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

Roger Klein
Rutgers University

Francis Vella
Georgetown University

March 2006
Preliminary Draft

Abstract

This paper employs a methodology based on conditional second moments to identify the impact of education in wage regressions where the level of education is treated as endogenous. This approach avoids the use of instrumental variables in a setting where instruments are frequently not available. Our methodology combines the presence of heteroskedasticity, in either or both the wage and education equations, with the imposition that the relationship between the unscaled error terms is constant. 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 5.8 percent to 8.8 percent.

1 Introduction

The impact of education on earnings has important implications for education policy and individual investment decisions in human capital and, accordingly, its estimation has become an important objective of empirical labor economics. However as the level of education is generally chosen by the individual, unobservable factors that influence education, 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 primarily focussed on the use of instrumental variables (IV) and involve the creative use of data to construct instruments or specific data transformations to eliminate unobserved heterogeneity in data comprising repeated observations on the same unit. An example of the latter is the use of deviations from unit specific means as instruments for education. When using panel data the definition of the unit is an individual (for example, Griliches 1978), while others have employed the family or twins as the unit of focus (see, for example, Ashenfelter and Krueger 1994). An example of the former is when IV is employed via the use of a defined policy "shock" (see, for example, Angrist and Krueger 1991). Both of these IV 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. These various moment conditions often produce different estimates as they are frequently identifying the education effect from different parts of the underlying 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 increasing concern that inference based on such estimation is not reliable when the moments are "weak" (see, for example, Staiger and Stock 1999). This paper adopts a different approach and focuses on conditional second moments. That is, we impose that the conditional correlation coefficient, CCC, between the unobservables factors influencing the individual's wage and education is constant after conditioning out the exogenous variables. That is, the relationship between the unscaled error terms across equations is invariant to the observable characteristics.

While this assumption is attractive for its economic content it is insufficient on its own to identify the education effect. However, when it is combined with the presence of heteroscedasticity in the wage and/or education equations the education effect is identified. We employ such an approach 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 affect earnings and thus it is difficult to find appropriate exclusion restrictions to act as instruments. Also, unlike other countries, there have not been any 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 engaged in an ongoing discussion about how much of the financial burden of higher education should be borne by the student. Clearly understanding the return to education is crucial to this debate. Also, while we focus on the Australian evidence the identification approach we outline is appropriate to other settings and other economic problems. The next section describes the estimation procedure while section 3 describes the data and estimation results. Section 4 provides some concluding comments.

2 Model and Identification Strategy

To motivate the CCC assumption and to illustrate the identifying power of heteroscedasticity consider the following model:

$$w_i = X_i\beta + \gamma E_i + u_i, \ i = 1..N \tag{1}$$

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

$$u_i = S_u(X_i)u_i^* \tag{3}$$

$$v_i = S_v(X_i)v_i^* \tag{4}$$

where w and E denote the level of wage and education respectively; X is vector of exogenous variables; β, γ and δ are unknown parameters; u and v are heteroskedastic zero mean error terms generated through equations (3) and (4) where u^* and v^* are homoscedastic (unscaled) disturbances with a non zero correlation and the $S's$ denote unknown functions.

The endogeneity of E in (1) operates through a non-zero correlation between u and v. To purge this common component from (1) one can employ the "control function" approach to IV by including v in (1). The wage equation error for individual i, following the inclusion of v in (1), becomes:

$$\varepsilon_i = u_i - \frac{cov(u,v)}{var(v)}v_i$$

which is uncorrelated with education by construction. Note that

$$\frac{cov(u_i, v_i)}{var(v_i)} = \frac{cov(u_i, v_i)}{\sigma_u \sigma_v}\frac{\sigma_u}{\sigma_v} = \rho_{uv}\frac{\sigma_u}{\sigma_v}$$

which indicates that the return to the unobserved component in the wage equation depends on the correlation, ρ_{uv}, between u and v weighted by the ratio of the standard deviations of the errors. As the weighting factor does not vary across i, IV requires an exclusion restriction as the mapping from v_i and u_i is linear. Thus even though the correlation coefficient is constant the model requires an exclusion to identify γ. However, consider where the distribution of u_i and v_i depend on X_i. The new wage equation error term, conditional on X_i, is:

$$\begin{aligned}
\varepsilon_i &= u_i - \frac{cov(u_i, v_i|X_i)}{var(v_i|X_i)}v_i \\
&= u_i - \frac{cov(S_{ui}(X_i)u_i^*, S_{vi}(X_i)v_i^*)|X_i)}{var(S_{vi}(X_i)v_i^*|X_i)} \\
&= u_i - \left[\frac{cov(u_i^*, v_i^*)}{var(v_i^*)}\right]\left[\frac{S_{ui}(X_i)S_{vi}(X_i)}{S_{vi}^2(X_i)}\right]v_i \\
&= u_i - \rho_{uv}^*\frac{S_{ui}(X_i)}{S_{vi}(X_i)}v_i
\end{aligned}$$

since $\rho_{uv}^* = \frac{cov(u_i^* v_i^*|X_i)}{var(v_i^*|X_i)} = \frac{cov(u_i^* v_i^*)}{var(v_i^*)}$under the CCC assumption[1]. This has the feature that the mapping from u_i to v_i, captured by the term $\rho_{uv}^*\frac{S_{ui}(X_i)}{S_{vi}(X_i)}$,is a function of X_i and this non-linearity in the mapping identifies the model even in the absence of an exclusion restriction.

The above shows that when the conditional variances are known it is straightforward to construct the appropriate control function. Strong assumptions regarding the conditional variances are similar to exclusion restrictions in that they impose information in estimation. Thus, ideally the CCC assumption would be employed while as little as possible is imposed on the variances. We do so by employing the procedure proposed by Klein and

[1]As we have separated the variances of the errors into a part that is constant and one that depends on X it is natural to assume that σ_{v^*} and σ_{u^*} are both equal to 1 which means $\rho_{uv}^* = \frac{cov(u_i^* v_i^*)}{var(v_i^*)} = \frac{cov(u_i^* v_i^*)}{\sigma_{v^*}\sigma_{u^*}}$ which is the reason we refer to the coefficient as a correlation coefficient.

Vella (2005). First, to facilitate estimation, we impose the following index restrictions on the variances:[2]

$$Var(u_i|X_i) = S_{ui}^2 = S_u^2(X_i) = S_u^2(X_i\theta_u) \quad (5)$$
$$Var(v_i|X_i) = S_{vi}^2 = S_v^2(X_i) = S_v^2(X_i\theta_v) \quad (6)$$

where the $\theta's$ are unknown parameters and the same $X's$ appear in each of the conditional variances as well as the conditional means. We treat the $S's$ as unknown functions. Defining $\widehat{v}_i = E_i - X_i\widehat{\delta}$, the equation to be estimated is:

$$w_i = X_i\beta + \gamma E_i + A(X_i)\widehat{v}_i + \varepsilon_i$$

where $A(X_i) = \rho_{uv}^* [S_{ui}/S_{vi}]$ and the matrix $M \equiv [X : E : (S_{ui}/S_{vi})\widehat{v}]$ is of full rank providing the ratio (S_{ui}/S_{vi}) is not constant. The interaction between the $X_i's$ and $\widehat{v}_i$ identifies the model noting that the interaction must take the form S_{ui}/S_{vi}.

To illustrate our identifying assumptions consider their implication for our empirical investigation. By definition, v captures the level of education the individual attains in excess of what is predicted by his/her characteristics. If v^* captures unobserved ability, the v reflects the level of additional education which is observed after it is scaled up by S_{vi}. This scaling up reflects that individuals with certain $X's$ receive more for their v^*. For example, certain types of schools, or individual characteristics, may magnify the impact of unobserved ability on schooling attainment. Thus the contribution to the schooling equation for each individual is no longer v_i^* but $S_{vi}v_i^*$. Assume a similar process occurs in the wage equation. That is, individuals with a specified level of unobserved ability receive a different return on the unobservables depending on their characteristics and/or the characteristics of the job. For example, it is generally accepted that union membership decreases wage variance within the union sector. Thus while the individual's endowment is u_i^* the contribution to the wage is $S_{ui}u_i^*$.

The CCC assumption imposes a constant relationship between v_i^* and u_i^*. Thus a specified level of unobserved ability v^* merits a return which is the product of a price ρ_{uv}^* which is common to all individuals, independent of their $X's$, and the ratio of scaling factors S_{ui}/S_{vi} which varies across

[2] Klein and Vella (2005) provide proofs of identification of the coefficient of the endogenous regressor for the cases where the index restriction is imposed and also when it is not.

individuals on the basis of their $X's$. In the absence of heteroskedasticity the return to v is constant and given by $\rho^*_{uv}\frac{\overline{\sigma}_u}{\overline{\sigma}_v}$ where $\overline{\sigma}_u = \sqrt{\frac{1}{N}\sum_{i=1}^N \widehat{u}_i^2}$ and $\overline{\sigma}_v = \sqrt{\frac{1}{N}\sum_{i=1}^N \widehat{v}_i^2}$. However when both the schooling and wage equations have heteroskedastic errors the value to be augmented to the wage equation is $\rho^*_{uv}\left[S_{ui}/S_{vi}\right]v_i^*$ highlighting that the ρ^*_{uv} which captures the underlying value of the unobservables, conditional on X, has remained constant across individuals although the return will vary by X_i.

The CCC assumption thus appears very reasonable and could arise in the context of many models of educational attainment. For example, models of human capital accumulation would reasonably require that the underlying level of skill was constantly rewarded while the observed value could depend on the schooling and work environments. Alternatively, models based on the screening hypothesis clearly impose the same type of mechanism. That is, for certain values of S_{vi} the individuals are ranked by their endowment of v_i^*.

The model to be estimated is captured in equations (1),(2), (5) and (6) and our estimation strategy is the following. We first estimate the reduced form by OLS to obtain the residuals $\widehat{v}$. We then estimate θ_v through a variant of the semi-parametric least squares procedure discussed by Ichimura (1993) by regressing $\widehat{v_i}^2$ on a index representation of X_i.[3] We then estimate S_{vi} as $\sqrt{\widehat{E}[\widehat{v}_i^2|X_i\widehat{\theta}_v]}$ where $\widehat{E}$ is a non parametric expectation estimated by appropriately chosen kernel methods. With $\widehat{v}_i$ and $\widehat{S}_{vi}$ we can construct the following control function equation:

$$w_i = X_i\beta + \gamma * E_i + \rho^*_{uv}S_{ui}\frac{\widehat{v}_i}{\widehat{S}_{vi}} + \varepsilon_i \tag{7}$$

where ε is a zero mean error term uncorrelated with the regressors and noting that it is necessary to estimate S_{ui}. Noting that $S_{ui} = S_u(X_i\theta_u) = \sqrt{(w_i - X_i\beta + \gamma * E_i)^2}$, we proceed by estimating β, γ and ρ^*_{uv} at the same time as we estimate θ_u and S_{ui} non-parametrically in the same manner as $\widehat{\theta}_v$ and $\widehat{S}_{vi}$. Klein and Vella (2005) show that the resulting estimates of β, γ and ρ^*_{uv} are $\sqrt{N}$ consistent and asymptotically normal.[4] It is also possible to

[3]The procedure differs from the case considered in Ichimura (1993) since the dependent variable is the residual (squared) from an earlier regression.

[4]In estimating the model it is necessary to impose the both constant conditional correlation coefficient and index assumptions. The manner in which this is done is described in Klein and Vella (2005).

estimate a "GLS" counterpart of this model by dividing the above equation by our estimate of S_{ui}. The simulation evidence in Klein and Vella (2005) indicated that this form of the estimator worked best in finite samples and this we report below the estimates from:

$$\frac{w_i}{\widehat{S_{ui}}} = \beta \frac{X_i}{\widehat{S_{ui}}} + \gamma * \frac{E_i}{\widehat{S_{ui}}} + \rho^*_{uv} \frac{\widehat{v}_i}{\widehat{S}_{vi}} + \frac{\varepsilon_i}{\widehat{S_{ui}}}. \tag{8}$$

The general approach 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 a special case of this estimator. 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 for 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 and we examine the determinants of wages for a sample of 5371 working individuals. We focus on the wage determination process conditional on working and do not address the endogeneity of the working decision.

Estimating the impact of education in the Australian context is an interesting problem. Vella and Gregory (1996) discuss how the Australian Federal Government actively encouraged the increased participation in the educational process on the basis that the returns to education merited the increased investment. 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. Second, previous papers have supported the conjecture that education is endogenous to wages in this market. Third, Vella and Gregory (1996) provide evidence that the individual's background characteristics directly influenced wages making it is difficult to apriori assign background characteristics the role of instruments. The model we estimate has the following form:

$$\begin{aligned} wage &= \beta_0 + \sum_{j=1}^{3} \beta_{1j} * family + \sum_{j=1}^{2} \beta_{2j} * parent's\ labor\ market + \\ &\quad \beta_3 * siblings + \sum_{j=1}^{6} \beta_{4j} * state\ of\ school + \sum_{j=1}^{2} \beta_{5j} * school\ type + \\ &\quad \beta_7 * Married + \beta_8 * Australian\ Born + \beta_9 * Male + \\ &\quad \beta_{10} * Age + \beta_{11} * Age^2 + \gamma * school + error_1 \end{aligned}$$

$$\begin{aligned} school &= \delta_0 + \sum_{j=1}^{3} \delta_{1j} * family + \sum_{j=1}^{2} \delta_{2j} * parent's\ labor\ market + \\ &\quad \delta_3 * siblings + \sum_{j=1}^{6} \delta_{4j} * state\ of\ school + \sum_{j=1}^{2} \delta_{5j} * school\ type + \\ &\quad \delta_7 * Married + \delta_8 * Australian\ Born + \delta_9 * Male \\ &\quad \delta_{10} * Age + \delta_{11} * Age^2 + error_2 \end{aligned}$$

where the $family$ variables denote which parents were present in the household when the individual was aged 14; $parent's\ labor\ market$ captures whether the individual's mother was employed and whether the father experienced period of unemployment; $siblings$ denotes number of siblings; $state\ of\ school$ and $school\ type$ are dummy variables indicating the region and type of school the individual attended; $Australian$ denotes Australian Born 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. We employ the same exogenous variables for the indices in the conditional variances.

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 average educational attainment. Other variables may also be source of heteroskedasticity. For example, while attendance at Roman Catholic or 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 heteroscedasticity in the wage equation. That is, in some instances the effect of the variable on wages may vary across individual. However, it is also true that 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 do well in the labor market.

The summary statistics for our sample are reported in Table 1. In Table 2 we report, under the heading **School,** the estimates for the schooling equation. The White standard errors 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 half a year of education less than foreign born individuals. Family composition has relatively large and statistically significant effects on the level of schooling. 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 Private schools has a very large and statistically significant positive effects 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 increase educational attainment by 1.4 years as opposed to attendance at a Government School. Attendance at a Catholic school increases educational attainment by .7 years. The regional variables indicate some differences across States noting that the coefficient for the Australian Capital Territory reflects that those attending school there appear to obtain clearly higher levels of education. Finally, males acquire .25 years of schooling less than females.

In column 2 of Table 2 we report the coefficients from the semi-parametric regression of the residuals squared from the schooling equation against the same exogenous variables. As these coefficients involve normalizations and the marginal effects depend on the form of the $\widehat{S}_{vi}$ function, in column 1 of Table 3 we report the marginal effect of each of these variables noting that the marginal effects are for $\widehat{S}_{vi}$ while the model is estimated with $\widehat{S}_{vi}^2$ as the dependent variable.[5] Before we focus on these estimates note that the NR^2 from the regression of the $\widehat{S}_{vi}^2$ on the exogenous variables is 308.58 which 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.

Noting that the average value of the $\widehat{S}_{vi}$ is 2.03 the marginal effects in Column 1 of Table 3 are quite large for several of the variables. For example, attendance at a Private school has a large reduction in the variance. While Australian Born individuals have lower education levels, they also have a lower variance. There is also a large reduction associated with both having both parents present in the household. The remaining variables which appear to have large effects are those capturing the state in which the student attended school recalling that the control group is the most populous state New South Wales.

In Table 4 we now report the estimates for the wage equation. The standard errors are reported in parentheses. First consider the estimation of the OLS equation noting again that we employ age and age squared in place of

[5] The marginal effects for the indicator functions were evaluated by taking the average difference when each individual in the sample is evaluated with and without that particular characteristic. For age and number of the siblings each individual was evaluated at the actual values and then with 1 year of additional education and 1 additional sibling.

the more conventionally employed experience and experience squared. We do this to avoid the potential endogeneity of experience. A number of features are worth noting from the OLS estimates in Table 4. There is some evidence that the background variables have direct influence on the wage level although the statistical evidence is not strong. The number of siblings 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. While both of the school type variables increase the level of education it appears that only the *Catholic* variable has a direct influence on wages. 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. Finally, the point estimate for the education effect is 5.8 percent.

Now focus on the GLS estimates which are reported alongside the OLS estimates in Table 4. Before we focus on the coefficient of primary interest it is valuable to note that the estimates across the two columns for the exogenous variables are generally quite similar. There is some reduction in statistical significance for many coefficients but this does not lead to any drastic reversals in substantive conclusions. The key feature of this column is the estimate of the education coefficient. While the OLS estimate was 5.8 percent we see that the CF estimate is 8.8 percent. Moreover, the coefficient is statistically significant at the 5%. This estimate indicates that the return to education obtained through our procedure is notably higher than that indicated through OLS estimation. We return to a discussion of this below.

The coefficient on the control function is -.16 and this negative coefficient indicates that the unobservables that are correlated with wages are negatively correlated with education. Also, the coefficient is statistically different from zero at the 10 percent level indicating that education is endogenous to wages. This is consistent with the results of Vella and Gregory (1996) who interpreted such a result as a "penalty" to educational over achieving. That is, factors which increased one's education level above what was predicted by their background characteristics received less for their incremental increase in education than those who were predicted to obtain that level. Recall that as the penalty to overachievement is given by $\rho^*_{uv}\frac{S_{ui}}{S_{vi}}$ the penalty is greater for individuals with characteristics associated with lower variances in the schooling equation and higher variances in the wage equation. Also, the contribution of this control function is not small. As the mean value of S_{ui} is .38 the contribution of the control function at the means of these variances is .033. This implies that for every year of over education the individual is penalized by about 3.3 percent.

To further explore this in column 2 of Table 4 we report the estimates from the semi-parametric least squares regression of the wage residuals squared on the exogenous variables. In the final column of this Table we report the marginal effects which are implied by these estimates. 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. As the average of $\widehat{S}_{ui}$ is .38 it is clear that a higher wage variance is associated with many of the individual characteristics and school type. For example, attendance at a Private school greatly increases the variance of wages while males have a more disperse wage distribution. Finally there is variation in the variance of wage in different regions. Note however, it seems that the variability in S_u is more difficult to explain with the conditioning variables than that in S_v. It is also valuable to examine how the ratio $\frac{S_u}{S_v}$ varies in responses to movements in the exogenous variables and this is done in the Table 3. Recalling that the mean of this ratio in the sample is approximately .19 we can see that generally the partial effects in Table 3 are small. The one exception appears to be the Northern Territory state variable which indicates that those in this state who acquire more education than predicted obtain relatively less for their investment than those in other states.

One should note that 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. Accordingly, it is important to consider the underlying factors driving such a result. First of all, however, while the CF estimate appears high it 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 motivation. 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 lead to higher wages through the high returns to education. However, while the individual's wage increase substantially through the increased investment we see that the actual return is somewhat lower 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.

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. That is, since many of them are marginal in the least squares estimation, in terms of statistical significance, 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 to estimate the following model:

$$\begin{aligned} wage_i &= \beta_0 + \sum_{j=1}^{6} \beta_{1j} * state\ of\ school + \sum_{j=1}^{2} \beta_{2j} * school\ type + \\ &\quad \beta_3 * Married + \beta_4 * Australian\ Born + \beta_5 * Male + \\ &\quad \beta_6 * Age + \beta_7 * Age^2 + \theta * schooling + error_3. \end{aligned}$$

The IV estimate for the return to schooling from the above equation, where the background characteristics operate as instruments, produced an estimate of θ of .10 with a standard error of ?. However, if one was to include siblings and mother employed in the wage equation, as is suggested by our empirical results, the point estimate for education decreases to .045 and is not statistically significant from zero. This indicates that while the presence of parents at age 14 affects the education level and not the wage level directly, it is ineffective in identifying the education effect on wages. This may reflect the group affected by the instrument (see Imbens and Angrist 1994). An inspection of the estimates indicates that the key variable is siblings in that when it is employed as an instrument the estimate increases drastically. However, our results indicate this is not a valid instrument. This result highlights the value of our approach in that we do not need to impose such restrictions.

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.

Table 1: Summary Statistics

Variable	**Mean**	**Std**
Hourly wage	20.279	9.663
Years of Education	12.484	2.226
Australian	.777	
Both Parents	.826	
Father Only	.016	
Mother Only	.088	
Siblings	2.741	1.990
Father Unemployed	.095	
Mother Employed	.518	
Catholic School	.154	
Private School	.101	
New South Wales	.312	
Victoria	.241	
Queensland	.209	
South Australia	.089	
Western Australia	.098	
Tasmania	.027	
North Terriorty	.008	
ACT	.022	
Married	.543	
Age	38.467	11.54
Male	.534	

Table 2: Education Equation

Variable	**School**	S_v^2	**Variable**	**School**	S_v^2
Constant	10.038	—	**Tasmania**	-.276	-3.357
	(.309)			(.170)	
Australian	-.517	-3.155	**Nth Terr**	.119	-5.508
	(.073)			(.285)	
Both parents	.743	-2.529	**ACT**	1.265	-3.384
	(.115)			(.183)	
Father Only	.187	1.148	**Married**	.067	-.732
	(.246)			(.065)	
Mother Only	.361	-1.902	**Male**	-.247	-.049
	(.144)			(.056)	
No. of Siblings	-.166	.343	**Age**	.162	1
	(.016)			(.016)	
Father Unemployed	-.186	.051	**Age Squared**	-.002	-.006
	(.095)			(.0002)	
Mother Employed	.089	1.332			
	(.058)				
Catholic School	.706	1.042			
	(.082)				
Private School	1.418	-2.859			
	(.088)				
Victoria	.040	-.989			
	(.079)				
Queensland	-.345	-2.674			
	(.079)				
South Aust.	-.086	-1.148			
	(.107)				
Western Aust.	.057	-1.700			
	(.105)				

Table 3: **Heteroskedasticity Partial Effects**

Variable	$\frac{\partial S_v}{\partial X}$	$\frac{\partial S_u}{\partial X}$	$\frac{\partial (S_u/S_v)}{\partial X}$	Variable	$\frac{\partial S_v}{\partial X}$	$\frac{\partial S_u}{\partial X}$	$\frac{\partial (S_u/S_v)}{\partial X}$
Constant				**Tasmania**	-.165	.006	-.009
Australian	-.161	.005	.003	**Nth Terr**	-.258	.010	.067
Both parents	-.125	.004	.004	**ACT**	-.166	.006	.013
Father Only	.057	-.002	-.013	**Married**	-.036	.001	.015
Mother Only	-.009	.003	.027	**Male**	-.002	0	.026
No. of Siblings	.017	-.001	-.001	**Age**	.049	-.002	-.006
Father Unemployed	.002	0	.004				
Mother Employed	.065	-.002	-.012				
Catholic School	.052	-.002	-.005				
Private School	-.138	.005	.030				
Victoria	-.050	.002	.015				
Queensland	-.133	.004	.010				
South Aust.	-.057	.002	-.003				
Western Aust.	-.085	.003	.018				

Table 4: **Wage Equation**

Variable	OLS	GLS	S_u^2	Variable	OLS	GLS	S_u^2
Constant	1.079	.809		**Tasmania**	-.083	-.059	31.244
	(.072)	(.189)			(.036)	(.039)	
Australian	.005	.026	13.126	**Nth Terr**	-.039	-.134	-43.816
	(.014)	(.016)			(.091)	(.102)	
Both parents	-.012	-.019	8.669	**ACT**	.090	.060	3.691
	(.022)	(.026)			(.039	(.045)	
Father Only	.001	-.001	10.864	**Married**	.042	.045	12.018
	(.048)	(.049)			(.012)	(.012)	
Mother Only	-.014	-.015	-20.382	**Male**	.105	.112	28.970
	(.028	(.030)			(.011)	(.012)	
No. of Siblings	-.011	-.007	-.865	**Age**	.051	.044	
	(.003)	(.004)			(.003)	(.005)	
Father Unemployed	-.009	.002	-4.634	**Age Squared**	-.0005	-.0005	
	(.018)	(.020)			(.0000)	(0)	
Mother Employed	.025	.025	6.214	**School**	.058	.088	
	(.011)	(.011)			(.005)	(.017)	
Catholic School	.054	.031	-.195	ρ		-.163	
	(.015)	(.018)				(.092)	
Private School	.025	-.009	16.806				
	(.019)	(.032)					
Victoria	-.057	-.058	-12.275				
	(.014)	(.015)					
Queensland	-.090	-.076	3.061				
	(.015)	(.015)					
South Aust.	-.118	-.119	9.885				
	(.020)	(.020)					
Western Aust.	-.071	-.070	-10.952				
	(.019)	(.020)					

References

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

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

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

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

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

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

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

[8] Klein, R., and F.Vella (2005): "Estimating a Class of the Triangular Simultaneous Equations Model Without Exclusion Restrictions," unpublished mansuscript.

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

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

[11] 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.

[12] Staiger, R. and J.Stock (1999): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 68, 1055-1096.

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