

A discusión

A PARAMETRIC CONTROL FUNCTION APPROACH TO ESTIMATING THE RETURNS TO SCHOOLING IN THE ABSENCE OF EXCLUSION RESTRICTIONS: AN APPLICATION TO THE NLSY*

Lidia Farré, Roger Klein and Francis Vella**

WP-AD 2008-16

Corresponding author: L. Farré. Departamento de Fundamentos del Análisis Económico. Universidad de Alicante. Campus de San Vicente, 03080 Alicante, Spain. Phone: +34 965 90 34 00 (ext 3224) / Fax: +34 965 90 38 98 / E-mail: lidiafarre@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Diciembre 2008
Depósito Legal: V-5150-2008

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

* Farré thanks the Spanish Ministry of Education (Grant SEJ 2005-02829/ECON) for financial support.

** L. Farré: Universidad de Alicante. R. Klein: Rutgers University. F. Vella: Georgetown University.

**A PARAMETRIC CONTROL FUNCTION APPROACH TO
ESTIMATING THE RETURNS TO SCHOOLING IN THE
ABSENCE OF EXCLUSION RESTRICTIONS:
AN APPLICATION TO THE NLSY**

Lidia Farré, Roger Klein and Francis Vella

ABSTRACT

We estimate the return to education using a sample drawn from the National Longitudinal Survey of Youth 1979 (NLSY79). Rather than accounting for the endogeneity of schooling through the use of instrumental variables we employ a parametric version of the Klein and Vella (2006a) estimator. This estimator bypasses the need for instruments by exploiting features of the conditional second moments of the errors. As the Klein and Vella (2006a) procedure is semi-parametric it is computationally demanding. We illustrate how to greatly reduce the required computation by parameterizing the second moments. Accounting for endogeneity increases the estimate of the return to education by 5 percentage points, from 7.6% to 12.7%.

JEL Classification: J31, C31.

Keywords: Return to education, heteroskedasticity, endogeneity.

1 Introduction

Perhaps the most commonly explored "treatment effect" in the empirical economics literature is the impact of an individual's educational attainment level on his/her level of earnings. The popularity of these investigations reflects two considerations. First, and most importantly, the implications of human capital investment, at both the individual and aggregate level, are of significant economic interest and importance. Second, the endogeneity of educational choices to wages is clearly understood to bias the OLS estimates of the return to education due to the possibility of reverse causation, unobservable factors and/or measurement error. To account for the endogeneity of education in the estimation of wage equations a number of strategies have been employed. While they are too great in number to allow a detailed description here, they are generally based on instrumental variables estimation (see, for example, Angrist and Krueger 1991, Duflo 2001, Lochner and Moretti 2004, and Carneiro, Heckman and Vytlačil 2005).¹

A feature of the more interesting of the various IV approaches is that they exploit some innovative variation in the conditional mean of the education level which is exogenous to wages. An alternative strategy is to impose restrictions on the conditional second moments. The first paper to employ such a methodology is Vella and Verbeek (1997) who provide a rank order IV procedure. Rummery et al (1999) employ this strategy to estimate the returns to schooling for Australian youth. The rank order IV procedure first allocates observations into different subsets defined by some observed characteristics. Within each of these subsets observations are ordered on the basis of some measure of unobserved heterogeneity responsible for the endogeneity of school-

¹For a detailed survey see Card (1999).

ing. The effect of education on wages is identified by comparing individuals in one subset with the individuals in similar areas of the distribution of the unobserved heterogeneity in other subsets. The rank order IV procedure requires heteroskedasticity in at least one equation and requires that it is not related to the heteroskedasticity in the other equation. Hogan and Rigobon (2002) also study the returns to education and use the identifying moments proposed by Rigobon (1999). Although the Rigobon procedure is a GMM estimator, the approach is similar to rank order IV in that it assumes the heteroskedasticity is a function of a particular variable(s) but that the covariance of the errors across equations is not.

While the Vella and Verbeek (1997) and Rigobon (1999) estimation strategies are attractive in that they provide an identifying source in the absence of exclusion restrictions, their value to empirical work is limited by the limited error structures they can account for. A far more general error structure is allowed for in Klein and Vella, hereafter KV, (2006a) in that the heteroskedasticity in both equations can be functions of the same variables provided the correlation coefficient for the unscaled unobservables in the model is constant. This identification strategy is a potentially useful device for many models in which exclusion restrictions are not available and the assumptions of the alternative heteroskedasticity based estimators are not satisfied.

The identification results in KV (2006a) are based on non parametric and semi parametric representations of the heteroskedasticity. This flexible treatment of the heteroskedasticity is theoretically attractive as it indicates that identification is not reliant on very specific forms of heteroskedasticity. KV (2006a) also provide an estimation strategy which is consistent with this flexibility and this is employed in the simulation evidence in KV (2006a) and the empirical investigation of the returns of

schooling for Australian youth reported in KV (2006b). While the estimation strategy employed in those papers is attractive for its treatment of heteroskedasticity, this lack of structure creates computational demands which complicate estimation. In this paper we parameterize the KV (2006a) estimator thereby making it simpler to implement and thus more readily applied to problems with a large number of explanatory variables. Note however, that while we parameterize the estimator to simplify estimation we rely on the identification results in the more general setting discussed in KV (2006a).

We estimate the return to education using a sample of individuals from the National Longitudinal Survey of Youth 1979 (NLSY79). This survey contains information on individuals living in the US aged 15 to 22 years in 1979. Data on these respondents were annually collected until 1994 and biannually subsequently. We estimate the return to education for the most recent wave of the survey in 2004. These data represent an interesting object of study as they have been used in other empirical investigations of the return to schooling and this allows a comparison of our estimates with those using alternative identifying restrictions. Our results suggest that schooling is endogenous and the adjusted impact of schooling is 12.7% in contrast to the OLS estimate of 7.6%.

In the next section we describe the KV identification strategy and the associated estimation procedure. Section 3 describes the data and estimation results and conclusions follow in Section 4. Concluding comments are provided in Section 5.

2 The Model

2.1 Model and Identification

In presenting the model of interest and to motivate the estimation procedure we closely follow the discussion of KV (2006a). Consider the following triangular model for wages and education:

$$W_i = X_i\beta_0 + \beta_1 E_i + u_i, \quad i = 1, \dots, N \quad (1)$$

$$E_i = X_i\delta_0 + v_i, \quad (2)$$

where W_i and E_i denote the wage and education level of individual i ; and X_i denotes a vector of exogenous variables such that $E[u|X] = E[v|X] = 0$. Endogeneity of E_i arises through the possible correlation between u_i and v_i . This correlation renders the OLS estimates of the β 's inconsistent. As the same X 's appear in (1) and (2), and we impose no restrictions on the parameter vectors β and δ , there are no available instruments.

To identify the model KV assume the presence of heteroskedasticity and impose an additional restriction. More explicitly, let $S_u^2(X)$ and $S_v^2(X)$ denote the conditional variance functions for u and v and assume:

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

where u^* and v^* are homoskedastic error terms. The additional imposed restriction is that the conditional correlation coefficient between these homoskedastic error terms

is constant.² That is:

$$E[u^*v^*] = E[u^*v^*|X] = \rho. \quad (3)$$

KV note that β_1 can be consistently estimated using a control procedure which removes the component of u which is correlated with v . This is done by including a consistent estimate of v in equation (1) making the new error term in (1):

$$\varepsilon = u - \rho v,$$

where $\rho = cov(u, v)/var(v)$. Note, critically, that in the absence of heteroskedasticity ρ is not a function of X . Thus the inclusion of v_i without exclusion restrictions does provide any variation which cannot be fully explained by E and X and the model is not identified. However, KV note that when the distribution of the error terms does depend on X , we can condition on X making the new error term in (1):

$$\varepsilon = u - A(X)v,$$

where $A(X) = \rho_0 S_u(X)/S_v(X)$ and $\rho_0 = [cov(u, v|X)/(S_v(X)S_u(X))]$. $A(X)$ is now a non linear function of X and this non linearity in $A(X)$ is a source of identification provided one can impose the appropriate structure in estimation. KV show that this can be done by imposing (3). This gives the following controlled regression:

$$W_i = X_i\beta_0 + \beta_1 E_i + \rho_0 \frac{S_u(X)}{S_v(X)} v_i + \varepsilon_i, \quad i = 1, \dots, N \quad (4)$$

²KV (2006a) show that this constant conditional correlation assumption is consistent with a number of data generating processes. While the economic implications of the assumption are dependent on the circumstance under investigation it is useful to note that the assumption is generated by a range of processes.

where ε_i is a zero mean error term. Note that the main features of this estimation equation are the following. First, with either or both S_u and S_v non constant the model is identified. Second, identification requires $\frac{S_u(X)}{S_v(X)}$ is not a constant implying that the form of heteroskedasticity must vary across equations. Finally, as both v_i and $S_v(X)$ are straightforward to estimate, the difficulty arises in the estimation of $S_u(X)$.

KV (2006a) show that for several error structures it is possible to consistently estimate $A(X)$. In the return to schooling context it may arise if, for instance, both wages and education depend on unobserved ability a^* . For example, assume the impact of a^* differs in the two equations. Moreover, assume the impact of a^* depends on a component that is a function of X , and a random component. Denote the components dependent on X as $a_1(X)$ and $a_2(X)$, for the wage and education equations respectively, and let ε_1 and ε_2 be the corresponding random components. If we assume that unobserved ability enters the wage and education equations as a multiplicative function of these components we get:

$$u = a_1(X)a^*\varepsilon_1 \text{ and } v = a_2(X)a^*\varepsilon_2.$$

With this form of error structure the appropriate control function has the form in (4).

2.2 Estimation

KV (2006a) provide an estimator for the above model without making any assumptions regarding S_u and S_v . While KV (2006b) employ that proposed estimator the computational difficulties associated with estimating these functions, particularly S_u ,

reduces the attractiveness of the procedure. Accordingly, we now outline how to estimate the model while treating the S functions as known functions but with unknown parameters. To do this we specify the following forms:³

$$S_{ji}^2 = \exp(\alpha_{1j}(Z_{ji}\theta_j)), \quad j = u, v$$

where Z_j is the vector of variables considered to be responsible for the heteroskedasticity in the respective equations.⁴ Although we employ the above functions in estimation it is straightforward to explore alternative forms. We also experimented in the empirical work with:

$$S_{ji}^2 = \exp(\alpha_{1j}(Z_{ji}\theta_j) + \alpha_{2j}(Z_{ji}\theta_j)^2), \quad j = u, v$$

but found that the two approaches gave almost identical estimates for the unknown coefficients in (1).

Given this parameterization of S the estimation procedure we employ is the following:

- i) Regress E on X to get \hat{v} (i.e. a consistent estimate of v).
- ii) Estimate α_{1v} and θ_{jv} through non linear least squares using $\ln(\hat{v}^2)$ as the dependent variable. With these estimates we compute the standard error of the reduced form as $\hat{S}_{vi} = \sqrt{\exp(\hat{\alpha}_{1v}(Z_{vi}\hat{\theta}_v))}$.

³For the sake of exposition we present the specifications of the S'_j s that were used in the empirical work. Note that one could use alternative parameterizations of these functions.

⁴KV (2006a) allow for $X = Z$ and this is the specification employed in KV (2006b). However, while there might be overlap between X and Z it seems reasonable in practice to allow them to differ. Note, however, including variables in Z which do not appear in X is not a source of identification. That is, while we allow for the error distribution to be a function of Z we maintain the assumption that $E[u|Z] = E[v|Z] = 0$.

iii) With these estimates we proceed to the final step. This can be conducted in two ways.

a) First given that we assume a form for S_u we can estimate the model parameters as the solution to following non linear least squares problem:

$$\min_{\beta, \rho, \alpha_{1u}, \theta_u} \sum_{i=1}^N \left(W_i - X_i \beta_0 - \beta_1 E_i - \rho \left(\sqrt{\exp \alpha_{1u}(Z_{ui} \theta_u)} \right) * \frac{\hat{v}_i}{\hat{S}_{vi}} \right)^2. \quad (5)$$

b) While the approach in (a) produces consistent estimates it requires the estimation of S_u through the minimization of a least squares problem related to W . This is somewhat problematic as one is trying to uncover S_u by examining variations in u . An alternative to (a) is to estimate α_{1u} and θ_u in S_u in the similar manner as is done for the education equation. For a given value of β , say β_c , we define the residual $u(\beta_c)$. Using this value of $u(\beta_c)$ we regress $u(\beta_c)^2$ on $Z_{ui} \theta_{cu}$ where we also use candidate values for θ_{cu} . From this regression we compute $\hat{S}_u(\beta_c)$ as $\sqrt{\widehat{\alpha_{1u}}(Z_{ui} \theta_{cu})}$ and estimate ρ_c as:

$$\min_{\rho_c} \sum u(\beta_c) - \rho_c \frac{\hat{S}_u(\beta_c)}{\hat{S}_v} \hat{v}_i.$$

We search over β_c , θ_{cu} and ρ_c to get the final estimates.

While this latter procedure worked very well in this context we found that in general it is useful to employ one additional step. With the final estimates of β , which we denote β_f , from this last optimization problem we define the residual $u_{if} = W_i - X_i \beta_{0f} - \beta_{1f} E_i$. We then use u_{if}^2 to get $\hat{S}_u(\beta_f)$ in precisely the same way as in step (ii) above. Once we have $\hat{S}_u(\beta_f)$ we can regress W_i on X_i , E_i and $\frac{\hat{S}_u(\beta_f)}{\hat{S}_v} \hat{v}_i$ to get the estimates. This final step has the advantage that it separates the estimation of the β' s from the estimation of S_u . Note, however, that in this particular example it

gave almost identical estimates.

3 Results

We estimate the effect of education on earnings using a sample of male and female respondents in the National Longitudinal Survey of Youth (NLSY79). We estimate the return to education on the most recent (2004) wave noting that the respondents are 40 and 47 years old. In the core sample of the survey 4161 individuals satisfy our sample selection criteria.⁵

The NLSY79 is an attractive data source for estimating the return to schooling as it contains detailed family background information and a large array of cognitive ability tests. Card (1999) argues that adding such controls in the wage equation substantially reduces the ability bias in the measured return to education. However, despite the wealth of information contained in the survey it is difficult to find exogenous sources of variation for schooling to employ as instruments. For example, an identification strategy based on changes in the minimum school-leaving age is not valid due to the lack of educational reforms while the sample was enrolled at high school (Oreopoulos 2008).

Some studies have used various proxies of the costs of school attendance (e.g. distance to the nearest school, average local tuition and the local unemployment rate in the area of residence of the respondent at the school going age) to identify the effect of education on earnings. Carneiro and Lee (2006) and Chen (2008) using samples of

⁵The NLSY79 contains 3 subsamples. A core sample aimed to be representative of the US population. A second subsample that contains a disproportionally large percentage of disadvantage non-black and non-hispanic respondents. The third subsample contains individuals in the military service.

male respondents from the NLSY79 obtain IV estimates of the return to education between 13 and 15 percent. While the resulting IV estimates are larger than the OLS estimates, which is consistent with the general consensus regarding the impact of endogeneity, some authors argue against the validity of such instruments due to the non random assignment of households to schools (see, for example, Cameron and Taber 2004).

The extensive set of family background measures collected in the NLSY79 have also been employed as instruments. However Card (1999) argues that IV estimates based on family background characteristics are systematically higher than the corresponding OLS estimates and probably contain a bigger upward ability bias. This is supported by Blackburn and Neumark (1995) which reports an IV estimate (9.6 percent) notably higher than the OLS estimate (4.2 percent).

We now focus on our approach. Our measure of earnings, W , is the log of the hourly wage and our schooling measure, E , is the years of education. The variables contained in X are as shown below. We discuss below our choice of the variables that enter the heteroskedastic index, Z . The model is the following:

$$\ln(wage_i) = \beta_0 + \beta_1 School_i + \beta_2 married_i + \beta_3 NE + \beta_4 W + \beta_5 NC + \beta_6 city + \beta_7 siblings + \beta_8 Mwork14 + \beta_9 Hispanic_i + \beta_{10} Black_i + \beta_{11} male + \beta_{12} Feduc + \beta_{13} Meduc + \beta_{14} South14 + \beta_{15} city14 + \beta_{16} age + \beta_{17} IQ + u_i = \beta X^W + u_i$$

$$School_i = \delta_0 + \delta_1 siblings + \delta_2 Mwork14 + \delta_3 Hispanic_i + \delta_4 Black_i + \delta_5 male + \delta_6 Feduc + \delta_7 Meduc + \delta_8 South14 + \delta_9 city14 + \delta_{10} age + \delta_{11} IQ + v_i = \delta X^S + v_i$$

Table 1 describes the variables employed and Table 2 provides their summary

statistics. Note that the wage equation contains variables, namely the geographical indicators in 2004 and the individual's marital status, which do not appear in the education equation. These variables do not identify the model as IV requires variable(s) in the education equation which do not appear in the wage equation.

Table 3 displays the OLS estimates and standard errors from the estimation of the education model. The estimates are consistent with those in the existing schooling literature. Parental education and ability (as measured by the AFQT standardized by age and gender) have an important positive effect on years of education. In contrast, respondents in larger families accumulate less human capital. There is also evidence of a schooling gap in favor of females. Consistent with Cameron and Heckman (2001) we also find a small positive education gap for the minority groups after controlling for family background.

The KV procedure requires that at least one of the equations' error terms are heteroskedastic. Using the estimates from Table 3 we examine the presence of heteroskedasticity in the schooling equation. The statistic for the White test is 264.21 and that for the Breusch-Pagan, using all the explanatory variables in the model, is 86.75. These values reject the null hypothesis of homoskedastic errors.

We now focus on the estimation of S_v^2 . An examination of the results for the heteroskedasticity tests suggested that the variables responsible for the heteroskedasticity are the Hispanic indicator, some of the geographical indicators and the IQ measure. The regional result is consistent with that of Rummery, Vella and Verbeek (1999) where it is argued that if distance to school influences the likelihood of school attendance, as suggested by Card (1995a), then differences in the distribution of educational institutions within regions can produce also pronounced differences in the

variances of regional educational attainment. The result related to Hispanics captures the heterogenous nature of the group which identifies itself as Hispanic while the IQ effect captures that the level of education varies within individuals who have similar levels of ability. Though we suspect that some of the variables in the schooling model may affect the error variance we do not have strong arguments to exclude others from the heteroskedastic index. Accordingly in estimating the determinants of the conditional variance for the education equation we use all variables which appeared in the conditional mean (i.e. $Z_v = X$).⁶

The non linear least squares estimates of S_v^2 are reported in Table 4. The standard errors alongside the estimated coefficients are calculated from 1000 bootstrap replications with random replacement. Given that we have assumed an exponential form for S_v^2 and that the coefficient on the index is positive and statistically significant, we can directly interpret the sign of the coefficients in the table. The coefficient on the IQ measure is positive and statistically significant. This reflect that more able students have a larger set of educational alternatives and thus the variance of schooling levels increases along the ability distribution. The estimate for the living in the South at age 14 indicator is also statistically significant and negative suggesting a lower dispersion in schooling levels among individuals living in the South of the country during their early teens.

We now turn to the estimation of the wage equation. In addition to the variables which appear in the education equation we include some additional variables, such as the geographical indicators in 2004 and the individual's marital status, which are

⁶We estimated the model for alternative forms of heteroskedasticity. In particular we estimated S_v^2 including a quadratic term for the heteroskedastic index. In an alternative specification we included in the index only the geographic indicators, the IQ measure and the Hispanic indicator. Our main results were unaffected by these alternative specifications.

considered to influence wages. Before considering the adjusted estimates we report the OLS estimates, and their standard errors, in columns (1) and (2) in Table 5. The primary feature of interest in these columns is the estimated impact of education on earnings which is .076. The magnitude of this coefficient is in line with the previously reported OLS estimates of Kane and Rouse (1995), Cameron and Taber (2004) and Chen (2008), which use the NLYS.

In implementing the strategy described above to estimate equation (4) it is necessary to specify the variables entering the index $Z_{ui}\theta_u$. Although we experimented with different choices for the variables in Z_{ui} , including one specification which uses all the variables that enter the conditional mean of the wage, we focus our most detailed discussion on our preferred specification which included only a few variables in the index. To allow for differences in the variance of wages due to economic conditions across regions the index underlying the heteroskedasticity in the wage equation includes the geographic indicators in 2004. We also include the age of the respondent to account for the disparity across individuals in terms of wage growth.⁷

Table 5 presents the estimates of the coefficients in the wage equation obtained from estimating (4) using the method denoted (iiib) in section 2.2. We refer to these figures as CF estimates and they, along with their reported standard errors, are displayed in columns (3) and (4).⁸ Before we focus on the estimated impact of education on wages we highlight a number of the interesting features of this table. First, the estimates for the exogenous variables for the OLS and the CF procedures

⁷The main results are unaffected under alternative specifications of S_u^2 . However when all the exogenous variables in the wage equation enter the heteroskedastic index the coefficients inside the index are erratically estimated.

⁸The reported standard errors for all the parameters estimated in the second step of the CF procedure are based on 1000 bootstrapped replications of the estimator.

are generally quite similar. Both estimates provide evidence of a small marriage premium and a gender differential of almost 30% in favor of males. Some of the regional variables such as the indicator for living in a city and in the North Eastern region in 2004 are positive and statistically significant. Also, the indicator for living in a city at age 14 years is statistically significant and positive while living in the South of the country at age 14 years has a negative effect on the 2004 level of wages. The two specifications also provide evidence of a wage penalty for blacks. Finally there seems to be evidence of an ability premium as captured by the positive and statistically significant coefficient on the ability measure.

The key feature of the columns of this table, however, is the difference in the estimate of the education coefficient. While the OLS estimate was 7.6 percent the CF estimate is 12.7 percent. Moreover while there is some loss in statistical significance, in comparison to the OLS estimate, the coefficient is statistically significant at conventional levels of testing. Finally the estimate of the correlation coefficient, ρ , is negative and statistically significant, indicating that education is clearly not exogenous.

Our results suggest, as is frequently found in this literature, that the OLS estimate tends to be below the estimate obtained after controlling for the endogeneity of education. This may reflect the sizeable measurement error in the education variable or the heterogeneity in the return to education in the population (see for example Angrist and Krueger, 1991; Card 1995a and 1999; Harmon and Walker, 1995; Kling 2001 and Cameron and Taber 2004). While measurement error is partially responsible for the downward bias in the OLS estimate the negative sign of the estimated correlation coefficient, -0.190 , may also capture a negative relationship between unobserved fac-

tors across equations. For example, workers with higher levels of earnings ability have a higher opportunity cost of attending school and might leave school sooner to take up a job. The OLS estimate may also be subject to discount-rate bias if individuals who leave school earlier have higher returns to schooling but discount their future earnings more than individuals who stay longer (see Lang, 1993 and Card, 1994). Accordingly the observed average return to education is lower than the true value. Thus our findings are in line with the results in previous studies. Our estimate of the return to education falls within the range of estimates reported in the surveys by Card (1999, 2001), where most estimates of the return to schooling after adjusting for the endogeneity of education are between 8 percent and 13 percent per school year.

The non linear least squares estimates of S_u^2 and corresponding standard errors are in Table 6. These estimates indicates that the variance of wages is significantly larger in the Western states of the country. The other variables included in the index do not seem to affect the variance of the unobservables in the model. Note that the absence of heteroskedasticity in the wage equation does not threaten our identification strategy as this requires heteroskedasticity in either equation.

4 Conclusions

This paper uses a parametric version of the Klein and Vella (2006a) control function estimator for triangular systems with no exclusion restrictions to study the impact of endogenous schooling levels on wages. In this particular setting there is sufficient heteroskedasticity to identify the schooling effect and the identifying restriction appears reasonable. The results suggest that schooling is endogenous and the adjusted impact of schooling is 12.7 percent in contrast to the OLS estimate of 7.6 percent.

References

- [1] Lochner, L. and E. Moretti (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, 94: 155-89.
- [2] Angrist J. and A.B. Krueger (1991): "Does compulsory school attendance affect schooling and earnings." *Quarterly Journal of Economics*, 106: 979-1014.
- [3] Blackburn, M.L. and D. Neumark (1995): "Are OLS estimates of the return to schooling biased downward? Another Look." *The Review of Economics and Statistics*. Vol. 77, No. 2 (May, 1995): 217-30.
- [4] Cameron S. and J. Heckman (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy*, Vol 109, No. 3: 455-99.
- [5] Cameron S. and C. Taber (2004): "Estimation of Educational Borrowing Constraints Using Returns to Schooling." *Journal of Political Economy* 112: 132-82
- [6] Card, D. (1995a): "Using Geographic Variation in College Proximity to Estimate the Returns to Schooling." in *Aspects of Labour Market Behavior: Essays in Honor of John Vanderkamp*, eds. L.N: Christofiedes et al. Toronto: University of Toronto Press: 201-21.
- [7] Card, D. (1995b): "Earnings, Schooling and Ability Revisited." in Solomon W. Polachek (Eds), *Research in Labor Economics* (Greenwich, Connecticut: JAI Press, 1995): 23-48.
- [8] Card, D. (1999): "The causal effect of education on earnings." in Ashenfetler, O. and Card, D. (eds.), *Handbook of Labor Economics*, vol 3A, chap. 30, Amsterdam: Elsevier Science/North-Holland.
- [9] Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (September): 1127-60.
- [10] Carneiro, P., H. Heckman and E. Vytlacil (2005): "Understanding what Instrumental Variables estimate: estimating average and marginal returns to schooling." working paper, The University of Chicago.
- [11] Carneiro, P. and S. Lee (2006): "Changes in College Enrollment and Wage Inequality: Distinguishing Price and Composition Effects." UCL Working Paper

- [12] Chen, S.H. (2008): "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity." *The Review of Economics and Statistics*, forthcoming
- [13] Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91(4):795-813.
- [14] Harmon, C. and I. Walker (1995): "Estimates of the economic return to schooling for the United Kingdom." *American Economic Review* 85: 1278-86.
- [15] Hogan, V. and R. Rigobon (2002): "Using Heteroscedasticity to Estimate the Returns to Education." NBER Working Papers 9145.
- [16] Kane, T. and C. Rouse (1995): "Labor-Market Returns to Two- and Four-Year College." *American Economic Review*, 85(3): 600-14.
- [17] Klein, R. and F. Vella (2006a): "Estimating a Class of Triangular Simultaneous Equations Models Without Exclusion Restrictions." IZA working paper 2378.
- [18] Klein, R. and F. Vella (2006b): "Estimating the Return to Endogenous Schooling Decisions for Australian Workers via Conditional Second Moment." IZA working paper 2407.
- [19] Kling, Jeffrey R. (2001): "Interpreting Instrumental Variable Estimates of the Returns to Schooling." *Journal of Business and Economic Statistics* 19: 358-64.
- [20] Lang, Kevin (1993): "Ability Bias, Discount Rate Bias, and the Return to Education." Mimeo, Boston University, 1993.
- [21] Oreopoulos, P. (2008): "Should We Raise the Minimum School Leaving Age to Help Disadvantaged Youth? Evidence from Recent Changes to Compulsory Schooling in the United States." in "An Economic Framework for Understanding and Assisting Disadvantaged Youth", Jonathan Gruber (ed), NBER.
- [22] Rigobon, R. (1999): "Identification through heteroskedasticity", *Review of Economics and Statistics*, 85, 777-92.
- [23] Rubb, S. (2002): "Overeducation in the labor market: a comment and re-analysis of a meta-analysis." *Economics of Education Review*, Vol 22 (6): 621-629.
- [24] Rummery, S., F. Vella and M. Verbeek (1999): "Estimating the Returns to Education for Australian Youth via Rank-Order Instrumental Variables." *Labor Economics*, 6: 777-92.

- [25] Vella, F. and R.G. Gregory (1996): "Selection Bias and Human Capital Investment: Estimating the rates of Return to Education for Young Males." *Labour Economics*, 1996 Vol 3: 197-219.
- [26] Vella, F. and M. Verbeek (1997): "Using rank order as an instrumental variable: an application to the return to schooling." CES Discussion Paper 97.10, K.U. Leuven.

Appendix

Table 1: Variable definition:

$\ln(wage)$	ln of hourly wage
<i>school</i>	years of education completed
<i>married</i>	indicator for being married in 2004
<i>NE</i>	indicator for living in a North Eastern state in 2004
<i>West</i>	indicator for living in a Western state in 2004
<i>NC</i>	indicator for living in a North Central state in 2004
<i>city</i>	indicator for living in a city in 2004
<i>siblings</i>	number of siblings
<i>Mwork14</i>	indicator for whether the mother of i works when i is 14
<i>Hispanic</i>	indicator for being Hispanic
<i>Black</i>	indicator for being Black
<i>male</i>	indicator for being male
<i>Feduc</i>	years of education completed by the father
<i>Meduc</i>	years of education completed by the mother
<i>S14</i>	indicator for living in the South at age 14
<i>city14</i>	indicator for living in a city at age 14
<i>Age</i>	age of i
<i>IQ</i>	Score obtained in the Armed Forces Qualifying Test (a measure of cognitive ability)

Table 2: Summary Statistics:

	mean	S.D.
$\ln(wage)$	2.767	0.662
<i>school</i>	13.515	2.347
<i>married</i>	0.593	
<i>NE</i>	0.161	
<i>West</i>	0.265	
<i>NC</i>	0.186	
<i>city</i>	0.746	
<i>siblings</i>	3.563	2.496
<i>Mwork14</i>	0.564	
<i>Hispanic</i>	0.170	
<i>Black</i>	0.265	
<i>male</i>	0.500	
<i>Feduc</i>	11.129	3.026
<i>Meduc</i>	11.085	3.829
<i>S14</i>	0.342	
<i>city14</i>	0.795	
<i>Age</i>	43.255	2.183
<i>IQ</i>	0.195	0.915
<i>Nobs</i>	4161	

Table 3: Schooling Equation (OLS estimates):

	estimates	S.D.
<i>siblings</i>	−0.049	(0.013)
<i>Mwork14</i>	0.021	(0.060)
<i>Hispanic</i>	0.737	(0.093)
<i>Black</i>	1.168	(0.084)
<i>male</i>	−0.261	(0.059)
<i>Feduc</i>	0.063	(0.011)
<i>Meduc</i>	0.083	(0.014)
<i>S14</i>	0.114	(0.065)
<i>city14</i>	0.056	(0.075)
<i>age</i>	0.009	(0.013)
<i>IQ</i>	1.302	(0.037)
constant	11.270	(0.606)
Test for Heteroskedasticity (statistics)		
White	263.21	
Breush-Pagan	86.75	

Table 4: Estimating the form of heteroskedasticity (Education)

	estimates	S.D.*
<i>siblings</i>	−0.020	(0.021)
<i>Mwork14</i>	0.101	(0.092)
<i>Hispanic</i>	0.043	(0.158)
<i>Black</i>	0.023	(0.133)
<i>male</i>	0.108	(0.091)
<i>Feduc</i>	−0.024	(0.016)
<i>Meduc</i>	−0.029	(0.022)
<i>S14</i>	−0.187	(0.093)
<i>city14</i>	0.115	(0.105)
<i>age</i>	0.020	(0.019)
<i>IQ</i>	0.670	(0.064)
<i>constant</i>	−0.332	(0.787)
<i>alpha0</i>	−0.043	(0.067)
<i>alpha1</i>	0.801	(0.041)

*The standard errors are obtained from 1000 bootstrap replications with random replacement

Table 5: Wage Equation (OLS and CF estimates)

	OLS	S.D.	CF	S.D.*
<i>married</i>	0.087	(0.018)	0.087	(0.020)
<i>NE</i>	0.070	(0.034)	0.069	(0.034)
<i>West</i>	0.050	(0.033)	0.051	(0.034)
<i>NC</i>	-0.046	(0.032)	-0.045	(0.030)
<i>City</i>	0.037	(0.022)	0.038	(0.020)
<i>siblings</i>	0.001	(0.003)	0.004	(0.004)
<i>Mwork14</i>	0.019	(0.018)	0.019	(0.018)
<i>Hispanic</i>	0.072	(0.030)	0.033	(0.033)
<i>Black</i>	-0.043	(0.027)	-0.103	(0.037)
<i>male</i>	0.274	(0.018)	0.285	(0.018)
<i>Feduc</i>	0.001	(0.006)	-0.003	(0.004)
<i>Meduc</i>	0.009	(0.004)	0.005	(0.005)
<i>S14</i>	-0.051	(0.030)	-0.058	(0.028)
<i>city14</i>	0.052	(0.023)	0.047	(0.023)
<i>age</i>	0.006	(0.004)	0.006	(0.004)
<i>IQ</i>	0.146	(0.013)	0.075	(0.029)
<i>educ</i>	0.076	(0.005)	0.127	(0.020)
ρ_0			-0.190	(0.066)
<i>constant</i>	1.112	(0.190)	0.532	(0.289)

*The standard errors are obtained from 1000 bootstrap replications with random replacement

Table 6: Estimating the form of heteroskedasticity (Wages)

	estimates	S.D.*
<i>age</i>	0.016	(0.018)
<i>NE</i>	0.187	(0.119)
<i>West</i>	0.278	(0.108)
<i>NC</i>	0.080	(0.100)
constant	−3.227	(0.761)
alpha0	−0.359	(0.024)
alpha1	0.919	(0.008)

*The standard errors are obtained from 1000 bootstrap replications with random replacement