Estimating the Return to Endogenous Schooling Decisions for Australian Workers via Conditional Second Moments *

Roger Klein
Rutgers University

Francis Vella
Georgetown University

October 2006

Abstract

This paper employs conditional second moments to identify the impact of education in wage regressions where education is treated as endogenous. This approach avoids the use of instrumental variables in a setting where instruments are frequently not available. We employ this methodology to estimate the returns to schooling for a sample of Australian workers. We find that accounting for the endogeneity of education in this manner increases the estimated return to education from 6 percent to 10 percent.

---

1 Introduction

The impact of education on earnings has important implications for education policy and individual investment decisions in human capital. Accordingly, estimating this impact has become an important objective in empirical labor economics. However the level of education is generally chosen by the individual and unobserved factors that influence this choice, such as ability and motivation, are also likely to have a direct affect on earnings. Due to this potential simultaneity, accounting for the endogeneity of schooling is an integral part of estimating how earnings respond to educational investments (for a discussion see Card 2001).

Recent innovations in this literature have generally focussed on instrumental variables (IV) estimation. One popular approach is the creative construction of variables to employ as instruments for schooling. One well known example is the use of a policy "shock" such as a change in the compulsory schooling laws (see, for example, Angrist and Krueger 1991). A second popular approach is to exploit repeated observations on the same unit to eliminate unobserved heterogeneity via appropriate data transformations. This includes panel data where the basis of the unobserved heterogeneity is the individual (for example, Griliches 1978) and family or twins data where the family effect is seen as the unobserved heterogeneity (see, for example, Ashenfelter and Krueger 1994). It is well known that such estimators which exploit the repeated observations aspect of the data have an IV interpretation.

These IV type procedures are based on orthogonality conditions involving the sample first moments of wages and education or the sample moments of differences in wages and education. For example, when one uses multiple observations on the same unit, identification of the education effect is obtained by assuming that appropriately differenced data will produce an observation with a wage equation error which is orthogonal to the differenced education level. The policy shock, on the other hand, imposes orthogonality between the "instrument" or policy shock and the wage equation error. That is, the policy shock affects the education decision but does not directly affect wages. Using different types of policy shocks on the same data may produce different estimates of the impact of education as they may identify the impact from different parts of the population (see Imbens and Angrist 1994).

While estimates based on the conditional first moments possess desirable statistical features when the moment conditions are both "valid" and informative, there is concern that in many empirical examples there are no available instruments. There is also concern that inference based on such estimation is not reliable when the moments are "weak" (see, for example, Staiger and Stock 1997). Accordingly, we adopt an approach which bypasses the need to find instruments. Instead, we employ a methodology which relies on the presence of heteroskedasticity in the model as the basis for identification.

We employ such an approach, described in detail below, to estimate the returns to schooling for a sample of Australian workers. Previous research

on the returns to schooling in Australia has indicated that the individual's background features often directly affect earnings making it difficult to find appropriate exclusion restrictions to act as instruments. Also, there have been few relevant policy shocks which might be exploited to generate variation in educational attainment and there exist limited data sets which allow one to control for unit specific endogeneity. Moreover, the Australian tertiary educational system is currently the focus of an ongoing discussion about how much of the financial burden of higher education should be borne by the student. Clearly a reliable estimate of the return to education is crucial to this debate. The next section describes the estimation procedure while section 3 describes the data and estimation results. Section 4 provides some concluding comments. Finally, note that while this paper focusses on the Australian evidence for the returns to education the identification approach we outline is appropriate to other settings and other economic problems.

2 Model and Identification Strategy

Consider the following simultaneous (triangular) model for wages and education:

$$W_i = X_i \beta_o + \beta_1 E_i + u_i, \ i = 1..N \tag{1}$$

$$E_i = X_i \delta_o + v_i \tag{2}$$

where the X-variables are uncorrelated with the errors u and v in both wage (W) and education (E) equations. Note that the same $X's$ are allowed to appear in both equations. We assume that the u and v are correlated and thus OLS on the wage equation does not produce consistent estimates of the parameter β_1.

To consistently estimate β_1 a control procedure, which is equivalent to IV, is frequently employed where an estimate of v is incorporated into the wage equation with a constant impact. To explain this procedure and our subsequent modification, for notational convenience, we will not distinguish v from its estimate (residual) in what follows. Define:

$$\varepsilon \equiv u - \lambda_o v, \ \lambda_o \equiv \left[cov\left(u, v \right) / \sigma_v^2 \right].$$

The parameter λ_o is the population value or limiting value in large samples of the OLS coefficient obtained by regressing u on v. By construction, once the impact of v is removed from u, the resulting ε component is uncorrelated with v. This observation motivates estimating the controlled regression:

$$W_i = X_i \beta_o + \beta_1 E_i + \lambda_o v_i + \varepsilon_i$$

where we control for endogeneity by including v in the wage equation.[1] Without exclusion restrictions, however, the variables on the right-hand-side are linearly dependent and the matrix of these variables does not have full col-

---

[1]In practice the residual is used in place of v.

umn rank. Accordingly, it is not possible to identify the wage equation parameters.

Note that we would be able to identify the coefficients of interest if the impact of the control was not constant (i.e dependent on X) and if it were possible to estimate this impact. Accordingly, suppose that the joint distribution of u and v depends on X. In particular, let the conditional variances of the errors depend on X. Denoting $S_v^2(X)$ as the conditional variance function for v, define:

$$\varepsilon \equiv u - A(X)v, \ A(X) \equiv \left[cov\left(u,v|X\right)/S_v^2\left(X\right) \right].$$

For each value of X, ε is now defined by removing the impact of v on u. Therefore, conditioned on X, u is uncorrelated with v by construction. Similar to the discussion above, we will refer to $A(X)$ as the impact of the control. However, unlike the case above, this impact may now depend on X and therefore may not be constant. Thus, provided it is possible to consistently estimate the variable impact $A(X)$, the model can be identified without exclusion restrictions.

Below we will discuss several error structures where the control has a variable impact property (VIP) and for which Klein and Vella (2006) show that it is possible to consistently estimate $A(X)$. We first note that the VIP does not necessarily hold even when the joint distribution of u and v depends on X. For example, with ε as a random error component consider

the additive error structure below with the conditional variances of u and v depending on X:

$$u = \lambda_o v + \varepsilon, \quad E\left(v|X\right) = E\left(\varepsilon|X\right) = E\left(v\varepsilon|x\right) = 0.$$

In this model, $A(X)$ is the constant λ_o even though both conditional variance functions may depend on X.

We now turn to several error structures under which the VIP holds and it is possible to consistently estimate $A(X)$. We first provide the statistical models and then provide an economic interpretation of one that is reasonable in the context of the returns to schooling. Denote $S_u^2\left(X\right)$ and $S_v^2\left(X\right)$ as the conditional variance functions for u and v respectively. Then with:

$$u \equiv S_u(X)u^* \text{ and } v \equiv S_v(X)v^*$$

consider the following additive structure for the unscaled errors.

$$
\begin{aligned}
u^* &= \lambda_o v^* + \varepsilon^*, \\
E\left(u^*|X\right) &= E\left(v^*|X\right) = 0; \ cov\left(v^*, \varepsilon^*|X\right) = 0.
\end{aligned}
$$

For this case, it follows that:

$$A(X) = \alpha_o S_u\left(X\right)/S_v\left(X\right),$$

a variable impact which, as discussed below, can be consistently estimated. Alternatively, with ε_1 and ε_2 being mean-zero error components that are independent of X, consider the multiplicative error structure:

$$u = \alpha_1(X)\,\omega^*\varepsilon_1 \;\; ; \;\; v = \alpha_2(X)\,\omega^*\varepsilon_2,$$

where ε_1 and ε_2 are independent of the common error component, ω^*. The conditional second moment for the common error component, ω^*, may or may not depend on X. In this case, with ρ_o as the correlation between ε_1 and ε_2 :

$$A(X) \equiv \left[cov\,(u, v|X)\,/S_v^2\,(X) \right] = \rho_o \left[\frac{S_u(X)}{S_v(X)} \right],$$

a form identical to that in the previous example.

For both of the above error structures, the conditional correlation:

$$\rho_o \equiv cov(u, v|X = x)/\left[S_u\,(X)\,S_v\,(X) \right]$$

is constant. Whenever this condition holds, it follows in general that:

$$A(X) \equiv \left[cov\,(u, v|X)\,/S_v^2\,(X) \right] = \rho_o \left[\frac{S_u(X)}{S_v(X)} \right].$$

Klein and Vella (2006) show that under the above constant correlation condition (CCC) it is possible to estimate $A(X)$ consistently in a semiparametric formulation without specifying functional forms for the conditional variance

8

functions. Provided that the ratio of these functions depends on X, a condition typically satisfied when heteroskedasticity is present and conditional variance functions are not proportional to each other (for all X values), then the model is identified without exclusion restrictions.

Before briefly indicating the nature of the estimation method it is instructive to interpret one of the above error structures in the context of the current application. Assume that wages and education each depend on unobserved ability, a^*. Further, assume that the impact of ability is not constant in at least one of the equations. Let this impact differ in these two equations and let it consist of a potentially predictable or estimable component that depends on X and a random component that cannot be estimated. Denote $a_1(X)$ and $a_2(X)$ as the predictable impacts for wage and education equations respectively and let ε_1 and ε_2 be the corresponding unpredictable components. Under these conditions, unobserved ability enters wage and education equations as:

$$W \quad : \quad u = a_1(X) a^* \varepsilon_1$$

$$E \quad : \quad v = a_2(X) a^* \varepsilon_2.$$

With the components satisfying the conditions above, the control has a variable impact that depends on the ratio of conditional variance functions.

Although it is possible to estimate $A(X)$ nonparametrically, it is more practical to impose an index structure in order to obtain reasonable esti-

mates at moderate sample sizes. Namely, assume that conditional variance functions both depend on linear indices:

$$S_v^2 (X) = S_v^2 (X\alpha_o); \quad S_u^2 (X) = S_v^2 (X\gamma_o).$$

With $\hat{E}$ as a nonparametric conditional expectations estimator, we employ Semiparametric Least-Squares (see Ichimura 1993), to estimate the parameters of the v-index. Namely, subject to the usual normalization on α_o : [2]

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{N} \left[\hat{v}_i^2 - \hat{E} \left(\hat{v}_i^2 \mid X\alpha \right) \right]^2 /N,$$

where $\hat{v}_i^2$ is a consistently estimated squared residual. It then follows that the conditional variance function for v is given as:

$$\hat{S}_v^2 (X) \equiv \hat{E} \left(\hat{v}_i^2 \mid X\hat{\alpha} \right).$$

The other conditional variance function must be estimated simultaneously along with other parameters of interest. To describe the estimation method, define:

$$u_i (\beta) \equiv W_i - X_i \beta_0 - \beta_1 E_i$$

---

[2]With X_1 as a continuous variable and with X not containing a column of ones:

$$\alpha = [1; \ \alpha_2].$$

Subject to this normalization (much like setting the variance of the error in a probit model to 1), the α_2 are identified. In a probit model, we are able to identify probability functions under the normalization that the variance is 1. Similarly, here conditional variance functions are identified once we know the normalized parameters above.

where the $\beta's$ are arbitrary parameter values. Notice that at the true parameter values that $u_i\left(\beta_o\right) = u_i$, the wage error. Next, define a "variance-type" function:

$$S_{ui}^2\left(\beta, \gamma\right) \equiv E\left[u_i^2\left(\beta\right) | X_i\gamma\right].$$

Notice that at the true parameter values:

$$S_{ui}^2\left(\beta_o, \gamma_o\right) = S_u^2\left(X_i\right),$$

the true conditional variance function for u_i Replacing the true expectation E above with the nonparametric estimator $\hat{E}$, we obtain the feasible estimator:

$$\hat{S}_{ui}^2\left(\beta, \gamma\right) \equiv \hat{E}\left[u_i^2\left(\beta\right) | X_i\gamma\right].$$

We then consider the following controlled nonlinear model:

$$W_i = X_i\beta_o + E_i\beta_{1o} + \hat{A}_i\left(\beta, \gamma\right)v_i + error.$$

Parameter estimates are then "essentially" obtained by selecting γ and β parameters to minimize the sum of squared residuals.[3] Klein and Vella (2006) establish that the resulting estimator is consistent and asymptotically distributed as normal at the usual $\sqrt{N}$ parametric rate.

---

[3]Here, "essentially" means that as shown in Klein and Vella (2006), several different nonparametric expectations operators are required to obtain identification, one based in the u-index and the other on both u and v indices. For details the reader is referred to the identification discussion in Klein and Vella (2006).

The general approach employed here is related to other estimators used in this context. The rank order estimator of Vella and Verbeek (1997), applied to the returns to education in Rummery et al (1999), is also identified via heteroskedasticity. The Vella and Verbeek (1997) estimator requires that rank order of the individual's residual, in a given subset of the data, is relevant rather than the value of the residual itself. They then construct an instrumental variables procedures based on this premise. This is a special case of the approach employed here in that while Vella/Verbeek assume that S_{ui} and S_{vi} are uncorrelated this is not imposed here. Another related paper is Hogan and Rigobon (2004) who form an alternative structure for $A(X)$ in that they focus on conditional covariances. That is, they assume that some variable is related to the variances of the education equation error but does not directly determine wages. They also do not estimate the model in the control function manner but follow the procedure outlined in Rigobon (1999). A closely related procedure to Rigobon (1999) is proposed by Lewbel (2004) although the issue of the returns to schooling is not addressed there.

3   Data and Model Specification

To estimate the impact of schooling on earnings, we employ the 2001 wave of "The Household, Income and Labour Dynamics in Australia (HILDA) Survey". These data contain labor market and background information on a sample of Australians. We examine the determinants of wages for a sample of

5070 working individuals living in the five most populous states in Australia.[4] We focus on the wage determination process conditional on working and do not address any sample selectivity issues induced by the working decision.

Estimating the impact of education in the Australian context is an interesting problem. Arguing that the high returns to education merited the additional investment, the Australian Federal Government actively encouraged individuals to seek additional education (Vella and Gregory 1996). There is also a very active ongoing debate regarding who should bear the cost of such investment. That is, there are proposals to shift an increasing share of the cost of tertiary education onto the students undertaking the investment.

An important contribution to this debate would be a reliable estimate of the return to schooling. However, previous papers have supported the conjecture that education is endogenous to wages thereby invalidating the evidence from studies which do not account for the endogeneity. However, obtaining estimates which do account for the endogeneity is complicated by the lack of instruments in the Australian context. For example, Vella and Gregory (1996) provide evidence that the individual's background characteristics directly influenced wages, making it is difficult to assign background characteristics the role of instruments. Leigh and Ryan (2005) employ a strategy similar to Angrist and Krueger (2005) but find that the instruments do not provide an estimate of the return to education with tight bounds in

---

[4]We exclude observations from the less populous states and regions to avoid estimation issues related to small cell sizes.

the Australian context. Miller et al (1995) employ a sample of twins to esti-
mate the returns to schooling but as the data do not contain a direct measure
of wages it is necessary to impute them. Given this evidence we employ the
estimation strategy outlined in the previous section where we employ all the
variables in the schooling equation as regressors in the wage regression and
identify the schooling effect via heteroskedasticity.

The model we estimate has the following form:

$$
\begin{aligned}
wage \;=\; & \beta_0 + \beta_{1j} * both\ parents + \beta_2 * siblings + \\
& \sum_{j=1}^{2} \beta_{3j} * parent's\ labor\ market + \beta_{4j} * private\ school + \\
& \sum_{j=1}^{4} \beta_{5j} * state\ of\ school + \beta_6 * Married + \beta_7 * Australian\ Born + \\
& \beta_8 * Male + \beta_9 * Years\ in\ Aust + \beta_{10} * Tenure + \beta_{11} * Tenure^2 + \\
& \beta_{12} * Age + \beta_{13} * Age^2 + \beta_{14} * school + error_1
\end{aligned}
$$

$$
\begin{aligned}
school \;=\; & \delta_0 + \delta_{1j} * both\ parents + \delta_2 * siblings + \\
& \sum_{j=1}^{2} \delta_{3j} * parent's\ labor\ market + \delta_{4j} * private\ school + \\
& \sum_{j=1}^{4} \delta_{5j} * state\ of\ school + \delta_6 * Australian\ Born + \delta_7 * Male \\
& + \delta_8 * Years\ in\ Aust + \delta_9 * Age + \delta_{10} * Age^2 + error_2
\end{aligned}
$$

14

where *both parents* denotes that both parents were present in the household when the individual was aged 14; *siblings* denotes the number of siblings; *parent's labor market* captures whether the individual's mother was employed and whether the father experienced period of unemployment; *state of school* and *private school* are dummy variables indicating the region of the school attended and whether the individual attended a private school; *Australian* denotes Australian Born and *Years in Aust* captures the number of years immigrants have been in Australia; *Tenure* denotes years of tenure with the same employer; and the remaining variables are self explanatory. The dependent variables are the log of the hourly wage rate and the number of years of schooling. Note that we employ additional, work related, variables in addition to marital status, in the wage equation which do not appear in the education equation. These do not identify the model as conventional IV procedures require a variable(s) in the education equation which does not appear in the wage equation. We employ the same exogenous variables for each of these indices for conditional variances that appear in the corresponding conditional means.

Before presenting the empirical evidence consider why heteroskedasticity may arise in the schooling equation. Rummery, Vella and Verbeek (1999) argue that one source might be the regional variables. For example, consider, as in Card (1995), where the distance to the nearest school influenced the schooling decision. In that case various geographical allocations of schools within a region may produce not only different levels of schooling but also

drastically different variances in regional educational attainment. Other variables may also be source of heteroskedasticity. For example, while attendance at private schools generally increases educational attainment there is a large degree of heterogeneity across these schools in Australia and it seems implausible that one would expect the same impact irrespective of the quality. In this way, the varying marginal effect of school type, due to the heterogeneity of schools, on educational attainment might lead to heteroskedasticity. It is also possible that the school effect may vary across individual. For example, reconsider the effect of attendance at a Private School. Such schools may increase educational attainment because they are able to provide a superior quality of education. However, they may also be able to focus more on students which require more remedial help ensuring that each individual attains a minimum level of education. Students with such specific demands may do less well in an alternative system and this would generate a greater variance.

Similar logic applies to the presence of heteroskedasticity in the wage equation. That is, in some instances the effect of the variable on wages may vary across individual. However, many individual characteristics may directly influence the variance of the error. For example, attendance at certain School types could either decrease or increase the variance of the wage error by not only ensuring a minimum level of quality of education for the less talented individual but by also assisting the more talented individuals to do well in the labor market. Also, as individuals obtain greater work experience it is likely that some will do better than others in terms of wage growth.

The summary statistics for our sample are reported in Table 1. In Table 2 we report the estimates for the schooling equation. The absolute value of the *t statistics* for the estimates are reported in parentheses. As expected, a number of the individual's background characteristics have a statistically significant impact on the level of acquired schooling. Australian Born individuals acquire approximately one year of education less than relatively recently arrived foreign born individuals but this difference decreases as the number of years in Australia increases. This latter effect probably reflects when the individual arrived in Australia. Having both parents present when the individual was aged 14 increases educational attainment by about half of one year. While this may reflect parental guidance or household stability it is also likely to capture income effects. The presence of siblings also decreases educational attainment. Attendance at Catholic or other types of Private schools has a very large and statistically significant positive effect on the level of education obtained recalling that the excluded group is those attending a Government financed school. The *Private School* coefficient indicates that attendance at such a school increases educational attainment by .87 years as opposed to attendance at a Government School. The regional variables indicate some differences across States. Finally, males acquire .3 years of schooling less than females.

As the coefficients in the index for the heteroskedastic function have no immediate interpretation we do not report them here. We note, however, that several of them, including the continuous age variables were statisti-

cally significant. Moreover, the NR^2 from the regression of the residuals squared on the all of the exogenous explanatory variables employed in the schooling equation produced a value of 263.39. This test value strongly rejects the null hypothesis of homoskedastic errors. Moreover, several of the groups of variables, as well as some individually, appeared to be a source of heteroskedasticity.

In Table 3 we report the estimates for the wage equation. The absolute value of the t $statistics$ are reported in parentheses. First consider the OLS equation as a number of its features are worth noting. There is some evidence that the background variables have direct influence on the wage level when the endogeneity of education is not taken into account. The number of siblings, for example, appears to directly decrease the wage. This may be explained by the quality of education one obtains in the presence of several siblings if there are trade-offs with quality as well as quantity as indicated in Table 2. There is also a positive relationship between the level of earnings and whether or not the individual's mother worked but the effect is only relatively weak in terms of statistical significance. Attendance at a Private school increases wages by 5.5 percent and this effect would be statistically significant in the absence of the endogeneity problem. There is also evidence of a small marriage premium and gender differential of 10 percent in favor of males. The regional variables are statistically significant but this most likely reflects the higher cost of living in the control group NSW noting that most people are likely to be living in the state in which they attended school.

18

Finally, the point estimate for the education effect is 6 percent.

Now focus on the CF estimates which are reported alongside the OLS estimates in Table 3. Once again the absolute value of the *t statistics* appear in parentheses noting that the manner in which they are estimated accounts for the nature of the estimation process. Before we focus on the coefficient of primary interest, it is valuable to note that while the estimates across the two columns for the exogenous variables are generally quite similar there are some important differences. More importantly these differences have implications for what may or may not be reasonable instruments. More explicitly, this column indicates that two possible instruments may be the number of siblings or school type although each of these was statistically significant in the OLS estimation. The key feature of this column, however, is the estimate of the education coefficient. While the OLS estimate was 6 percent we see that the CF estimate is 10 percent. Moreover, while there is some loss in statistical significance, in comparison to the OLS estimate, the coefficient is statistically significant at levels far below 5%. This estimate indicates that the return to education obtained through our procedure is notably higher than that indicated through OLS estimation.

The increase in the return to schooling is due to the control function having an estimated coefficient of -.203. This negative coefficient indicates that the unobservables correlated with wages are negatively correlated with education. This finding is consistent with the results of Vella and Gregory (1996) who interpreted such a result as a "penalty" to educational over achieving.

That is, the unobserved factors which increased an individual's education level above what was predicted by his or her background characteristics have a negative impact on the wage level. Thus individuals who obtain a level of education above what is predicted by the model received less for their incremental increase in education than those who were predicted to obtain that level. As the penalty to overachievement is given by $\rho_o \frac{S_{ui}}{S_{vi}}$ the penalty is greater for individuals with characteristics associated with lower variances in the schooling equation and higher variances in the wage equation.

While the return to education is somewhat offset by this "penalty" operating through the control function, the large increase in the point estimate indicates that the return to education is still high and markedly higher than the OLS estimate. This finding is consistent with the general finding that accounting for the endogeneity of education leads to an increase in the estimated return to education in general and also, more specifically, in the Australian context. Vella and Gregory (1996), for example, find that accounting for the endogeneity of education greatly increased the estimated return for Australian youth.

To interpret our results suppose the unobserved education residual captures unobserved ability. Thus an individual who has a high level of motivation attains a higher than expected, on the basis of his/her characteristics, level of education and this leads to higher wages through the high returns to education. However, although the individual's wage increase substantially through the increased investment we see that the actual return is reduced due

to the fact that the individual is perceived to have over achieved. Moreover, the level of penalty depends on the individual's characteristics. The impact of this penalty process is that the return to education is greater than what is revealed in an OLS regression where the penalty has been internalized.

As with the schooling equation there is some indication of heteroskedasticity in the wage equation. The NR^2 for the regression of the residuals squared on the exogenous variables produced a value of 110.78 which indicates there is less evidence of heteroskedasticity in this equation than in the schooling equation. However, while there is less heteroskedasticity in the wage equation this does not affect the validity of the estimator as heteroskedasticity in only one of the equations is sufficient.

Before proceeding to considering IV estimates of the return to schooling with these data it is also worth highlighting that the coefficient ρ_o is statistically different from zero at the 5 percent level. This indicates that education is endogenous to wages. This is an important result as it implies that the identifying moment is sufficiently informative to identify that education is endogenous to earnings.

The empirical investigation graphically highlights the issue associated with this uncertain choice of instruments. Many of the background variables have statistical significance levels which makes it unclear whether they can be employed as instruments. Many of them are marginal in the CF estimation in terms of statistical significance. Consequently, neglecting pre-test issues, it is possible that they can be employed as instruments. Accordingly, employing

the same data we re-estimated the model where we used the same first step, shown in Table 2, and in the second we employed the predicted value as an instrument for education.

Given the estimates in Table 3 from the control function estimator it appears that the background variables plus school type are potential instruments. This is interesting since the OLS estimates in column 1 of this Table indicated that the number of siblings and school type were not valid instruments as they had a direct effect on wages. Accordingly, we first employed Both Parents, Father Unemployed and Mother Employed as the only instruments. Doing so provided an estimate of the schooling effect of .063 with a *t-statistic* of 2.42. Not surprisingly, given the small increase in the education coefficient in comparison to the OLS estimate, the *t-statistic* on the residual is -.12 indicating that one could not reject exogeneity of schooling.[5]

On the basis of the results in the column headed CF, however, it seems that both Siblings and Private School are both potential instruments. Accordingly, we excluded each separately from the wage equation to estimate the schooling effect. Including Siblings as an instrument with the other backgrounds while including the school type variable in the wage equation produced an estimate of .090 with a *t-statistic* of 6.76 for the schooling variable. The coefficient on the residual is -2.19 which rejects exogeneity. Including Private School as an instrument, while keeping siblings in the wage equa-

---

[5]The estimated education coefficient was obtained via IV, while the t-statistic for the residuals was obtained via the control function version of IV.

tion, produced an estimate of .102 (t=8.48) and a *t-statistic* on the residual of -3.49. Finally using all the background variables plus siblings and Private school produced an estimate of .101 (t=10.46) and a *t-statistic* on the residual of -4.14.

Continuing to neglect pre-test issues, these IV estimates provide support for our approach. First, the OLS estimates illustrate the inappropriateness of trying to identify instruments from a regression that does not account for the endogeneity. Second, while several of the background variables appeared to be valid instruments they were unable to identify the endogeneity of the schooling variable. Third, the estimates based on schooling type and number of siblings were able to identify the presence of endogeneity. However, without our approach being first employed one not have known if these were valid instruments. Finally, the resulting IV estimate provided an estimate of the return to schooling which is almost identical to the estimate based on our control function procedure. As this IV estimate is not based on the CCC assumption this appears to provide strong support for our approach in this setting. Overall, the evidence is quite suggestive that the return to schooling for the individuals in this data set is underestimated in OLS estimation and accounting for the bias in OLS estimation leads to an estimate of around 10 percent.

4 Conclusions

This paper provides an alternative approach to identifying and estimating the returns to education in a model where the level of education is endogenous to wages. The identification strategy is based on combining the presence of heteroskedasticity in the model with the assumption that the relationship between the errors, conditional on the exogenous variables, is constant. For a sample of Australian workers we find a large increase in the return to schooling in comparison to what is found when the endogeneity is not accounted for. More explicitly, we obtain an estimate of 10 percent in contrast to the OLS estimate of 6 percent. We also find that using our approach we are able to identify variables which would otherwise have been excluded as instruments. Using these variables as instruments provides an estimate of the return to schooling which is almost identical to our estimate thereby providing support for our identifying strategy.

Table 1: Summary Statistics

Variable	Mean	Std
Hourly wage	20.538	10.823
Years of Education	12.470	2.222
Australian	.778	
Both Parents	.829	
Siblings	2.727	1.977
Father Unemployed	.095	
Mother Employed	.520	
Private School	.255	
New South Wales	.312	
Victoria	.257	
Queensland	.221	
South Australia	.094	
Western Australia	.104	
Married	.544	
Male	.536	
Age	38.488	11.522
Years in Australia	4.503	(10.573)
Tenure	6.488	(7.604)

Table 2: Education Equation

Variable		Variable	
Constant	10.254	Years in Australia	-.029
	(30.358)		(5.400)
Age	.192		
	(11.008)		
Age^2	-.003		
	(11.418)		
Both Parents	.539		
	(7.155)		
No. of Siblings	-.176		
	(11.146)		
Father Unemployed	-.158		
	(1.620)		
Mother Employed	.116		
	(1.963)		
Private School	.871		
	(13.198)		
Victoria	.075		
	(.982)		
Queensland	-.305		
	(3.794)		
South Aust.	.079		
	(.681)		
Western Aust.	-.082		
	(.820)		
Australian Born	-1.018		
	(8.473)		
Male	-.313		
	(5.439)		

Table 3: Wage Equation

Variable	OLS	CF	Variable	OLS	CF
Constant	.947	.573	Male	.097	.121
	(12.387)	(2.615)		(8.414)	(8.264)
Age	.051	.042	Years in Australia	.004	.005
	(14.591)	(7.909		(4.955)	(3.606)
Age^2	-.0005	-.0005	Tenure	.012	.011
	(13.176)	(6.425)		(5.605)	(3.803)
Both Parents	-.003	-.026	Tenure^2	-.0001	-.0001
	(.206)	(1.247)		(2.145)	(1.163)
No. of Siblings	-.006	-0.001	School	.060	.100
	(2.235)	(.142)		(21.757)	(5.260)
Father Unemployed	.002	.006	ρ		-.203
	(.123)	(.292)			(2.213)
Mother Employed	.017	.017			
	(1.447)	(1.327)			
Private School	.055	.018			
	(4.075)	(.798)			
Victoria	-.034	-.042			
	(2.257)	(2.560)			
Queensland	-.081	-.073			
	(5.130)	(4.294)			
South Aust.	-.121	-.130			
	(5.691)	(5.838)			
Western Aust.	-.052	-.064			
	(2.577)	(2.955)			
Australian Born	.089	.121			
	(3.734)	(3.527)			
Married	.036	.040			
	(2.781)	(2.955)			

References

[1] Angrist, J., and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics,* 106, 979-1014.

[2] Ashenfelter.,O. and A.Krueger (1994): "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review*, 84, 1157-1173.

[3] Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Returns to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp*, eds. L. N. Christofides et al. Toronto: University of Toronto Press. 201-221.

[4] Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica,* 69, 1127-1160.

[5] Griliches, Z. (1977): "Estimating the Return to Schooling: Some Econometric Problems," *Econometrica,* 45, 1-22.

[6] Hogan, V., and R.Rigobon (2004): "Using Unobserved Supply Shocks to Estimate the Returns to Education," unpublished paper.

[7] Ichimura, H, (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single index models" *Journal of Econometrics*, 58, 71-120.

[8] Imbens, G., and J.Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica,* 62, 467-476.

[9] Klein, R., and F.Vella (2005): "Estimating a Class of the Triangular Simultaneous Equations Model Without Exclusion Restrictions," *IZA* working paper 2378.

[10] Leigh, A., and C.Ryan (2005): "Estimating Returns to Education: Three Natural Experiments Techniques Compared", *ANU CEPR* Discussion paper 493.

[11] Lewbel, A. (2004): "Identification of Heteroskedastic Endogenous Models or Mismeasured Regressor Models," unpublished manuscript.

[12] Miller, P.,C.Mulvey and N.Martin (1995): "What do Twin Studies Reveal about the Economic Returns to Education? A Comparison of Australian and U.S. Findings," *American Economic Review*, 85, 586-599.

[13] Rigobon, R. (1999): "Identification through heteroskedasticity," *Review of Economics and Statistics,* 85, 777-792.

[14] Rummery, S., F.Vella and M.Verbeek (1999): "Estimating the Returns to Education for Australian Youth via Rank-Order Instrumental Variables," *Labour Economics,* 6, 491-507.

[15] Staiger, R. and J.Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 68, 1055-1096.

[16] Vella, F. and M.Verbeek (1997): "Rank Order as an Instrumental Variable" unpublished manuscript