Average and Marginal Returns to Upper Secondary Schooling in Indonesia

Pedro Carneiro†

Michael Lokshin

Cristobal Ridao-Cano

Nithin Umapathi

November 27, 2015

Abstract:

This paper estimates average and marginal returns to schooling in Indonesia using a semi-parametric selection model. Identification of the model is given by exogenous geographic variation in access to upper secondary schools. We find that the return to upper secondary schooling varies widely across individuals: it can be as high as 50 percent per year of schooling for those very likely to enroll in upper secondary schooling, or as low as -10 percent for those very unlikely to do so. Average returns for the student at the margin are substantial, but they are also well below those for the average student attending upper secondary schooling.

JEL Code: J31

Key words: Returns to Schooling, Marginal Return, Average Return, Marginal Treatment Effect

† Pedro Carneiro is at University College London, Institute for Fiscal Studies, and Centre for Microdata Methods and Practice. Nithin Umapathi, Michael Lokshin, and Cristobal Ridao-Cano are at the World Bank. We thank Martin Ravallion for very useful comments. Carneiro and Umapathi gratefully acknowledge the financial support from the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001) and the hospitality of the Development Economic Research Group of the World Bank. Carneiro gratefully acknowledges the support of ESRC-DFID (RES-167-25-0124), the European Research Council through ERC-2009-StG-240910-ROMETA and ERC-2009-AdG-249612. These are the views of the authors and do not reflect those of the World Bank, its Executive Directors, or the countries they represent. Correspondence by email to mlokshin@worldbank.org

1. Introduction

The expansion of access to secondary schooling is at the center of development policy in most of the developing world. Analyzing the effects of such expansions requires knowledge of the impact of education on earnings for those affected by the expansions. In contrast with the standard model, much of the recent literature on the returns to schooling emphasizes that returns vary across individuals, and are correlated with the amount of schooling an individual takes (e.g., Card, 2001, Carneiro, Heckman and Vytlacil, 2011). In terms of the traditional Mincer equation, $Y = a + bS + u$ (where Y is log wage and S is years of schooling), b is a random coefficient potentially correlated with S. This has dramatic consequences for the way we conduct policy analysis.

In this model one could define multiple average returns of interest, which are substantially different from each other. The individual at the margin between two levels of schooling may have very different returns from all the infra-marginal individuals. Standard instrumental variables estimates of the returns to schooling estimate the Local Average Treatment Effect (or LATE; Imbens and Angrist, 1994), which may or may not be close to the return to the marginal person (who is more likely to be affected by the expansion of secondary schooling than anyone else in the economy). Furthermore, different policies may affect different groups of individuals.

This paper studies the returns to upper secondary schooling in Indonesia in a setting where b varies across individuals and it is correlated with S (which in this paper is a dummy variable indicating whether an individual enrolls in upper secondary school or not). We find that the return to upper secondary schooling for the *marginal* person (who is indifferent between going to secondary schooling or not) is substantial, but much lower than the returns for the *average* person enrolled in upper secondary schooling (14.2% vs. 26.9% per year of schooling). Finally, we simulate what would happen if distance to upper secondary schooling was reduced by 10% for everyone in the sample, and we estimate that the return to upper secondary schooling for those induced to attend schooling by such an incentive is 14.2%.

When evaluating marginal expansions in access to school, the relevant quantities are the returns and costs for the marginal student, not the returns and costs for the average student. In spite of the importance of this topic, there are hardly any estimates of average and marginal returns to schooling in developing countries. Two exceptions using Chinese data are Heckman and Li (2004) and Wang, Fleisher, Li and Li (2011).

We estimate a semi-parametric selection model of upper secondary school attendance and wages using the method of local instrumental variables (Heckman and Vytlacil, 2005). Our data comes from the Indonesia Family Life Survey. Carneiro, Heckman and Vytlacil (2011) use a similar model to estimate the returns to college in the US. Although they examine a different country, time period, and level of schooling, they also find that the returns to education vary across individuals in the US, and that the return to education for the marginal student is well below the return to college for the average student (see also Carneiro and Lee, 2009, 2011). These papers document, across very different environments, how important it is to account for heterogeneity in the returns to schooling.

This paper also proposes a methodological innovation. In the presence of multiple control variables, the construction of various parameters (average returns for different groups of individuals) using the framework of Heckman and Vytlacil (2005) requires the estimation of conditional densities, where the conditioning set is of high dimensionality. These estimators are notoriously difficult to implement. We use instead a simulation method that avoids such a high dimensional non-parametric estimation problem (in contrast, Carneiro, Heckman and Vytlacil, 2010, 2011, impose a restrictive index assumption to reduce the dimensionality of the problem).

Since schooling is endogenously chosen by individuals, we require an instrumental variable for schooling. We use as the instrument the distance (in kilometers) from the community of residence to the nearest secondary school (see also Card, 1995). Distance takes the value zero if there is a school in the community of residence. This variable is a strong determinant of enrolment in upper secondary school. One could be concerned that the forces driving the location of schools and parents are correlated with wages, implying that distance is an invalid instrument. Below we discuss this problem in detail.

In addition, we are not able to reliably measure distance to school at the time of the relevant schooling decision, and use current distance instead. One major drawback of this approach is that schooling is likely to be correlated with migration to more urban areas, which are areas where distance to school is smaller.

We control for several family and village characteristics, namely father's and mother's education, an indicator of whether the community of residence was a village, religion, whether the location of residence is rural, province dummies, and distance from the village of residence to the nearest health post. In order for our instrumental variable to be valid our assumption would have to be that if we take two individuals with equally educated parents, with the same religion, living in a village at age 12 which is located in an area that is equally rural, in the same current province, and at the same current distance of a health post, then current distance to the nearest secondary school is uncorrelated with direct determinants of wages other than schooling. We discuss under what circumstances these assumptions are more likely to hold, and also discuss potential consequences of deviations from these assumptions.

Our instrumental variable estimates of the returns to schooling are higher than the returns to schooling for Indonesia estimated in Duflo (2000), with the qualification that the dataset, the instrumental variable, and the time period are not the same. Pettersson (2010) finds similar rates of return using the same year and same data as us, but a different sample and a different instrument variable.

The standard errors in our instrumental variables estimates greatly exceed those of the standard least squares estimates, but this is typical in the literature on the returns to education, and in that sense our paper does not differ than many other papers on this topic. This imprecision transpires to our semi-parametric estimates since they rely essentially on an instrumental variable method. In spite of this, we strongly reject the null hypothesis that there is no selection on returns to education in our data, which justifies our procedure and the emphasis we place on heterogeneous returns.

This paper proceeds as follows. Section 2 discusses the data. Section 3 reviews the econometric framework. Section 4 presents our empirical results. Section 5 concludes.

2. Data

We use data from the third wave of the Indonesia Family Life Survey (IFLS) fielded from June through November, 2000. For a detailed description of the survey see Strauss, Beegle, Sikoki, Dwiyanto, Herawati and Witoelar (2004). In the appendix we list the main variables we use.

The IFLS is a household and community level panel survey that has been carried out in 1993, 1997 and 2000. The sample was drawn from 321 randomly selected villages, spread among 13 Indonesian provinces containing 83% of the country's population. The sub-sample we use consists of males aged 25-60 who are employed, and who have reported non-missing wage and schooling information. We consider salaried workers, both in the government and in the private sector. We exclude females from the analysis because of low labor force participation, and we exclude self-employed workers because it is difficult to measure their earnings. The dependent variable in our analysis is the log of the hourly wage. Hourly wages are constructed from self-reported monthly wages and hours worked per week. The final sample contains 2608 working age males.

In our empirical model we collapse schooling into two categories: i) completed lower secondary or below, and ii) attendance of upper secondary or higher. While this division groups together several levels of schooling, it simplifies the model and is standard in the literature (e.g., Willis and Rosen, 1979). The transition to upper secondary schooling is of interest in the Indonesian context given its current effort to expand secondary education. We present both the return to upper secondary schooling, and an annualized version of this parameter, obtained by dividing the estimated return by the difference in average years of schooling completed by those with lower secondary or less, and those with upper secondary or more. Upper secondary schooling corresponds to 10 or more years of completed education. In order to compare our estimates with the literature (say, Duflo, 2000), in Appendix B we also present least squares (OLS) and IV estimates of returns using a continuous education variable, corresponding to years of completed schooling.

The control variables in our models are indicator variables for age, indicators for the level of schooling completed by each of the parents (no education, elementary education, secondary education, and an indicator for unreported parental education), an indicator for whether the individual was living in a village at age 12, indicators for the province of residence, an indicator of rural residence, and distance (in kilometers) from the office of the head of the community of residence to the nearest community health post.

Our instrumental variable for schooling is the distance (in kilometers) from the office of the community head to the nearest secondary school (i.e., of all the schools in each community, we take the one that is closest to the office of the head). The distance is self-reported by the community head in the Service Availability Roster of the IFLS.[1]

Table 1 presents descriptive statistics for the main variables used in our analysis. It shows that individuals with upper secondary or higher levels of education have, on average, 108% higher wages than those with lower education. They have 7.79 more years of schooling. They are younger than those without and upper secondary education. They are more likely to have better-educated parents, to have lived in towns or cities at age 12, and to live closer to upper secondary schools, than those with less education.

3. Theoretical Framework

3.1. A Semi-Parametric Selection Model

This section of the model follows Heckman and Vytlacil (2005). We repeat part of the presentation in that paper because it lays out the empirical model we use, and provides the basis for discussing a new approach to estimating some of our parameters.

We consider a standard model of potential outcomes applied to schooling, as in Willis and Rosen (1979) or Carneiro, Heckman and Vytlacil (2010, 2011). Consider a model with two schooling levels:

$$\begin{aligned} Y_1 &= \alpha_1 + X\beta_1 + U_1 \\ Y_0 &= \alpha_0 + X\beta_0 + U_0 \end{aligned} \qquad (1)$$

$$S = 1 \text{ if } Z\gamma - U_s > 0 \qquad (2)$$

Y_1 are log wages of individuals if they have upper secondary education and above, Y_0 are log wages of individuals if they do not have upper secondary education, X is a vector of observable characteristics which affect wages, and U_1 and U_0 are the error terms. Z is a vector of characteristics affecting the schooling decision.

In theory, agents decide whether to enroll or not in upper secondary schooling based on the expected net present value of earnings with and without upper secondary

[1] We would have liked to use instead the distance between the community of residence in childhood and the nearest school in childhood. It is in theory possibly to construct such a variable because there should be information about the opening date for all schools in the sampled communities. Unfortunately, many of these dates are missing, and using this information would result in a drop in our sample size of more than 50%, and in hopelessly imprecise estimates. Our assumption is that current residence and current school availability are good approximations to the variables we need (as in Card, 1995). We show below that this measure of distance to school is a good predictor of upper secondary school attendance, and discuss in detail how likely is it that our assumption holds, and the consequences of its invalidity.

schooling, and costs, which can be financial or not. There can be liquidity constraints. There is heterogeneity and we expect agents with the highest returns to upper secondary schooling ($Y_1 - Y_0$) to be more likely to enroll in higher levels of schooling. Costs and returns to schooling can be correlated. Willis and Rosen (1979) and Carneiro, Heckman and Vytlacil (2011) show how we can approximate the schooling decision just described with a model such as the one in equation (2).

It is convenient to rewrite equation (2) as:

$$S = 1 \text{ if } P(Z) > V \qquad (3)$$

$P(Z) = F_{U_S}(Z\gamma)$ and $V = F_{U_S}(U_S)$ and F_{U_S} is a cumulative distribution function of U_s . V is uniformly distributed by construction. This is an innocuous transformation given that U_S can have any density, but it is very convenient, as shown below.

Observed wages can be written as:

$$Y = SY_1 + (1-S)Y_0 \qquad (4)$$

And the return to schooling can be written as:

$$Y_1 - Y_0 = \alpha_1 - \alpha_0 + X(\beta_1 - \beta_0) + U_1 - U_0 \qquad (5)$$

Notice that the return to schooling varies across individuals with different X's and different U_1, U_0. This is an important feature of this framework and of our paper, which emphasizes heterogeneity in returns (and the distinction between the returns for average and marginal individuals).

In order to credibly identify the parameters of the model in equations (1) and (2), it is important that standard IV-type assumptions are satisfied. In particular, we require that Z is independent of (U_1, U_0) given X, and that Z is correlated with S (see Heckman and Vytlacil, 2005, for the full set of assumptions).

In practice, we will appeal to an even stronger assumption: that X and Z are independent of U_1, U_0, U_S. This stronger assumption is quite standard in empirical applications of a selection model of the type described here. We discuss the advantages of using this stronger assumption in the empirical section (see also Carneiro, Heckman and Vytlacil, 2011). One way to make this assumption more palatable in practice is to interpret the coefficients on X in the wage equations as capturing not only the impact of those variables, but also the impacts of changes in unobservables on wages, as they are projected into the *X*. Nevertheless, we will assume that the remaining unobservables are not only orthogonal to *X*, but they are fully independent of *X*.

The marginal treatment effect (MTE) is the central parameter of our analysis. In the notation of our paper it can be expressed as:

$$\begin{aligned} MTE(x,v) &= E(Y_1 - Y_0 \mid X = x, V = v) \\ &= \alpha_1 - \alpha_0 + x(\beta_1 - \beta_0) + E(U_1 - U_0 \mid X = x, V = v) \end{aligned} \tag{6}$$

The MTE measures the returns to schooling for individuals with different levels of observables (*X*) and unobservables (*V*), and therefore it provides a simple characterization of heterogeneity in returns.

To be specific, suppose that *X* is maternal education. $(\beta_1 - \beta_0)$ could be positive or negative depending on whether children with more educated mothers have higher or lower than average returns to schooling. The first case would indicate that maternal schooling and child schooling are complementary inputs in the production of skill (which eventually feeds into wages), while the second case would say that they are substitutes.

Similarly, one possible interpretation of *V* is as the negative of unobserved ability: individuals with high values of *V* (or low ability) are less likely to enroll in school than those with low values of *V*. Then, $E(U_1 - U_0 \mid X = x, V = v)$ would tell us how the returns to schooling varied with unobserved ability. If individuals with high ability also had higher returns, then this function should be declining in *V*.

In addition, Heckman and Vytlacil (2005) show how to construct several parameters of interest as weighted averages of the MTE. For example:

$$\begin{aligned} ATE(x) &= \int MTE(x,v) f_{V|x}(v \mid x) dv \\ ATT(x) &= \int MTE(x,v) f_{V|x}(v \mid x, S = 1) dv \\ ATU(x) &= \int MTE(x,v) f_{V|x}(v \mid x, S = 0) dv \end{aligned} \tag{7}$$

where *ATE(x)* is the average treatment effect, *ATT(x)* is average treatment on the treated, *ATU(x)* is average treatment on the untreated (conditional on *X*=*x*), and $f_{V|X}(v|x)$ is the density of *V* conditional on *X*.[2] The MTE can be used to build many other parameters.

[2] Notice that $f_{V|x}(v|x) = 1$, because $v|x$ is uniformly distributed by assumption. Heckman and Vytlacil (2005) do not use exactly this representation of the parameters. For example, they write: $ATT(x) =$

A less standard parameter but equally (if not more) important is the policy relevant treatment effect (PRTE), introduced in the literature by Heckman and Vytlacil (2001b). It measures the average return to schooling for those induced to change their enrolment status in response to a specific policy (see also the related parameter of Ichimura and Taber, 2000). Obviously, this parameter depends on the policy being evaluated.

Consider a determinant of enrolment Z, which does not enter directly in the wage equation. The policy shifts Z from Z=z to Z=z'. We can write the PRTE as:

$$PRTE(x) = \int MTE(x,v) f_{V|x}(v|x, S(z)=0, S(z')=1) dv$$

3.2. Estimating the *MTE*

Assuming that the unobservables in the wage (1) and selection (2) equations are jointly normally distributed, the MTE could be estimated using a standard (parametric) switching regression model (see Heckman, Tobias and Vytlacil, 2001). Assume:

$$U_0, U_1, U_s \sim N(0, \Omega) \qquad (8)$$

where Ω represents the variance and covariance matrix. Under this assumption:

$$MTE(x,v) = E(Y_1 - Y_0 \mid X = x, V = v) = (\alpha_1 - \alpha_0) + x(\beta_1 - \beta_0) + \left(\frac{\sigma_{U_S,1}}{\sigma_{U_S}} - \frac{\sigma_{U_S,0}}{\sigma_{U_S}}\right)\Phi^{-1}(v) \qquad (9)$$

where $\sigma^2_{U_S}$ denotes variance of U_s, σ^2_i variance of U_i with i = 0,1, $\sigma^2_{U_S,i}$ covariance between U_s and U_i, $\sigma^2_{i,j}$ the covariance between U_i and U_j and Φ is the c.d.f. of the standard normal. Therefore MTE can be constructed by estimating parameters $\alpha_1, \alpha_0, \beta_1, \beta_0, \gamma$ and the matrix Ω.

One advantage of imposing restriction (8) is that the resulting model is well studied, and can be readily estimated using a variety of statistical packages. However, this model relies on strong assumptions about the distribution of the error terms in equations (1-2), which can be unattractive in several circumstances. For example, these restrictions impose that the MTE is a linear function of U_s ($U_S = F^{-1}{}_{U_S}(V)$), as in equation (9).

$\int MTE(x,v) h_{TT}(v|x) dv$, where $h_{TT}(v|x)$is a parameter weight (in this case, the parameter for TT). Our representation is equivalent since $h_{TT}(v|x)$ in their paper can be shown to be equal to $f_{V|x}(v|x, S=1)$.

To relax these restrictions, and allow for more flexible functions for the MTE, we use the method of local instrumental variables that imposes no distributional assumptions on the unobservables of the model (Heckman and Vytlacil, 2000), besides the assumption that X and Z are independent of U_1, U_0, U_S (which is also important for credible estimation of the model under normality assumptions).

This is a two step procedure. The first step of this procedure is to estimate a regression of the outcome Y, on X and P. We can write it as:

$$\begin{aligned} E(Y \mid X, P) &= E\left[\alpha_0 + X\beta_0 + S(\alpha_1 - \alpha_0) + SX(\beta_1 - \beta_0) + U_0 + S(U_1 - U_0) \mid X, P\right] \\ &= \alpha_0 + X\beta_0 + P(\alpha_1 - \alpha_0) + PX(\beta_1 - \beta_0) + E(U_1 - U_0 \mid S = 1, X, P)P \qquad (10) \\ &= \alpha_0 + X\beta_0 + P(\alpha_1 - \alpha_0) + PX(\beta_1 - \beta_0) + K(P) \end{aligned}$$

K(P) is a function of P, which we want to be flexible. Therefore, we will estimate it using a non-parametric procedure, such as, for example, local linear regression.

Once this regression is estimated, notice that, taking the derivative of (10) with respect to P we get the MTE:

$$MTE(x, v) = \frac{\partial E(Y \mid X, P)}{\partial P} \Big|_{X=x, P=v} = X(\beta_1 - \beta_0) + K'(P) \qquad (11)$$

Therefore , the local instrumental variables estimator of Heckman and Vytlacil (2005) for the model of equations (1) and (3) just requires running a regression of Y on X and P and taking the derivative of the estimated regression function with respect to P. Notice that the regression in (10) is partially linear, where X and XP are partially linear, and the function *K(P)* is nonparametrically estimated.

V can take values from 0 to 1, which means that the MTE is defined over the whole unit interval. However, in practice it is only possible to estimate the MTE over the observed support of P, since we will not be able to estimate *K'(P)* for values of P that are not observed (unless we impose a functional form which allows some extrapolation). In our data the support of P is almost the full unit interval, so we are able to estimate the MTE close to its full support. However, this will not be true in all applications of this method, which may mean estimating the MTE only over small ranges of values for V.

In fact, if we had assumed that Z is independent of (U_1, U_0) given X, instead of full independence between (Z,X) and (U_1, U_0), it would be difficult to estimate the MTE over a large support, as emphasized in Carneiro, Heckman and Vytlacil (2011). The reason is that, in that case, we have to consider estimation of the whole model conditioning non-parametrically in X. And for each value of *X* it is only possible to estimate the MTE over the support of *P* conditional on *X*, which usually will be much smaller than the unconditional support of *P* (see Carneiro, Heckman and Vytlacil, 2011). The whole method can become impractical in that case, which is why we rely instead on the assumption of full independence of (*Z,X*) and (U_1, U_0), which is common in empirical applications of selection models (and it allows us to use the full support of *P*).

Equations (10) and (11) can be estimated using standard methods. In particular, we use the partially linear regression estimator of Robinson (1988) to estimate (β_1, β_0), which entails two steps. The first step is a set of non-parametric regressions *Y*, and of each element of *X* and *XP* (the variables entering linearly in the model) on *P* (the variable entering non-parametrically in the model). We use local linear regression to estimate these regressions. Then we save the residuals of all these regressions. Finally, we regress the residualized outcome on the residualized *X* and *XP*, to estimate β_0 and $(\beta_1 - \beta_0)$.

For the second step of this method, which involves estimating *K(P)*, we start by computing the residual $R = Y - \left[\alpha_0 + X\beta_0 + PX\left(\beta_1 - \beta_0\right)\right]$. *K(P)* (and *K'(P)*) is estimated using a non-parametric regression of *R* on *P* (we use locally quadratic regression; Fan and Gijbels, 1996, suggest using a local polynomial of order *n+1* if the goal is to estimate a derivative of order *n*). Notice that (α_1, α_0) cannot be identified separately from *K(P)*.

A simple test of heterogeneity and selection on unobserved characteristics is a test of whether *K'(P)* is flat (or of whether *E(Y |X, P)* is nonlinear in *P*). If *K'(P)* is flat (if it does not depend on P) then heterogeneity is not important, or individuals do not select on it.

One important limitation of this procedure, and of the program evaluation literature relying on the propensity score, is that we never observe *P*, but we need to estimate it. This means that *P* will have estimation error. The main consequence of this, as argued by Abadie and Imbens (2012) in the context of matching, is that one needs to adjust the estimated standard errors of the treatment to account for this estimation error.

With finite samples we may also worry about measurement error bias, although it is not clear how it affects our estimates. In non-linear models the usual attenuation intuition fails, as discussed in Chesher (1991), and the literature reviewed in Schennach (2013).

3.3 Average Marginal Returns to Education

Economic decisions involve comparisons of marginal benefits and marginal costs. Therefore it is important to estimate the average returns to schooling for individuals at the margin between enrolling or not. They would be those who are the most likely to change their upper secondary schooling decision in response to a change in education policy.

The definition of who is marginal depends on the policy being considered. This is made clear in Carneiro, Heckman and Vytlacil (2010, 2011), who focus on three particular definitions of individuals at the margin:

$$i)\left|P-V\right|<\varepsilon,\;\; ii)\left|Z\gamma-U_s\right|<\varepsilon,\;\; iii)\left|\frac{P}{U}-1\right|<\varepsilon.$$

ε is a small positive number. These parameters are defined by taking the limit as ε goes to zero. They correspond to three different marginal policy changes.

The three policy changes considered are (i) a policy that increases the probability of attending college (P) by an amount α, so that $P_\alpha = P_0 + \alpha$; (ii) a policy intervention that has an effect similar to a shift in one of the components of Z, say Z^k, so that $Z_\alpha^k = Z^k + \alpha$ and $Z_\alpha^j = Z^j$ for $j \neq k$; (iii) and a policy that changes each person's probability of attending college by the proportion (1+ α), so that $P_\alpha = (1+\alpha)P_0$.

In this paper we estimate the average marginal returns to upper secondary schooling in Indonesia according to the definition of marginal in ii) above, although we could have chosen a different one. The MTE provides a general characterization of heterogeneity in returns and from it we can construct various other parameters.

Carneiro, Heckman and Vytlacil (2010) show how it is possible to write the average marginal treatment effect (or AMTE) as a weighted average of the MTE:

$$AMTE(x) = \int MTE(x,v) f_{X,V}(v||Z\gamma - U_s| < \varepsilon, x)dv \qquad (12)$$

3.4 IV vs Average and Marginal Returns

Heckman (2011) provides a discussion of structural and program evaluation approaches to evaluating policy. He argues that IV estimates conflate the definition and the identification of the parameter: IV estimates identify the Local Average Treatment Effect (LATE) corresponding to the instrument being used.

The methods developed by Heckman and Vytlacil (2005) allow you to use exactly the same data and identifying assumptions that one would use in a standard IV setup (Vytlacil, 2002), but estimate a wider range of parameters. Notice that, in parallel with the MTE, IV estimates can (and should) always be constructed and presented.

Although the pointwise standard errors for the MTE are surely higher than the IV standard errors, the precision with which one can estimate several other parameters, which are weighted averages of the MTE, is of a similar order as the precision of the IV estimates, which are also a weighted average of the MTE (Heckman and Vytlacil, 2005).

In particular, we will be able to estimate, conditional on observing enough support, parameters such as the average treatment effect, or treatment on the treated, or the policy relevant treatment effect for a particular policy question (Heckman and Vytlacil, 2001). And even with limited support one can estimate different versions of the AMTE just mentioned above, by computing the corresponding weights.

In practice, however, applications of selection models do impose some simple structure in the equations that they add to the standard IV models, mainly because of data limitations. This is true even when using semi-parametric methods, as in this paper. One should make these functions flexible, and examine robustness to different specifications.

3.5 Estimating vs. Simulating the Weights: A New Procedure

So far this section has shown how to recover the MTE from the data, and how to construct economically interesting parameters as weighted averages of the MTE. Heckman and Vytlacil (2005) and Carneiro, Heckman and Vytlacil (2010, 2011) provide formulas for the necessary weights in equations 7 and 12, conditional on *X*:

$$
\begin{aligned}
& f_{V|X}(v) = 1 \\
& f_{V|X}(v \mid X, S=1) = \frac{1 - F_{P|X}(v \mid X)}{E(P \mid X)} \\
& f_{V|X}(v \mid X, S=0) = \frac{F_{P|X}(v \mid X)}{E(P \mid X)} \qquad\qquad (13) \\
& f_{V|X}(v \mid X, S(z)=0, S(z')=1) = \frac{F_{P|X}(v \mid X) - F_{P'|X}(v \mid X)}{\int \left[F_{P|X}(v \mid X) - F_{P'|X}(v \mid X)\right] dv} \\
& f_{V|X}(v \mid X, |Z\gamma - V| < \varepsilon) = \frac{f_{P|X}(v \mid X) f_{U_S|X}\left[F^{-1}_{U_S|X}(v \mid X)\right]}{E\left[f_{U_S|X}(Z\gamma \mid X)\right]}
\end{aligned}
$$

where $f_{P|X}(p \mid X)$ and $F_{P|X}(p \mid X)$ are respectively the p.d.f and the c.d.f. of *P* conditional on *X*, $f_{U_S|X}(u_S \mid X)$ and $F_{U_S|X}(u_S \mid X)$ are respectively the p.d.f and the c.d.f. of U_S conditional on *X*, and $F_{P'|X}(p \mid X)$ is the c.d.f. of *P* conditional on *X* when *Z=z'*.

In practice it is difficult to implement these formulas since they involve estimation of conditional density and distribution functions, such as $f_{P|X}(p \mid X)$ and $F_{P|X}(p \mid X)$, and *X* is generally a high dimensional vector (there are 28 variables in *X* in our empirical work). It is impractical to estimate these functions, even with enormous amounts of data.

Therefore, Carneiro, Heckman and Vytlacil (2010, 2011) aggregate *X* into an index, namely $I = X(\beta_1 - \beta_0)$. They then proceed by estimating conditional densities and distributions of *P* with respect to *I*, which requires conditioning only on one variable. This makes the whole procedure feasible. But there is no theoretical basis for this aggregation. It is very much ad-hoc, which makes it quite unattractive. The only reason to implement it is because it is essential to get a dimensionality reduction in the problem, and this seemed to be a natural one, although as good as many others.

In this paper we use an alternative procedure, which avoids making this aggregation, and sidesteps the problem of estimating a multidimensional conditional density function.

Notice that the selection equation relates *S*, *X*, *Z*, and *V* (which is uniform by construction). Our idea is that, using the estimated parameters, we can simulate (instead of estimating) the following objects:

$$f_{V|X}(v|S=1,x), f_{V|X}(v|S=0,x), f_{V|X}(v||Z\gamma - U_S| < \varepsilon, x)$$

For example, in order to simulate $f_{V|X}(v|S=1,x)$ all we need to do is to draw many values of u_S (which is assumed to be logistic) for each value of *x*, and select all the cases where the selection equation (2) predicts that *S=1*. Then all we need to do is to estimate the average value of the MTE for the simulated population for whom *S=1*. This simulation procedure is simple, and its steps are described in detail in Appendix A.

4. Empirical Results

4.1 Is Distance to School a Valid Instrument?

To account for the potential endogeneity of the schooling decision we instrumented schooling with the distance to the nearest secondary school.[3] In order for it to be a valid instrument distance to school needs to satisfy two conditions: i) it should affect the probability of school enrolment and ii) it should have no direct effect on adult wages.

We show that condition i) is satisfied. Condition ii) is controversial. There are two main issues to discuss. First, families and schools may not randomly locate across locations in Indonesia. Second, we can only use distance measured from the current municipality of residence, and not distance measured at the time of the schooling decision. Although papers such as Carneiro, Heckman and Vytlacil (2011) have been able to measure distance at the time of the schooling decision, this is not always the case in the literature. The first papers using this instrumental variable (Card, 1993, 1995) only have contemporaneous distance measures, exactly as in our paper, although they are able to observe individuals at a much younger age than we do. The main problem of this approximation is that educated individuals may move to more urban areas which also have more schools, so the first stage relationship could have the causality reversed.

It is instructive to examine the consequences of these two issues in a simple model. We start by looking at a fixed coefficient model with no heterogeneity in returns (in terms of equation (1), $\beta_1 = \beta_0$ and $U_1 = U_0$), since it provides us with clear and intuitive results. We then discuss briefly what could happen in a random coefficients model like the one we have in this paper, although the problem there is much less clear.

Take the following model:

$$y = \tau S + u$$

$$S = \sigma d^* + v$$

$$d = \rho S + \delta d^* + \varepsilon \qquad (14)$$

where y is the outcome (wages), S is schooling (upper secondary education), d^* is distance at the time of the secondary school decision (which is unobserved), and d is current distance (measured at the time of the outcome). There is no heterogeneity in any

[3] Distance to the nearest school has been used by Card (1995), Kane and Rouse (1995), Kling (2001), Currie and Moretti (2003), Cameron and Taber (2004) and Carneiro, Heckman and Vytlacil (2011).

of the parameters of this model, namely in τ. We expect that $\tau > 0$ (positive return to schooling) and $\sigma < 0$ (negative effect of distance to school on attendance of upper secondary school). In principle, $\rho < 0$, capturing the idea that schooling can induce migration towards a job in a large city where schools are abundant. δ measures inertia and it is probably between 0 and 1, and ε are other migration shocks. Migration to larger cities (or lower *d*) and wages should be correlated: $Cov(u, \varepsilon) < 0$.

In this model τ is the sole parameter of interest, measuring the return to education. Let τ^* be the instrumental variable estimate of τ when we use *d* as the instrument. Then:

$$plim\tau^* = \frac{Cov(y,d)}{Cov(S,d)} = \tau + \rho\frac{Cov(u,S)}{Cov(S,d)} + \delta\frac{Cov(u,d^*)}{Cov(S,d)} + \frac{Cov(u,\varepsilon)}{Cov(S,d)} \quad (15)$$

In this model there are three sources of potential bias for τ. To start with, $Cov(u, d^*)$ could be different from (smaller than) zero. For example, Carneiro and Heckman (2002) and Cameron and Taber (2004) show that individuals living closer to universities in the US have higher levels of cognitive ability and come from better family backgrounds. In Indonesia, those with better educated parents are also located closer to secondary schools.

In addition, if regions where schools are abundant are also regions where other infrastructure is abundant, we may confound the impact of school availability on wages with the impact of infrastructure on wages (see Jalan and Ravallion, 2002). This will be true unless labor is perfectly mobile, which is unlikely to be the case in Indonesia. However, as argued in Duflo (2004), perhaps the response of other (private or public) infrastructure to school construction and to a better skilled workforce is very slow.

It is possible that school location is exogenous after we account for a detailed set of individual and regional characteristics, namely: age, parental education, religion, an indicator for rural residence at age 12, dummies for the province of residence, and distance to the nearest health post. We also show that removing these regional controls hardly affects our results, indicating that this problem may be unimportant in our setting. Under the assumption that these are rich enough controls, $\delta\frac{Cov(u,d^*)}{Cov(S,d)} = 0$. Otherwise: $\delta\frac{Cov(u,d^*)}{Cov(S,d)} > 0$ (since $Cov(S, d) < 0$).

Then, we expect that $Cov(u, \varepsilon)$ could be negative, if individuals moving to more urban locations (with lower *d*) have higher wages. This would mean that $\frac{Cov(u,\varepsilon)}{Cov(S,d)} > 0$. We also expect $Cov(u, S) > 0$ if those with high levels of ability also have high levels of schooling, resulting in $\rho \frac{Cov(u,S)}{Cov(S,d)} > 0$ (since, as mentioned above, it is likely that $\rho < 0$).

Again, it is possible that, even if these terms are positive, they are small, if the set of controls we include in the model are rich enough to capture the sources of simultaneity. For example, variables such as rural or urban location at age 12, combined with variables such as distance to the nearest health post in the current residence, and province of residence, may account for the most relevant migration patterns for each individual. In that case, endogenous migration from rural to urban areas is hopefully controlled for (so that $\rho = 0$ and $Cov(u, \varepsilon) = 0$), and the reverse causality problem minimized.[4]

If τ is a random coefficient with mean $\bar{\tau}$ then we can write the wage equation as:

$$y = \tau S + u = \bar{\tau} S + u + (\tau - \bar{\tau}) S$$

It has already been mentioned repeatedly in this paper (and in the literature) that, in this case, there can be many potentially interesting parameters of interest. However, to simplify the discussion, suppose we are interested in estimating $\bar{\tau}$. Then, with IV we get:

$$plim\tau^* = \bar{\tau} + \rho \frac{Cov(u + (\tau - \bar{\tau})S, S)}{Cov(S,d)} + \delta \frac{Cov(u + (\tau - \bar{\tau})S, d^*)}{Cov(S,d)} + \frac{Cov(u + (\tau - \bar{\tau})S, \varepsilon)}{Cov(S,d)}$$

Now it is much harder to sign each of these three bias terms, especially if there is some correlation between S and $(\tau - \bar{\tau})$. Nevertheless, if, as above, we can assume that the rich set of observable variables we include in the regressions are enough to make $\rho = 0$ and $Cov(u + (\tau - \bar{\tau})S, \varepsilon) = 0$, then, if d^* can be assumed to be independent of all the unobservables in the model (conditional on the controls), we can still estimate the MTE (even though, even in this case, $\frac{Cov(u+(\tau-\bar{\tau})S,d^*)}{Cov(S,d)}$ is not likely to be equal to zero).

Below we will report MTE estimates that are valid under the assumptions that u is independent of Z and X, $\rho = 0$ and $Cov(u + (\tau - \bar{\tau})S, \varepsilon) = 0$. In order to gain some insight about the sensitivity of our estimates to violations of one of these assumptions, we

[4]We checked whether distance to the nearest secondary school was correlated with pre-secondary educational outcomes of each individual (elementary school completion, grade repetition, work in school), which are correlated with the early ability of the child. In Table A1 in the appendix we regress each early schooling variable on distance, and our results show no correlation between distance and these variables.

present simulations illustrating potential changes to the IV estimate as ρ becomes increasingly different from zero.

Table 2 shows that distance to the nearest secondary school is a strong predictor of enrolment in secondary school. We run a logit regression where the dependent variable is an indicator taking value 1 if an individual ever attended upper secondary school. Regressors include distance to the nearest secondary school and all the control variables mentioned above. The table displays marginal effects of each variable on the probability of enrolling in upper secondary education. We include as a control the distance to the nearest health post, as a proxy for location characteristics. Unlike distance to school, distance to health post does not predict school enrollment. Children of highly educated parents are more likely to attend upper secondary school than children of parents with low levels of education. Catholics and Protestants are more likely to attend secondary school than Muslims (the omitted category). Children in rural areas are less likely to attend upper secondary school than children in urban areas.

This model is fairly flexible in the sense that the impact of distance on secondary school attendance varies with X. In particular, we interact distance to school with age (which, for a fixed year, also captures cohort), religion, parental education, and rural residence. It is useful to estimate such a rich model for two related reasons. First, because of its flexibility. Second, by allowing the impact of the instrument to vary with the variables in X we are able to use extra variation in the instrument. As a result, the standard errors in the IV estimates and in the selection model are smaller than if we just used a simpler model without these interactions. Therefore, the estimates in this paper come from this model, while estimates of a simpler model without interactions are shown in the Appendix. Average derivatives are computed at the mean value of the X variables.

Table 3 also displays p-values for the test of the null hypothesis that distance to school does not affect upper secondary school attendance. We perform a joint test on all coefficients involving distance. We reject that distance to school does not determine upper secondary school attendance.

In spite of this, one could worry that the instruments are weak. The F-statistic for a specification where distance to the nearest school is used as the only instrument is 5.86 ($F(1,303)$), and it is equal to 2.22 ($F(13,303)$) in the case where we use interactions between distance and the controls. However, the instrument used in the estimation of the semi-parametric selection model in the paper is P, the estimated propensity score, which is obviously a very strong instrument (F(1,303)=27), even after including all the controls mentioned above (below we mention IV results with very flexible specifications of the outcome equation, which minimize the danger that results are driven by nonlinearities of the logit function). In addition, the consequences of weak instruments for the estimation of selection models such as the ones we present in this paper are not well studied.

4.2 Standard Estimates of the Returns to Schooling

In order to more easily make a comparison between our data and estimates and those in the literature we start by presenting standard OLS and IV estimates of the returns to schooling. Throughout the paper schooling takes two values: 0 for less than upper secondary, and 1 for upper secondary or above. We use the log hourly wage in 2000 as our dependent variable. The full set of controls consists of: age (or cohort), parental education, religion, an indicator for whether the individual was living in a city or in a village at age 12, an indicator for whether the individual lived in a rural area at age 12, dummies for the province of residence, and distance to the nearest health post.

We present ordinary least squares (OLS) and IV results. This is shown in Table 4. Recall from table 1 that individuals with upper secondary schooling or above have on average 13.13 years of schooling, while those with less than upper secondary have on average 5.34 years of schooling. The difference between the two groups is 7.79 years of schooling. Using this figure to annualize the returns to upper secondary education we have an OLS estimate of 9% and an IV estimate of 12.9% (without annualizing returns we have OLS and IV estimates of 70.5% and 100% respectively).[5]

These estimates are higher than (but of comparable magnitude to) those in Duflo (2001), although we use more recent data. Petterson (2010) finds a return of 14% using the same data as we do, but a different sub-sample and instrument.

As in most of the literature, our IV estimates of the return to education are larger than OLS estimates. However, it is well known that this depends on the instrument, a point which is explored in Cameron and Taber (2006). More importantly, as discussed in

[5] With a more flexible model where we include all two way interactions between the controls the IV point estimate is 0.184 with a standard error of 0.076.

Heckman and Vytlacil (2005), OLS does not necessarily correspond to the average return to schooling for any group in the population.

In section 4.1 we discussed several potential problems with the instrument, in particular the possibility that endogenous migration can lead to reverse causality from schooling to the instrument, current distance to school. It is interesting therefore to examine what could happen as ρ and $Cov(u, \varepsilon)$ in equation (14) become increasingly different from zero. In Appendix C we describe these simulations in detail. Both sets of simulations show that, when we deviate from the assumptions of the model, the implied IV estimate falls (as predicted from calculations based on the fixed coefficient model). There is more sensitivity of this estimate to $Cov(u, \varepsilon)$ than to ρ.

Appendix table A2 presents OLS and IV estimates where we use years of schooling as the main explanatory variable (as opposed to upper secondary schooling). The OLS estimate of the return to a year of schooling is 9.6%, while the IV estimate is 15.7%. In appendix table A3 we also present IV estimates of returns for models where we do not interact the instrument with the variables in X. The point estimate is smaller than the one in Table A2, and the standard error is larger, but the main pattern remains: the IV estimate is much higher than the OLS estimate. In a model with heterogeneous returns, it is not surprising that the instrumental variable is sensitive to the choice of instrument. In appendix table A4 we present results were we omit regional dummies from the model. Our IV estimate is very similar to the ones in tables A2 and A3. Finally, in Appendix table A5 we go back to estimating annualized returns to upper secondary education, presenting estimates using different sets of instruments, and different ways of annualizing the returns. There is some variation across columns, but this is natural in a context where different instruments are used to estimate a model of heterogeneous returns, since different instruments lead to different parameters (Heckman, Urzua, and Vytlacil, 2006).

It is important to mention the role of experience. As in most of the literature, our modeling of experience is unsatisfactory. Experience affects wages and is endogenously determined. In addition, experience is likely to affect the returns to education.

In our data, we cannot even observe work experience, and therefore include age (and its square) in the model instead. Although the IV estimates we just presented do not allow the returns to schooling to depend on age, the semi-parametric model we estimate below allows for an interaction between schooling and age, which we comment on below.

4.3 Average and Marginal Treatment Effect Estimates

We start with the semi-parametric model. The first step is to construct *P*, for which we use a parametric model. We take the predicted probability of ever attending upper secondary school from a logit regression of upper secondary school attendance on the *X* and *Z* variables of section 3. Table 3, discussed above, reports the coefficients of the logit model. All variables work as expected.

It is only possible to identify the MTE over the support of *P*. Therefore, we need to examine the density of *P* for individuals who attend upper secondary school or above, and those who do not. This is done in Figure 1, which shows the distributions of the predicted propensity score (*P*) for these two groups. The supports for these two distributions overlap almost everywhere, although the support at the tails is thin for low values of *P* among those with upper secondary school or above. We construct the MTE as described in Section 2. In order to estimate *K(P)* we run a local quadratic regression of *R* on *P*, using a Gaussian kernel and a bandwidth of 0.27.[6] The implied *MTE*(*x,v*) is computed by calculating the slope on the linear term of the local quadratic regression (the coefficients on *X* in the outcome equations are presented in Appendix Table A6).

Figure 2 displays the estimated MTE (which we evaluate at the mean values of the components of *X*). The MTE is monotonically decreasing for all values of *V*. Returns are very high for individuals with low values of *V* (individuals who are more likely to enroll in upper secondary school or facing low costs). The figure demonstrates substantial heterogeneity in the return to schooling, which ranges from 34% for individuals with V around 0.1 to 13% for those with V close to 0.5, and becomes negative for those with values of V close to 1. The fact that returns are the lowest for individuals who are least

[6] The bandwidth is determined by leave-one-out cross validation, although below we also present estimates with much lower values for the bandwidth. We focus on equation (10), and estimate this equation for a very large set of bandwidths with values between 0 and 1. We pick the bandwidth that minimizes the mean squared error. To compute the mean squared error for each value of the bandwidth we estimate equation (10) N times, with N being the sample size, omitting one observation at a time. We then use the model to predict the value of the omitted observation and compute the corresponding residual, which we then square. Finally, we average the squared residuals across all N repetitions. The reason why we pick a local quadratic polynomial is because Fan and Gijbels (1996) suggest that if we want to estimate the n^{th} derivative of a function, say *K(P),* then we should use a polynomial of order *n+1*.

likely to go to school is consistent with a simple economic model where agents sort into different levels of schooling based on their comparative advantage.

All our confidence intervals are estimated using the bootstrap. In each bootstrap iteration we perform every single step of the estimation procedure. To be precise, we first draw bootstrap samples from the raw data. We use a block-bootstrap procedure, where the block is the village of residence (the cluster). Then, for each sample, we proceed from the first step, which is the estimation of P, and we proceed all the way towards the last step, which is the estimation of the treatment parameters. We perform 250 replications, and using the bootstrap replications we can construct Highest Posterior Density (HPD) 95% confidence intervals for the MTE and the treatment parameters (which are the shortest intervals in the distribution of the parameter that hold 95% of the data).[7]

Unfortunately the confidence intervals on our estimated MTE are quite wide. As mentioned above, this problem had appeared before in our discussion of standard IV estimates, and the standard errors of our estimates are not much larger than those reported in the literature (Card, 2001). However, it is still possible to reject that the MTE is flat, and therefore, that there is selection in returns and that our concern with heterogeneity is important. Appendix Table A7 tests and rejects that adjacent segments of the MTE are equal (see Carneiro, Heckman and Vytlacil, 2011).

Once we have established that selection in returns is important, one way to obtain smaller standard errors is to estimate a parametric model. This would require us to consider heterogeneity and selection in returns using a less flexible model, but which delivers much more precise estimates. This is exemplified in figure 3, which shows that the standard errors improve dramatically when we estimate the MTE assuming joint normality of (U_1, U_0, U_S). The shape of the MTE is declining as before, although the normality assumption does not allow the MTE to have a flat section as in Figure 2, so the MTE is declining everywhere, again taking negative values for very high values of V. In principle, it would be possible to consider more flexible parametric models.

Table 4 presents average returns to upper secondary schooling for different groups of individuals. The return to upper secondary school for a random person (ATE) is 13.8%.

[7] The fact that our estimate of P has estimation error is reflected in the standard errors. However, we assume that any measurement error type bias disappears as the size of the sample grows, and therefore, measurement error does not affect the consistency of our estimates.

The return for those individuals who were enrolled in upper secondary schooling (ATT) is considerably higher, at 21.8%. The return that individuals who did not go to upper secondary school would experience had they gone there (ATU) is 8.1%. Average parameters are estimated with the assumption of full support (although figure 1 shows a very small lack of support in the left tail of the distribution of *P*). Estimates of the return to the marginal student (AMTE) are robust to the lack of full support (Carneiro, Heckman and Vytlacil, 2010, 2011). The return to the marginal student is 14.0%, well below the return to the average student in upper secondary school (21.8%).[8]

Finally, the last line of Table 4 (PRTE) reports the average return for those induced to attend upper secondary school by a particular policy shift: a 10% reduction in distance to an upper secondary school. This is the parameter needed to understand the impacts of such an education expansion. By coincidence, it is remarkably similar to the MPRTE.

In the appendix we show that results are similar but more imprecise when we do not interact *Z* and *X* in the selection equation. This is reassuring, and shows the importance of using a more flexible model for the precision of our estimates. We also present estimates of treatment effects for much lower values of the bandwidth, which show some sensitivity in the point estimates for ATE, TT, and TUT (as expected), especially for very low bandwidths, but little sensitivity for the two policy parameters, MPRTE and PRTE.[9]

4.4. Estimates by Age Group

The individuals in our sample have between 25 and 60 years of age, which is a very large age range. It is plausible that the returns to schooling vary with the age of the individual, either because of genuine age effects, or because of cohort effects. Distinguishing the two is well known to be a very difficult problem.

We include an age polynomial in the model, which accounts for age profiles, and we allow the age profile to vary with schooling (by interacting the age polynomial with upper secondary school attendance). Nevertheless, we cannot reject that the age profiles

[8] These are apparently high returns, but they are not out of line with other estimates for Indonesia by Duflo (2001) and Petterson (2010). A review of the studies on the returns to secondary and higher education in developing countries by Psacharopuolos and Partinos (2004) further indicates that private returns to secondary and higher education range from around 18 percent in non-OECD Asian countries to almost 28 percent in countries of Sub-Saharan Africa. So, our results are very much within the range found in the literature for developing countries, even if they are high for developed country standards.

[9] See tables A3 , A8 and A9 which present the tables just reported, but for the case where we exclude these interactions, as well as figures A1, A2 and A3. Table A10 reports results for much lower bandwidths.

are not affected by schooling (see Table A7 in the Appendix), although one could argue that our specification of age effects is restrictive.

The model laid out so far does not allow the MTE to vary with age (apart from its mean), since we assume the MTE is separable in observables and unobservables. It is possible that age or cohort have an impact not only on average wage profiles, but also on the MTE. In order to investigate this possibility we divide the sample into two groups: individuals with an age above 37, and individuals with age equal to or below 37 (which is the median age in our sample). Our main results are in Appendix Table A10, which shows ATT, ATE and ATU estimate for these two age groups. These results are nevertheless less robust than the ones reported before, since the instruments become weak predictors of schooling in the selection equation once we divide the sample in these two halves. Average returns to schooling are generally much higher for the older age group. Also, for the older age group the MTE is declining with *V*, whether we estimate the model using the semi-parametric LIV estimator, or a parametric normal selection model. For the younger age group the results are sensitive to which method is used.[10]

4.5. Comparison with Carneiro, Heckman and Vytlacil (2011)

As pointed out above, in this paper we introduce a procedure that relaxes the assumptions in previous applications of this method, namely Carneiro, Heckman and Vytlacil (2011). In order to compare the two methods we conducted a simple Monte Carlo simulation, for sample sizes like the ones used in this paper. We describe the simulation in detail in the Appendix D, and here we report the most important results.

We find that both procedures work relatively well in our simulation. However, the procedure in this paper performs better. For the three parameters we look at, the ATE, the

[10] Throughout the paper we have use data only on wage earners. This is standard in the literature, mainly because they have a well defined wage measure. However, one could also include self-employed individuals in the sample. This is what we do in Table A12 in the Appendix, which has 3 columns: one corresponding to our baseline results, one including only self-employed individuals, and a third one including both self-employed and wage earners in the sample. The point estimates are different across samples but the patterns are similar. We prefer to use our baseline specification in our main text not only because we have a more uniform wage measure for this subsample, but also because the instrumental variable is a stronger predictor of schooling than in the other samples (especially when compared with the sample including only self-employed, for which we do not have a strong first stage). Table A13 has our last sensitivity exercise, where we omit all post schooling controls from the model. There is very little change in the point estimates relative to the baseline model.

TT, and the TUT, the mean squared error is almost half as large for the procedure used in this paper as in the procedure in Carneiro, Heckman and Vytlacil (2011).

5. Conclusion

Indonesia has an impressive record of educational expansion since the 1970s. The enrollment rates are nearly universal for elementary schooling and are around 75% for secondary education. There is an ongoing effort to extend universal education attainment to the secondary level. And although enrollment in secondary education continues to rise we find striking inequality in returns to education. Individuals who are more likely to be attracted by educational expansions at the upper secondary level (marginal) have lower average returns than those already attending upper secondary schooling. In this paper we document a large degree of heterogeneity in the returns to upper secondary schooling in Indonesia. We estimate the return to upper secondary education to be 7 to 8 percentage points higher (per year of schooling) for the average than for the marginal student.

Therefore, efforts aimed at educational expansion will attract students with lower levels of returns. However, returns are still fairly high for the marginal person. It is difficult to know what would happen to returns if there were large education expansions, because our framework does not allow us to say anything about the equilibrium impacts of such policies. Our estimates indicate that the quality of the average student would probably decline, and if demand is downward sloping, we may also expect the price of skill to decline. However, this is outside the scope of our model.

What is behind such a large inequality in the returns to schooling? There is a growing body of literature that argues that human capital outcomes later in life (including the ability to learn) are largely influenced by what happens early in life (e.g., Carneiro and Heckman, 2003). It is therefore important for the design of schooling policy to determine whether the inequality in secondary schooling outcomes can be remedied at earlier stages, for example during early childhood, or during the elementary school years.

References

Abadie, A. and G. Imbens (2012), "Matching on the Propensity Score", working paper

Bjorklund, A. and R. Moffitt (1987), "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69:42-49.

Cameron, S. and C. Taber (2004), "Estimation of Educational Borrowing Constraints Using Returns to Schooling", *Journal of Political Economy*, part 1, 112(1): 132-82.

Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling ", *Aspects of Labour Economics: Essays in Honour of John Vanderkamp*, edited by Louis Christofides, E. Kenneth Grant and Robert Swindinsky. University of Toronto Press.

Card, D. (1999), "The Causal Effect of Education on Earnings," Orley Ashenfelter and David Card, (editors), Vol. 3A, *Handbook of Labor Economics,* Amsterdam: North-Holland.

Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5): 1127-60.

Card, D. and T. Lemieux (2001), "Can Falling Supply Explain the Rising Return to College For Younger Men? A Cohort Based Analysis," *Quarterly Journal of Economics* 116: 705-46.

Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post-secondary Schooling," *Economic Journal* 112(482): 705-34.

Carneiro, P. and J. Heckman (2003), "Human Capital Policy," in *Inequality in America: What Role for Human Capital* Policy, J. Heckman, A. Kruger and B. Friedman eds, MIT Press.

Carneiro, P., J. Heckman and E. Vytlacil (2010), "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin", *Econometrica*.

Carneiro, P., J. Heckman and E. Vytlacil (2011), "Estimating Marginal Returns to Education ", *American Economic Review*, 101(6).

Carneiro, P. and S. Lee (2009), “Estimating Distributions of Potential Outcomes using Local Instrumental Variables with an Application to Changes in College Enrolment and Wage Inequality”, Journal of Econometrics.

Carneiro, P. and S. Lee (2011), “Trends in Quality Adjusted Skill Premia in the US: 1960 to 2000”, *American Economic Review*, 101(6).

Chesher, A. (1991), “The Effect of Measurement Error”, *Biometrika*, 78, 451.

Currie, J. and E. Moretti (2003), Mother's Education and the Intergenerational Transimission of Human Capital: Evidence from College Openings, *Quartely Journal of Economics* , 118:4.

Duflo, E. (2001), “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment”, *American Economic Review*, 91(4), 795-813.

Duflo, E. (2004), “The medium run effects of educational expansion: evidence from a large school construction program in Indonesia”, Journal of Development Economics, 74, 163-197.

Fan, J. and I. Gijbels (1996), Local Polynomial Modelling and its Applications, New York, Chapman and Hall.

Griliches, Z. (1977), “Estimating the Return to Schooling: Some Persistent Econometric Problems”, Econometrica.

Heckman, J. (2010), “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Public Policy”, *Journal of Economic Literature*, 48(2), 356-98.

Heckman, J. and X. Li (2004), “Selection Bias, Comparative Advantage, and Heterogeneous Returns to Education: Evidence from China in 2000”, *Pacific Economic Review*.

Heckman, J., S. Urzua and E. Vytlacil (2006), “Understanding What Instrumental Variables Really Estimate in a Model with Essential Heterogeneity”, *Review of Economics and Statistics*.

Heckman, J. and E. Vytlacil (1999), Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects, *Proceedings of the National Academy of Sciences*, 96, 4730-4734.

Heckman, J. and E. Vytlacil (2001a), “Local Instrumental Variables,” in C. Hsiao, K. Morimune, and J. Powells, (eds.), *Nonlinear Statistical Modeling: Proceedings of* *the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honorof Takeshi Amemiya* , (Cambridge: Cambridge University Press, 2000), 1-46.

Heckman, J. and E. Vytlacil (2001b), "Policy Relevant Treatment Effects", *American Economic Review Papers and Proceedings*.

Heckman, J. and E. Vytlacil (2005), "Structural Equations, Treatment, Effects and Econometric Policy Evaluation," *Econometrica*, 73(3):669-738.

Ichimura, H. and C. Taber (2000), "Direct Estimation of Policy Impacts", working paper.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects,' *Econometrica*, 62(2):467-475.

Jalan, J. and M. Ravallion (2002), "Geographic poverty traps? A micro model of consumption growth in rural China", *Journal of Applied Econometrics*, 17, 329-346.

Kane, T. and C. Rouse (1995), "Labor-Market Returns to Two- and Four-Year College", *American Economic Review*, 85(3):600-614.

Kling, J. (2001), "Interpreting Instrumental Variables Estimates of the Returns to Schooling", *Journal of Business and Economic Statistics*, 19(3), 358-364.

Psacharopoulos, G. and H. Patrinos (2004), "Returns to Investment in Education: A Further Update", *Education Economics*, 12(2), 111-134.

Pettersson, G. "Do supply-side education programs targeted at under-served areas work? The impact of increased school supply on education and wages of the poor and women in Indonesia." PhD dissertation (Draft), Department of Economics, University of Sussex

Robinson, P. (1988), "Root-N-Consistent Semiparametric Regression", Econometrica, 56(4), 931-954.

Schennach, S. (2013), "Measurement Error in Nonlinear Models – A Review", in Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, eds. D. Acemoglu, M. Arellano, E. Dekkel.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati and F. Witoelar (2004), "The Third Wave of the Indonesia Family Life Survey (IFLS): Overview and Field Report", March 2004. WR-144/1-NIA/NICHD.

Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", Econometrica, 70(1), 331-341.

Wang, X., B. Fleisher, H. Li and S. Li, "Access to Higher Education and Inequality: the Chinese Experiment", working paper, Ohio State University.

Willis, R. and S. Rosen (1979), "Education and Self-Selection,"Journal of Political Economy, 87(5):Pt2:S7-36.

Table 1: Sample statistics for the treatment and control groups

	Upper secondary or higher (Treatment group)	*Less than upper secondary (Control group)*
	N = 1085	*N = 1523*
Log hourly wages	8.198	7.481
Years of education	13.133	5.341
Distance to school in km	1.053	1.564
Distance to health post in km	0.889	1.079
Age	37.058	38.675
Religion Protestant	0.050	0.022
Catholic	0.029	0.009
Other	0.062	0.043
Muslim	0.860	0.927
Father uneducated	0.130	0.383
…elementary	0.503	0.507
...secondary and higher	0.330	0.061
...missing	0.020	0.037
Mother uneducated	0.201	0.425
…elementary	0.484	0.406
...secondary and higher	0.204	0.022
...missing	0.098	0.133
Rural household	0.240	0.476
North Sumatra	0.057	0.063
West Sumatra	0.047	0.058
South Sumatra	0.048	0.032
Lampung	0.016	0.027
Jakarta	0.181	0.095
Central Java	0.085	0.163
Yogyakarta	0.092	0.054
East Java	0.121	0.180
Bali	0.056	0.038
West Nussa Tengara	0.050	0.048
South Kalimanthan	0.040	0.020
South Sulawesi	0.035	0.035

Source: Data from IFLS3. Sample restricted to males aged 25-60 employed in salaried jobs in government and private sectors. Hourly wages constructed based on self-reported monthly wages and hours.

Table 2: Upper school decision model – Average Marginal Derivatives

	Coef	Average Derivative
Dist. to secondary school in km	-0.123***	-0.0300**
	(0.040)	*(0.0127)*
Age	0.077*	0.0130
	(0.044)	*(0.0090)*
Age Squared	-0.096*	-0.0162
	(0.055)	*(0.0111)*
Protestant	0.730***	0.1382***
	(0.264)	*(0.0484)*
Catholic	1.211***	0.2123**
	(0.395)	*(0.0890)*
Other religions	0.245	0.0552
	(0.363)	*(0.0878)*
Fathers education elementary	0.766***	0.1342***
	(0.127)	*(0.0217)*
Father higher education	1.835***	0.3769***
	(0.178)	*(0.0320)*
Mother education elementary	0.443***	0.0852***
	(0.123)	*(0.0230)*
Mother higher education	1.851***	0.3730***
	(0.237)	*(0.0418)*
Rural	-0.593***	-0.1143***
	(0.110)	*(0.0276)*
Distance to health post in km	-0.017	0.0000
	(0.040)	*(0.0083)*
Location fixed effect	Yes	
Mean of dependent variable	0.416	
	(0.010)	
Test for joint significance of instruments: Chi-square/p-value	9.42/0.0021	

Note: This table reports the coefficients and average marginal derivatives from a logit regression of upper secondary school attendance (a dummy variable that is equal to 1 if an individual has ever attended upper secondary school and equal to 0 if he has never attended upper secondary school but graduated from lower secondary school) on several variables. Type of location is controlled for using province dummy variables. A dummy variable for missing parental education is included in the regressions but not reported in the table. The first column presents coefficients of logit where only distance to school is used an IV. In the second column average derivatives (computed at the average values of X) are presented and instruments include distance to secondary school and interactions with all the Xs. Reference categories are Muslim, not educated. Standard errors (in parenthesis) are robust to clustering at the community level, with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table 3: Annualized OLS and IV estimates of the return to upper secondary schooling

	OLS	IV
Upper secondary (annualized)	0.090***	0.129***
	(0.005)	*(0.048)*
Age	0.052***	0.048**
	(0.019)	*(0.020)*
Age Squared	-0.042*	-0.037
	(0.023)	*(0.025)*
Protestant	0.182**	0.142
	(0.084)	*(0.104)*
Catholic	0.059	0.001
	(0.189)	*(0.202)*
Other religions	0.109	0.097
	(0.126)	*(0.125)*
Fathers education elementary	0.135***	0.091
	(0.048)	*(0.070)*
Fathers education secondary or higher	0.215***	0.101
	(0.067)	*(0.153)*
Mother's education elementary	-0.052	-0.080
	(0.048)	*(0.060)*
Mother's education secondary or higher	-0.031	-0.128
	(0.078)	*(0.136)*
Rural household	0.111**	0.152**
	(0.045)	*(0.068)*
Distance to health post in km	-0.023	-0.020
	(0.018)	*(0.017)*
Location controls	YES	YES
Mean of dependent variable	7.779	
	(0.932)	
Number of observations	2,608	2,608
Test for joint significance of instruments: F-stat/p-value	2.22/0.00	
R2	0.210	0.190

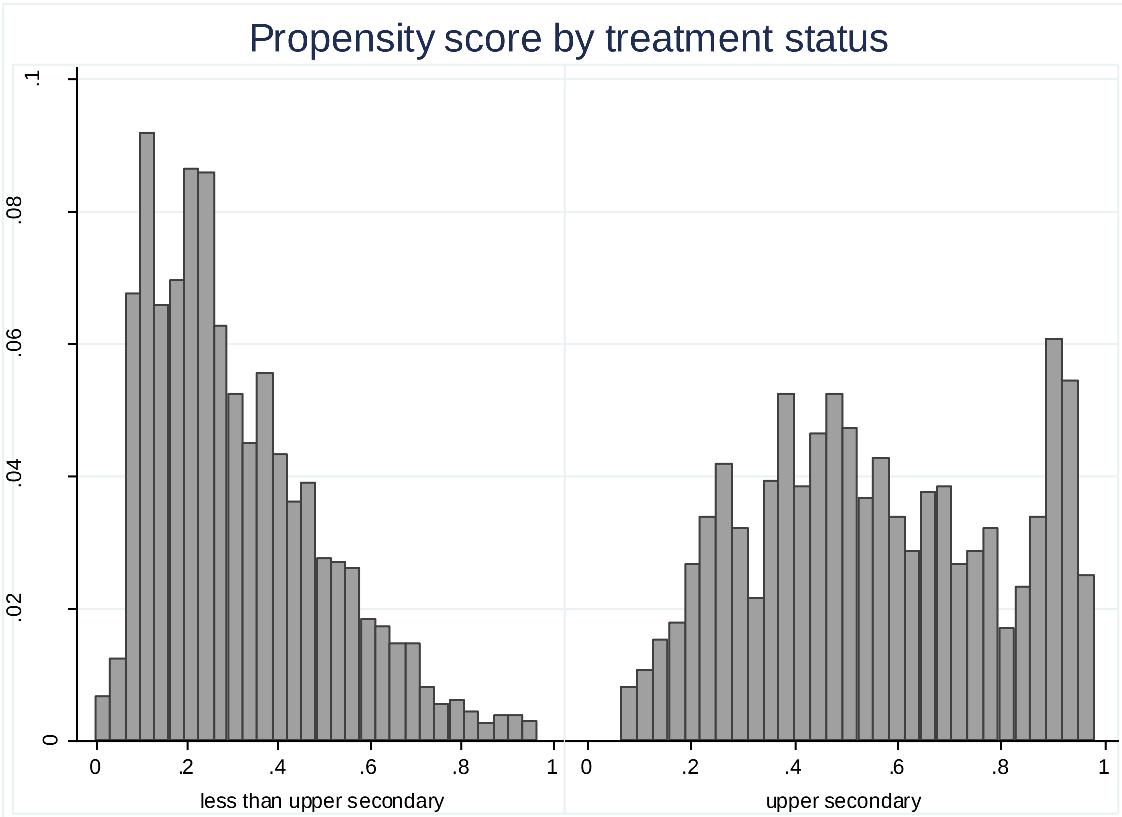
Note: This table reports the coefficients for OLS and 2SLS IV for regression of log of hourly wages on upper school attendance (a dummy variable that is equal to 1 if an individual has ever attended upper secondary school and equal to 0 if he has never attended upper secondary school but graduated from lower secondary school), controlling for parental education, religion and location. Excluded instruments are distance to secondary school and interactions with parental education, religion and age. Type of location is controlled using province dummies. A dummy variable for missing parental education is included in the regressions but not reported in the table. Reference categories are Muslim for religion, and not educated for education.Standard errors (in parenthesis) are robust to clustering at the community level with significance at *** $p<0.01$, ** $p< 0.05$, * $p<0.1$ indicated.

Table 4: Estimates of Average Returns to Upper Secondary Schooling

Parameter	Non parametric Estimate	Normal selection model
ATT	0.218	0.203
	(-0.002, 0.353)	(-0.015,0.246)
ATE	0.138	0.067
	(0.025, 0.254)	(-0.016,0.199)
ATU	0.081	-0.029
	(-0.121, 0.326)	(-0.082,0.175)
MPRTE	0.140	0.104
	(-0.049, 0.338)	(-0.017, 0.246)
PRTE	0.145	0.100
	(-0.053, 0.337)	(-0.015,0.251)

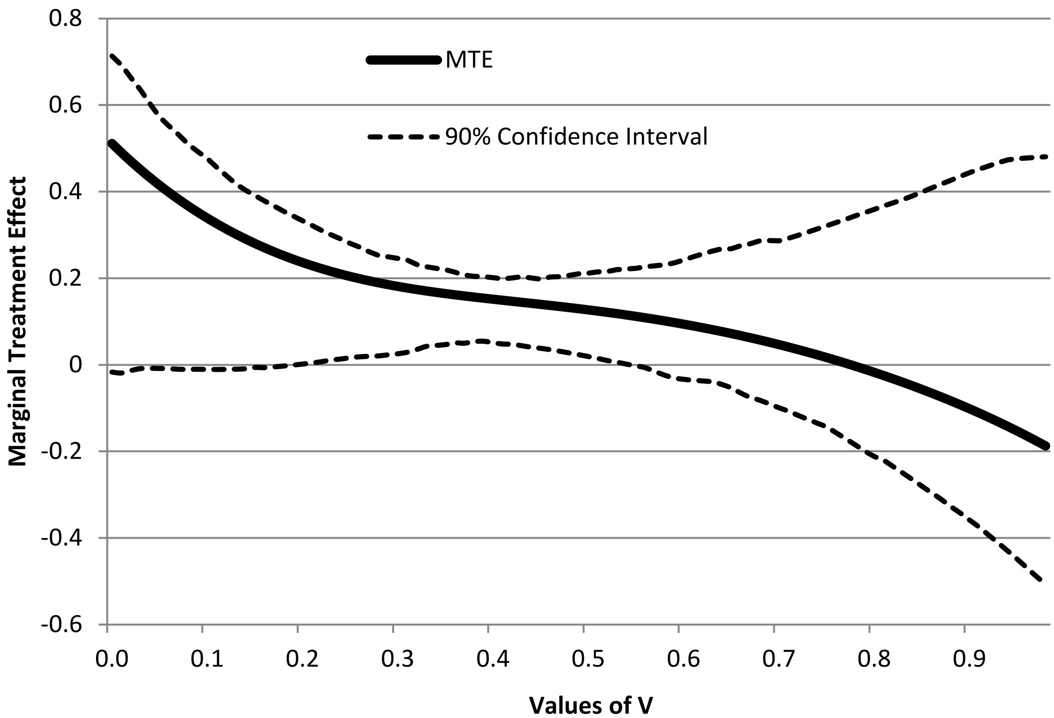
Note: This table presents estimates of various returns to upper secondary school attendance for the semi-parametric and normal selection models: average treatment on the treated (ATT), average treatment effect (ATE), treatment on the untreated (ATU), marginal policy relevant treatment effect (MPRTE), and the policy relevant treatment effect (PRTE) corresponding to a 10% reduction in distance to upper secondary school. Returns to upper school are annualized to show returns for each additional year. Bootstrapped Highest Posterior Density 95% intervals are reported in parentheses.

Figure 1: Propensity score (*P*) support for each schooling group *S* = 0 and *S* = 1



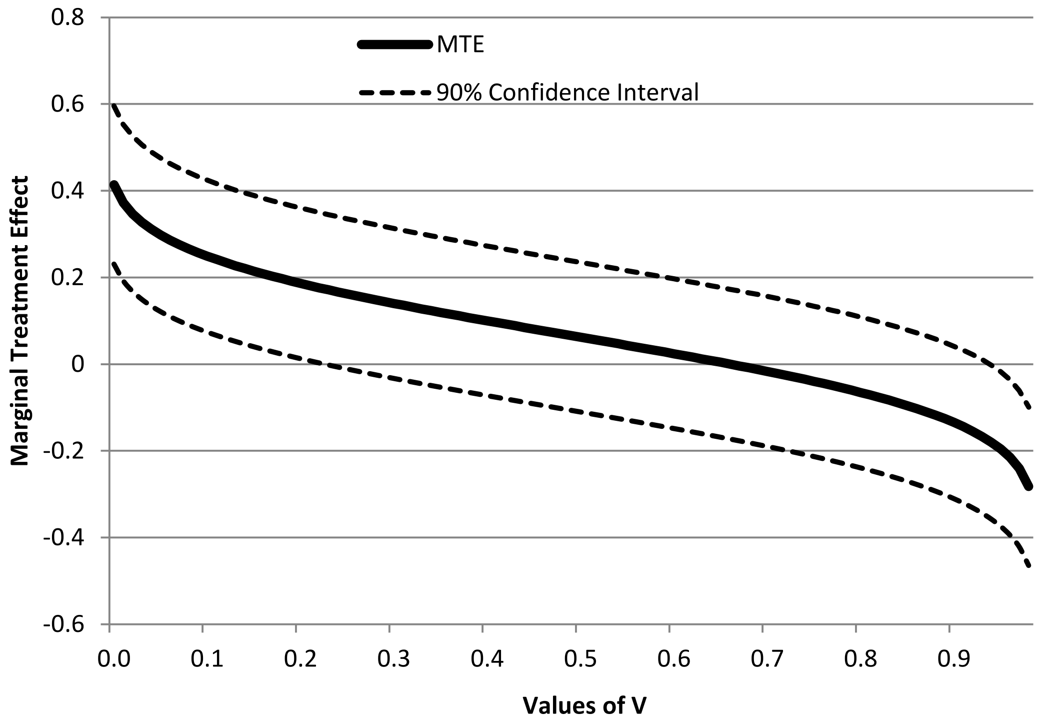
Note: *P* is estimated probability of going to upper secondary school. It is estimated from a logit regression of upper school attendance on Xs, distance to school, interactions of X and distance to school (Table 4).

Figure 2: Marginal treatment effect with 90% Confidence Interval – Semi-parametric regression estimates



Note: To estimate the $E(Y_1-Y_0|X, U_s)$ function we used a partial linear regression of log wages on X and K(P) ,with a bandwidth of 0.27. X includes age, age squared, religion, parental education, rural and province dummy variables. 90% confidence interval constructed using 250 boostrap repetitions. Values of V on the x-axis.

Figure 3: MTE with 90% Confidence Interval – Parametric normal selection model estimates



Note: Parametric MTE estimated using a switching regression model with normally distributed errors.

Appendix A

Simulation-based approach for estimating average treatment effects in equations 7 and 12.

Step 1: Estimate MTE(x, v) as described in section 3.

Step 2: For each individual in the sample construct the corresponding *P(Z)*. Then, since $V \sim Unif\,[0,1]$, take a grid with *n* values uniformly spaced between 0 and 1 (recall that we assumed that *V* was independent of *X* and *Z*). Assign all *n* values of *V* in the grid to each individual. Since there are 2608 individuals in the sample, we can then create a simulated dataset of size $2608^{*}n$ (we use *n*=1000). Evaluate the *MTE(x,v)* for each value of *X* and each value of simulated *V*.

Step 3: In this simulated dataset both *X* and *V* are observed for all 2608^{*}n observations. In addition, we have estimates of *MTE(x,v)* for each of them. Therefore it is trivial to construct the following quantities:

$$ATE = \iint MTE(x,v) f_{X,V}(x,v)dxdv$$

$$ATT = \iint MTE(x,v) f_{X,V}(x,v|S=1)dxdv = \iint MTE(x,v) f_{X,V}(x,v|P>V)dxdv$$

$$ATU = \iint MTE(x,v) f_{X,V}(x,v|S=0)dxdv = \iint MTE(x,v) f_{X,V}(x,v|P \leq V)dxdv$$

by respectively averaging the *MTE* for everyone in the simulated sample, for those who have *P>V*, and for those with *P≤V*.

Step 4: There is one parameter that remains to be estimated: the AMTE. The version of the AMTE we use in this paper defines marginal individuals as those for whom:

$$|Z_\gamma - U_s| < \varepsilon$$

Carneiro, Heckman and Vytlacil (2010) show that this is equivalent to estimating the average return to schooling for those induced to enroll in upper secondary schooling when one of the components of *Z*, say the intercept, changes my a marginal amount. This is exactly what we do in our simulations: we change the intercept of the selection equation marginally and we see which members of our simulated dataset change their schooling decision. Finally, we average the MTE for this group.

Appendix B

Table A1: Regression of elementary education experiences on distance to school

	Failed grade	**Number of repeats**	**Worked**
Dist. to nearest secondary school in km	0.007	0.011	-0.001
	(0.007)	*(0.008)*	*(0.005)*
Mean of dependent variable	0.227	0.258	0.087
	(0.427)	*(0.6158)*	*(0.285)*
Number of observations	2,248	2,244	2,250
R2	0.041	0.043	0.043

Note: Sample restricted to males with the repeated grade information non-missing. The individual and family controls include age, age squared, religion, fathers and mother's schooling levels completed, distance to local health outpost, rural and province dummies. All regressions include individual and family controls, and location fixed effects. Standard errors (in parenthesis) are robust to clustering at the community level, with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table A2: OLS and IV estimates of the return to a year of schooling

	OLS		First stage		IV	
	Coef	**se**	**Average Marginal Derivative**	**se**	**Coef**	**Se**
Years of education	0.096***	*0.005*			0.157***	*0.037*
Age	0.058***	*0.017*	0.027	*0.078*	0.055***	*0.018*
Age Squared	-0.047**	*0.022*	-0.062	*0.098*	-0.042*	*0.022*
Muslim						
Protestant	0.084	*0.082*	2.033	*0.381*	-0.037	*0.118*
Catholic	0.003	*0.152*	2.196	*0.856*	-0.117	*0.149*
Other religions	0.055	*0.121*	0.987	*0.754*	0.002	*0.128*
Father uneducated						
… elementary	0.062	*0.048*	1.759	*0.228*	-0.049	*0.080*
… secondary or higher	0.135**	*0.067*	3.627	*0.312*	-0.083	*0.144*
Mother uneducated						
… elementa	-0.086*	*0.046*	1.000	*0.216*	-0.147**	*0.063*
… secondary or higher	-0.119	*0.078*	3.173	*0.344*	-0.316**	*0.145*
Rural household	0.149***	*0.044*	-1.146	*0.301*	0.234***	*0.073*
Distance to health post in km	-0.020	*0.015*	0.037	*0.084*	-0.015	*0.013*
Location controls			Yes			
Dist to nearest sec school			-0.298***	*0.102*		
Number of observations	2,608				2,608	
Test for joint significance of instruments: F-Stat/p-value			3.62/0.000			
R2	0.260				0.204	

Note: This table reports the coefficients for OLS and 2SLS IV for regression of log of hourly wages on years of schooling controlling for parental education, religion and location. We report average marginal derivatives for the first stage equation. Excluded instruments are distance to secondary school and interactions with parental education, religion, age and distance to health center. Type of location is controlled using province dummies. A dummy variable for missing parental education is included in the regressions but not reported in the table. Standard errors (in parenthesis) are robust to clustering at the community level, with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table A3: IV estimates of the return to a year of schooling without distance and X interactions

	IV		First stage	
	coef	**se**	**coef**	**se**
Years of education	0.144***	0.053		
Age	0.056***	0.017	0.036	0.077
Age Squared	-0.043*	0.022	-0.072	0.096
Muslim				
Protestant	-0.011	0.141	2.050***	0.380
Catholic	-0.091	0.164	2.229**	0.906
Other religions	0.014	0.128	0.839	0.778
Father uneducated				
… elementary	-0.025	0.102	1.800***	0.231
… secondary or higher	-0.036	0.198	3.525***	0.316
… education missing	-0.034	0.109	0.353	0.444
Mother uneducated				
… elementary	-0.134*	0.073	0.973***	0.215
… secondary or higher	-0.274	0.185	3.180***	0.331
… education missing	-0.183***	0.063	0.367	0.301
Rural household	0.215**	0.091	-1.144***	0.302
Distance to health post in km	-0.016	0.013	0.007	0.082
W Java				
N Sumatra	0.114	0.088	-0.615	0.500
W Sumatra	0.282**	0.112	-0.704	0.476
S Sumatra	0.137	0.125	0.667	0.476
Lampung	-0.044	0.108	0.149	0.477
Jakarta	-0.077	0.078	0.752*	0.421
C Java	0.051	0.091	-0.937*	0.498
Yogyakarta	-0.303***	0.100	1.128**	0.570
E Java	-0.007	0.066	-0.300	0.411
Bali	-0.197	0.159	1.027	0.946
W Nusa Tenggara	-0.176	0.107	0.715	0.839
S Kalimantan	0.298***	0.114	1.726***	0.540
S Sulawesi	0.032	0.097	0.226	0.702
Dist to nearest sec school			-0.244***	0.072
Number of observations	2,608			
Test for joing significance of instruments: F-stat/p-value			11.34/0.00	
R2	0.206			

Note: This table reports the coefficients for 2SLS IV for regression of log of hourly wages years of schooling, controlling for parental education, religion and location. Excluded instruments are distance to secondary school. Type of location is controlled using province dummies. Dummy variable for missing parental education is included in the regressions but not reported in the table. Reference categories are Muslim, and not educated. Standard errors (in parenthesis) are robust to clustering at the community level, with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table A4: IV estimates of the return to a year of schooling without regional dummies

	IV	
	coef	**Se**
Years of education	0.135***	0.034
Age	0.059***	0.018
Age Squared	-0.046**	0.022
Muslim		
Protestant	-0.032	0.100
Catholic	-0.153	0.154
Other religions	-0.109	0.091
Father uneducated		
… elementary	-0.006	0.077
… secondary or higher	-0.004	0.141
… education missing	-0.002	0.107
Mother uneducated		
… elementary	-0.074	0.057
… secondary or higher	-0.190	0.131
… education missing	-0.156***	0.060
Rural household	0.227***	0.072
Distance to health post in km	-0.008	0.014
Number of observations	2,608	
Test for joing significance of instruments: F-stat/p-value	4.08/0.00	
R2	0.22	

Note: This table reports the coefficients for 2SLS IV for regression of log of hourly wages years of schooling, controlling for parental education, religion and location. Excluded instruments are distance to secondary school. Type of location is controlled using province dummies. Dummy variable for missing parental education is included in the regressions but not reported in the table. Reference categories are Muslim, and not educated. Standard errors (in parenthesis) are robust to clustering at the community level, with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table A5: IV estimates of the return to upper secondary schooling – Additional sensitivity results

	Basic	**No Z*X Interactions**	**IV using P**
Upper Secondary	0.129***	0.282*	0.095**
(annualized by 7.79)	*(0.048)*	*(0.144)*	*(0.047)*
Upper Secondary	0.112***	0.244*	0.082**
(annualized by 8.99)	*(0.042)*	*(0.125)*	*(0.041)*

Note: This table reports the coefficients for OLS and 2SLS IV for regression of log of hourly wages on upper school attendance (a dummy variable that is equal to 1 if an individual has ever attended upper secondary school and equal to 0 if he has never attended upper secondary school but graduated from lower secondary school), controlling for parental education, religion and location. Excluded instruments are distance to secondary school and interactions with parental education, religion and age in the first column (our main specification in the paper), distance to secondary school in the second column, and the estimated propensity score in the third column. Type of location is controlled using province dummies. A dummy variable for missing parental education is included in the regressions but not reported in the table. Reference categories are Muslim for religion, and not educated. We present estimates normalized by 7.79 (first line), the average difference in years of schooling between those with and without upper secondary school attendance, and 8.99 (second line), the estimated coefficient of an IV regression of years of completed education on upper secondary school attendance (using our basic specification). Standard errors (in parenthesis) are robust to clustering at the community level with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table A6: Outcome equation: Partial linear regression estimates

	Coeffients	Standard Errors
Age	0.070*	*0.042*
Age Squared	-0.076	*0.051*
Protestant	-0.022	*0.368*
Catholic	-0.816	*0.634*
Other religions	0.786*	*0.406*
Father with elementary education	0.042	*0.192*
… secondary or higher	0.103	*0.675*
… education missing	0.425	*0.292*
Mother with elementary education	-0.144	*0.156*
… secondary or higher	-1.570*	*0.938*
… education missing	-0.173	*0.170*
Rural household	0.288*	*0.161*
Distance to health post in km	-0.016	*0.030*
N Sumatra	0.333	*0.214*
W Sumatra	0.177	*0.218*
S Sumatra	0.233	*0.309*
Lampung	0.253	*0.294*
Jakarta	-0.248	*0.233*
C Java	0.071	*0.153*
Yogyakarta	-0.127	*0.301*
E Java	-0.071	*0.149*
Bali	-1.022**	*0.478*
W Nusa Tenggara	-0.267	*0.325*
S Kalimantan	0.013	*0.451*
S Sulawesi	-0.434	*0.274*
N Sumatra	-0.550	*0.465*
S Sumatra	-0.134	*0.595*
C Java	-0.197	*0.415*
Yogyakarta	-0.127	*0.602*
E Java	0.326	*0.357*
Bali	1.660*	*0.898*
W Nusa Tenggara	0.192	*0.711*
S Kalimantan	0.367	*0.860*
W Sumatra*P	0.465	*0.535*
Lampung*P	-0.993	*0.839*
Jakarta*P	0.394	*0.452*
S Sulawesi*P	0.979	*0.598*
Age*P	-0.069	*0.097*
Age Squared*P	0.124	*0.121*
Protestant*P	0.130	*0.639*
Catholic*P	1.171	*0.931*
Other religions*P	-1.261*	*0.703*
Father with elementary*P	0.053	*0.605*
Father with secondary/higher*P	0.002	*1.280*
Father education missing*P	-1.322	*0.942*
Mother with elementary*P	0.187	*0.393*
Mother with secondary/higher*P	1.977	*1.433*
Mother education missing*P	0.109	*0.458*
Rural *P	-0.275	*0.362*
Distance to health post*P	0.037	*0.082*
Number of observations	2,608	
R2	0.080	

note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$ The table presents the coefficients on X and $P^{*}X$ from the Robinson's (1988) double residual semi-parametric regression estimator. The logit estimated pscore (P) enters the equation nonlinearly according to a non-binding function and estimated using a gaussian kernel regression with bandwidth equal to 0.27.

Table A7: Test for heterogeneity in returns: compare adjacent sections of the semi-parametric MTE

Ranges of U_S for *LATE*^{j}	(0,0.1)	(0.1,0.2)	(0.2,0.3)	(0.3,0.4)	(0.4,0.5)	(0.5,0.6)	(0.6,0.7)	(0.7,0.8)	(0.8,0.9)
Ranges of U_S for *LATE*$^{j+1}$	(0.1,0.2)	(0.2,0.3)	(0.3,0.4)	(0.4,0.5)	(0.5,0.6)	(0.6,0.7)	(0.7,0.8)	(0.8,0.9)	(0.9,1)
Difference in LATEs	-0.078	-0.039	-0.013	-0.012	0.00	0.005	-0.014	-0.024	-0.04
p-value	0.00	0.00	0.00	0.00	0.597	0.759	0.005	0.00	0.00

Note: In order to compute the numbers in this table we construct groups of values of *Us* and average the MTE within these groups, where U_S^{Lj} and U_S^{Hj} are the lowest and highest values of *Us* defined for interval *j*. Then we compare the average MTE across adjacent groups and test whether the difference is equal to zero using the bootstrap with 250 replications. Take, for example, the first column of the table. In the first line we show the average value the MTE takes when *X* is fixed at its mean and *V* takes values between 0 and 0.1, while the second line corresponds to values of *V* between 0.1 and 0.2. The third line shows the difference between the first two lines, and the fourth line reports the p-value of a test of whether this difference is equal to zero. We reject equality in almost all columns of the table at the 5% significance level. Therefore, we reject that the MTE is flat, even with the large standard errors shown in figure 2.

Table A8: Testing for equality of LATEs over different Intervals of MTE – Model Without Interactions between X and Z

$$\left(\mathbf{H_0: LATE^j}\left(\mathbf{U_S^{L_j}, U_S^{H_j}}\right) - \mathbf{LATE^{j+1}}\left(\mathbf{U_S^{L_{j+1}}, U_S^{H_{j+1}}}\right) = \mathbf{0}\right)$$

Ranges of U_S for *LATE*^{j}	(0,0.1)	(0.1. 0.2)	(0.2,0.3)	(0.3,0.4)	(0.4,0.5)	(0.5,0.6)	(0.6,0.7)	(0.7,0.8)	(0.8,0.9)
Ranges of U_S for *LATE*$^{j+1}$	(0.1. 0.2)	(0.2,0.3)	(0.3,0.4)	(0.4,0.5)	(0.5,0.6)	(0.6,0.7)	(0.7,0.8)	(0.8,0.9)	(0.9,1)
Difference in LATEs	-0.078	-0.04	-0.014	-0.012	-0.010	-0.011	-0.012	-0.014	-0.014
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: In order to compute the numbers in this table we construct groups of values of *Us* and average the MTE within these groups, where U_S^{Lj} and U_S^{Hj} are the lowest and highest values of *Us* defined for interval *j*. Then we compare the average MTE across adjacent groups and test whether the difference is equal to zero using the bootstrap with 250 replications.

Table A9: Estimates of Average Returns to Upper Secondary Schooling with 95% confidence interval – Model Without Interactions between X and Z

Parameter	*Non parametric Estimate*	*Normal selection model*
ATT	0.217 (-.1, 0.525)	0.198* (-0.041,0.438)
ATE	0.13 (-0.06, 0.32)	0.065 (-0.099, 0.231)
ATU	0.07 (-0.227, 0.365)	-0.028 (-0.217, 0.160)

Note: This table presents estimates of various returns to upper secondary school attendance for the semi-parametric and normal selection models: average treatment on the treated (ATT), average treatment effect (ATE), treatment on the untreated (ATU), and marginal policy relevant treatment effect (MPRTE). Returns to upper school are annualized to show returns for each additional year. Bootstrapped 95% confidence interval are reported in parentheses, with significance at *** p<0.01, ** p< 0.05, * p<0.1 indicated.

Table A10: Estimates of Average Returns to Upper Secondary Schooling for different bandwidths

Bandwidth	0.05	0.10	0.15
ATT	0.800	0.456	0.347
	(0.0632, 0.8018)	(0.026,0.555)	(-0.001,0.454)
ATE	0.155	0.062	0.090
	(-0.183, 0.425)	(-0.109,0.318)	(-0.050,0.296)
ATU	-0.311	-0.22	-0.094
	(-0.584, 0.509)	(-0.388,0.434)	(-0.297,0.337)
MPRTE	0.106	0.119	0.130
	(0.068, 0.815)	(0.030,0.575)	(-0.013, 0.449)
PRTE	0.109	0.125	0.136
	(0.075, 0.870)	(0.048,0.578)	(-0.036,0.444)

Note: This table presents estimates of various returns to upper secondary school attendance for the semi-parametric and normal selection models: average treatment on the treated (ATT), average treatment effect (ATE), treatment on the untreated (ATU), marginal policy relevant treatment effect (MPRTE), and the policy relevant treatment effect (PRTE) corresponding to a 10% reduction in distance to upper secondary school. Returns to upper school are annualized to show returns for each additional year. Bootstrapped Highest Posterior Density 95% intervals are reported in parentheses.

Table A11: Estimates of Average Returns to Upper Secondary Schooling – by age group

Parameter	Non parametric Estimate	Normal selection model
	Respondents younger than the median age of 37 years old	
ATT	0.006	0.003
	(-0.230, 0.280)	(-0.078, 0.242)
ATE	0.078	-0.046
	(-0.072, 0.198)	(-0.087, 0.162)
ATU	0.141	-0.088
	(-0.107, 0.304)	(-0.140, 0.149)
	Respondents older than the median age of 37 years old	
ATT	0.291	0.184
	(0.025, 0.435)	(0.001, 0.264)
ATE	0.057	0.071
	(-0.086, 0.288)	(-0.007, 0.267)
ATU	-0.080	0.006
	(-0.344, 0.387)	(-0.054, 0.305)

Note: This table presents estimates of various returns to upper secondary school attendance for the semi-parametric and normal selection models: average treatment on the treated (ATT), average treatment effect (ATE), and treatment on the untreated (ATU). Returns to upper school are annualized to show returns for each additional year. Bootstrapped Highest Posterior Density 95% intervals are reported in parentheses.

Table A12: Estimates of Average Returns to Upper Secondary Schooling for different subsamples

Sample	Only Wage Earners	Only Self-Employed	Both
IV	0.129	0.212	0.106
	(0.048)	(0.142)	(0.070)
ATT	0.218	0.218	0.213
	(-0.002, 0.353)	(-0.106,0.419)	(0.002,0.342)
ATE	0.138	0.099	0.104
	(0.025, 0.254)	(-0.204,0.422)	(-0.018,0.271)
ATU	0.081	0.066	0.051
	(-0.121, 0.326)	(-0.344,0.493)	(-0.127,0.296)
MPRTE	0.140	0.192	0.143
	(-0.049, 0.338)	(-0.134, 0.399)	(0.021,0.382)
PRTE	0.145	0.174	0.141
	(-0.053, 0.337)	(-0.123,0.417)	(0.002,0.342)

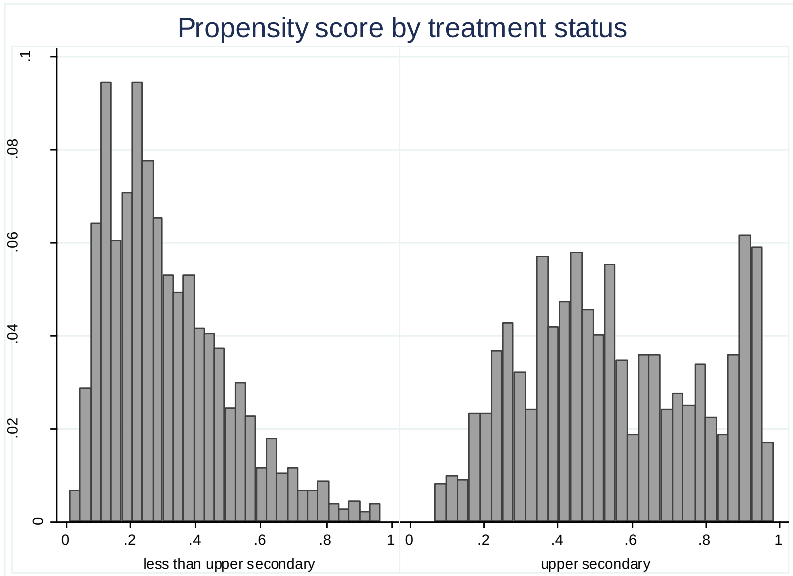
Note: This table presents estimates of various returns to upper secondary school attendance for the semi-parametric and normal selection models: average treatment on the treated (ATT), average treatment effect (ATE), treatment on the untreated (ATU), marginal policy relevant treatment effect (MPRTE), and the policy relevant treatment effect (PRTE) corresponding to a 10% reduction in distance to upper secondary school. Returns to upper school are annualized to show returns for each additional year. Bootstrapped Highest Posterior Density 95% intervals are reported in parentheses. In the case of the IV estimates we report standard errors in parenthesis.

Table A13: Estimates of Average Returns to Upper Secondary Schooling Omitting Post-Schooling Controls

IV	0.116
	(0.045)
ATT	0.228
	(-0.015, 0.383)
ATE	0.145
	(0.018, 0.277)
ATU	0.085
	(-0.179, 0.315)
MPRTE	0.137
	(-0.015, 0.383)
PRTE	0.150
	(-0.015, 0.383)

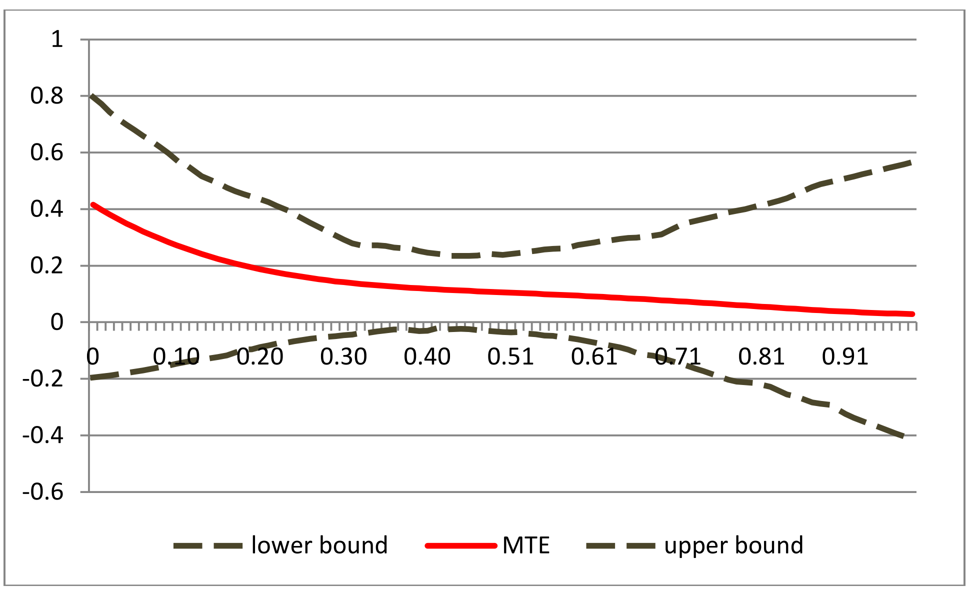
Note: This table presents estimates of various returns to upper secondary school attendance for the semi-parametric and normal selection models: average treatment on the treated (ATT), average treatment effect (ATE), treatment on the untreated (ATU), marginal policy relevant treatment effect (MPRTE), and the policy relevant treatment effect (PRTE) corresponding to a 10% reduction in distance to upper secondary school. Returns to upper school are annualized to show returns for each additional year. Bootstrapped Highest Posterior Density 95% intervals are reported in parentheses. In the case of the IV estimates we report standard errors in parenthesis.

Figure A1: Propensity score (*P*) support for each schooling group *S* = 0 and *S* = 1 (without distance and Xs interactions)



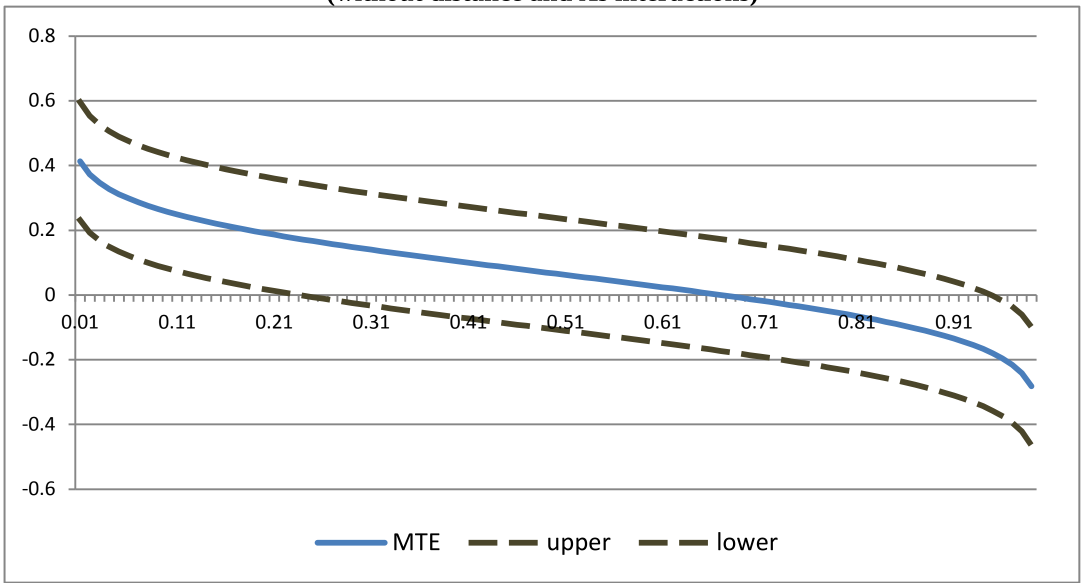
Note: *P* is estimated probability of going to upper secondary school. It is estimated from a logit regression of upper school attendance on Xs, distance to school (Table A2).

Figure A2: Marginal treatment effect with 90% Confidence Interval – Semi-parametric regression estimates (without distance and Xs interactions)



Note: To estimate the $E(Y_1-Y_0|X, U_s)$ function we used a partial linear regression of log wages on X and K(P) ,with a bandwidth of 0.27. X includes age, age squared, religion, parental education, rural and province dummy variables. 90% confidence interval constructed using 250 boostrap repetitions. Values of V on the x-axis.

Figure A3: MTE with 90% Confidence Interval – Parametric normal selection model estimates (without distance and Xs interactions)



Appendix C - Simulating the consequences of violations of instrument validity

Take the following model of section 4.1:

$$y = \tau S + u$$

$$S = \sigma d^* + v$$

$$d = \rho S + \delta d^* + \varepsilon \quad (14)$$

where y is the outcome (wages), S is schooling (upper secondary education), d^* is distance at the time of the secondary school decision (which is unobserved), and d is current distance (measured at the time of the outcome).

This is a possible representation of the IV model that we estimate in the paper, which also includes control variables X. τ may be a random coefficient. In this appendix we briefly examine what could potentially happen to the IV estimates we present in the paper if $\rho \neq 0$, if $Cov(u, \varepsilon) \neq 0$, or both.

We start by examining examine what could happen as ρ becomes increasingly different from zero, and we now present some simple simulations which illustrate what could happen. We assume that $Cov(u, \varepsilon) = 0$.

As mentioned above, for a small subsample, we have estimates of both d and d^*. For this subsample, we estimate equation (14) and, after adding the control variables used in the rest of the paper, we obtain an estimate of ρ of -0.13, with a 95% confidence interval going from -0.3 to 0.04. Since we do not expect $\rho > 0$, in our simulation we consider values for this parameter between -0.3 and 0.

The simulation works in a very simple way. Take 10 different values of ρ equally spaced between -0.03 and -0.3 (including the extremes of this interval) we construct a new pseudo instrument: $\tilde{d} = d - \rho S$ (we understand that we could have chosen to divide this by δ, but since it is just a constant, we can ignore it). This gives us ten different values for $\tilde{d}$ which we can use to generate ten different IV estimates, which go from 0.128 when $\rho = -0.03$ to 0.051 when $\rho = -0.30$. The original IV estimate is 0.129 and all the simulated IVs are below it, as predicted by the simple model we wrote above. The entire set of simulated values is as follows:

ρ	0	-0.03	-0.06	-0.09	-0.12	-0.15	-0.18	-0.21	-0.24	-0.27	-0.30
$\hat{\tau}_{IV}$	0.129	0.129	0.125	0.115	0.101	0.085	0.070	0.060	0.054	0.052	0.051

The second set of simulations attempts to examine what happened to our IV estimate if the assumption that $Cov(u, \varepsilon) = 0$ was violated. In order to conduct these simulations in the simplest possible way we start by adding an additional equation to the system defined by (14):

$$\varepsilon = \pi u + \epsilon$$

where $Cov(u, \epsilon) = 0$.

Then we can write:

$$u = y - \tau S$$

$$d = \rho S + \delta d^{*} + \pi u + \epsilon = (\rho - \pi\tau)S + \delta d^{*} + \pi y + \epsilon \; (15)$$

As above, we estimate π by running a regression of d on S, d^{*}, and y. π is the coefficient on y. Then, assuming now that $\rho = 0$, and that τ equals the current IV estimate (since we need to plug a value for this parameter), we compute the following alternative instrument for ten different values of π between -0.01 and -0.1 (since -0.01 is the lower limit of the estimated 95% confidence interval for this parameter when we estimate equation (15)).

The set of simulated values is as follows:

π	0	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.07	-0.08	-0.09	-0.10
$\hat{\tau}_{IV}$	0.129	0.115	0.100	0.086	0.071	0.058	0.046	0.034	0.023	0.013	0.005

Appendix D - Comparison with Carneiro, Heckman and Vytlacil (2011) – Monte Carlo Simulation

In this appendix we compare the procedures in this paper and the one in Carneiro, Heckman and Vytlacil (2011). The MTE is estimated exactly in the same way in the two procedures, so there is no difference there. The only difference is in the calculation of the treatment parameters, as weighted averages of the MTE.

We simulate the model described in equations (1) and (2) of our paper. There are two regressors, which we call x_1 and x_2, and two instrumental variables, which we call z_1 and z_2. These variables are simulated to be normal and independent of each other, with means equal to -2, 2, -1, and 1, and with variances equal to 4, 4, 9 and 9, respectively.

In order to simulate the unobservables of the model, we start by constructing three independent standard normal random variables, e_1, e_2 and e_3. Then:

$$U_1 = 0.012e_1 + 0.01e_2$$

$$U_0 = -0.05e_1 + 0.02e_3$$

$$V = -e_1$$

Finally:

$$Y_1 = 0.24 + 0.8x_1 + 0.4x_2 + U_1$$

$$Y_0 = 0.02 + 0.5x_1 + 0.1x_2 + U_0$$

$$S = 1 \; if \; 0.2 + 0.3z_1 + 0.1z_2 - V > 0$$

Given these parameters, we simulate 1000 datasets with 2000 observations each, and then estimate ATE, TT and TUT for each of them, using the procedure in our paper, and the procedure in Carneiro, Heckman and Vytlacil (2011). The table below compares the results.

In each line of the table we have a different treatment parameter. In the first column, labeled True, we present the true value of each parameter. The second column, labeled CHV, shows the mean and standard deviation of the estimates of each parameter using the method in Carneiro, Heckman and Vytlacil (2011). The third column, labeled MSE-CHV, shows the corresponding mean squared error. The fourth column, labeled CLNR, shows the mean and standard deviation of the estimates using the method developed in this paper. Finally, the last column, labeled MSE-CLNR, shows the corresponding standard error.

	True	CHV	MSE-CHV	CLNR	MSE-CLNR
ATE	0.2200	0.2036 (0.0600)	0.0039	0.2246 (0.0356)	0.0013
TT	0.2558	0.2380 (0.0676)	0.0049	0.2584 (0.0487)	0.0024
TUT	0.1842	0.1704 (0.0685)	0.0049	0.1908 (0.0500)	0.0025