A Parametric Control Function Approach to Estimating the Returns to Schooling in the Absence of Exclusion Restrictions: An Application to the NLSY*

Lídia Farré Roger Klein Francis Vella

Abstract

An innovation which bypasses the need for instruments when estimating endogenous treatment effects is identification via conditional second moments. The most general of these approaches is Klein and Vella (2010) which models the conditional variances semiparametrically. While this is attractive, as identification is not reliant on parametric assumptions for variances, the nonparametric aspect of the estimation may discourage practitioners from its use. This paper outlines how the estimator can be implemented parametrically. The use of parametric assumptions is accompanied by a large reduction in computational and programming demands. We illustrate the approach by estimating the return to education using a sample drawn from the National Longitudinal Survey of Youth 1979. Accounting for endogeneity increases the estimate of the return to education from 6.8% to 11.2%.

JEL: J31, C31

Keywords: return to education, heteroskedasticity, endogeneity

*L. Farré, Institut d'Anàlisi Econòmica, Campus UAB, 08193, Bellaterra, Barcelona (Spain), email: lidia.farre@iae.csic.es. R. Klein, Rutgers University, email: klein@economics.rutgers.edu. F. Vella, Georgetown Universisty, email: fgv@georgetown.eud. We are grateful to the editor and two anonymous referees for helpful comments. Farré acknowledge the support of the Barcelona GSE Research Network and the Government of Catalonia, and the financial support from the Spanish Ministry of Science (grant ECO2008-05721/ECON).

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 earnings. The popularity of these investigations reflects two considerations. First, 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 understood to bias the OLS estimates of the return to education. This endogeneity is typically attributed to factors such as reverse causation, the correlation between the unobservable factors which determine education level and wages, and/or the presence of 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 (IV) estimation (see, for example, Angrist and Krueger 1991, Duflo 2001 and Heckman, Urzua and Vytlacil 2006).[1]

A feature of these various IV approaches is that they exploit the existence of a variable(s) which is responsible for some variation in the conditional mean of the education level, but is exogenous and unrelated to wages. While this typically involves the use of an exclusion restriction other techniques, such as those based on the comparison of twins and siblings, also have an IV interpretation. An alternative strategy is to exploit the variation in the conditional error variances while imposing restrictions on other conditional second moments. The first paper to employ such a methodology is Vella and Verbeek (1997) who provide a rank order IV procedure.

[1] For a detailed survey see Card (1999).

Rummery et al (1999) use this strategy to estimate the return to schooling for Australian youth. This procedure assigns observations into different subsets on the basis of some observed characteristic. Within each subset the observations are then ranked by their value of the reduced form education residual as this represents a measure of the unobserved heterogeneity responsible for the endogeneity of schooling. The effect of education on wages is then identified by comparing the wages and education of individuals in one subset with those of individuals of similar rank in the other subsets. This method identifies the schooling effect provided there is heteroskedasticity in at least one equation and that the heteroskedasticity in that equation is not correlated with the heteroskedasticity in the other. Hogan and Rigobon (2002) also study the return to education and employ the approach proposed by Rigobon (1999). Although the Rigobon procedure is based on the method of moments estimation, the identifying restriction imposed is similar to rank order IV in that it assumes the heteroskedasticity in one of the equations is a function of a particular variable(s) but that the covariance of the errors across equations is not.

The Vella and Verbeek (1997) and Rigobon (1999) strategies provide an identifying source in the absence of exclusion restrictions. However, their value to empirical work is limited due to the nature of the error structures they allow. Klein and Vella (2010), hereafter KV(2010), provide an estimator for a more general error structure that allows the heteroskedasticity in both equations to be functions of the same variables provided the correlation coefficient for the underlying homoskedastic error terms across equations is constant. This is potentially useful 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(2010) are based on nonparametric and semiparametric representations of the heteroskedasticity and this is theoretically appealing as the results are not reliant on specific forms of the heteroskedasticity. KV(2010) also provide an estimation strategy which is employed in their simulation evidence and in the empirical investigation of the return to schooling reported in Klein and Vella (2009), hereafter KV(2009).

Due to the nonparametric nature of the KV(2010) estimator the programming and computation requirements are demanding. The contribution of this paper is to adapt the estimator to a parametric setting thereby making it more easily to implement. We illustrate how the components of the model treated nonparametrically in KV(2010) can be parameterized in a flexible manner. We also outline an appropriate estimation procedure. We stress that the objective of this paper is implementation. We do not provide any new theoretical results and rely on the KV(2010) identification results in the more general setting.

We illustrate the procedure by estimating the return to education using a sample of individuals from the 2004 wave of the National Longitudinal Survey of Youth 1979 (NLSY79). This survey contains information on individuals living in the US aged 14 to 22 years in 1979. 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 strategies. Our results suggest that schooling is endogenous and the adjusted impact of one additional year of schooling on wages is 11.2% in contrast to the OLS estimate of 6.8%. This result is consistent with the range of estimates which have been obtained for these data using IV techniques.

In the next section we discuss the KV(2010) identification strategy in the return to schooling context. We discuss the features of the data exploited in estimation and assign them an economic interpretation in the return to education framework. We also outline how a parametric version of this procedure can be implemented. Section 3 presents the data and estimation results. Some concluding comments are presented in Section 4.

2 The Model

2.1 Model and Identification

We begin by recasting the KV(2010) approach in the returns to schooling framework. We state the assumptions of the model and the implications of the identifying restriction. The model has the following triangular form:

$$w_i = x_i\beta_0 + \beta_1 educ_i + u_i, \ i = 1, ..., n \tag{1}$$

$$educ_i = x_i\delta_0 + v_i, \tag{2}$$

where w_i and $educ_i$ denote the wage and the years of education of individual i; and x_i denotes a vector of exogenous variables such that $E[u_i|x_i] = E[v_i|x_i] = 0$. The endogeneity of $educ_i$ arises through the possible correlation between the error terms u_i and v_i which renders the OLS estimates of the $\beta's$ inconsistent. The same $x_i's$ appear in (1) and (2), and we impose no restrictions on the parameter vectors β and δ. Accordingly there are no available instruments to estimate (1).

To identify the model we assume the presence of heteroskedasticity and impose an additional restriction on the correlation between the error terms. More explicitly, let $S_u^2(x_i)$ and $S_v^2(x_i)$ denote the conditional variance functions for u_i and v_i and assume:

$$u_i = S_u(x_i)u_i^* \text{ and } \ v_i = S_v(x_i)v_i^*, \tag{3}$$

where u_i^* and v_i^* are homoskedastic error terms. Moreover, assume that either $S_u(x_i)$ and/or $S_v(x_i)$ are not constant and the ratio $S_u(x_i)/S_v(x_i)$ is not constant across i. The key identifying restriction is that the conditional correlation coefficient between the homoskedastic error terms, u_i^* and v_i^* is constant.[2] That is:

$$E[u_i^* v_i^*] = E[u_i^* v_i^* | x_i] = \rho_0. \tag{4}$$

Before focusing on how (3) and (4) can be combined to identify the parameters in (1) consider the interpretation of this error structure in the wage/schooling setting. It seems reasonable to view u_i^* and v_i^* as correlated measures of unobserved ability. Thus equation (3) suggests that the contribution of unobserved ability to wages and schooling depends on the individual's socioeconomic factors. These factors may be time invariant or capture considerations which may evolve over the individual's life cycle (see, for example, Cunha and Heckman 2007). For instance, the effect of an individual's unobserved ability on wages and schooling might be influenced by the type of school he/she attended and his/her family background. KV(2009) argue that when distance to school is the sole determinant of educational attainment the location of schools within a region would produce not only differences in the average regional

[2] KV (2010) 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.

level of education but also in the regional variance in educational attainment. To illustrate this they consider two regions, one where all the individuals are equidistant from schools and another where all individuals are at varying distances from the nearest school. In this case it is not clear which region will have the highest average attendance but clearly in the region where all the individuals are equidistant the variance will be lower. KV(2009) argue that heteroskedasticity may also arise from other factors. For example, having a working mother may increase educational attainment for some children as it may reflect a positive family attitude towards professional achievement. In contrast, it may also reflect that the family has income concerns and the mother needs to work for pecuniary reasons. Including an indicator for whether the individual's mother works in the schooling equation will capture the average effect of having a working mother but clearly there will be some variance in this effect.

A similar logic underlies the presence of heteroskedasticity in the wage equation. As workers obtain work experience it is possible that some perform better than others in term of wage growth. This introduces heteroskedasticity as a function of the individual's age. Moreover, regional variables may reflect factors like the cost of living and while these variables capture the average effect of living in a city, it is likely that the heteroskedasticity in the wage residuals will also be related to these variables due to different living expenses within the same region.

While the presence of heteroskedasticity is largely an empirical issue the imposition of (4) is a restriction with economic implications. As u_i^* captures the level of unobserved ability in the schooling equation and v_i^* in the wage equation, the assumption regarding the constant value of ρ_0 indicates that after conditioning out the role of the $x_i's$ the return to unobserved ability is constant. Note that this does not imply that unobserved ability is not rewarded differently, in either the wage or schooling equations, depending on the individual's socioeconomic background. In fact, this is the process which is captured in (3) and provides the variation which identifies the return to education. The restriction in (4) indicates that once the influence of these socioeconomic factors is accounted for the return to unobserved ability is constant. This seems a reasonable assumption.

The assumption that these homoskedastic error terms capture unobserved ability suggests that ρ_0 should be positive. However a common finding is that the OLS estimate of the return to education is smaller than the IV estimate and this is only compatible with a negative correlation between the error terms. KV(2009) propose an error structure that satisfies the identification assumption in (4) and is consistent with a negative estimate of ρ_0 even in the presence of ability positively affecting both earnings and schooling. They argue that the effect of ability on wages and schooling has a predictable and an unpredictable component. That is:

$$u_i = a_u(x_i)a_i^* e_i^u \text{ and } \ v_i = a_v(x_i)a_i^* e_i^v, \tag{5}$$

where $a_u(x_i)$ and $a_v(x_i)$ capture the predictable impact while e_i^u and e_i^v the unpredictable. In this specification ρ_0 measures the correlation between e_i^u and e_i^v. This error structure allows ability to positively affect wages and schooling, while its interaction with the e's may produce a negative correlation across equations. To consider how a negative value of ρ_0 might arise we follow the interpretation of Vella and Gregory (1996) who assume that e_i captures some unobservable factors capturing "over achievement". Accordingly the impact of e_i^v on schooling is positive while that of e_i^u on wages is negative. This is consistent with the empirical finding that the return to over education is lower than the average return to education (Dolton and Vignoles, 2000; Groot and van den Brink, 2000 and Rubb, 2002). As it is not possible to identify an individual's level of over education, empirical models assume a constant rate of return on the total years of education. Under this specification the wage equation error term captures the "over education penalty". Note that while this situation produces a negative estimate of ρ_0, it does not imply that the wage decreases as the level of e_i^u increases as a higher value of e_i^u is associated with a higher wage resulting from the rise in education.

Finally, it is important to remark on some features, which are of potential empirical interest, that are not incorporated in the model in equations (1)-(4). The first feature it does not allow for is the heterogeneity in the returns to education. This type of heterogeneity is captured in the random coefficient models of Heckman and Vytlacil (1998) and Wooldridge (2003). In those models the returns to education are assumed to be a function of observed and unobserved characteristics that affect schooling choices.[3] Using instrumental variables they obtain consistent estimates of the average return to education assuming a constant correlation coefficient between the unobserved factors affecting education and its returns. Our specification is related to this model as the return to education is constant but there is heterogeneity in the returns to unobserved factors. For models where the returns to education are heterogenous it is not immediately clear which parameter our procedure estimates.[4] A second restrictive feature of (1)-(4) is that education enters the wage equation

[3] Note that our model allows the return to education to depend on observed covariates. Column (3) in Table A2 investigates this possibility in our sample.

[4] In models where the heterogeneity in the return to education is a function of observable characteristics it may be possible to adapt our estimator by explicitly modelling the process. However, when the heterogeneity is due to unobservables it is not clear how one would adjust our procedure nor interpret the estimated parameter.

linearly. An alternative specification would be to allow for a non linear relationship between wages and schooling, but this would represent a substantial departure from our setting. One possibility would be to extend the nonparametric instrumental variables procedure of Newey et al (1999). However, including non linear education terms requires including additional control terms to accounts for their endogeneity and it is not clear how this could be incorporated in the present estimation strategy.

We now discuss the strategy employed to estimate the model in (1)-(4). KV(2010) note that the unknown parameters in the wage equation can be consistently estimated using a control procedure which removes the component of u_i which is correlated with v_i. This is done by including a consistent estimate of v_i in equation (1) and making the new error term in (1):

$$\varepsilon_i = u_i - \lambda v_i,$$

where $\lambda = cov(u_i, v_i)/var(v_i)$. Note, critically, that in the absence of heteroskedasticity λ is not a function of x_i. Thus the inclusion of v_i in (1) in the absence of exclusion restrictions does not provide any variation which can not be fully explained by $educ_i$ and x_i, and the model is not identified. However, KV(2010) note that when the distribution of the error terms does depend on x_i, we can condition on x_i making the new error term in (1):

$$\varepsilon_i = u_i - A(x_i)v_i,$$

where $A(x_i) = \rho_0 S_u(x_i)/S_v(x_i)$ and $\rho_0 = [cov(u_i, v_i|x_i)/(S_v(x_i)S_u(x_i))]$. $A(x_i)$ is now a non linear function of x_i and this non linearity in $A(x_i)$ is a source of identification provided one can impose the appropriate structure in estimation. KV(2010) show that this can be done by imposing (4). Accordingly the parameters in (1) can be estimated from the following controlled regression:

$$w_i = x_i\beta_0 + \beta_1 educ_i + \rho_0 \frac{S_u(x_i)}{S_v(x_i)} v_i + \varepsilon_i,\ i = 1, ..., n \tag{6}$$

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 $S_u(x_i)$ and $S_v(x_i)$ non constant the model is identified. Second, identification requires $\frac{S_u(x_i)}{S_v(x_i)}$ is not constant implying that the form of heteroskedasticity must vary across equations. Finally, as both v_i and $S_v(x_i)$ are straightforward to estimate, the difficulty arises in the estimation of $S_u(x_i)$.

2.2 Empirical Strategy

KV(2010) provide an estimator for the above model without imposing any structure on $S_u(x_i)$ and $S_v(x_i)$. While KV(2009) employ that proposed estimator in their empirical investigation the computational difficulties associated with estimating these functions, particularly $S_u(x_i)$, may discourage the use of the procedure. Before proceeding to the parametric version it is useful to outline the KV(2010) estimation process and highlight where the computational demands arise.[5]

As is highlighted below the approach in KV(2010) requires the sequential estimation of v_i and $S_v(x_i)$ and then the joint estimation of the $\beta' s$ and $S_u(x_i)$ using the estimated values of v_i and $S_v(x_i)$. As v_i is the reduced form error, its estimate, $\widehat{v_i}$, comes directly from the OLS estimate of the education equation. The estimation of $S_v(x_i)$ is obtained as the square root of the nonparametric estimate of $E[\widehat{v}_i^2|x_i]$

[5] A formal description of the Klein and Vella (2010) estimator is provided in the Appendix.

noting this is further simplified by imposing that the $x_i's$ enter the $S_u(x_i)$ and $S_v(x_i)$ functions in index form. Thus, obtaining estimates of v_i and $S_v(x_i)$ is straightforward in the semiparametric case.

The computational demands associated with the KV(2010) procedure arise in the estimation of the main equation. First, the nonparametric nature of the $S_u(x_i)$ function and the non linearity inherent in estimating the parameters in (6) requires estimating this nonparametric function multiple times in each round of each iteration of the optimization problem. This is exacerbated by two additional issues. First, the identification proof in the semiparametric case requires that the optimization problem generating the wage equation comprises two criterion functions which must be separately and jointly minimized. One of these functions involves conditioning on the index which enters the $S_u(x_i)$ function while the other requires jointly conditioning on both the index that enters $S_u(x_i)$ and also that which enters $S_v(x_i)$. The presence of these nonparametric expectations not only increases computation but also requires the use of the appropriate forms of bias reduction. KV(2010) employ two bias reduction methods. First, rather than use higher order kernels, which are known to have poor finite sample behavior, they employ local smoothing. This requires that the bandwidth in the employed kernel is able to vary depending on where in the data the density is estimated. Second, KV(2010) employ a two step procedure where in the first step the objective function is trimmed on the basis of the $x_i's$ to get initial estimates. Using these consistent estimates one constructs the estimated indices which appear in the $S_u(x_i)$ and $S_v(x_i)$ functions and the objective function is trimmed again on the basis of these indices. The use of local smoothing increases computation as it requires estimates of the pilot densities to estimate the varying bandwidth. The use of the x_i and index trimming means that each step has to be estimated twice.

While the above estimation approach is feasible, and implemented in both KV(2010) for simulation evidence and in KV(2009) for empirical evidence, it is likely that the required computation and programming discourages practitioners from adopting the semiparametric approach. Accordingly, we now outline how to estimate the model while treating both $S_u(x_i)$ and $S_v(x_i)$ as known functions of an index with unknown parameters. This reduces the degree of computation for several reasons. First, the need to estimate nonparametric functions is eliminated. Second, the absence of these unknown functions eliminates the need to condition on two indices simultaneously. Finally, the absence of the nonparametric estimation means we do not require bias reducing methods and the multiple steps. These considerations combine to greatly reduce the computational burden.

To bypass the semiparametric estimation we specify the following form for the heteroskedastic functions:

$$S_{ji}^2 = \exp(z_{ji}\theta_j), \qquad j = u, v, \tag{7}$$

where z_{ji} is the vector of variables considered to produce the heteroskedasticity in the respective equations[6] and θ_j is a vector of unknown parameters. In what follows we will refer to $z_{ji}\theta_j$ as the heteroskedastic index. Although we employ the above functions in estimation it is straightforward to explore alternative forms. In the next section we investigate the sensitivity of our estimates of primary interest to different

[6]KV(2010) allow $x_i = z_i$ and this is the specification employed in KV(2009) although it seems reasonable to allow them to differ in practice. Note that including variables in z_i which do not appear in x_i does not identify the education coefficient in the same way as excluded variables can be used as an instruments. However, they do provide a source of identifying power in as much they are able to explain movements in the variances.

parameterizations of the heteroskedasticity.

Given the parameterization of S_{ji}^2 in (7) the estimation procedure is the following:

i) Regress $educ_i$ on x_i to obtain a consistent estimate of the residual which we denote $\widehat{v}_i$.

ii) Estimate θ_v through non linear least squares using $ln(\widehat{v}_i^{\,2})$ as the dependent variable. Compute the standard deviation of the reduced form error as $\widehat{S}_{vi}=\sqrt{\exp(z_{vi}\widehat{\theta}_v)}$.

iii) Using $\widehat{v}_i$ and $\widehat{S}_{vi}$ it is possible to estimate the wage equation parameters in two ways.

a) First given an assumed form for $S_u(x_i)$ estimate the model parameters as the solution to the following non linear least squares problem:

$$\min_{\beta,\rho_0,\alpha_{1u},\theta_u} \sum_{i=1}^{n} \left(w_i - x_i\beta_0 - \beta_1 educ_i - \rho_0 \left(\sqrt{\exp(z_{ui}\theta_u)} \right) * \frac{\widehat{v}_i}{\widehat{S}_{vi}} \right)^2 . \tag{8}$$

b) While the approach in (a) produces consistent estimates it requires the estimation of $S_u(x_i)$ through the minimization of a least squares problem related to w_i. This requires uncovering $S_u(x_i)$ though the observed variation in $\widehat{u}_i$. An alternative to (a) is to estimate θ_u in $S_u(x_i)$ in the similar manner as is done for the education equation. For a given value of β, say β_c, we define the residual $u_i(\beta_c)$. Using this value of $u_i(\beta_c)$ we regress $ln(u_i(\beta_c)^2)$ on $z_{ui}\theta_{cu}$ where we also use candidate values for θ_{cu}. From this regression we compute $\hat{S}_{ui}(\beta_c)$ as $\sqrt{\exp(z_{ui}\theta_{cu})}$ and estimate ρ_{0c} as:

$$\min_{\rho_{0c}} \sum_{i=1}^{n} \left(u_i(\beta_c) - \rho_{0c} \frac{\hat{S}_{ui}(\beta_c)}{\widehat{S}_{vi}} \widehat{v}_i \right)^2 . \tag{9}$$

The final estimates of β_c, θ_{cu} and ρ_{0c} are those that minimize (9) and are obtained through a standard iterative procedure.

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 $u_{if} = w_i - x_i\beta_{0f} - \beta_{1f}educ_i$. We then use u_{if}^2 to get $\hat{S}_{ui}(\beta_f)$ in precisely the same way as in step (ii) above. Once we have $\hat{S}_{ui}(\beta_f)$ we can regress w_i on x_i, $educ_i$ and $\frac{\hat{S}_{ui}(\beta_f)}{\widehat{S}_{vi}}\widehat{v}_i$ to get the estimates. This final step has the advantage that it separates the estimation of the $\beta's$ from the estimation of S_u.

To evaluate the performance of this approach in a controlled setting we applied it to the simulated data in KV(2010). Note that in obtaining the estimates we specify the correct parametric form of heteroskedasticity as employed by KV(2010) but estimate the unknown parameters which control the degree of heteroskedasticity. Table A1 in the appendix reports the results from this exercise and compares them with the simulation results in KV(2010). As expected Table A1 suggests that the parametric form of the estimator works well and there are efficiency gains from imposing the correct parametric assumptions.

3 Results

3.1 Previous Studies

To illustrate the utility of this approach we focus on an empirical application. We estimate the effect of education on earnings using a sample of male and female respondents in the 2004 wave of the National Longitudinal Survey of Youth (NLSY79). In the core sample of the survey 4081 individuals report valid information to estimate our wage-schooling model and this is the sample we employ.[7]

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 the inclusion of such controls in the wage equation substantially reduces the ability bias in the measured return to education. However, despite the wealth of information 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 individuals in the sample were enrolled at high school (Oreopoulos 2008).

To identify the effect of education on earnings previous studies have employed various proxies of the costs of school attendance such as the 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. Using these sources of variation Carneiro and Lee (2008) and Chen (2008) obtain IV estimates of the return to education between 13 and 15 percent in a sample of males from the NLSY79. While these 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 various family background measures in the NLSY79 have also been employed as instruments. However Card (1999) shows that IV estimates based on family back-

[7] The NLSY79 contains 3 subsamples. A core sample aimed to be representative of the US population. A second supplemental sample designed to oversample Hispanics, blacks and disadvantaged whites. The third subsample contains individuals in the military service.

ground 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).[8] It is also unclear which type of family background characteristics can be expected to affect education but not wages. Given the difficulty of finding an appropriate instrument in the NLSY79 context the identification strategy outlined above seems particularly useful. A comparison of our results with those in previous studies using the NLSY79 also seems of interest since, despite the concerns regarding instrument validity, there is some consensus regarding the presence and the effect of the endogeneity bias.

3.2 Estimation Results

In estimating the model in (1)-(4) we use as a measure of earnings, w_i, the log of the hourly wage in 2004[9] and as a measure of schooling, $educ_i$, the years of education completed. The explanatory variables are those commonly employed in the estimation of schooling and wage equations and capture the individual's background and some features of the school type and location. Table 1 describes the variables employed and Table 2 provides their summary statistics. The OLS estimates for the education and wage equations are reported in Tables 3 and 5 respectively. We include in the wage equation variables that describe the family background of the respondent (e.g. whether the mother worked when the respondent was 14 and parental education)

[8] The instruments employed are the number of siblings, the number of younger siblings, the birth order percentile among siblings, the mother's high grade, the father's high grade, dummies for the presence of magazines or newspapers in the home while growing up, a dummy for living with both parents at age 14, and a dummy for living with one parent and step-parent at age 14.

[9] We delete the observations with extremely low or extremely high wages (i.e. below the 1% and above the 99% of the hourly wage distribution).

and an standardized measure of ability available in the NLSY79, denoted AFQT, to reduce the ability bias as suggested in Card (1999).[10] The wage equation also 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.

3.2.1 The Education Equation

We first discuss the OLS estimates of the educational model reported in Table 3. The estimates are consistent with those in the existing schooling literature. Parental education and the AFQT score have an important positive effect on years of education. In contrast, respondents in larger families obtain less years of education. 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 the family background.

The KV(2010) procedure requires at least one of the equations' error terms to be heteroskedastic. Using the estimates from Table 3 we examine the presence of heteroskedasticity in the schooling equation. The statistic for the White test is 244.06 and that for the Breusch-Pagan, using all the explanatory variables in the model, is 81.21. These values clearly reject the null hypothesis of homoskedastic errors.

The next step is to estimate the form of heteroskedasticity in the schooling equation, S_{vi}^2. An examination of the results for the heteroskedasticity tests suggested that the variables responsible for the heteroskedasticity are the Hispanic indicator,

[10] The ability measure in the NLSY79 corresponds to the Armed Force Qualification Test (AFQT) taken by all the respondents in 1980. This variable is standardized by gender and age.

some of the geographical variables and the AFQT score. The regional result is consistent with that of KV(2009) and Rummery et al (1999). The result related to the Hispanic indicator captures the heterogeneous nature of the group which identifies itself as Hispanic while the AFQT 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. $z_{vi} = x_i$).[11]

The non linear least squares estimates of $S_v^2(x_i)$ are reported in Table 4.[12] The standard errors of the estimated parameters in the wage equation, the $\beta's$, and those in both heteroskedastic indices, the $\theta_j's$, are obtained from 1000 bootstrap replications with random replacement. Moreover, in obtaining the standard errors we account for the multiple step nature of the procedure by re-estimating each of the steps for each replication. Given that we have assumed an exponential form for $S_v^2(x_i)$ we can directly interpret the sign of the coefficients for the variables inside the index displayed in the Table 4. The coefficient on the AFQT score is positive and statistically significant. This reflects that more able students have a larger set of educational alternatives and thus the variance of schooling levels is positively related to this measure of ability. 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

[11] Column (1) of Table A2 to A4 in the appendix shows the results obtained when the index includes only the geographic indicators, the AFQT and the Hispanic indicator. Our main results are unaffected by this alternative specification.

[12] Note that among all the explanatory variables in Table 4 that enter the heteroskedastic index, $z_{ji}\theta_j$, there is also a constant.

among individuals living in the South of the country during their early teens.

3.2.2 The Wage Equation

We now turn to the estimation of the primary equation. In addition to the variables in the education equation we include in the wage equation some additional variables, such as the geographical indicators in 2004 and the individual's marital status, which are considered to influence an individual current earnings level. Before considering the adjusted estimates we briefly discuss the OLS estimates in Table 5. The primary feature of interest is the estimated impact of education on earnings which is .068. The magnitude of this coefficient is in line with the previously reported OLS estimates in Kane and Rouse (1995), Cameron and Taber (2004) and Chen (2008), which use the same data.

In implementing our estimation strategy it is first necessary to specify the variables entering the index underlying the heteroskedasticity of the wage equation, $z_{ui}\theta_u$. Although we experimented with different choices for the variables in z_{ui}, including one with 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.[13] In this specification the heteroskedastic index contains a constant and the geographic indicators in 2004 to allow for differences in the variance of wages due to economic conditions across regions. We also include the age of the respondent to account for the disparity across individuals in terms of wage growth.

Table 5 presents the estimates of the coefficients in the wage equation obtained from estimating (6) using the method denoted (iiib) in section 2.2. We refer to these as

[13] See Table A2 to A4 in the appendix for a comparison of the results obtained under alternative specifications of the heteroskedasticity function.

the CF estimates and they, along with their reported standard errors, are displayed in columns (3) and (4). 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 are generally quite similar. Both estimates provide evidence of a small marriage premium and a gender differential of about 25% 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 is statistically significant and positive while living in the South of the country at age 14 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 AFQT score.

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 6.8 percent the CF estimate is 11.2 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_0, is negative and statistically significant, indicating that education is not exogenous.

Our results suggest, as is frequently found in this literature, that the OLS estimate is below the estimate obtained after controlling for the endogeneity of education (see for example Angrist and Krueger, 1991; Card 1995a, 1995b and 1999; Harmon and Walker, 1995; Kling 2001 and Cameron and Taber 2004). The OLS-IV gap may reflect a sizeable measurement error in the education variable, but the large size of our estimated correlation coefficient, -0.172, is also compatible with the "penalty" to educational over achievement suggested by Vella and Gregory (1996). Under this interpretation the error component of the education equation not only captures unobserved ability, thus the negative ρ_0 would suggest that the return to ability is negative, but it can also include some other factors (i.e. motivation) that lead an individual to obtain a level of education above what is predicted by the model. If the wage equation does not allow the return to education to be different from that to over education, the residuals in this equation will capture "the over education penalty". Therefore the correlation between the unobserved terms in the schooling and the wage equation will be negative, as the factors responsible for the over education are penalized in the wage equation. We conclude that 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 indicate that the variance of wages is significantly larger in the Western and North-Eastern states of the country. The other variables included in the index do not seem to affect the variance of the unobservables in the model.[14]

As our approach assumes certain parametric representations for the heteroskedastic processes it is important to examine the robustness of our results to alternative forms. Tables A2 to A4 in the appendix show the estimates obtained under alternative specifications of the heteroskedasticity function. The estimates of the return

[14] Note that the absence of heteroskedasticity in the wage equation would not invalidate our identification strategy as this requires heteroskedasticity in either equation.

to education range between 11.4% and 12.1%, slightly above that obtained under our preferred specification, 11.2%. However the different estimates are within the 95% confidence interval of each other. Thus we conclude that our main results are unaffected by the use of these alternative forms of heteroskedasticity.

Overall the results are very supportive of this empirical approach. First, the resulting estimate is in the range of the estimates obtained via the conventional IV approach for these data. This suggests that there is sufficient heteroskedasticity in the data to identify the return to education and that the identifying restriction regarding the correlation coefficient is reasonable in this context. Second, while there is some efficiency lost in the CF estimates compared to the OLS estimates, the increase in the standard errors is similar to that obtained when comparing OLS to IV estimates. Third, the coefficient on ρ_0 is reasonably precisely estimated which suggests the identifying restriction in this context is informative.

4 Conclusions

The objective of this paper is to provide a fully parametric procedure of the KV(2010) semiparametric estimator which can be employed to control for endogeneity in triangular systems in the absence of exclusion restrictions. A disadvantage of the KV(2010) approach is that it is restricted to a model where the endogenous regressor enters linearly and without any heterogeneous effects. An advantage is that for the model they consider KV(2010) provide a procedure that does not require instruments to enable estimation. Accordingly, the KV(2010) estimator seems to be suitable to address a wide range of empirical questions. In providing a fully parametric procedure we expect that many practitioners who would be discouraged by the computational demands of the semiparametric estimator will be more likely to employ this approach. In addition to discussing how the procedure can be implemented we apply our approach to study the impact of schooling on wages for a sample of individuals drawn from the NLSY79. Our results suggest that schooling is endogenous and the adjusted impact of schooling is 11.2 percent in contrast to the OLS estimate of 6.8 percent. This is consistent with studies that employ instrumental variables approaches in these data and find that education is endogenous and that OLS underestimates the return to education.

References

[1] Angrist J and Krueger AB (1991) Does compulsory school attendance affect schooling and earnings. Q J Econ. 106: 979-1014.

[2] Blackburn ML and Neumark D (1995) Are OLS estimates of the return to schooling biased downward? Another Look. Rev Econ Stat 77(2): 217-30.

[3] Cameron S and Heckman J (2001) The Dynamics of Educational Attainment for Black, Hispanic, and White Males. J Polit Econ 109 (3): 455-99.

[4] Cameron S and Taber C (2004) Estimation of Educational Borrowing Constraints Using Returns to Schooling. J Polit Econ 112: 132-82

[5] 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.

[6] Card D (1995b) Earnings, Schooling and Ability Revisited. in Solomon W. Polachek (Eds), Research in Labor Economics (Greenwich, Connecticut: JAI Press, 1995): 23-48.

[7] Card D (1999) The causal effect of education on earnings. in Ashenfelter, O. and Card, D. (eds.), Handbook of Labor Economics, vol 3A, chap. 30, Amsterdam: Elsevier Science/North-Holland.

[8] Card D (2001) Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. Économ 69: 1127–60.

[9] Chen SH (2008) Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity. Rev Econ Stat 90(2): 275-289.

[10] Cunha F and Heckman J (2007) The Technology of Skill Formation. Am Econ Rev P&P 97(2): 31-47.

[11] Duflo E (2001) Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. Am Econ Rev 91(4):795-813.

[12] Dolton P and Vignoles A (2000) The incidence and effects of overeducation in the UK graduate labour market. Econ Educ Rev 19(2): 179–198.

[13] Groot W and van den Brink HM (2000) Overeducation in the labor market: a meta-analysis. Econ Educ Rev 19(2): 149–158.

[14] Harmon C and Walker I (1995) Estimates of the economic return to schooling for the United Kingdom. Am Econ Rev 85: 1278-86.

[15] Heckman J, Urzua S and Vytlacil E (2006) Understanding Instrumental Variables in Models with Essential Heterogeneity. Rev Econ Stat 88 (3): 389-432.

[16] Heckman J and Vytlacil E (1998) Instrumental Variables Methods for the Correlated Random Coefficient Model. J Hum Resour 33(4): 974-987.

[17] Hogan V and Rigobon R (2002) Using Heteroscedasticity to Estimate the Returns to Education. NBER Working Papers 9145.

[18] Ichimura H (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single index models. J Économ 58: 71-120.

[19] Kane T and Rouse C (1995) Labor-Market Returns to Two- and Four-Year College. Am Econ Rev 85(3): 600-14.

[20] Klein R and Vella F (2010) Estimating a Class of Triangular Simultaneous Equations Models Without Exclusion Restrictions. J Économ 154: 154-164.

[21] Klein R and Vella F (2009) Estimating the Return to Endogenous Schooling Decisions for Australian Workers via Conditional Second Moment. J Hum Resour 44(4): 1047-65.

[22] Kling JR (2001) Interpreting Instrumental Variable Estimates of the Returns to Schooling. J Bus Econ Stat 19: 358-64.

[23] Newey W, Powell J and Vella F (1999) Nonparametric Estimation of Triangular Simultaneous Equation Models. Économ 67: 565-603.

[24] 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.

[25] Rigobon R (1999) Identification through heteroskedasticity. Rev Econ Stat 85:777-92.

[26] Rubb S (2002) Overeducation in the labor market: a comment and re-analysis of a meta-analysis. Econ Educ Rev. 22 (6): 621-629.

[27] Rummery S, Vella F and Verbeek M (1999) Estimating the Returns to Education for Australian Youth via Rank-Order Instrumental Variables. Labour Econ 6: 777-92.

[28] Vella F and Gregory RG (1996): "Selection Bias and Human Capital Investment: Estimating the rates of Return to Education for Young Males." Labour Econ 3: 197-219.

[29] Vella F and Verbeek M (1997) Using rank order as an instrumental variable: an application to the return to schooling. CES Discussion Paper 97.10, K.U. Leuven.

[30] Wooldridge JM (2003) Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. Econ Lett 79:185-191.

Table 1: Variable definition:

w	log of hourly wage
$educ$	years of education completed
$married$	indicator for being married in 2004
NE	indicator for living in a North Eastern state in 2004
W	indicator for living in a Western state in 2004
NC	indicator for living in a North Central state in 2004
$city$	indicator for living in a city in 2004
$siblings$	number of siblings
$Mwork14$	indicator for whether the mother of i works when i is 14
$Hispanic$	indicator for Hispanic
$black$	indicator for black
$male$	indictor for male
$Feduc$	years of education completed by the father of i
$Meduc$	years of education completed by the mother of i
$S14$	indicator for living in the South at age 14
$city14$	indicator for living in a city at age 14
age	age of i
$AFQT$	Score obtained in the Armed Forces Qualifying Test (a measure of cognitive ability)

Table 2: Summary Statistics:

	mean	S.E.	min	max
w	2.767	0.662	2.88	96.5
$educ$	13.515	2.347	0	20
$married$	0.593			
NE	0.161			
W	0.265			
NC	0.186			
$city$	0.746			
$siblings$	3.563	2.496	0	17
$Mwork14$	0.564			
$Hispanic$	0.170			
$black$	0.265			
$male$	0.500			
$Feduc$	11.129	3.026	0	20
$Meduc$	11.085	3.829	0	20
$S14$	0.342			
$city14$	0.795			
age	43.255	2.183	40	47
$AFQT$	0.195	0.915	2.181	-2.781
$Nobs$	4081			

Table 3: Schooling Equation (Conditional Mean):

	estimates	S.D.*
siblings	−0.051	(0.013)
*Mwork*14	0.020	(0.060)
Hispanic	0.720	(0.092)
black	1.164	(0.080)
male	−0.278	(0.059)
Feduc	0.068	(0.011)
Meduc	0.072	(0.014)
*S*14	0.104	(0.065)
*city*14	0.043	(0.072)
age	0.009	(0.014)
AFQT	1.401	(0.040)
constant	11.085	(0.614)
R^2	0.354	

Test for Heteroskedasticity (statistics)	
White	244.06
Breush-Pagan	81.21

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

Table 4: Schooling equation (Conditional Variance)

	estimates	S.D.*
siblings	−0.020	(0.017)
*Mwork*14	0.087	(0.079)
Hispanic	0.002	(0.130)
black	−0.021	(0.107)
male	0.050	(0.074)
Feduc	−0.023	(0.013)
Meduc	−0.026	(0.018)
*S*14	−0.145	(0.078)
*city*14	0.069	(0.090)
age	0.015	(0.016)
AFQT	0.544	(0.050)
constant	−0.141	(0.733)

*Standard errors adjusted for the presence of heteroskedasticity

Table 5: Wage Equation (Conditional Mean)

	OLS	S.D.	CF	S.D.*
educ	0.068	(0.0054)	0.112	(0.017)
ρ_0			−0.172	(0.061)
married	0.079	(0.017)	0.079	(0.016)
NE	0.072	(0.031)	0.070	(0.032)
W	0.036	(0.030)	0.037	(0.031)
NC	−0.023	(0.029)	−0.023	(0.029)
city	0.029	(0.019)	0.029	(0.019)
siblings	0	(0.004)	0.002	(0.003)
*Mwork*14	0.019	(0.016)	0.019	(0.016)
Hispanic	0.081	(0.027)	0.048	(0.031)
black	−0.022	(0.024)	−0.074	(0.031)
male	0.248	(0.016)	0.258	(0.016)
Meduc	0.006	(0.004)	0.002	(0.004)
Feduc	0.002	(0.003)	−0.002	(0.003)
*S*14	−0.054	(0.027)	−0.060	(0.027)
*city*14	0.054	(0.021)	0.051	(0.020)
age	0.061	(0.153)	0.064	(0.152)
age^2	−0.001	(0.002)	−0.001	(0.002)
AFQT	0.162	(0.012)	0.088	(0.025)
constant	0.034	(3.309)	−0.498	(3.303)
R^2	0.30		0.30	

*The standard errors are obtained from 1000 bootstrap replications with random replacement. The values of the White and Breush-Pagan test for the presence of heteroskedasticity in the wage equation are respectively 265.45 and 69.23.

Table 6: Wage equation (Conditional Variance)

	estimates	S.D.*
age	0.008	(0.017)
NE	0.206	(0.113)
W	0.220	(0.109)
NC	0.090	(0.094)
constant	−3.175	(0.720)

*Standard errors in parenthesis. The standard errors are obtained from 1000 bootstrap replications with random replacement

5 Appendix

5.1 A description of the Klein and Vella (2010) estimator

This section describes how to estimate the model in equations (1)-(4) using the two step semiparametric procedure in Klein and Vella (2010). First regress $educ_i$ on x_i to obtain $\widehat{\delta}$ and define the reduced form residuals as:

$$\widehat{v_i} = educ_i - x_i\widehat{\delta}.$$

KV(2010) assume the following single index structure for the conditional variance function:

$$S_{v_i}^2 \equiv E\left[v_i^2 \mid x_i\right] = E\left[v_i^2 \mid I_{vi}\left(\theta_{vo}\right)\right],$$

where $I_{vi}\left(\theta_{vo}\right) \equiv x_{1i} + x_{2i}\theta_{vo}$. The o subscript denotes the true parameter, x_1 is a continuous regressor and x_2 contains the remaining explanatory variables.[15] The unknown parameters in the index, θ_{vo}, are estimated by semiparametric least squares with $\widehat{v}_i^2$ as the dependent variable (see Ichimura, 1993) as:

$$\widehat{\theta} = \arg\min_{\theta} \sum \hat{\tau}_i \left[\hat{v}_i^2 - \hat{E}\left(\hat{v}_i^2 \mid I_{vi}\left(\theta_v\right)\right)\right]^2,$$

where $\hat{\tau}_i$ is a trimming function that restricts x_i to a compact set depending on sample quantiles. The estimator for the conditional variance function is then given as:

$$\hat{S}_{vi}^2 = \widehat{E}\left(\widehat{v_i}^2 \mid I_{vi}\left(\widehat{\theta_v}\right)\right),$$

[15] Note that a continuous regressor is not required when the conditional variances are estimated parametrically.

where $\widehat{E}$ is a non-parametric estimator for the indicated conditional expectation. KV(2010) show with a monte carlo investigation that the estimator of S_{vi} and also that of the parameters in the primary equation improves if the above process is repeated in a GLS step using the initial estimator of $\hat{S}_{vi}$.

For the wage equation the conditional variance function and the parameters of interest are estimated simultaneously. Once again an index restriction is imposed. The index for the wage equation heteroskedasticity is constructed as $I_{ui}(\theta_{uo}) \equiv x_{1i} + x_{2i}\theta_{uo}$. Note that for generality the same $x's$ are allowed to appear in both I_u and I_v although this is not necessary and not imposed in the empirical work. Similar to the education approach the wage equation parameters, $\pi = (\beta, \theta_u, \rho)$, could be estimated as follows:

$$\pi \equiv \arg\min_{\pi} \hat{Q}_1(\pi).$$

where:

$$u_i(\beta) \equiv (w_i - x_i\beta_0 - \beta_1 educ_i)\,; \;\hat{S}_{ui}(\beta, \theta_u)^2 \equiv \hat{E}\left[u_i^2(\beta) \mid I_{ui}(\theta_u)\right].$$

and :

$$\begin{aligned} \hat{A}_i(\pi) &\equiv \rho\left[\hat{S}_{ui}/\hat{S}_{vi}\right]; \\ \hat{Q}_1(\pi) &\equiv \frac{1}{N}\sum_i \hat{\tau}_i\left[w_i - x_i\beta_0 - \beta_1 educ_i - \hat{A}_i(\pi)\,\hat{v}_i\right]^2. \end{aligned}$$

KV(2010) show that minimizing $\hat{Q}_1(\pi)$ may not ensure that the correct minima is obtained. Namely, in minimizing the probability limit of $\hat{Q}_1(\pi)$, the set of potential minimizers is not sufficiently restricted to enable an identification argument. To ensure that the set of minimizers does satisfy an appropriate index restriction, the objective function is modified as follows. Let:

$$
\begin{aligned}
\hat{S}^{*2}_{ui}\left(\beta,\theta_u\right) &\equiv \hat{E}\left(u_i^2\left(\beta\right)|I_{ui}\left(\theta_u\right),\hat{I}_{vi}\right)\\
\hat{A}^*_i\left(\pi\right) &\equiv \rho\left[\hat{S}^*_{ui}/\hat{S}_{vi}\right]\\
\hat{Q}_2\left(\pi\right) &\equiv \frac{1}{N}\sum_i \hat{\tau}_i\left[w_i - x_i\beta_0 - \beta_1 educ_i - \hat{A}^*_i\left(\pi\right)\hat{v}_i\right]^2.
\end{aligned}
$$

Here, S^{*2}_{ui} captures the conditional expectation of the variance function when one conditions on its own index plus the additional index characterizing the data generating process. The final, combined objective function is then given as:

$$
\hat{Q}\left(\pi\right) \equiv \hat{Q}_1\left(\pi\right) + \hat{Q}_2\left(\pi\right).
$$

Denote Q, Q_1 and Q_2 as the limiting values (uniform probability limits) for the above objective functions. KV(2010) show that π_o, the vector of true parameter values, is a minimizer not only for Q but also separately for Q_1 and Q_2. The reader is referred to KV(2010) for the argument underlying the identification strategy.

As discussed above the semiparametric estimation procedure is associated with substantial computational and programming demands. In this paper we impose parametric assumptions on the conditional variance functions, $S^2_{v_i}$ and $S^2_{u_i}$, thereby reducing computational demands. Also, as noted in the text above, there is no need for the simultaneous conditioning on the two indices nor estimation of the trimming functions and this also reduces computational issues.

5.2 Simulation Evidence

KV(2010) conduct a monte-carlo exercise to analyze the finite sample performance of the semiparametric estimator. Below we re-estimate their simulated model using the estimator proposed in this paper and compare the results.

In the simulated model the same exogenous variables appear in the conditional means and the conditional variances of both endogenous variables. The two indices underlying the heteroskedasticity are also highly correlated and the same functional form for the heteroskedastictiy is employed in each equation. The model has the following form:

$$Y_{1i} = 1 + x_{1i} + x_{2i} + Y_{2i} + u_i$$
$$Y_{2i} = 1 + x_{1i} + x_{2i} + v_i$$
$$u_i = 1 + \exp(.2 * x_{1i} + .6 * x_{2i}) * u_i^*$$
$$v_i = 1 + \exp(.6 * x_{1i} + .2 * x_{2i}) * v_i^*$$
$$u_i^* = .33 * v_i^* + N(0,1) \text{ and } v_i^* \sim N(0,1).$$

The exogenous variables x_{1i} and x_{2i} are generated as standard normal random variables and then x_{2i} is transformed into a chi-squared variable with 1 degree of freedom. The simulation results in Table A1 are for a sample size of 1000 and 100 replications. The Table compares the results obtained using the parametric version of the estimator employed in this paper (column 3), to those in KV(2010) obtained with the semiparametric procedure (column 2). For comparison purposes, the first column reports the OLS estimates. To investigate the performance of the estimator when the parametric assumptions are misspecified, we perform two additional simulation exercises. In column 4 we assume that the conditional variance is only a function of x_1, while in column 5 we assume that is a second order polynomial in x_2.

Table A1: Simulation Results

	OLS	Semiparametric	Parametric		
		CF	CF	CF	CF
	(1)	(2)	(3)	(4)	(5)
constant	0.858 (.122)*	1.003 (.201)	0.964 (.115)	0.985 (.170)	0.943 (.126)
x_1	0.858 (.120)	1.003 (.201)	0.966 (.151)	0.977 (.172)	0.929 (.149)
x_2	0.866 (.121)	1.011 (.203)	0.969 (.116)	0.993 (.162)	0.951 (.126)
y_2	1.137 (.108)	.993 (.119)	1.036 (.108)	1.016 (.162)	1.057 (.121)
ρ		.298 (.110)	0.212 (.250)	1.583 (2.42)	0.130 (.278)

*Standard errors in parenthesis.

5.3 Alternative specifications of the functional forms

Table A2 displays the CF estimates of the wage equation obtained using alternative functional forms for the heteroskedasticity. Table A3 and A4 show the estimates of the heteroskedasticity function parameters in the wage and the schooling equations respectively. Column (1) in the tables report the estimates obtained using the specification in equation (7), but excluding some explanatory variables from the heteroskedastic index. In particular, the index only includes the variables that according to the results in Table 4 seem to be responsible for the variability in schooling levels (i.e. AFQT and regional indicators). In column (2) we report the estimates when the heteroskedastic indices for wages and schooling include the same explanatory variables that their respective models for the conditional mean. Finally, in column (3) we use the main specification in the paper but include in the conditional mean for wages interaction terms between the years of education and some explanatory variables (i.e. gender, race and AFQT).

The estimates obtained under the different specifications are between 11.4% and 12.1%, slightly above that obtained under our preferred specification 11.2%. Note that the different estimates are within the 95% confidence interval of each other. Thus we conclude that our main results are unaffected by the use of alternative functional forms for the heteroskedasticity. Moreover, we do not find evidence that the return of education varies across the socioeconomic characteristics considered.

Table A2: Wage Equation (Conditional mean)

	(1)	(2)	(3)
educ	0.114(0.018)*	0.120(0.020)	0.121(0.028)
ρ_0	−0.178(0.063)	−0.191(0.072)	−0.210(0.093)

*Standard errors in parenthesis. The standard errors are obtained from 1000 bootstrap replications with random replacement

	(1)	(2)	(3)
married	0.079(0.016)*	0.079(0.017)	0.078(0.017)
NE	0.070(0.032)	0.070(0.032)	0.072(0.032)
W	0.037(0.031)	0.037(0.030)	0.037(0.031)
NC	−0.022(0.029)	−0.023(0.029)	−0.022(0.029)
city	0.029(0.019)	0.029(0.020)	0.030(0.019)
siblings	0.002(0.003)	0.002(0.003)	0.003(0.004)
$Mwork14$	0.019(0.016)	0.019(0.016)	0.019(0.016)
Hispanic	0.048(0.031)	0.042(0.032)	−0.021(0.138)
black	−0.076(0.031)	−0.083(0.034)	−0.245(0.141)
male	0.259(0.016)	0.260(0.017)	0.287(0.099)
Feduc	−0.002(0.003)	−0.002(0.003)	−0.002(0.004)
Meduc	0.002(0.004)	0.002(0.004)	0.001(0.003)
$S14$	−0.060(0.027)	−0.060(0.027)	−0.060(0.027)
$city14$	0.051(0.020)	0.051(0.020)	0.051(0.020)
age	0.065(0.152)	0.060(0.152)	0.065(0.152)
age^2	−0.001(0.002)	−0.001(0.002)	−0.002(0.002)
AFQT	0.086(0.029)	0.079(0.028)	0.083(0.055)
$(educ * male)$			−0.002(0.007)
$(educ * AFQT)$			−0.001(0.006)
$(educ * hispa)$			0.004(0.011)
$(educ * black)$			0.012(0.010)
constant	−0.178(3.298)	−0.492(3.311)	−0.616(3.317)
R^2	0.30	0.30	0.30

*Standard errors in parenthesis. The standard errors are obtained from 1000 bootstrap replications with random replacement

Table A3: Wage Equation (Conditional Variance)

	(1)	(2)	(3)
married	.	−0.121(0.102)*	.
NE	−0.132(0.076)	0.192(0.161)	0.156(0.116)
W	0.028(0.090)	0.135(0.157)	0.197(0.108)
NC	0.503(0.042)	−0.006(0.149)	0.049(0.094)
city	.	0.038(0.104)	.
siblings	.	−0.004(0.018)	.
*Mwork*14	.	−0.057(0.084)	.
Hispanic	.	0.148(0.129)	.
black	.	0.051(0.117)	.
male	.	0.041(0.080)	.
Meduc	.	0.030(0.014)	.
Feduc	.	−0.009(0.021)	.
*S*14	.	0.009(0.139)	.
*city*14	.	0.175(0.121)	.
age	0.132(0.108)	−0.578(0.795)	0.013(0.017)
age^2	.	0.007(0.009)	.
AFQT	.	0.077(0.056)	.
constant	−0.052(0.084)	9.346(17.24)	−3.310(0.750)

*Standard errors in parenthesis. The standard errors are obtained from 1000 bootstrap replications with random replacement

Table A4: Schooling Equation (Conditional Variance)

	(1)	(2)	(3)
siblings	.	−0.020(0.017)*	−0.020(0.017)
*Mwork*14	.	0.087(0.080)	0.087(0.080)
Hispanic	0.132(0.108)	0.002(0.130)	0.002(0.130)
black	.	−0.021(0.107)	−0.021(0.107)
male	.	0.050(0.074)	0.050(0.074)
Feduc	.	−0.023(0.013)	−0.023(0.013)
Meduc	.	−0.026(0.183)	−0.026(0.018)
*S*14	−0.132(0.076)	−0.145(0.078)	−0.145(0.078)
*city*14	0.028(0.090)	0.069(0.090)	0.069(0.090)
age	.	0.015(0.016)	0.015(0.016)
AFQT	0.503(0.042)	0.544(0.050)	0.544(0.050)
constant	−0.052(0.084)	−0.141(0.733)	−0.141(0.733)

*Standard errors in parenthesis. The standard errors are obtained from 1000 bootstrap replications with random replacement