1 Outline

- Brief Review of Instrumental Variables Methods for Estimating the Returns to Schooling
- Identification via Conditional Second Moments
- Implementation
- Data and Results

2 Instrumental Variables Estimation of the Schooling Effect on Wages

- One of most common application of estimation with endogenous regressors is the estimation of the impact of schooling on earnings
- Partially because it is a very important policy issue
- Also an example where the potential endogeneity of the regressor has a straightforward story
- Huge literature on endogeneity of schooling in wage equations
- Accounting for this endogeneity generally increases schooling coefficient

- Common to account for endogeneity via instrumental variables
- Consider the following model of wages and schooling

$$\begin{aligned} w_i &= X_i\beta + \beta_1 E_i + u_i,\ i = 1..N && (1) \\ E_i &= X_i\delta + Z_i\theta + v_i && (2) \end{aligned}$$

  where w and E denote the level of wage and education respectively;

- X is vector of exogenous variables which affect wages and education;
- Z is a vector of variable(s) which affect education but not wages directly;

- Endogeneity is captured in the moment

$$E[u_i|v_i] \neq 0$$

which indicates that the unobservables that influence education levels are correlated with those influencing wages

- OLS estimates of the parameters from 1 are inconsistent.

- The wage equation parameters are not identifed without additional information.
- This can be incorporated in a number of different forms
- However the essential information being exploited is the following
-
$$E[u_i|X_i] = 0; E[u_i|Z_i] = 0 \tag{3}$$

- Note the imposition of the "identifying" moments is often contentious (i.e. no such variables as z exist).
- Many well known examples.

- While the

$$E[u_i|Z_i] = 0$$

restriction is frequently,often the other requirement, namely

$$cov(E_i, Z_i) \neq 0$$

is more contentious.

- This is the basis for the literature on weak instruments.
- This common occurence of weak instruments is partially what led us to investigate this possibility of employing alternative IV approaches.

- Another commonly approach is based on panel data estimation
- This approach also has an IV interpretation

- Both approaches are very popular but have their respective potential problems

- First, in the conventional IV, natural experiment, approach it is frequently difficult to identify instruments

- Second, in the repeated observations model it is frequently the case that the implicit "thought experiment is not convincing"

- Also, there is frequently a lack of data set or the data set which is available deletes the variables of interest in the necessary transformations

- Estimates based on the conditional first moments possess desirable statistical features when the moment conditions are both "valid" and informative.
- However, increasing concern about inference based on"weak instruments" (see, for example, Staiger and Stock 1997).
- This paper adopts a different approach and focuses on conditional second moments.

- 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.
- 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 either or both equations the education effect is identified.
- We employ such an approach to estimate the returns to schooling for a sample of Australian workers.

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

$$w_i = X_i\beta + \beta_1 E_i + u_i,\ i = 1..N \quad (4)$$
$$E_i = X_i\delta + v_i \quad (5)$$
$$u_i = S_u(X_i)u_i^* \quad (6)$$
$$v_i = S_v(X_i)v_i^* \quad (7)$$

- where w and E denote wage and education respectively; X is vector of exogenous variables; $\beta,\ \gamma$ and δ are unknown parameters

- u and v are heteroskedastic zero mean error terms where u^* and v^* are homoscedastic (unscaled) disturbances with a non zero correlation and the $S's$ denote unknown functions.

- Endogeneity of E occurs through a non-zero correlation between u and v.

- To purge this common component from (4) one can employ the "control function" approach to IV by including v in (4).
- The wage equation error then 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.

- Weighting factor does not vary across i and IV requires an exclusion restriction due to linear mapping.

- Thus even though the correlation coefficient is constant the model requires an exclusion to identify γ.

- Consider where the distribution of u_i and v_i depend on X_i.
- 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.

- 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.

- With known conditional variances it is straight-forward to construct the appropriate control function
- Strong assumptions regarding the conditional variances are "similar" to exclusion restrictions.
- Ideally the CCC assumption would be employed while as little as possible is imposed on the variances.

- To facilitate estimation, we impose the following index restrictions on the variances:

$$S_u^2(X_i) = S_u^2(X_i\alpha) \tag{8}$$
$$S_v^2(X_i) = S_v^2(X_i\gamma) \tag{9}$$

  where the α and γ are unknown parameters. We treat the $S's$ as unknown functions.

- Defining $\widehat{v}_i = E_i - X_i\widehat{\delta}$, we estimated:

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

  where $A(X_i) = \rho^*_{uv}\,[S_{ui}/S_{vi}]$ and $M \equiv [X : E : (S_{ui}/S_{vi})\widehat{v}]$ 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}.

- What error structures do we allow for?
- $$u \equiv S_u(X)u^* and\ v \equiv S_v(X)v^*$$
- The following additive structure for the unscaled errors.
$$\begin{aligned} u^* &= \lambda_o v^* + \varepsilon^*, \\ E(u^*|X) &= E(v^*|X) = 0;\ cov(v^*, \varepsilon^*|X) = 0. \end{aligned}$$
- Then:
$$A(X) = \alpha_o S_u(X)/S_v(X).$$

- 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.
- With ρ_o as the correlation between ε_1 and ε_2 :
- $$A(X) \equiv \left[cov\,(u, v|X)\,/S_v^2\,(X)\right] = \rho_o \left[\frac{S_u(X)}{S_v(X)}\right]$$

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

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

Thus

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

- In this application assume that wages and education each depend on unobserved ability, a^*.

- 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:

$$\begin{aligned} W &: u = a_1(X)\, a^* \varepsilon_1 \\ E &: v = a_2(X)\, a^* \varepsilon_2. \end{aligned}$$

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

- Implementation
- We employ Semiparametric Least-Squares (see Ichimura 1993), to estimate the parameters of the v-index.

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

where $\hat{v}_i^2$ is a consistently estimated squared residual.

- 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:

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

where the $\beta's$ are arbitrary parameter values.

- Define a "variance-type" function:

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

- Replacing the true expectation E above with the nonparametric estimator $\hat{E}$, we obtain the feasible estimator:

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

- We then consider the following controlled nonlinear model:

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

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

- The general approach is somewhat 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.
- Hogan and Rigobon (2004) 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.

- Data and Model Specification
- 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 5070 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 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:

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

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

- Why might heteroskedasticity 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.
- 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.
- While attendance at Roman Catholic or Private schools generally increases educational attainment there is a large degree of heterogeneity across these schools in Australia.
- Similar logic applies to the presence of heteroscedasticity in the wage equation.

- Table 2 reports the estimates for the schooling equation.

Variable	School	Variable	School
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)		

- A number of the individual's background characteristics are important
- Australian Born individuals acquire approximately half a year of education
- Family composition has relatively large and statistically significant effects

- Attendance at Catholic or Private schools has a very large and statistically significant positive effects
- The regional variables indicate some differences across States
- Finally, males acquire .313 years of schooling less than females.
- Strong evidence of heteroskedasticity

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)			

- 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. Thus, not a valid instrument.
- 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.

- Private school variable has a direct influence on wages.
- 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 .060.

- Now focus on the CF estimates
- 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.
- Two important differences are that now siblings and school type not statistically significant. Thus they can be used as instruments.
- The key feature of this column is the estimate of the education coefficient. CF estimate is 10 percent.

- The coefficient for the control function is -.20 and this negative coefficient indicates that the unobservables that are correlated with wages are negatively correlated with education.
- 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.
- 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.
- 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.

- The IV estimate for the return to schooling from the above equation, where the background characteristics operate as instruments, is .102 with a small standard error.
- However, if one was to exclude siblings and mother employed in the wage equation, as is suggested by our empirical results, the point estimate for education decreases to .06.
- 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.

- An inspection of the estimates indicates that the key variable is siblings and school type 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.
- We can also test the identfiying restrictions via this additional moment