

CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2003

Using Heteroscedasticity to Estimate the Returns to Education

Vincent Hogan, University College Dublin and Warwick University and
Roberto Rigobon, Sloan School of Management,
Massachusetts Institute of Technology

WP03/01

January 2003

**DEPARTMENT OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4**

Using Heteroscedasticity to Estimate the Returns to Education¹

Vincent Hogan
University College Dublin
and
Warwick University

Roberto Rigobon
Sloan School of Management
Massachusetts Institute of Technology

December 10, 2002

¹We thank Tom Stoker, Josh Angrist, Jim Poterba, Colm Harmon, Ian Walker and seminar participants in UCD, Warwick University and the Centre of Economics of Education at LSE, for helpful comments. We also thank Yu Zhu for help with the data. All errors remain our own. Comments welcome to vincent.hogan@ucd.ie or rigobon@mit.edu

Abstract

We apply a new estimator to the measurement of the economic returns to education. We control for endogenous education, unobserved ability and measurement error using only the natural heteroscedasticity of wages and education attainment. Our preferred estimate, 6.07%, is closer to the OLS estimate but smaller (and more precise) than the estimates typically reported by studies that use IV. Our results indicate that the biases generated by unobserved ability and measurement error tend to cancel each other out as suggested by Griliches (1977). We also present Monte Carlo evidence to show that the finite sample bias our estimator is small.

JEL codes: C30, I20, J31

Keywords: Identification; Returns to Education

Vincent Hogan
Department of Economics
UCD
Belfield, Dublin 4
Ireland

vincent.hogan@ucd.ie

Roberto Rigobon
MIT, E52-434
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER

rigobon@mit.edu

1 Introduction

In recent years there has been a resurgence of interest in the measurement of the economic returns to education. The issue is important to both governments and individuals contemplating investment in education and is clearly a major determinant of individual income. There has also been interest in the topic as a way of explaining the growing differences in income between more and less well educated workers (see for example Katz and Autor, 1999 and Card and Lemieux, 2000). Unfortunately the precise measurement of the economic returns to education has been plagued by difficulties in isolating the causal effect of education from the joint process of education and income. In this paper we make use of a new estimator to identify the returns to education using only the natural heteroscedasticity of the data.

Education is almost certainly an endogenous variable not least because individuals seek higher education in order to boost income. Similarly high ability individuals will tend to earn higher wages controlling for education level and will probably also attain a higher level of education. Griliches (1977) noted that while unobserved ability would tend to bias the OLS estimates of the return upwards, measurement error in the education variable would tend to bias estimates towards zero. He suggested that the biases may actually cancel out leaving OLS estimates a good guide to the true return to education.

In his survey of the literature Card (2001) notes that the typical response of the literature to the identification problem is to use IV with instruments often based on changes in the institutional structure of the education system (see DuFlo, 2001, for a recent example).¹ Staiger and Stock (1997) argue that many of these studies employ weak instruments, implying that estimates are even more imprecise than they may first appear and often not significantly different from the OLS estimate. Similarly Manski and Pepper (2000) obtain upper bounds on the returns to schooling that are below some IV point estimates, casting doubt on the validity of those estimates.

In this paper we present estimates of the return to education that control for the potential joint endogeneity of education and income using the natural heteroscedasticity of the data. Building on Rigobon (2000), we show that if we can split the sample into groups that have different covariance matrices then, under reasonable conditions, we can identify the structural parameters. Our estimates are much more precise than the usual IV estimates. This is because the data exhibits strong heteroscedasticity whereas the IV estimates are based on instruments that are only weakly correlated with education attainment.

Our estimates of the return to education suggest that it is close in magnitude to the OLS estimates, lending support to both the Griliches (1977) hypothesis. We also show that our results are quite robust and that the finite sample bias of the estimator is small. From a methodological point of view, this suggests that heteroscedasticity can be used more generally to solve the problem of identification when omitted variables, measurement error, and standard simultaneity issues arise in cross-sectional data.

The paper proceeds as follows. In section two we present a discussion of a simplified version

¹Belzil and Hansen (2002) adopt an alternative approach by using the non-linearity of the education choice function that results from of an intertemporal optimising model to identify the return to education.

of our technique by way of illustration. We also present estimates of the returns to education using some of the standard IV techniques. In section three we present a more realistic version of heteroscedasticity technique that allows for unobserved ability and measurement error; and it is this version that we take to the data. We also test the robustness of our estimator, applying it to a second dataset. Section four presents Monte Carlo simulations of the finite sample properties of the estimator. Section five concludes.

2 The identification Problem

Consider the model of education returns in (1) where X and Y are vectors consisting of a constant and the observable characteristics of individual i ; ε_i and η_i are random structural disturbances to (log) wages and education respectively; and β is the return to education. For future reference we will denote the *variances* of the structural disturbances as ν_ε and ν_η and we assume (for now) that $Cov(\varepsilon_i, \eta_i) = 0$.

$$\begin{aligned} wage_i &= \beta educ_i + \mu_1 X_i + \varepsilon_i \\ educ_i &= \alpha wage_i + \mu_2 Y_i + \eta_i \end{aligned} \tag{1}$$

The first equation is the standard Mincer (1974) equation which is consistent with viewing education as the accumulation of human capital or as the process of signaling innate ability. The second equation reflects the potential endogeneity of education attainment i.e. individuals may seek higher education in anticipation of resulting higher wages. This specification, while a simplistic description of actual behaviour does capture the possibility that education and wages may be simultaneously related.²

Note that the model of equation (1) excludes the possibility that unobserved ability may affect both education attainment and market wage and also the possibility that the variables are measured with error. We make these simplifying assumptions solely in order to illustrate our identification technique. The model that we take to the data in the next section controls for unobserved individual effects and measurement error.

The the covariance matrix - of the reduced form residuals will be given by

$$- = \frac{1}{(1 - \beta\alpha)^2} \begin{bmatrix} \nu_\varepsilon + \beta^2 \nu_\eta & \alpha \nu_\varepsilon + \beta \nu_\eta \\ . & \alpha^2 \nu_\varepsilon + \nu_\eta \end{bmatrix}$$

The problem of identification is that the reduced form covariance matrix – which we can always estimate – provides us with three equations in four unknowns ($\nu_\varepsilon, \nu_\eta, \alpha, \beta$). As is well known, OLS applied to (1) will lead to biased estimates if $\alpha \neq 0$. The usual approach is find some instrumental variable i.e. some element of Y that is not in X and is uncorrelated with ε_i .

²Hogan and Walker (2002) present a model of intertemporal education choice that allows for uncertainty in the return to education. They show that the resulting education choice function is highly non-linear – even for simple primitive functions.

Table 2 shows the results of OLS and IV estimation of model (1) for men using the UK Labour Force Survey (1993-2000).³ We measure schooling in terms of years in school (as opposed to credentials attained) and wage as the log of usual gross hourly earnings and the data is pooled across all the years.⁴ All the regressions include the usual demographic controls (age, squared, marital status, union membership) and region and year dummies.

As can be seen from the first column, the return to education as measured by OLS is 6.8% which is in line with other estimates for the UK (see Card, 2001). The second column shows the results when we instrument using the quarter of birth interacted with year, as in Angrist and Krueger (1991). The estimate of 5.2% is borderline significant (p-value of 0.075) and, as in Angrist and Krueger (1991), smaller in magnitude than the OLS estimate, but not significantly different from it.

[Tables 1 and 2 here]

The lack of precision of IV estimates of the return to education is not unusual. In fact most of the studies surveyed by Card (2001) present IV estimates of the return to education that are not significantly different from the OLS estimates.⁵ Furthermore as Staiger and Stock (1997) show, the degree of imprecision is greater than suggested by the IV standard errors. They show that when the $F - test$ from a regression of the endogenous variables on the instruments is less than 5, the instruments are weak and the IV estimates and confidence intervals are biased even in large samples. Clearly this is true of our example.⁶ Manski and Pepper (2000) calculate upper bounds for the true return to schooling and show that, in general, IV point estimates are very close to the upper bound and sometimes even above it. This suggests that some IV estimates can be biased upwards.

As an alternative to IV, we make use of the ‘‘Identification Through Heteroscedasticity’’ (IH) procedure.⁷ The idea is that we split the data into at least two groups that have different reduced form covariance matrices.⁸

$$\begin{aligned} \beta_1 &= \frac{1}{(1 - \beta\alpha)^2} \begin{bmatrix} \nu_\varepsilon^1 + \beta^2\nu_\eta^1 & \alpha\nu_\varepsilon^1 + \beta\nu_\eta^1 \\ \alpha^2\nu_\varepsilon^1 + \nu_\eta^1 \end{bmatrix} \\ \beta_0 &= \frac{1}{(1 - \beta\alpha)^2} \begin{bmatrix} \nu_\varepsilon^0 + \beta^2\nu_\eta^0 & \alpha\nu_\varepsilon^0 + \beta\nu_\eta^0 \\ \alpha^2\nu_\varepsilon^0 + \nu_\eta^0 \end{bmatrix} \end{aligned}$$

³Excluding women from the sample allows us to abstract from the issue of labour market participation. We also excluded self-employed, non-prime-aged, and those for whom hours of work or wages were missing.

⁴Table 1 provides the summary statistics for the main variables used.

⁵See especially Table 2 of Card (2001).

⁶IV estimates have been criticised on other grounds also. For example, quarter of birth will not be exogenous if it reflects family planning decisions which in turn may be correlated with parental background and family level unobservables (see Card, 2001).

⁷We present only an intuitive justification of the estimator here. The interested reader is referred to Rigobon (2000) for a formal discussion. The first reference to identification using changes in second moments is by Sewell Wright in the appendix to Philip Wright (1928). Similar techniques have been applied by several authors: King et al. (1994); Sentana and Fiorentini (2001); Klein and Vella (2000a,b); Rigobon and Sack (2002).

⁸Note that we include group dummies in both equations (in the matrices X and Y) so we are not using group as an instrument in the usual sense. We are relying on changes in the variance (not the mean) of the residuals to identify the model.

where the superscript $r \in [0, 1]$ indexes the group. Under the assumption that the coefficients are stable across groups, we now have six equations in exactly six unknowns $(\nu_\varepsilon^0, \nu_\eta^0, \nu_\varepsilon^1, \nu_\eta^1, \alpha, \beta)$, which we can solve for α and β .

After some algebraic manipulation to eliminate the structural variance terms, the system reduces to (2); two equations in α and β , the two unknowns of interest, where ω_{ij}^r is the (i, j) element of Σ^r .

$$\beta = \frac{\omega_{12}^r - \alpha\omega_{11}^r}{\omega_{22}^r - \alpha\omega_{12}^r} \quad \text{and} \quad \Sigma^r = \begin{bmatrix} \omega_{11}^r & \omega_{12}^r \\ \omega_{21}^r & \omega_{22}^r \end{bmatrix} \quad \forall r \in [0, 1] \quad (2)$$

We can see from (2) that this estimator is numerically identical to the OLS estimator when education is exogenous (i.e. $\alpha = 0$).

It is worth emphasising what exactly is needed in order to achieve identification in this example. The condition that number equations equals number of parameters is equivalent to an order condition. Rigobon (2000) shows that the necessary and sufficient (rank) condition for consistency is that the two reduced form covariance matrices are linearly independent.⁹ We also need the assumption that the coefficients on the endogenous variables do not vary across groups – a restriction that we can relax in an overidentified model. Under these circumstances we can be sure that the observed heteroscedasticity of the reduced form (for which we can test directly) must be due to heteroscedasticity in the structural form.

These considerations place some restrictions on how the sample can be split. For example, we might find that the variance of the reduced form was different for men and for women. We could not however, use this to identify the model as we do not believe that the returns to education are the same for men and for women. On the other hand, one could argue, for example, that the differing effects macroeconomic and industry specific shocks across regions would make the structural wage shock heteroscedastic while leaving the endogenous coefficients (representing the marginal effects of shocks) constant across regions.

One might expect that the methodology is very sensitive to the precise definition of the groups. In fact the model is quite robust to changes in the definition of the sample split. The estimates will still be consistent even if the sample split does not perfectly capture the variation in the structural shocks throughout the sample. For example, if the true difference in the covariance is between the south and the north, a London vs. non-London split will still give consistent (but less efficient) estimates. Only if the sample split is completely wrong, will the estimator be inconsistent (e.g. if the true split is north-south but we choose east-west). In this case the estimated covariance matrices will be multiples of each other and the systems of equations will lose rank.

While the sample split need not be perfect, equally it cannot be completely arbitrary. It must have economic content i.e. it must be possible to believe that structural shocks are heteroscedastic across groups. If the sample split were entirely arbitrary, the covariance matrices would be linearly related and the model would fail the rank condition.

We can test for this, in much the same way as we would test a linear system of equations for full

⁹The rank condition ensures that a real solution to (2) will exist. This solution will be a consistent estimates of α and β if the reduced form covariance matrices have been consistently estimated from the data.

rank, conditional on satisfying the order condition. For our estimator, the sample split allows us to satisfy the order condition (there are at least as many equations as unknowns). Once this requirement is met, we can test the rank condition by simply checking if the equations are independent.

One way to fix our intuition of the IH estimator is to think of it as a “probabilistic IV”. With standard IV we find some variable that will shift the education choice curve without affecting the position of wage curve, allowing us to trace out the wage curve (some policy reform for example). The intuition is similar for the IH estimator. Provided the covariance matrices are different across groups (and the coefficients on the endogenous variables are constant across groups), we know that for one group the variance in the structural education shock must be greater than for the other. Thus the cloud of disturbances will be elongated along the wage curve for that group relative to the other. In other words, an increase in the variance of education shocks increases the *probability* that the education curve will shift along the fixed wage curve allowing us to trace out the wage curve.¹⁰

3 Unobserved Ability and Measurement Error

We now extend the model of the previous section to the more realistic case where there is unobserved individual ability which may affect both education attainment and wage conditional on education and also where some or all of the variables may be measured with error. This is the version of the estimator that we take to the data. We also make use of variation in the data to over-identify the model and to test the robustness of the identification.

The model of education now becomes (3) where Z_i is the unobserved ability of person i , $educ_i$ is now interpreted as *observed* education which differs from true education by the (mean zero) random variable, u_i . All other variables are the same as before and the coefficient on unobserved ability in the education equation is normalised to unity.

$$\begin{aligned} wage_i &= \beta educ_i + \mu_1 X_i + \gamma Z_i - \beta u_i + \varepsilon_i \\ educ_i &= \alpha wage_i + \mu_2 Y_i + Z_i + u_i + \eta_i \end{aligned} \tag{3}$$

The addition of the unobserved disturbances complicates the analysis. Following much of the education literature we interpret Z here as being unobserved ability, but it clearly could represent any source of correlation in the structural unobservables.

It will become clear that our procedure can control for measurement errors in the wage (or any other) variable. We focus on measurement error in the education variable because it is particularly

¹⁰Our procedure bears some resemblance to Rank Order IV procedure of Vella and Verbeek (1997) and applied by them to the measurement of the returns to education in Rummery et. al. (1999). They also identify their model on the assumption that the distribution of the structural errors changes across groups. However, their method differs from ours in so far as they explicitly estimate the distribution functions of residuals in each group, by approximating them by linear functions. They then get consistent estimates of the model by comparing individuals from the two groups at similar positions in their respective distributions (i.e. at the same rank). It seems likely that the two methods will give similar answers when there are many observations and the distributions are approximately normal. However, our method avoids making any parametric assumptions about the distribution of the structural errors. It is sufficient that second moments exist and that they are linearly independent across groups – relatively weak conditions.

important in our application for two reasons. Firstly, errors are likely to arise in the education variable for reasons other than the usual coding and reporting errors. Education, when measured by time in school, will contain errors because the mapping from time in education to human capital is not uniform across persons. For example, most people take four years to complete a bachelor's degree, but some do it in three years and some take five.¹¹ So observed time in school will be a noisy signal of human capital. Secondly, it has been suggested since Griliches (1977) that the bias induced in OLS estimates by the errors in the education variable may offset the bias induced by the presence of unobserved ability.

A sufficient condition for identification of equation (3) is that the variance of the common shocks (Z and u) are constant across all groups, and we can identify enough groups. To see this, calculate the reduced form covariance matrix of the reduced form for every group r , where ν_z^r is the variance of the ability shock in group r , ν_u^r is the variance of the measurement error, and all other parameters have same the meaning as before.

$$\Sigma^r = \frac{1}{(1 - \beta\alpha)^2} \begin{bmatrix} \nu_\varepsilon^r + \beta^2\nu_\eta^r + (\beta + \gamma)^2\nu_z^r & \alpha\nu_\varepsilon^r + \beta\nu_\eta^r + (1 + \alpha\gamma)(\beta + \gamma)\nu_z^r \\ \alpha^2\nu_\varepsilon^r + \nu_\eta^r + (1 + \alpha\gamma)^2\nu_z^r + (1 - \alpha\beta)^2\nu_u^r & \end{bmatrix} \quad (4)$$

As it stands, with all observations in one group, the model is not identified as the covariance matrix provides three equations in seven unknowns ($\alpha, \beta, \gamma, \nu_\varepsilon^r, \nu_\eta^r, \nu_z^r, \nu_u^r$). If we split the sample, as in the last section, the model remains unidentified: each additional group provides a covariance matrix with three additional equations but also four new parameters ($\nu_\varepsilon^r, \nu_\eta^r, \nu_z^r, \nu_u^r$).

In order to identify the model we need to impose the further restriction that some of the shocks are homoscedastic across some of the groups. In what follows, we assume that the common (ability) shock and the measurement error are both homoscedastic throughout the entire sample i.e. $\nu_z^r = \nu_z \forall r$ and $\nu_u^r = \nu_u \forall r$.¹²

Now each extra group which we can identify will provide three more equations but with only two new parameters ($\nu_\varepsilon^r, \nu_\eta^r$). Define $\Sigma^r = \Sigma^r - \Sigma^r$ where Σ^r is the covariance matrix for the entire sample; ν_x represents the variance of a structural shock calculated over the entire sample; and $\Delta\nu_x = \nu_x^r - \nu_x$.

$$\Sigma^r = \frac{1}{(1 - \beta\alpha)^2} \begin{bmatrix} \Delta\nu_\varepsilon^r + \beta^2\Delta\nu_\eta^r & \alpha\Delta\nu_\varepsilon^r + \beta\Delta\nu_\eta^r \\ \alpha^2\Delta\nu_\varepsilon^r + \Delta\nu_\eta^r & \end{bmatrix}$$

Note that Σ^r has the same form as the covariance matrix from the previous section. Hence, the conditions for identification are the same - i.e. two independent Σ^r are enough to consistently estimate α and β . Because the matrices Σ^r are linear combinations of the estimable Σ^r , implying that three independent groups are sufficient to solve the problem of identification.

For each of the R groups in the sample, α and β solve (5); where ζ_{ij}^r is the (i, j) element of the

¹¹This raises the possibility that u_i is correlated with Z_i . We can simply include another common shock in (3) to account for this correlation. The identification will proceed as in the text.

¹²Note that we allow the average level of both common shocks to be free to vary across regions and years as any such variation will be picked up by the region and year dummies in the reduced forms.

respective matrix Σ^r ,

$$\beta = \frac{\zeta_{12}^r - \alpha\zeta_{11}^r}{\zeta_{22}^r - \alpha\zeta_{12}^r} \quad \forall r \in [1..R] \quad (5)$$

We treat the R equations (5) as a set of moment conditions and apply a GMM estimator with the weighting matrix determined by the number of individual observations occurring in each cell.¹³ The standard errors are computed by sampling (500 draws) from the estimated asymptotic distribution of the reduced form covariance matrixes.¹⁴

It should be clear at this point that we could have included any number of unobservables in equation (3) to account for any source of structural shocks, measurement error in any variables or any correlation between them. Providing we are prepared to believe that all of these shocks are homoscedastic across groups, they will be differenced out, and (5) will provide a consistent estimate of α and β . By the same token, however, we cannot distinguish between the difference source of shocks i.e. we cannot say whether ability bias is more important than measurement error. Nevertheless, we can say something about the relative importance of ability bias and measurement error on the one hand and simultaneous equation bias on the other hand, as our method provides an estimate of α .

From (5) we can see that the IH estimator works by taking differences in second moments across groups, whereas standard IV works by taking differences in means across the groups. If the wage shock were homoscedastic we would be unable identify the education equation. Similarly, heteroscedasticity of education shock allows us to trace out the wage curve, and identify the return to education – homoscedasticity of the education shock would prevent this. Furthermore, these assumptions of homoscedasticity and parameter stability are testable in an overidentified system, which will require more than three groups.

It is worth commenting on the economic, as distinct from the statistical, content of our identifying restrictions. Firstly, we have not relied on exclusion restrictions for identification. So any variable that may affect the wage is free also to affect education attainment. Secondly, we will allow wage shocks to be drawn from different distributions across space and time. Thus we allow for macroeconomic, regional or industry level shocks that affect the income distribution differently in different regions and years. Thirdly, in order to identify the wage equation and β the parameter of interest, we require some heteroscedasticity of the education shock. The source of this heteroscedasticity is not of direct concern. But it could arise, for example, if population density was different across regions. To the extent that proximity to school was an important determinant of education, regions that contain both large cities and rural areas would experience a large variation in education attainment.¹⁵ Fourthly, the homoscedasticity of common shocks seems perfectly reasonable – if they reflect ability and measurement error. To believe otherwise is to believe that there is some

¹³In fact we solve R equations of the form $(\zeta_{22}^r - \alpha\zeta_{12}^r)\beta - (\zeta_{12}^r - \alpha\zeta_{11}^r) = 0$ in order to avoid problems of instability of the numerical procedure when the denominator of (5) is close to zero.

¹⁴We have implicitly assumed that each ϵ^r is distributed independently of the others. This procedure could potentially underestimate the true standard errors, in the unlikely event that ϵ^r is negatively correlated across groups. We could explicitly account for any correlation by using a bootstrap and sampling from the empirical distribution of the residuals, but this was found to be far too computationally burdensome.

¹⁵Rummery et. al. (1999) make the same point.

region in which there is an unusually high number of low ability individuals and also an unusually high number of high ability individuals. Finally, as we use region-year groups, assuming that the structural parameters (α, β, γ) do not vary across groups is tantamount to assuming that tastes for education and its marginal benefits are constant across space and time. The former seems reasonable (and is implicitly a maintained hypothesis of most studies of education returns) but, given Katz and Autor (1999), the later does not. Because of the high degree of overidentification, we will relax this assumption below.

3.1 Results

We apply this model to the UK LFS data for men only. We divide the data into region and time cells. As there are 18 regions and 8 years in the LFS, this gives us 144 cells – a degree of overidentification which we exploit to test the robustness of the model. We estimate the model in two steps. First, we estimate the reduced form including the usual demographic control variables, region dummies, year dummies and their interactions.¹⁶ We then apply standard tests of the homoscedasticity of the *reduced form* residuals that reject the null hypothesis at all standard significance levels.¹⁷ This is a necessary but insufficient condition for our identification to hold. Note that it also suggests that our estimates of β may be more precise than our estimates of α , as the heteroscedasticity of the education shock seems stronger.

In the second step, we can calculate the reduced form matrices to give us a system of $R = 144$ non-linear simultaneous equations as in (5) which we solve by GMM. The results are shown in Table 3 together with the summary statistics of the distribution of the estimator calculated by sampling (500 draws) from the asymptotic distribution of the reduced form covariance matrices. The full density function of the estimate of β is plotted in figure 1.

[Table 3 here]

As can be seen, we have only reported the parameters of interest: α and β . The first column in Table 3 indicates the results from estimating the impact of wages on education (α) after controlling for ability, and measurement error. The second column is the estimate of the returns to schooling (β). The first row gives the point estimate of both parameters. The second and third rows show the mean and standard deviation, respectively, of the bootstrapped distribution of the estimator. The fourth row presents a quasi t-statistic for significance i.e. the ratio between the point estimate and the standard deviation. The fifth is the 95 percent confidence interval calculated from the 5th and 95th percentiles of the distribution of the estimator. The next two rows show the extrema of the distribution. Finally, the last row gives the percentage of the realizations that are negative.

The most striking observation is that estimated return to education (β) is much smaller than typically estimated in studies implementing IV estimators. In fact the IH estimate is close in

¹⁶The results of this first step are available from the authors upon request.

¹⁷We regress the squared residuals and their product on dummy variables that define the groups. F-Tests of zero slopes, produce test statistics of 2.11 for the wage residual, 5.27 for the education residual and 2.52 for the product of both. The critical value of $F(143, 70943)$ is 1.29 at the 1% significance level.

magnitude to, and insignificantly different from, the OLS estimate in Table 2 (p – value of 0.29). Our estimate seems to support the result of Angrist and Krueger (1991) that the return to education is quite small.

We can easily reject the hypothesis that the estimated return is insignificant from the reported t -statistics. It is also informative to observe the mass of realizations below zero, which we can interpret as the p -value of a test of parameter significance.¹⁸ Furthermore we can establish that the 95% confidence interval for β is [4.68%, 7.01%]; again in line with OLS estimates but well below (and much more precise than) most IV estimates. This provides evidence in support of the hypothesis of Griliches (1997) that measurement error and ability bias approximately cancel out, leaving OLS estimates close to the true value in practice.

Note also that our bootstrap procedure is implicitly checking the rank condition i.e. the linear independence of structural covariance matrices. If the system had been underidentified then there would have been a continuum of solutions to (5). Thus, the standard deviations computed from the bootstrap would have been very large.

We also have an estimate of α , the education choice coefficient. Most studies of the return to education do not identify α , but an estimate of it comes naturally here. It is less precisely estimated than β , is insignificant and, perhaps surprisingly, negative. Clearly the true process by which individuals make their choices in education is much more complicated than the simple linear model assumed here, so we should be cautious in interpreting this estimate. Nevertheless the negative point estimate and the overall insignificance have interesting implications. Firstly, the negative point estimates suggest that income effects dominate substitution effects. However, this is not unreasonable as it may appear once we realise that the result is conditional on the level of ability. In other words, controlling for an individual’s innate ability, giving her a higher wage may well induce her to leave school earlier. This result suggests that education may be a substitute for, rather than a complement to, ability. Secondly, the fact that we cannot reject the hypothesis that $\alpha = 0$ suggests that simultaneity is not the most important source of bias in the OLS estimate of the return to education. Further exploration of these issues would be interesting topics for future research.

3.2 Robustness

We can use the degree of overidentification to relax some of the identifying assumptions. Specifically, we can allow the return to education to be different in each of the eight years spanned by the data, and allow both common shocks (Z_i and u_i) to be heteroscedastic across years, but homoscedastic among regions. This is equivalent to estimating the model separately for each year.

In Table 4 we report the estimates of the returns to education generated by this exercise, where we use the same method as above to generate the distribution of the estimator. We report only the estimates of β and the summary statistics of the bootstrapped distribution of the estimator for each year, in the same manner as Table 3.

¹⁸It is possible that quasi t -test could be misleading if, for example, the distributions were not normal and had large standard errors, but no realizations on one side of zero.

Not surprisingly these results are less precise than those in Table 3. For all years other than 1998 and 1999, we can reject the hypothesis the true return to education is zero. The point estimates, however, are not significantly different from the estimate using all the data. Although, as can be seen there is some variation in the returns to education over time.

[Table 4 here]

Figure 2 graphs the point estimate of β over time together with the two standard error bands. Notice that the hypothesis that all the estimates are equal to the estimate using all cells cannot be rejected – lending support to the overidentifying restrictions.

[Figure 2 here]

We also apply our estimator to another dataset, namely the UK Family Expenditure Survey (FES), originally used by Harmon and Walker (1995). In Table 5 we report the results of OLS, IV and IH estimation of the returns to education.¹⁹ The IH estimate (using 324 region-time cells) is less than IV but now it is significantly different than the OLS estimate (p – value 0.056). Again this is in line with Griliches (1977) conjecture.

The IH estimate of the return to education is also lower using the FES than using the LFS. This may be explained by longer sample period covered FES. Figure 3 shows the IH estimate of the return over time using the FES. The annual returns are not estimated precisely, reflecting the small cell size. As we can see from a comparison of Figure 2 and Figure 3, the return over the period 1993-1995 estimated using the FES is not significantly different from that estimated using LFS. Thus the smaller overall estimate of the return to education from FES compared to LFS (4.45% vs. 6.07%) appears to be caused by a weak upward trend in the return to education during the 1990s. Similarly for FES, $\hat{\beta}_{IH}$ is less than $\hat{\beta}_{OLS}$. For brevity we do not report detailed results for α , but its point estimate is negative and insignificantly different from zero.

4 Finite Sample Properties of the IH Estimator

We present Monte Carlo simulations of the finite sample performance of this estimator and compare it with the OLS estimator. We show that the IH estimator is not only unbiased but is estimated with precision. The details are contained in the appendix, we summarize the process and results here. In order to reduce the dimensionality of the problem we concentrate only on one omitted variable bias, which could be interpreted either as the unobservable ability or the measurement error.

We start the simulation process by choosing arbitrary but plausible values for the structural parameters i.e. α , β , and γ in equation (3). For each combination of the “true” structural parameters,

¹⁹Our IV estimate is lower than in the original Harmon and Walker (1995) study and consequently not significantly different from the OLS estimate. This probably reflects the fact that the sample we use (1978-1995) is later than theirs (1978-1986) and so contains proportionately less individuals affected by the change in the school leaving law. When we use a sample covering the same period, our IV estimate is 11.35 with a standard error of 3.11, which is closer to theirs.

we generate 100 different artificial datasets that have second moment properties similar to the real (LFS) data so that the OLS estimates are all (asymptotically) identical.

In other words, for each set of parameters (α , β , and γ) we choose $(\nu_\varepsilon^r, \nu_\eta^r, \nu_z)$ so that (i) the overall (unconditional) covariance matrix to equal the overall (unconditional) covariance matrix of the real data (-); (ii) the individual heteroscedastic variances $(\nu_\varepsilon^r, \nu_\eta^r)$ match, roughly, the conditional reduced form heteroscedasticity (-^r) observed in the LFS data. Using these parameters, we generate samples of 500 observations for each region-time group (500 being the average cell size in the LFS). Hence, for each cell we compute a “small sample” covariance matrix that has unconditional variance similar to the LFS data, and conditional variation across groups also consistent with the data. We repeat this procedure 100 times. We then apply the OLS and IH estimators to each artificial dataset and report their distributions relative to the true values of the parameters.

When we conducted the simulation we allowed β to take on values between 0.05 and 0.225 which more or less covers the range of returns to education reported in the literature.²⁰ Lacking any concrete idea as to what the value of α should be, we simply allowed it to take on values from -5 to $+5$. Similarly, we have no strong priors on γ , we allow it to take on values in the interval -1 to 2 .²¹

Figures 4 and 5 illustrate these results graphically for β , the return to education parameter.²² In both, the true value of the parameter is denoted by the height of the light grey bar, with each bar representing a different structural form, i.e. a different combination of α, β and γ . In figure 4 the corresponding value of the OLS estimator is denoted by the black dot, with the vertical black bar denoting the 95 percent confidence interval of the distribution of the OLS estimates over the 100 data sets. Clearly, the OLS estimator is extremely biased changing very little even as the structural form changes dramatically. In fact the OLS estimates are almost identical across all the artificial datasets. This is the case (almost) by construction as we created these datasets so that they would all have second moments close to the true data. Asymptotically the OLS estimator should be identical in all simulations.

In contrast figure 5 shows that the IH estimator is close to the true value and estimated, largely with high precision. The exceptions represent those α , β , and γ combinations where the rank condition is close to failure. The IH point estimate is denoted by the horizontal line, with the vertical black line indicating the spread of the distribution of IH estimate i.e. the 95 percent confidence interval. Note that the small sample biases is negligible for small values of the true β . For larger values the bias is usually negative, but still small.

[Figures 4 and 5 here]

Note that in our simulations the IH estimates will be statistically different from OLS for true values of β that are sufficiently far from the OLS estimates - basically, IH tracks the true coefficient better than OLS. From a methodological point of view, this suggests that heteroscedasticity can be

²⁰See table 2 of Card (2001), for a summary of the results of 11 major studies of returns to education using IV methods.

²¹Notice that the errors in variable problem is a special case in this simulation - i.e. when $\gamma = -\beta$.

²²We present only a graphical summary of the results on the Monte Carlo simulation. An appendix presenting the detailed results is available upon request.

used to solve problem of identification when omitted variables, measurement error, and standard simultaneity issues arise in cross-sectional data.

5 Conclusions

In this paper we estimated the returns to education controlling for the endogeneity of education, unobserved ability and measurement error, using only the natural heteroscedasticity observed in the data. In essence the “Identification through Heteroscedasticity” (IH) estimator uses the random shocks to perform a role similar to standard instruments, generating an exogenous change in one variable allowing us to identify the effect on the other. Unlike, IV however, the IH method does not rely on exclusion restrictions nor on the use of natural experiments.

We applied the method to Labour Force Survey data from the UK and showed that the return to education over the sample period was approximately 6.1% for men. This estimate is close in magnitude to the OLS (6.8%) and much more precise than the returns estimated by the usual IV techniques. This could be due to the fact that IV estimates usually utilise natural experiments that fall foul of the “weak instruments” critique of Staiger and Stock (1997) whereas the IH estimator performs better because it makes use of the naturally strong heteroscedasticity in the data.

The closeness of the OLS and IH estimates suggests that the biases induced by unobserved ability and measurement error appear to offset each other. This lends support to the hypothesis of Griliches (1977). We also noted that standard simultaneity between wage and education (once we control for ability) did not appear to be important. We checked the robustness of our method by applying it to an alternative dataset (UK Family Expenditure Survey). The results were quite similar to those for the Labour Force Survey.

We also presented Monte Carlo simulations to show that the finite sample bias of the IH estimator is small. Furthermore, the Monte Carlo exercise shows that the IH estimator is not biased towards the OLS estimate, indicating that the closeness between the IH and the OLS estimates is not the result of small sample bias.

In our particular application, we required the structural model to be linear, parameters to be stable, and some of the shocks to be homoscedastic. In principle, the IH estimator can be extended to allow for the possible heterogeneity of returns across individuals, and also to account for the discrete nature of education choices. This is beyond the scope of the present paper and left for future research. Finally, our results suggest that Identification Through Heteroscedasticity is a viable alternative strategy for the solution of the identification problem in the presence of omitted variables, measurement errors, and simultaneity.

References

- [1] Angrist, Joshua D. and Krueger, Alan B. (1991), “Does Compulsory School Attendance Affect Schooling and Earnings”, *Quarterly Journal of Economics*, vol. 106, pp. 979-1014
- [2] Belzil, Christian and Hansen, Jorgen (2002), “Unobserved Ability and the Return to Schooling”, *Econometrica*, forthcoming.
- [3] Card, David (2001), “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems”, *Econometrica*, vol. 69, no. 5, pp. 1127-1160.
- [4] DuFlo, Esther (2001), “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment”, *American Economic Review*, vol. 91, pp. 795-813.
- [5] Griliches, Zvi (1977), “Estimating Returns to Schooling: Some Econometric Problems”, *Econometrica*, vol. 45, pp. 1-22.
- [6] Harmon, Colm and Walker, Ian (1995), “Estimates of The Economic returns to Schooling in the United Kingdom”, *American Economic Review*, vol. 85, pp. 1278-1286.
- [7] Hogan, Vincent and Walker, Ian (2002), “Education Choice Under Uncertainty”, mimeo, University College Dublin.
- [8] Katz, Lawrence and Autor, David (1999), “Changes in the Wage Structure and Inequality”, in Orey Ashenfelter and David Card, editors, *Handbook of Labor Economics*, vol. 3A. North Holland, Amsterdam and New York.
- [9] King, Mervyn, Enrique Sentana and Sushil Wadhvani (1994), “Volatility and Links between National Stock Markets”, *Econometrica*, vol. 62, pp. 901-933.
- [10] Klein, Roger W. and Vella, Francis (2000a), “Employing Heteroscedasticity to Identify and Estimate Triangular Semiparametric Models”, mimeo, Rutgers University.
- [11] Klein, Roger W. and Vella, Francis (2000b), “Identification and Estimation of the Binary Treatment Model Under Heteroscedasticity”, mimeo, Rutgers University.
- [12] Manski, Charles and Pepper, John (2000), “Monotone Instrumental Variables: with an Application to the Returns to Schooling”, *Econometrica*, vol. 68, pp. 997-1013.
- [13] Mincer, Jacob (1974), *Schooling, Experience and Earnings*, Columbia University Press, New York.
- [14] Rigobon, Roberto (2000), “Identification Through Heteroscedasticity: Measuring: ‘Contagion between Argentinian and Mexican Sovereign Bonds’”, Working Paper 7493, National Bureau of Economic Research.
- [15] Rigobon, Roberto and Sack, Brian (2002), “The Impact of Monetary Policy on Asset Prices”, Working Paper 8794, National Bureau of Economic Research

- [16] Rummery, S., Vella, F. and Verbeek, M. (1999), “Estimating the Returns to education for Australian Youth via Rank-Order Instrumental Variables”, *Labour Economics*, vol. 6, pp. 491-508.
- [17] Sentana, Enrique and Fiorentini, Gabriele (2001), “Identification, Estimation and testing of Conditional Heteroscedastic Fctor Models”, *Journal of Econometrics*, vol. 102, no. 2, pp. 143-164
- [18] Staiger, Douglas and Stock, James (1997), “Instrumental Variable Regression with Weak Instruments”, *Econometrica*, vol. 65, pp. 557-586.
- [19] 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.
- [20] Wright, Philip (1928), *The Tariff on Animal and Vegetable Oils*, MacMillan, New York

A Monte Carlo Simulations

To begin the simulation re-write expression (4) for - in vector form as (6). Note that absence of a the superscript “ r ” implies that the parameter is calculated over the whole sample i.e. with all cells aggregated together.

$$\begin{bmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{22} \end{bmatrix} = A \begin{bmatrix} \nu_\varepsilon \\ \nu_\eta \\ \nu_z \end{bmatrix} \quad (6)$$

$$A = \frac{1}{(1 - \beta\alpha)^2} \begin{bmatrix} 1 & \beta^2 & (\beta + \gamma)^2 \\ \alpha & \beta & (1 + \alpha\gamma)(\beta + \gamma) \\ \alpha^2 & 1 & (1 + \alpha\gamma)^2 \end{bmatrix}$$

Equation (6) implies that for each combination of parameters α , β , and γ there exists a unique set of variances of the structural shocks (over all cells aggregated together) that match the covariance matrix of the reduced form.

For each combination of parameters α , β , and γ , the simulation proceeds as follows:

1. Calculate the feasible set of structural variances. For every combination of parameters $\{\alpha, \beta, \gamma\}$ we solve (6) for the vector of structural variances $\{\nu_\varepsilon, \nu_\eta, \nu_z\}$ where $\{\omega_{11}, \omega_{12}, \omega_{22}\}$ are taken from the reduced form covariance matrix of the actual (LFS) data. If any of the variances in $\{\nu_\varepsilon, \nu_\eta, \nu_z\}$ is negative or if the matrix A is not invertible for a particular combination of $\{\alpha, \beta, \gamma\}$ then that combination is inconsistent with the original data and is excluded from further consideration.
2. Given the identifying restriction $\nu_z = \nu_z^r$, compute the set of $\{\nu_\varepsilon^r, \nu_\eta^r\}$ to match the variation of the reduced-form covariance matrix across the region-time cells. In other words for every $r \in [1..144]$, solve (7) for $\{\nu_\varepsilon^r, \nu_\eta^r\}$ given $\{\omega_{11}^r, \omega_{12}^r, \omega_{22}^r\}$ from the data

$$\begin{bmatrix} \omega_{11}^r \\ \omega_{12}^r \\ \omega_{22}^r \end{bmatrix} = A \begin{bmatrix} \nu_\varepsilon^r \\ \nu_\eta^r \\ \nu_z^r \end{bmatrix} \quad (7)$$

3. Use the set of cell specific variances $\{\nu_\varepsilon^r, \nu_\eta^r\}$ and the overall sample variance $\{\nu_\varepsilon, \nu_\eta\}$ to compute the standard deviation of the parameters $\{\nu_\varepsilon^r, \nu_\eta^r\}$ across the cells of the artificial data. We then assume that these variances are themselves parameters that are distributed normally across the cells (with mean $\nu_\varepsilon, \nu_\eta$). Take 144 draws from this distribution and assign one $\{\nu_\varepsilon^r, \nu_\eta^r\}$ pair to to each cell in the dataset.²³

²³Alternatively, we can randomly assign from the actual set of $\{\nu_\varepsilon^r, \nu_\eta^r\}$ that we calculated above. However, the actual distribution of $\{\nu_\varepsilon^r, \nu_\eta^r\}$ has thicker tails than the normal distribution. As the IH estimator works off differences in variances, using the normal rather than actual distribution, biases the procedure against the IH estimator. When we implemented this procedure the IH estimates were even less biased than reported in the text.

4. For each of the 144 cells, take 500 draws from the distribution of the structural residuals, assumed independently normally distributed with mean zero and variance $\{\nu_\varepsilon^r, \nu_\eta^r, \nu_z\}$. Use parameters $\{\alpha, \beta, \gamma\}$ and equation (3) to calculate the implied reduced form residuals. Then, for each cell we compute Σ^r , the covariance matrix of these residuals. By construction, these will be close to the covariance matrices of the real data – but not identical due to the randomisation process in step 3.
5. The final step is to compute the OLS and IH estimators on the artificial data.

The process (steps 1-5) is repeated 100 times for each set of parameters $\{\alpha, \beta, \gamma\}$ and we report the summary statistics for the distribution of the estimators.

Table 1: UK Labour Force Survey (Men)

Variable	Definition	Mean	Stn. Dev
age	age at interview	40.61	9.28
wage	usual real wage (stg. £ per hour)	9.71	6.53
educ	age left school - 5	12.28	2.69
union	=1 if union member	0.49	0.59
health	=1 if health can inhibit work	0.07	0.26
nonwhite	ethnic background (=1 if nonwhite)	0.04	0.19
cohab	=1 if cohabiting	0.09	0.29
married	=1 if married	0.71	0.45

1. Statistics are calculated for the pooled cross section (1993-2000)
2. Sample Size = 70,953

Table 2: OLS and IV Estimates of the Return to Education using UK Labour Force Survey (1993-2000)

Dependent Variable: logwage		
Method	(1) OLS	(2) IV2
Years of education/(100)	6.78 (0.06)	5.21 (2.93)
\bar{R}^2	0.21	0.21
F-test of Instruments (df_1, df_2)	-	1.46 (24,70730)
Sample Size	70,953	70,953

1. Standard errors are in parentheses
2. Regressions include a constant, quadratic in age, race, health, marital status, union membership plus region and year dummies
3. IV1: Interaction of quarter of birth and year as in Angrist and Krueger (1991)

Table 3: IH Estimates using UK LFS (1993-2000)

	α	β
Point Estimate	-0.0844	0.0607
Mean of Distribution	0.0933	0.0586
St.Dev. of Distribution	0.2214	0.0068
Point / St.Dev	-0.38	8.92
95% C.I. ²	[-0.2964, 0.4308]	[0.0468, 0.0701]
Maximum	0.7055	0.0770
Minimum	-0.5770	0.0364
Mass Below Zero	31.34%	0.00%

1. Standard errors calculated from 500 Monte Carlo draws from the distribution of $x^r \sim N(x^r, \frac{2}{N^r} \hat{r} \hat{r}')$ where $x^r = vec(-r)$
2. Calculated directly from the simulated distribution of the estimator

Table 4: Estimates of Return Across Years using LFS

	1993	1994	1995	1996
Point Estimate	0.0462	0.0355	0.0701	0.0543
Mean of Distribution	0.0455	0.0359	0.0683	0.0534
St. Dev. of Distribution	0.0140	0.0120	0.0152	0.0123
Point / St.Dev.	3.30	2.95	4.62	4.43
Maximum	0.0825	0.0722	0.1123	0.0896
Minimum	0.0011	0.0044	0.0259	0.0198
Mass Below Zero	0.00%	0.00%	0.00%	0.00%

	1997	1998	1999	2000
Point Estimate	0.0628	0.0251	0.0578	0.0940
Mean of Distribution	0.0586	0.0263	0.0492	0.0919
St.Dev. of Distribution	0.0260	0.0284	0.0321	0.0190
Point / St.Dev.	2.41	0.89	1.80	4.94
Maximum	0.1504	0.0927	0.1466	0.1463
Minimum	-0.0202	-0.0681	-0.0873	0.0241
Mass Below Zero	1.20%	17.17%	7.58%	0.00%

1. Standard errors calculated from 500 Monte Carlo draws
from the distribution of $x^r \sim N(x^r, \frac{2}{Nr} \hat{r} \hat{r}')$ where $x^r = vec(-r)$

Table 5: Estimates of the Return to Education using UK Family Expenditure Survey (1978-1995)

Dependent Variable: logwage

Method	(1) OLS	(2) IV	(3) IH
Years of education/(100)	6.03 (0.12)	7.29 (0.27)	4.45 (0.81)
\bar{R}^2	0.21	0.21	-
F-test of Instruments (df_1, df_2)	-	24.36 (2,24240)	-
Sample Size	24,266	24,266	24,266

1. Standard errors are in parentheses
2. Regressions include a constant, quadratic in age, plus region, year dummies and their interactions
3. IV: School leaving law as in Harmon and Walker (1995)

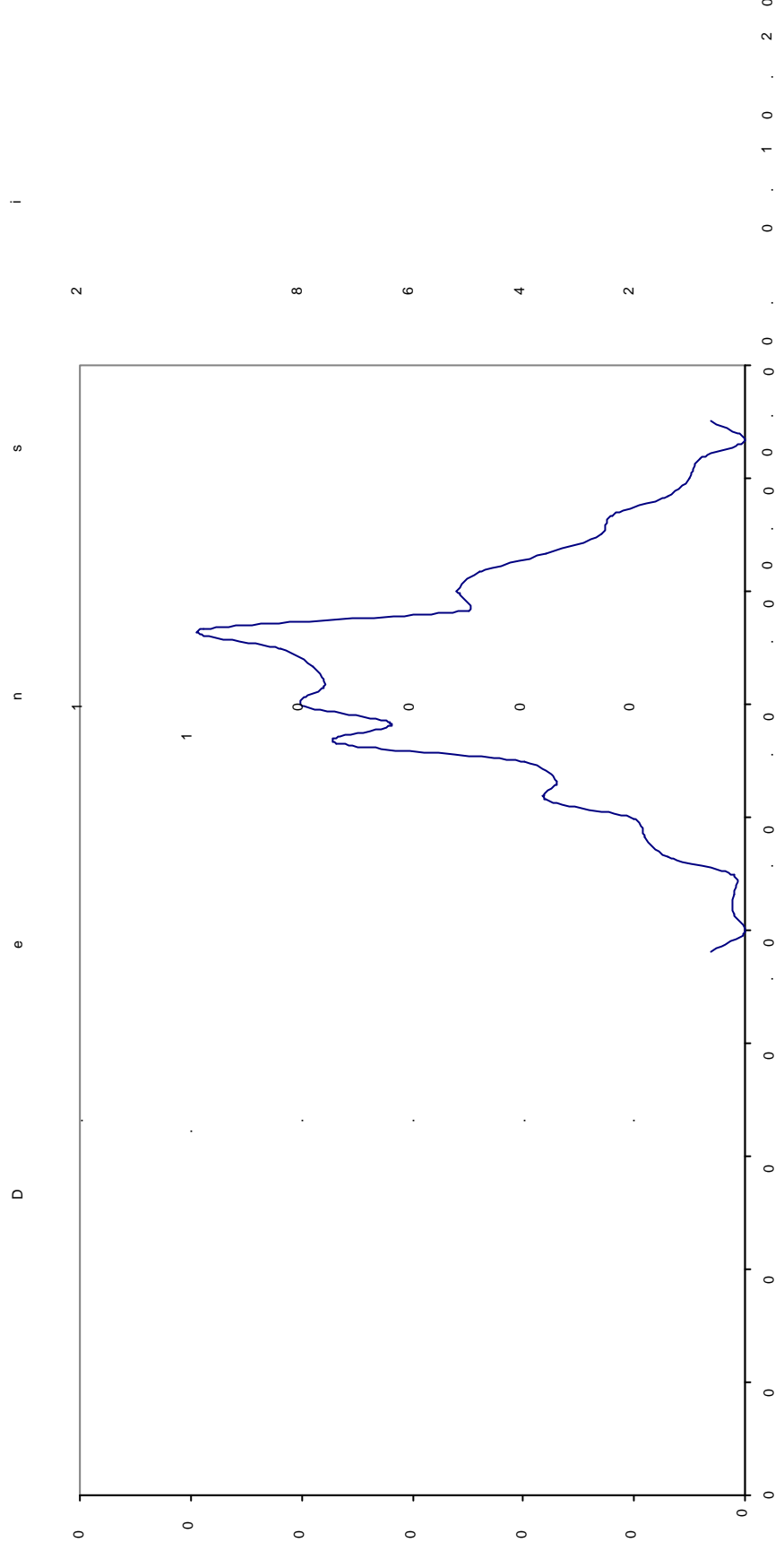


Figure 1:

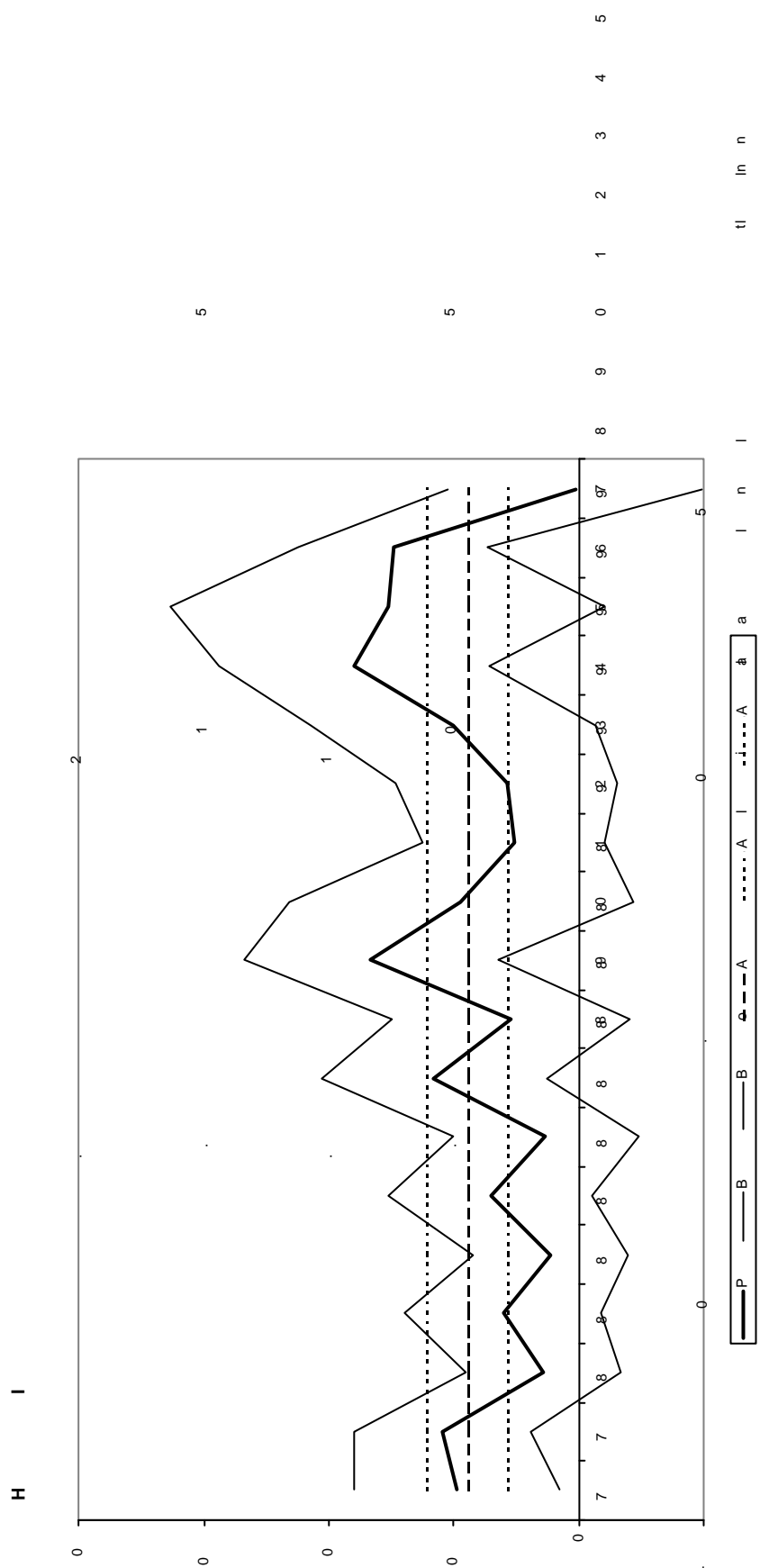
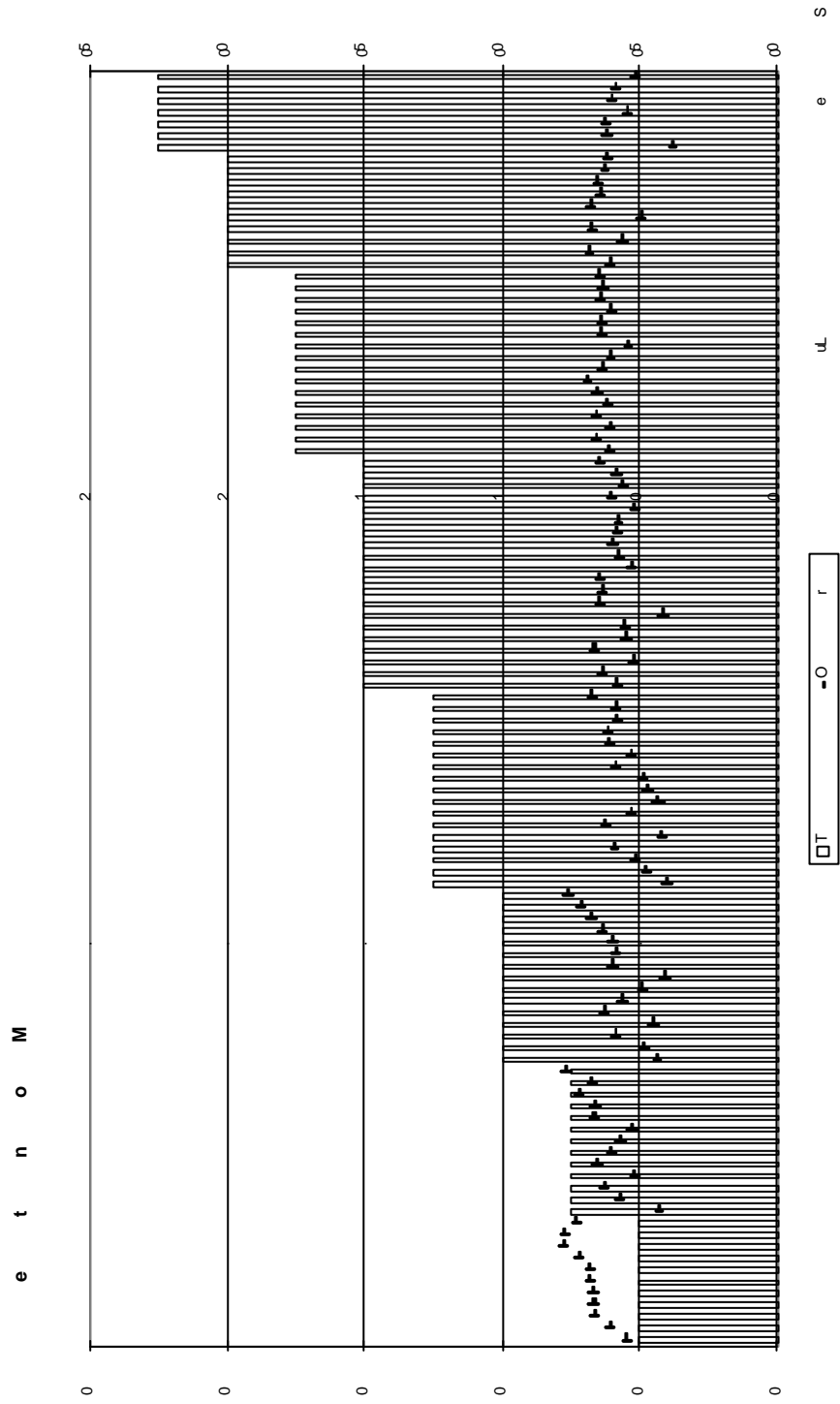
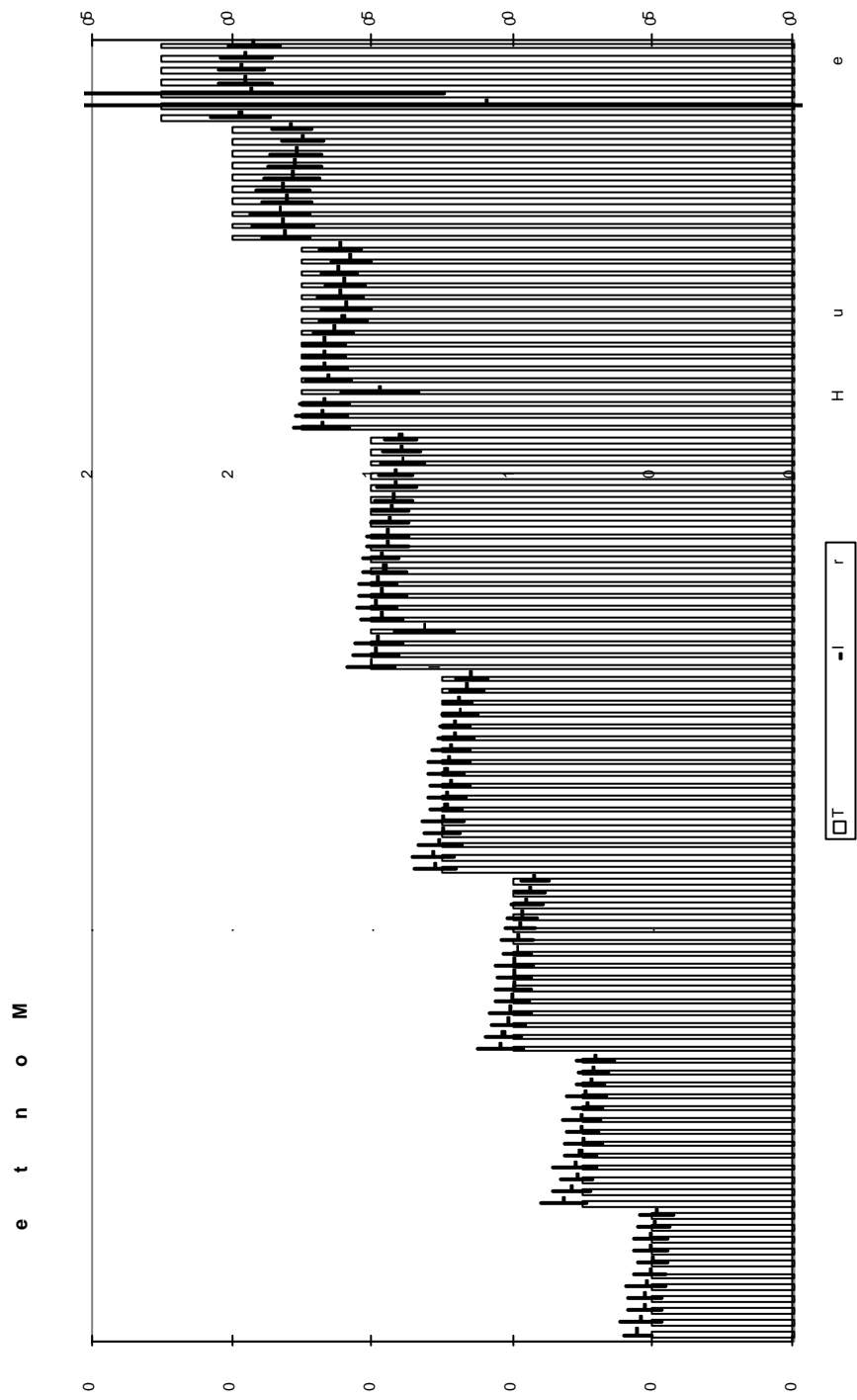


Figure 3:



24
Figure 4:



25
Figure 5: