Correlation versus Causation and the Apparent External Benefits of Education

John Gibson[†]
University of Waikato

Abstract

There is a widely reported positive correlation between education and the amount of volunteer time given. Many researchers assume this is a causal relationship, so increased volunteer work has been added to the list of 'external benefits' of education. These 'external benefits' make public subsidies to higher education seem more justifiable. This study uses data from a specially collected sample of New Zealand identical twins to test the relationship between education and volunteer labour supply, holding unobservable family effects constant. Multiple measurements of education levels were also collected so that any bias in the results due to measurement errors could be corrected. The results show that once family unobservables are controlled for, extra education significantly reduces the amount of volunteer time donated. This reversal in the results is a reminder of the lesson that correlation does not imply causation.

JEL: D64, I20, J22
Keywords: Education, Labour supply, Measurement error, Twins, Volunteer work

[†] Department of Economics, University of Waikato, Private Bag 3105, Hamilton, New Zealand. Fax: (64-7) 838-4331. E-mail: jkgibson@waikato.ac.nz.

I. Introduction

One of the first lessons taught in most econometrics and statistics classes is that correlation does not imply causation. As Kendall and Stuart (1961, p. 279) explain: "[a] statistical relationship, however strong and however suggestive, can never establish causal connection: our ideas of causation must come from outside statistics, ultimately from some theory or other." Nevertheless, correlation is often confused with causation. For example, in debates about the funding of tertiary education in New Zealand it is claimed that: "higher educational attainment … [brings] … social benefits in terms of better public health, lower crime, the environment, parenting, political and community participation, and social cohesion" (Edlin, 1998, p. 10). Although some vaguely defined theories underlie these links, such as the 'socialisation' of young people who remain at school reducing their risk of criminality, the claimed effects of education seem to be mainly based on an inspection of empirical correlations.

In this study I examine the link between education and community participation – specifically, the supply of volunteer labour. There is a widely reported correlation between education and the amount of volunteer time given (Bowen, 1977; Dye, 1980; van Dijk and Boin, 1993; Freeman, 1997). This correlation also holds in New Zealand, with 2.1 hours per week of volunteer work supplied by people with post-secondary school qualifications but only 1.6 hours supplied by those with lower levels of education (Statistics NZ, 1998).[1] Many researchers assume that this correlation is a causal relationship, so increased volunteer work has been added to the list of 'external benefits' of education (Wolfe, 1994).[2] These external benefits can double the estimated social rate of return to investment in education (Haveman and Wolfe, 1984), making public subsidisation of tertiary education seem more acceptable.[3] Hence, to avoid wasted public investment, it is important to

establish whether increased volunteer work (and other external benefits) is actually *caused* by education, rather than just reflecting some correlation.

Previous studies of the link between schooling and volunteer work ignore the possibility that unobservable family effects (e.g., a 'work ethic') cause both a higher level of educational attainment and a higher supply of volunteer time. The higher volunteering rate by the children of parents who also volunteer suggests that family unobservables are important determinants of volunteer activity (Smith and Baldwin, 1974; Segal, 1993). The most likely explanation for these intra-family correlations in volunteer activity is family socialisation through example and value modelling, because the correlations persist even after controlling for the transmission of socioeconomic status from parents to children (Janoski, 1995). It is also well-known that family unobservables are important determinants of educational attainment (Mayer, 1997), most likely because a child's early years – when their basic social, academic and language skills are developed – are almost completely monopolised by family. Families, and especially parents, imbue their children with certain values, such as conformity to rules, diligence, motivation, and the need for initiative, all of which contribute to the subsequent educational achievements of the child (Morgan, Alwin and Griffin, 1979).

If family unobservables create the correlation between schooling and volunteer work, providing individuals with extra schooling may not lead to any extra volunteer work being done. Indeed, extra schooling could even reduce volunteer labour supply, if the higher opportunity cost of time for more highly educated people outweighs any socialisation in favour of a volunteer ethos. Thus, volunteer work may not be an external benefit of education, possibly weakening the case for public subsidies by removing a frequently-used example of an external benefit. Moreover, evidence that the link between schooling and volunteer work is just a correlation may reduce our confidence in some of the other claimed external benefits of education, such as reduced criminal activity.

This study uses data from a specially collected sample of identical twins to test the relationship between schooling and volunteer work. The advantage of using data on twins is that unobservable family effects can be held constant so that the structural effect of schooling on volunteer labour supply can be identified. A feature of the sample is that independent reports of schooling were made by siblings, so any bias in the results due to measurement error in self-reported schooling can be corrected. Although twins data have been used to answer questions about the role of unobservable genetic and family effects in wage equations, they have not previously been used in the study of volunteer labour supply.[4]

The next section of the paper outlines the conceptual framework, which allows the relationship between schooling and volunteer labour supply to be observed without the confounding effect of family unobservables. Section III discusses the survey of twins that provides the data used in the analysis, paying particular attention to the measurement error properties of these data. Estimation results and related analyses are reported in Section IV, while the final section discusses the broader implications of the findings.

II. Conceptual Framework

The conceptual framework views an individual's supply of volunteer hours as a function of observable components, varying by family (**X**) and by individual (**Z**), and unobservable components, varying by family (m) and by individual (e). The observable family characteristics that are likely to be relevant, according to previous empirical models of volunteer labour supply, are the age and gender of the individual (Patterson,1982; van Dijk and Boin, 1993).[5] Ethnicity may also be a relevant observable characteristic because Census estimates show ethnic differences in volunteer participation rates (Statistics NZ, 1998a). In addition to

schooling, the main observable characteristic varying by individuals that is likely to affect volunteer hours is employment status (Freeman, 1997).

The value of using data on twins is that these data can help in dealing with the family unobservables, m in ways that are not possible with ordinary samples. Consider a model of the hours of volunteer labour supply of the twins in the ith set, h_{1i} and h_{2i}:

$$h_{1i} = a\,\mathbf{X}_i + b\,\mathbf{Z}_i + m_i + e_{1i} \tag{1}$$

and

$$h_{2i} = a\,\mathbf{X}_i + b\,\mathbf{Z}_i + m_i + e_{2i} \tag{2}$$

where the parameters are assumed to be identical for the two twins, who are randomly numbered as either 1 or 2. There are two ways of dealing with the unobservable family effects: either explicitly model them as a function of the observables, or else treat them as fixed effects and eliminate them by differencing. Neither of these treatments are available with an ordinary cross-sectional sample.

When the unobservable family effects are explicitly modelled, the correlation between m_i and the observables is represented by:

$$m_i = g\mathbf{Z}_{1i} + g\mathbf{Z}_{2i} + d\mathbf{X}_i + w_i \tag{3}$$

where the common g shows that the correlation is the same with each twin's observables, and ω_i is an uncorrelated, random error. Substituting (3) into (1) and (2) gives reduced form volunteer labour supply equations in terms of just observables and individual random errors:

$$h_{1i} = [a + d]\mathbf{X}_i + [b + g]\mathbf{Z}_{1i} + g\mathbf{Z}_{2i} + e'_{1i} \tag{4}$$

and

$$h_{2i} = [a + d]\mathbf{X}_i + g\mathbf{Z}_{1i} + [b + g]\mathbf{Z}_{2i} + e'_{2i} \tag{5}$$

where $e'_{1i} = w_i + e_{1i}$ and $e'_{2i} = w_i + e_{2i}$. The influence of unobservable family effects on an individual's volunteer labour supply operates via the coefficient vectors $d$and g In an ordinary sample it would be impossible to untangle the structural effect of schooling, b from the apparent effect of schooling that is really caused by unobservable family effects, g because the estimated coefficient on schooling is the sum of the structural effect and the family unobservable effect, [b+ g. But with data on twins it is possible to subtract the coefficient on sibling's schooling from the coefficient on own schooling, i.e., [b+ g- g in order to identify the structural effect of schooling. Intuitively, the volunteer labour supply of one twin depends on the schooling level of the other twin because that schooling level is an indicator of the unobserved family (and genetic) background that is common to the two twins.

When the unobservable family effects are treated as fixed effects and eliminated by subtracting equation (2) from equation (1), the resulting estimator is:

$$h_{1i} - h_{2i} = b\left(\mathbf{Z}_{1i} - \mathbf{Z}_{2i}\right) + e_{1i} - e_{2i} \,. \qquad (6)$$

Equation (6) gives a simple but powerful framework for measuring the effect of education on volunteer labour supply, without any confounding effect of family unobservables. However, care must be taken in using equation (6) because the increase in the noise-to-signal ratio caused by differencing the data makes attenuation bias (i.e., bias toward zero) a potentially serious problem (Ashenfelter and Krueger, 1994).

If extra education causes extra volunteer labour to be supplied, the coefficient b on the schooling variable will be positive in all of the above equations.[6] But if the positive correlation between education and volunteer labour supply is just due to family unobservables raising both of these variables, the coefficient b will be positive in equation (1) and (2) but will be less

positive or even negative in equations (4), (5) and (6). For example, if families who would otherwise have a high volunteer labour supply are also more likely to educate their children, the coefficient g on sibling's education will be positive. Thus, when the family unobservables component measured by g is removed, the structural effect of schooling identified from equations (4) and (5) will be less than the reduced form effect of schooling in equations (1) and (2). Hence, this framework shows that usual estimates of a volunteer labour supply equation (or of a wage equation) suffer omitted variables bias because they do not include controls for family unobservables (Ashenfelter and Krueger, 1994).

III. Data Collection and Description

A nationwide postal survey of adult twins was carried out in November and December 1994. The twins contacted were either listed on the Multiple Birth Register of the NZ Multiple Birth Association, or had replied to newspaper and radio publicity seeking volunteers. To increase the participation of some groups who were under-represented in the Multiple Birth Register (Maori and Pacific Island people, and prime age white males), university students were hired to identify eligible twins, gain their consent for participating, and deliver and retrieve the questionnaires. However, there were no personal interviews conducted. This extra surveying took place in December and January and was limited to the Auckland and Wellington metropolitan areas. The sample size was 253 individuals, comprised of identical twins, fraternal twins, and the singletons resulting when only one twin out of a pair completed the survey. This analysis concentrates on the 75 sets of identical twins where both had completed their schooling.

Table 1 contains descriptive statistics for the sample and comparisons, where available, with the population. The sample is younger and better qualified than the population, women are

over-represented, and there is a higher labour force participation rate. However, the ethnic balance of the sample is similar to the population. A feature of the questionnaire was the attention paid to years of schooling, with twins reporting their own and their sibling's schooling.[7] Own-reported years of education are slightly higher than sibling-reported years. For over one-third of the sample, self-reported years do not equal sibling reports.

The average member of the sample supplied slightly more hours of volunteer labour than is usual amongst the population, while the participation rates in volunteer activities were similar in both the sample and the population.[8] The relationship between education levels and volunteer hours that is apparent in the population is also apparent in the sample. The average volunteer labour supply of those twins without post-school qualifications was 1.6 hours per week (the same as the Census estimate), while for the more highly educated members of the sample it was 3.1 hours per week (this difference is statistically significant, t=1.91). Volunteer hours by sample members with post-school qualifications exceed the population average of similarly qualified people (2.1 hours), so there is no reason to believe that the sample is unfavourable to the hypothesis that education raises volunteer labour supply.

Errors in Measures of Schooling

Correlations between own-reported and sibling-reported education levels provide a way of estimating the extent of measurement error in these data. One twin was randomly selected as the first in each pair, with the education level of the nth twin as reported by the mth twin denoted by S_n^m, m,n=1,2. In the classical model of measurement error $S_n^m = S_n + v_n^m$ where S_n is the true schooling level and v_n^m (m=1,2) are measurement errors that are uncorrelated with S_n (n=1,2) and with each other. Under these assumptions, the correlation between the two

reports of schooling, S_n^1 and S_n^2 equals the fraction of the variance in the reported measures of schooling that is due to the true variance:

$$\text{Var}(S_n)/[\text{Var}(S_n^1) \cdot \text{Var}(S_n^2)]^{1/2},$$

which is sometimes called the "reliability ratio".

Table 2A contains the correlations among self-reported and sibling-reported schooling levels. The estimates of the reliability ratio of years of schooling are 0.91 and 0.92. Hence, about nine percent of total variation in years of schooling is measurement error variance. A similar reliability ratio, of 0.91, is reported by Behrman, Rosenzweig and Taubman (1996) for a sample of twins born in Minnesota, while reliability ratios are 0.93 and 0.91 (depending on whether twin 1 or twin 2 is the reference) in the sample of twins used by Ashenfelter and Rouse (1998). The estimated reliability ratio suggests that the OLS coefficient in a bivariate regression of volunteer hours on years of schooling would be attenuated to only 91-92 percent of its true value. This reliability ratio also suggests that the coefficient in a fixed effects estimator like equation (6) could be attentuated to only 75 percent of its true value, given the correlation of 0.66 between S_1^1 and S_2^2 – the schooling level of each twin in a pair.[9]

One objection to estimating measurement error from the correlation between own- and sibling-reported years of schooling is that discrepancies may just reflect the less complete information that a sibling has. However, the presence of measurement error in own-reports can be tested for using estimated covariances.[10] For example, if there is no error in the own-report made by the first twin in a pair, then $S_1^1 = S_1$. The sibling-report is $S_1^2 = S_1 + v_1^2$, so the covariance between the two reports is $\text{cov}[(S_1 + v_1^2), S_1^1] = \text{var}(S_1) = \text{var}(S_1^1)$. Hence, an implication of there being no measurement error in own-reports is that the covariance between own- and

sibling-reports should equal the variance of own-reports. The covariances reported in panel B of Table 2 suggest that the own-reports of schooling do contain measurement error. The variance of S_1^1 (7.361) is about 13 percent higher than the covariance between S_1^2 and S_1^1 (6.509) and the variance of S_2^2 is about nine percent higher than $\text{cov}[S_2^1, S_2^2]$.

It is also possible to test the assumption that the measurement errors are uncorrelated. The usual instrumental variables estimator is inconsistent if there are correlated measurement errors (Ashenfelter and Krueger, 1994). A test for correlated errors is to see whether the covariance between self-reports (ΔS) and sibling-reports (ΔS) of the intra-pair schooling difference equals the covariance between the reports of the difference made by the first twin (ΔS^*) and those made by the second twin (ΔS^{**}). Comparing the estimated covariances in Table 3C with the theoretical moment matrix in Table 10 of Ashenfelter and Krueger (1994), it appears that the measurement errors v_1^1 and v_2^1 (or v_1^2 and v_2^2) have a correlation of 0.61.

In light of these results about measurement errors, each model will be estimated in three different ways: using methods that assume no measurement error, using methods that assume classical (i.e., uncorrelated) measurement error, and using methods that are valid even with correlated measurement error. The method used under correlated measurement errors takes averages of the two reports of schooling for each twin, $(S_1^1 + S_1^2)/2$ and $(S_2^2 + S_2^1)/2$. An advantage of averaging is that any fixed tendency for a twin to misreport schooling levels is eliminated because the two schooling reports made by that individual are placed on either side of the subtraction when the fixed effects estimator is defined.

IV. Estimation Methods and Results

There are three models to be estimated. The first model consists of equations (1) and (2) and allows no way of identifying the structural effect of schooling on volunteer labour supply. This model replicates the results that would be estimated from usual samples without controls for family unobservables. The second model consists of equations (4) and (5) and allows the structural effect of schooling to be identified by removing the family unobservables component that is measured by the coefficient g on sibling's schooling. The third model (fixed effects) consists of equation (6) and also allows the effect of schooling on volunteer labour supply to be observed without any confounding influence of family unobservables.

Estimation Methods

Two issues affect the choice of estimation method: the need for cross-equation restrictions on the coefficients in equations (1) and (2) (or (4) and (5)), and the problem of zero hours of volunteer work reported by 60 percent of the sample. One way to ensure that the cross-equation restrictions are met is to 'stack' the equations and estimate them by OLS. However, OLS will not be optimal because it ignores the information contained in the cross-equation correlations of the errors. The Seemingly Unrelated Regression (SUR) method does take account of these correlations, so it is the method used here. When there is uncorrelated measurement error, the unbiased alternative to SUR is three-stage least squares (3SLS).

The presence of zeros in dependent variables is often dealt with by using the Tobit model because of bias in the OLS estimator (Greene, 1981). However, whether the Tobit model is appropriate depends on what process causes the zeros, and on what the econometrician is interested in. If the zeros are caused only by infrequency of volunteering, and provided there is no relationship between volunteer frequency and the explanatory variables, OLS is consistent

while Tobit is not (Case and Deaton, 1998).[11] If the interest is in the unconditional expected value of the dependent variable, which takes into account the probability of being above the limit and the expected value conditional upon being above the limit, the Tobit coefficients should be multiplied by the fraction of non-limit observations (McDonald and Moffitt, 1980). A close approximation to the adjusted Tobit coefficients is given by the OLS coefficients estimated on the full sample of limit and non-limit observations,[12] so there is no need to use the Tobit model here and OLS-type estimators are used instead.[13]

Model 1: No Controls for Family Unobservables

Table 3 reports estimates of the effect of schooling on volunteer labour supply, without any controls for family unobservables. Each additional year of education appears to raise an individual's supply of volunteer labour by 0.3 hours per week, and this effect is statistically significant. Education appears to have a slightly more powerful effect in the column (ii) estimates, which correct for (uncorrelated) measurement error bias, although the use of the estimator that is robust to the presence of correlated measurement errors (column iii) lowers the coefficient on years of education to its initial OLS value. The results in Table 3 also suggest that volunteer hours are lower for males, for non-Pakeha, for people who are working, and for younger people, but in contrast to the results for schooling, none of these effects are statistically significant.

The results in Table 3 seem to offer support to the view that tertiary education creates large external benefits. According to the coefficient on years of schooling, someone who completes a three-year university degree can be expected to supply almost 50 hours more volunteer labour per year than a person with similar characteristics who did not go to university. The opportunity cost of time (the wage rate) for university graduates in this sample is estimated as

almost $21/hour (Gibson, 1998).[14] Therefore, someone with a three-year degree can be expected to do extra volunteer work worth approximately $1000 per year, compared with someone without a degree. This is equivalent to about ten percent of the gross annual income difference estimated by Maani (1997) for an employed person (averaged over females and males) going from Bursary-level education to a Bachelors degree. Because increased volunteer work is only one of many external benefits of education listed by Haveman and Wolfe (1984), this calculation suggests that the total value of external benefits of education might equal or even exceed the value of private benefits.

Model 2: Sibling's Schooling Levels as a Control for Family Unobservables

The truth of the claim just made, that on average a university degree causes a person to supply 50 hours more volunteer labour per year, depends on whether unobservable family effects are the source of the positive correlation between education and volunteer labour supply. The results in Table 4, which correspond to equations (4) and (5), suggest that unobserved family effects are the source of this correlation. In these results, the coefficient on sibling's education level in each twin's volunteer labour supply equation, g is always positive and statistically significant. Once unobservable family effects are controlled for by including sibling's years of education in the model, the coefficient on own years of education becomes quantitatively small and statistically insignificant.

The structural effect of schooling on volunteer labour supply (b) is given by the coefficient on own years of education minus the coefficient on sibling's years of education ($\hat{g}$). This structural effect is negative according to all estimates reported in Table 4, and is statistically significant once account is taken of measurement error bias. Each additional year of education appears to reduce an individual's supply of volunteer labour by between 0.5 and 0.9 hours per

week, depending on whether measurement errors are treated as uncorrelated (column (ii)) or correlated (column (iii)). According to these results, people who have a high level of education and donate a lot of volunteer time would have donated even more time if they had lower education levels. The cause of their volunteer activity is not their level of education but their basic values and beliefs, which are here called "family unobservables".

Model 3: Family Unobservables as Fixed Effects Eliminated by Differencing

The results for the fixed effects model also confirm that additional years of education reduce an individual's supply of volunteer labour (Table 5). According to the OLS results in column (i), the more highly educated twin in a pair does 0.8 hours per week less volunteer work than their sibling, for each year of education they have more than their sibling. This difference is statistically significant, in contrast to the effect of intra-pair differences in employment status which is the other variable in the model. Because the two twins in a pair share the same family and genetic background, the result in Table 5 is as close as we can get to finding the causal effect of education on volunteer labour supply without any confounding effects of family unobservables.

The OLS estimate in Table 5 is attenuated by the effect of measurement error, and once this is corrected for with the instrumental variables (IV) estimator in column (ii) it appears that the more highly educated twin in a pair does 1.3 hours per week less volunteer work than their sibling, for each year of schooling that they have more than their sibling. However, the IV estimator of the fixed effects model is inconsistent when measurement errors are correlated, so the most robust estimate in Table 5 is in column (iii), where the intra-pair difference in years of schooling is based on averages of the reports made by each twin and their sibling. This estimate suggests that the more highly educated twin does one hour less volunteer work per week, for each year of education that they have more than their sibling.

The results in Table 4 and Table 5 defy conventional wisdom because they suggest that tertiary education (and other forms of schooling) create external costs, in terms of reduced volunteer labour supply. In other words, extra education makes people more selfish with their time rather than more generous as is commonly assumed. Someone who completes a three-year university degree can be expected to supply 150 hours less volunteer labour per year compared to a person of similar characteristics who did not go to university. Given an opportunity cost of time for university graduates of $21/hour (see above), this reduction in volunteer labour supply is worth approximately $3000 per year. Thus, rather than volunteer labour being a valid example of an external benefit of education, it is in fact an example of the opposite – apparently, extra education can impose some types of external costs on the community.

One important question about the results in Tables 4 and 5 is whether the negative relationship between schooling and volunteer labour supply is more apparent for people who have only recently completed their education. The liberalisation of the New Zealand economy and the pressure to recoup tuition fees may possibly have made recent graduates less willing to supply volunteer labour. The model in column (iii) of Table 5 was augmented with an interactive dummy variable to test this hypothesis. Specifically, the slope coefficient was allowed to vary between those twins who were above age 32 years (two-thirds of the sample) and the rest of the sample, where this cut-off point was chosen to ensure that people who had completed their education after the liberalisation beginning in 1984 were in the younger group.[15] The coefficient on this interaction term was statistically insignificant ($p<0.39$), so the negative effect of schooling on volunteer hours does not seem to be just a feature of people whose schooling has been in the more market oriented, fees-paying environment.

Whether highly educated people substitute money for time, by making more generous charitable contributions even as they reduce their volunteer labour supply, cannot be established because the survey of twins did not include any questions on charitable donations of money. However, previous studies usually find that time and money contributions are complements rather than substitutes (Menchik and Weisbrod, 1987; Freeman, 1997). Thus, it is not clear that the $3000 reduction in annual donations of time volunteered by a university degree holder (calculated above) is offset by higher donations of money, so the effect of extra education on overall volunteer activity can be expected to be negative.

V. Summary and Implications

The widely-reported positive correlation between education and volunteer hours is just that – a c*orrelation* rather than a causal relationship. Both the level of education attained and the amount of volunteer time given are caused by unobservable family effects. Once these unobservable family effects are controlled for, the revealed effect of extra education is to reduce the supply of volunteer labour. Usual estimates of volunteer labour supply equations suffer omitted variables bias because they do not include controls for family unobservables but the current study includes controls 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. Indeed, it appears that giving a person extra education

reduces their hours of volunteer work below what it would otherwise be, thereby offsetting the value of any genuine external benefits of education. Therefore, the case for public subsidies to higher levels of education may be weakened by these results. The reversal of the widely reported correlation between education and the amount of volunteer time also raises doubts about the interpretations placed on the correlations between education and other socially desirable actions. In particular, it is quite possible that the correlation between education and lower risk of criminal activity just reflects the effect of family unobservables influencing both education and criminality. The unobservable factors that make some people unable to resist the temptations of crime may be the same factors that make people unsuccessful in attaining high levels of education. Thus, extra education may not cause any reduction in the risk of criminal activity.

The findings of the current study are based on a small, somewhat unusual sample, so it would be worthwhile repeating the analysis on a larger sample. Whether this sample would need to be identical twins, rather than just siblings, depends on the assumption made about the unobservable family effects. If these omitted family effects are partially of genetic origin, and if there is no within-family variation in the other omitted environmental variables, then identical twins will identify models that cannot be identified just using siblings (Chamberlain, 1977). However, genetic effects are probably less important to volunteering than they are to wages (the so-called "ability bias") and there is little reason to assume *a priori* that there is no within-family variation in the omitted environmental variables. Thus, it would be appropriate to repeat the analysis with a sample of siblings to verify that family unobservables are the cause of the positive correlation between education and volunteer labour supply.

This paper has a more general message – beyond the study of education and volunteering – that the factors left out of econometric models because they are not easily observable are often the factors that matter to the behaviour that society seeks to influence. A similar conclusion is reached by Mayer (1997) who finds that the influence of parental incomes on children's outcomes (behavioural problems, risk of dropping out of school) is overstated in models that only control for observed characteristics. Unobserved attributes like enthusiasm and dependability, which influence parent's incomes, also influence their children's well-being. Consequently, giving extra money to parents may not help their children because it is the unobservable attributes of parents that matter most to children's outcomes. Studies such as these demonstrate the importance of separating causality from correlation if econometric results are to be a reliable guide for public policies.

Acknowledgements

My thanks to 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.

References

Ashenfelter, O. and Krueger, A. (1994), "Estimates of the economic return to schooling from a new sample of twins", *American Economic Review* 84(5): 1157-1173.

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

Behrman, J., Rosenzweig, M. and Taubman, P. (1996), College choice and wages: estimates using data on female twins", *Review of Economics and Statistics* 78(4): 672-685.

Blau, F., and Ferber, M. (1992), *The Economics of Women, Men, and Work* Englewood Cliffs, New Jersey: Prentice-Hall.

Bowen, H. (1977), *Investment in Learning: The Individual and Social Value of American Higher Education* Washington: Jossey-Bass.

Case, A. and Deaton, A. (1998), "Large cash transfers to the elderly in South Africa", *Economic Journal* 108(Sept): 1330-1361.

Chamberlain, G. (1977), "Are brothers as good as twins?", in P. Taubman (ed) *Kinometrics: Determinants of Socioeconomic Success Within and Between Families* New York: North-Holland, pp. 287-297.

Dye, R. (1980), "Contributions of volunteer time: some evidence on income tax effects", *National Tax Journal* 33(1): 89-93.

Edlin, B. (1998), "Dumbing down the nation for a $50m 'saving'", *The Independent* 29 July p.10.

Freeman, R.B. (1997), "Working for nothing: the supply of volunteer labor", *Journal of Labor Economics* 15(1): S140-S166.

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

Greene, W.H. (1981), "On the asymptotic bias of the Ordinary Least Squares estimator of the Tobit model", *Econometrica* 49(2): 505-513.

Haveman, R. and Wolfe, B. (1984) "Schooling and economic well-being: the role of nonmarket effects", *Journal of Human Resources* 19(3): 377-407.

Janoski, T. (1995), "Pathways to voluntarism: family socialization and status transmission models", *Social Forces* 74(1): 271-292.

Kendall, M.G. and Stuart, A. (1961), *The Advanced Theory of Statistics* (vol. 2), New York: Charles Griffin.

McDonald, J.F. and Moffitt, R.A. (1980), "The uses of Tobit analysis", *Review of Economics and Statistics* 62(3): 318-321.

Maani, S. (1997), *Investing in Minds: The Economics of Higher Education in New Zealand* Wellington: Institute of Policy Studies.

Mayer, S. (1997), *What Money Can't Buy: Family Income and Children's Life Chances* Cambridge: Harvard University Press.

Menchik, P. and Weisbrod, B. (1987), "Volunteer labor supply", *Journal of Public Economics* 32(2): 159-183.

Miller, P., Mulvey, C. and 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. and Griffin, L. (1979), "Social origins, parental values, and the transmission of inequality", *American Journal of Sociology* 85(1): 156-166.

Patterson, H. (1982), "Voluntary work in Australia", *Australian Bulletin of Labour* 8(2): 95-103.

Robertson, G. (1996), "Relative benefits of tertiary education", *New Zealand Education Review* 1(9):10-22.

Scobie, G.M. and Duncan, A. (1995), "Financing New Zealand's tertiary education: How much subsidy?", *Agenda* 2(2):211-222.

Segal, L.M. (1993), *Four Essays on the Supply of Volunteer Labor and Econometrics* Unpublished Ph.D dissertation, Northwestern University.

Smith, D., and Baldwin, B. (1974), "Parental socialisation, socioeconomic status, and volunteer organisation participation", *Journal of Voluntary Action Research* 3(3): 59-66.

Statistics NZ. (1998), "Highest Qualification Gained and Sex by Number of Hours of Unpaid Work in Different Household for Population Aged 15 Years and Over" *1996 Census of Population and Dwellings* (Unpublished table supplied by Statistics NZ, Wellington).

Statistics NZ. (1998a), *Unpaid Work: 1996 Census of Population and Dwellings* Wellington: Statistics New Zealand.

Van Dijk, J., and Boin, R. (1993), "Volunteer labor supply in the Netherlands", *De Economist* 141(3): 402-418.

Wolfe, B. (1994), "External benefits of education", in T. Husen and N. Postlethwaite (eds) *The International Encyclopaedia of Education* Oxford: Pergamon, pp. 2208-2212.

Table 1: Descriptive Statistics for the Sample

Variable	Means (standard deviations in parentheses)	
	Identical Twins[a]	Population[b]
Hours of volunteer work per person per week	2.5 (5.2)	1.9[c]
Participation rate in volunteer work	0.40 (0.49)	0.41[c]
Age	39.49 (14.58)	44
Male	0.31 (0.46)	0.48
Ethnicity (1=Pakeha, 0=other)	0.84 (0.37)	0.84
Labour force participation rate	0.74 (0.44)	0.66
No school qualifications	0.25 (0.43)	0.37
Self-reported years of education	13.35 (2.5)	…
Sibling-reported years of education	13.26 (2.5)	…
Self-reported equals sibling- reported years of education	0.61 (0.49)	…
Sample size	150	…

[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 hours of unpaid work outside the household in the previous four weeks.

Table 2: Correlation and Covariance Matrices For Schooling Variables

A. Correlations

		S_1^1	S_1^2	S_2^2	S_2^1
Sibling 1's own-reported schooling	(S_1^1)	1.000			
Sibling 1's sib-reported schooling	(S_1^2)	0.906	1.000		
Sibling 2's own-reported schooling	(S_2^2)	0.661	0.664	1.000	
Sibling 2's sib-reported schooling	(S_2^1)	0.680	0.583	0.924	1.000

B. Covariance of the Levels

		S_1^1	S_1^2	S_2^2	S_2^1
Sibling 1's own-reported schooling	(S_1^1)	7.361			
Sibling 1's sib-reported schooling	(S_1^2)	6.509	7.005		
Sibling 2's own-reported schooling	(S_2^2)	4.275	4.192	5.684	
Sibling 2's sib-reported schooling	(S_2^1)	4.374	3.662	5.223	5.626

C. Covariance of the Differences

		ΔS ¢	ΔS ¢¢	ΔS^*	ΔS^{**}
Self-reported difference $(S_1^1 - S_2^2)$	ΔS ¢	4.496			
Sibling-reported difference $(S_1^2 - S_2^1)$	ΔS ¢¢	3.167	5.306		
Difference reported by twin 1 $(S_1^1 - S_2^1)$	ΔS^*	3.936	4.099	4.240	
Difference reported by twin 2 $(S_1^2 - S_2^2)$	ΔS^{**}	3.726	4.374	3.795	4.305

Table 3: Effects of Schooling on Volunteer Hours
(No Controls for Family Unobservables)

	SUR (i)	3SLS[a] (ii)	SUR on average own-education[b] (iii)
Own years of education	0.318	0.354	0.297
	(2.08)	(2.02)	(1.79)
Male	-0.726	-0.805	-0.770
	(0.90)	(0.99)	(0.94)
Ethnicity	0.986	1.003	1.003
(1=Pakeha, 0=other)	(1.39)	(1.45)	(1.40)
Age	0.022	0.023	0.022
	(0.96)	(0.99)	(0.94)
Employment status	-1.161	-1.176	-1.069
(1=working, 0=otherwise)	(1.25)	(1.23)	(1.14)
Intercept	-2.896	-3.364	-2.673
	(1.54)	(1.62)	(1.34)
R^2	0.07	0.07	0.07

Notes: Numbers in ( ) are *t*-statistics, calculated from heteroscedastically-robust standard errors that also take account of cross-equation correlations in the errors. The dependent variable is the number of hours of volunteer work per week. The sample size is 150.

[a]Own-education is treated as endogenous and each sibling's report of the other sibling's education is added to the list of exogenous variables.

[b]Average own education is equal to $(S_i^i + S_i^j)/2$.

Table 4: Effects of Schooling on Volunteer Hours, Controlling for Family Unobservables

	SUR (i)	3SLS[a] (ii)	SUR on averages of schooling reports[b] (iii)
Own years of education	0.041	-0.231	-0.028
	(0.24)	(1.09)	(0.18)
Sibling's years of education	0.401	0.652	0.485
	(2.34)	(3.10)	(3.05)
Male	-1.197	-1.575	-1.234
	(1.47)	(1.79)	(1.52)
Ethnicity	0.911	1.015	0.924
(1=Pakeha, 0=other)	(1.26)	(1.42)	(1.27)
Age	0.024	0.027	0.026
	(1.02)	(1.07)	(1.11)
Employment status	-0.899	-0.463	-0.895
(1=working, 0=otherwise)	(1.01)	(0.48)	(1.03)
Intercept	-4.706	-4.826	-5.011
	(2.12)	(2.00)	(2.10)
Implied β (structural	-0.361	-0.883	-0.513
effect of schooling)	(1.23)	(2.31)	(1.92)
R^2	0.12	0.12	0.13

Notes: Numbers in ( ) are *t*-statistics, calculated from heteroscedastically-robust standard errors that also take account of cross-equation correlations in the errors. The dependent variable is the number of hours of volunteer work per week. The sample size is 150.

[a]Own-education and sibling's education are treated as endogenous and each sibling's report of the other sibling's education is added to the list of exogenous variables.

[b]Average own education is equal to $(S_i^i + S_i^j)/2$ and average sibling's education is $(S_j^i + S_j^j)/2$.

Table 5: Fixed-Effects Estimates of the Effect of Schooling on Volunteer Hours

	OLS	Classical errors IV[a]	OLS on averages
	(i)	(ii)	(iii)
$S_1^1 - S_2^2$	-0.832	-1.293	
	(2.25)	(2.27)	
$[(S_1^1 + S_1^2)/2] - [(S_2^2 + S_2^1)/2]$			-0.960
			(2.52)
ΔEmployment status[b]	-0.212	0.225	-0.369
	(0.14)	(0.14)	(0.24)
R^2	0.09	0.09	0.11

Notes: Numbers in ( ) are *t*-statistics. Models do not include an intercept. The sample size is 75.

[a]Instrumented for using $S_1^2 - S_2^1$.

[b]Equals one if Twin 1 in a pair is in paid work while Twin 2 is not, equals minus one if Twin 2 is in paid work while Twin 1 is not, and equals zero if both twins in a pair are in the same work state.

Notes

[1] These estimates are based on a specially prepared table of highest qualification gained by number of hours of unpaid work outside the household, using 1996 Census of Population data. The calculations use the midpoints of bracketed volunteer hours (e.g. 1-4 hours in the last four weeks) and assume a value of 90 for the open-ended bracket ("60 or more hours"). The average is calculated over the entire population aged 15 years and over, not just over those who participated in volunteer work. The volunteer work covered by the Census, and throughout the rest of the paper, includes both formal activities (e.g., fund-raising for a charity, coaching children's sports teams) and informal activities (e.g., doing the gardening for an elderly neighbour).

[2] These authors do not consider the point that if people supply volunteer work rather than market work, there may be no net increase in social output, and hence no net benefit to society. For more on this point, see Blau and Ferber (1992, p. 59).

[3] For example, Roberston (1996) claims that the student loans scheme makes graduates less willing to do volunteer or community work because the pressure of debt repayments is so high. Proponents of the view that external benefits justify the existing pattern of public subsidies to tertiary education ignore the fact that many other unsubsidised activities, such as on-the-job training, also bring external benefits (Scobie and Duncan, 1995).

[4] Recent examples of wage equations estimated with twins data include Ashenfelter and Rouse, 1998; Behrman, Rosenzweig and Taubman, 1996; Miller, Mulvey and Martin, 1995; and Ashenfelter and Krueger, 1994.

[5] Note that these are elements of **X** rather than **Z** because the observations in this study are identical twins.

[6] The coefficient is unchanged if the order of the twins is reversed when forming the first-differenced volunteer hours and years of schooling variables.

[7] Years of schooling was calculated from a series of questions about age at primary school entry, age at high school exit, and full-time equivalent years of study since leaving high school.

[8] The questions on volunteer work in the survey of twins were based on the questions used in the Census of Population.

[9] This calculation uses an equation presented on page 1163 of Ashenfelter and Krueger (1994).

[10] I am grateful to Orley Ashenfelter for suggesting this point.

[11] The 1991 Census asked about volunteer work in the previous week and found a participation rate of 19.2 percent while the 1996 Census asked about volunteer work in the previous four weeks and found a participation rate of 41.2 percent. Thus, infrequency of volunteering is likely to be a major reason for the presence of zeros.

[12] Greene (1981) shows that approximate Tobit coefficients can be obtained by dividing OLS coefficients by the sample proportion of non-limit observations.

[13] Tobit estimates are available from the author on request.

[14] This estimate comes from a sample-selectivity corrected wage equation, for someone with 16 years of education and of average age (44), calculated separately for Pakeha/non-Pakeha and male/female and then averaged according to the proportion of these groups in the working age population. The coefficients of the wage equation are reported in column (1) of Table 4.6 in Gibson (1998).

[15] The results are the same if cut-offs of age 35 or age 40 are used.