1.47)	0.054
Intercept	- 3.008 (1.38)	...	-3.389 (1.56)	...
Implied β (structural effect of schooling)	...	...	-0.130 (1.88)	-0.048
Pseudo-R^2	0.117		0.152	
LR (slopes=0) test	$c^{2}_{(9)} = 24.34$		$c^{2}_{(10)} = 32.77$	

Notes: Numbers in ( ) are *t*-statistics, calculated from robust standard errors that are corrected for the within-pair correlations in disturbances. Dependent variable is a binary variable taking a value of 1 if the person did volunteer work in the week preceding the survey (*n*=63) and 0 if they did not (*n*=107).

[a] The probability impact is the change in the probability of volunteering, given a unit change in the independent variable (evaluated at the mean for continuous variables and as a switch from 0 to 1 for dummy variables).

Table 4: Probit Estimates of the Effect of Education on the Probability of Volunteering (Identical Twins)

	Omitting Family Effects		Including Family Effects	
	Probit coefficients	Probability impact [a]	Probit coefficients	Probability impact [a]
Own school Years	0.144 (2.98)	0.055	0.048 (1.01)	0.018
Sibling's school years	...	...	0.169 (3.80)	0.065
Male (=1)	-0.661 (2.19)	-0.234	-0.743 (2.40)	-0.260
White (=1)	0.762 (1.91)	0.258	0.706 (1.66)	0.242
Age	0.010 (0.99)	0.004	0.013 (1.19)	0.005
Married (=1)	-0.219 (0.73)	-0.084	-0.174 (0.57)	-0.067
Employed (=1)	-0.538 (1.55)	-0.209	-0.497 (1.43)	-0.194
ln (annual household income)	0.010 (0.05)	0.004	-0.050 (0.22)	-0.019
No. of adults in household	0.039 (0.25)	0.015	0.076 (0.46)	0.029
No. of children in household	0.147 (1.44)	0.056	0.140 (1.31)	0.054
Intercept	- 2.913 (1.29)	...	-3.403 (1.49)	...
Implied β (structural effect of schooling)	...	...	-0.121 (1.66)	-0.047
Pseudo-R^2	0.113		0.156	
LR (slopes=0) test	$c^2_{(9)} = 22.33$		$c^2_{(10)} = 30.01$	

Notes: Numbers in ( ) are *t*-statistics, calculated from robust standard errors that are corrected for the within-pair correlations in disturbances. Dependent variable is a binary variable taking a value of 1 if the person did volunteer work in the week preceding the survey (*n*=60) and 0 if they did not (*n*=90).

[a] The probability impact is the change in the probability of volunteering, given a unit change in the independent variable (evaluated at the mean for continuous variables and as a switch from 0 to 1 for dummy variables).

Table 5: Effects of Schooling on Volunteer Hours, Controlling for Family Unobservables With Selection Effects and Fixed Effects Models

	OLS (i)	OLS (ii)	Fixed Effects[a] (iii)
Own school years	0.215	-0.249	-0.964
	(1.19)	(1.08)	(1.81)
Sibling's school years		0.732	
		(2.14)	
Male (=1)	-1.927	-2.108	
	(2.23)	(2.29)	
White (=1)	1.674	1.454	
	(2.56)	(2.10)	
Age	0.060	0.065	
	(1.40)	(1.47)	
Married (=1)	-0.883	-0.722	-0.814
	(0.75)	(0.63)	(0.42)
Employed (=1)	-0.616	-0.471	-0.170
	(0.63)	(0.50)	(0.14)
ln (annual household income)	0.919	0.708	-0.339
	(1.27)	(0.97)	(0.35)
No. of adults in household	0.058	0.153	0.301
	(0.11)	(0.31)	(0.43)
No. of children in household	0.128	0.107	-0.202
	(0.63)	(0.50)	(0.50)
Intercept	-12.924	-14.573	
	(1.84)	(1.94)	
Implied β (structural effect of schooling)		-0.982	
		(1.86)	
R^2	0.078	0.140	0.081
Sample size	170	170	85

Notes: Numbers in ( ) are *t*-statistics, calculated from heteroscedastically-robust standard errors that are corrected for the within-pair correlations in the disturbances. The dependent variable in columns (i) and (ii) is the number of hours of volunteer work per week.

[a] The dependent variable is the difference in weekly hours of volunteer work between the twin randomly numbered #1 and the twin randomly numbered #2 in a set, while the explanatory variables are the intra-pair differences in the variable listed in the first column.

Acknowledgements

My thanks to three anonymous referees, the New Zealand Econometric Study Group, and especially Rainer Winkelmann, for helpful comments, and to Joy Brown, John Kirkland, the New Zealand Multiple Birth Association, and Massey University for providing access to the Multiple Birth Register used to gather the data. The financial support of the New Zealand Treasury and the University of Waikato is gratefully acknowledged.

Appendix 1

Detecting Collinearity Within the Data Matrix

Traditional methods for detecting collinearity often cannot detect relationships amongst more than two explanatory variables and are unable to ascertain which coefficient estimates are degraded (Judge, et al., 1982). Procedures suggested by Belsey, et al. (1980) are designed to overcome these weaknesses. For the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, if $\mathbf{A}$ is the matrix whose columns are the eigenvectors of $\mathbf{X'X}$ then $\mathbf{A'X'XA} = \boldsymbol{\Lambda}$ where Λ is a diagonal matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ along the diagonal. It can be shown (Fomby, et al., 1984, p. 286) that $(\mathbf{X'X})^{-1} = \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A'} = \sum_{i=1}^{k} \lambda_i^{-1}\mathbf{a}_i\mathbf{a}_i'$ where $\mathbf{a}_i$ is the eigenvector associated with eigenvalue λ_i and $\mathbf{A}=(\mathbf{a}_1,\mathbf{a}_2,..,\mathbf{a}_k)$. Thus, the covariance matrix of the least-squares estimator is $\mathrm{cov}(\mathbf{b}) = \sigma^2(\mathbf{X'X})^{-1} = \sigma^2\sum_{i=1}^{k} \lambda_i^{-1}\mathbf{a}_i\mathbf{a}_i'$ and the variance of an individual coefficient estimate b_j is:

$$\mathrm{var}(b_j) = \sigma^2\left(\frac{a_{j1}^2}{\lambda_1} + \frac{a_{j2}^2}{\lambda_2} + \cdots + \frac{a_{jk}^2}{\lambda_k}\right)$$

where a_{ji}^2 are the squares of the elements of $\mathbf{A}$ and σ^2 is the error variance from the regression of which b_j is a part. Hence, three factors contribute to imprecise coefficient estimates: a large error variance and small eigenvalues combined with large squared elements of **A** (which are variance weights). This observation forms the basis of the Belsey et al. detection statistic. If the **X** matrix is scaled to unit length, $\mathrm{tr}(\mathbf{X'X}) = k = \sum \lambda_i$ and collinearity problems may be occur when the ‘condition index’ $\sqrt{l_1/l_i}$ (where λ_1 is the largest eigenvalue) exceeds 30. In addition, for coefficient estimates to be degraded, two or more coefficients must have more than half of their variance associated with a ‘small’ eigenvalue (i.e., one where the condition index exceeds 30). According to Judge et al. (1982, p.623), the proportion of the variance of each coefficient associated with a particular eigenvalue can be calculated (from the unscaled **X** matrix) as $\frac{a_{ji}^2}{l_i} \Big/ \sum_{i=1}^{k} \frac{a_{ji}^2}{l_i}$.

Appendix Table 1: Proportions of the Variance of Each Explanatory Variable Associated with Each Condition Index

Eigenvalue (l_i)	Condition Index $\sqrt{(l_1/l_i)}$	Own school years	Sibling's school years	Male	White	Age	Married	Employed	ln (annual household income)	No. adults in household	No. children in household
7.5238	1.00	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.00	0.00
0.8133	3.04	0.15	0.16	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
0.7006	3.28	0.67	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02
0.4024	4.32	0.07	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.77
0.2093	6.00	0.03	0.01	0.00	0.00	0.09	0.00	0.00	0.01	0.77	0.02
0.1741	6.57	0.06	0.01	0.01	0.03	0.00	0.03	0.06	0.34	0.07	0.00
0.0952	8.89	0.01	0.00	0.16	0.00	0.17	0.42	0.10	0.00	0.00	0.07
0.0599	11.23	0.00	0.00	0.75	0.04	0.05	0.13	0.07	0.06	0.00	0.03
0.0112	25.94	0.00	0.00	0.08	0.80	0.02	0.00	0.05	0.21	0.14	0.05
0.0106	26.64	0.00	0.01	0.00	0.13	0.00	0.41	0.72	0.38	0.00	0.03

Note:

Belsey et al. (1980) suggest that parameter estimates are degraded when the variances of two or more coefficients have at least one half of their magnitude associated with a condition index exceeding 30.

Notes

[1] Freeman also uses another specification with hourly earnings, but that restricts the estimating sample to those who are employed which results in a loss of over one-quarter of the observations in the current data set.

[2] For the identical twins and same-sex fraternal twins used here, age and gender do not vary by individual, so can be considered family characteristics.

[3] In Freeman's model, the number of household earners was used rather than the number of adults, and a dummy for large cities was included. Neither of these variables were available in the current data.

[4] There are no Census or national sample estimates of average years of schooling in New Zealand so this comparison is in terms of secondary school qualifications. For a full description of the data, see Gibson (1998).

[5] This difference is statistically significant at the $p<0.03$ level.

[6] Comparing covariances of the reports of the intra-pair schooling difference with the theoretical moment matrix in Table 10 of Ashenfelter and Krueger (1994) suggests that the measurement errors have a correlation of 0.54.

[7] This was implemented with the "cluster" option in *Stata* v. 6.0.

[8] This estimate comes from a sample where males and females are pooled, which differs from the split-sample specification used by Freeman (1997). However, when separate slopes are allowed for males and females the effect of education on volunteering does not differ between the two groups ($p<0.26$), so in light of the small size of the sample, males and females are pooled and the effect of gender enters the model just as an intercept dummy variable.

[9] Using averages of the two reports, the probit coefficient on own school years is 0.028 and on sibling's school years is 0.157.

[10] I am grateful to an anonymous referee for raising this point.