WORKING PAPERS IN ECONOMICS & ECONOMETRICS

NONPARAMETRIC BOUNDS ON RETURNS TO EDUCATION IN SOUTH AFRICA: OVERCOMING ABILITY AND SELECTION BIAS

Martine Mariotti
Research School of Social Sciences and School of Economics
The Australian National University
Canberra, ACT 0200, Australia.
E-Mail: martine.mariotti@anu.edu.au

Juergen Meinecke
School of Economics
The Australian National University
Canberra, ACT 0200, Australia
E-mail: juergen.meinecke@anu.edu.au

Working Paper No: 510
ISBN: 0 86831 510 9

December 2009

NONPARAMETRIC BOUNDS ON RETURNS TO EDUCATION IN SOUTH AFRICA: OVERCOMING ABILITY AND SELECTION BIAS

Martine Mariotti and Jürgen Meinecke[1]

Our objective is to estimate the average treatment effect (ATE) of education on earnings for African men in South Africa. Estimation of the ATE in our data is difficult because of omitted ability bias and a high degree of sample selection due to low labor force participation. Manski and Pepper (2000) suggest is a promising nonparametric identification strategy but it only helps with the problem of omitted ability bias. We propose an extension of their identification strategy to deal with the sample selection problem.

Accounting for ability and selection bias, we compute upper bounds on the ATE for the years 1995 and 2000. We estimate an upper bound of 12.64 percent in 1995 and 10.68 percent in 2000. Compared to parametric estimation our bounds are informative: The OLS returns to schooling equal 15.59 percent in 1995 and 15.31 percent in 2000. Our results suggest that many parametric estimates are severely upwards biased, which results from unobserved heterogeneity.

Keywords: returns to schooling, partial identification, nonparametric estimation.

1. INTRODUCTION

One of the policies of Apartheid in South Africa (1950–1994) was the unequal distribution of education across race groups such that whites have historically attained high levels of education while Africans have attained low levels. An important consequence of the educational distribution was unequal opportunities for employment in skilled occupations, with Africans forced to work in lower skilled occupations because of their lower educational attainment (Mariotti 2009). As a result, a post Apartheid adjustment to education policy was called for in order to level the playing field with the expectation that higher education leads to higher incomes through employment in more skilled occupations.

However, there is little consensus on just how much higher education contributes to higher incomes. A number of studies have attempted parametric estimations of the returns to education in post–Apartheid South Africa. A summary of these results shows a wide distribution of returns with some returns as high as 100 percent for secondary and tertiary education. A possible explanation for the large spread in the point estimates is that the data are sensitive to the parametric methodology used. The lack of suitable instruments further contributes to the difficulty in pinning down a reliable point estimate.

Research School of Social Sciences and School of Economics, The Australian National University, Canberra, ACT, 0200, Australia.

martine.mariotti@anu.edu.au

School of Economics, The Australian National University, Canberra, ACT, 0200, Australia juergen.meinecke@anu.edu.au

[1]We thank seminar participants at ESAM 2009, the ANU School of Economics seminar, the RSSS brown bag seminar, UNSW and La Trobe University for valuable comments. We are grateful to Farshid Vahid, Timothy Hatton and Bob Gregory for suggestions. Guochang Zhao provided excellent research assistance.

Manski and Pepper's (2000) nonparametric identification strategy is the starting point for our estimation. Manski and Pepper propose two mild monotonicity assumptions with which they are able to impose an upper bound on the average treatment effect of education one earnings. First, the assumption of monotone treatment response (MTR) states that the human capital earnings function is weakly increasing in education. This is implicit in any Mincerian wage regression and should be uncontroversial. Second, monotone treatment response (MTS) concerns how people select themselves into education. Roughly speaking, it means that high ability people select themselves into higher levels of education.

The MTS assumption is credible whenever people can freely select themselves into education. If there exist institutional restrictions (e.g., borrowing constraints, discrimination, household bargaining) that prevent highly able people to access high levels of education then the MTS assumption could be violated. In South Africa, for African males, this might be the case. In addition, this group has a low labor force participation rate of around 50%. The determinants that drive selection into education and employment are partly unobserved and are likely to break the MTS assumption.

We therefore propose an extension to Manski and Pepper's framework to account for sample selection into employment. We introduce the weakest set of assumption that enables us to identify the ATE of education on earnings for the population of African males in South Africa.

The data that we use come from the October Household Survey and Income and Expenditure Survey of 1995 and the Labor Force Survey and Income and Expenditure Survey of 2000. Using our modified nonparametric estimator, we find an upper bound on the average treatment effect of a high–school degree of 12.64 percent (per year of high–school) in 1995 and 10.68 percent in 2000. Compared to standard parametric estimates, our bound is informative. The OLS estimates of the return to education for 1995 and 2000 are 15.59% and 15.31%.

2. LITERATURE REVIEW

One of the secondary benefits of the fall of Apartheid in 1994 has been an improvement in data collection that has provided researchers the opportunity to document the transformation of individual social and economic characteristics. In particular, a large body of work has documented changes in the return to education since the fall of Apartheid. The results are consistent in the racial hierarchy of returns, in that, firstly, Africans persistently earn a higher return for higher levels of education and, secondly, that higher levels of education earn a higher return across all race groups relative to no education.[1] However, there is a large variation in the quantification of the returns with several studies finding surprisingly high returns.

Thomas (1996) provides a brief, useful account of the state of education in South Africa by 1991. Dividing the 1991 population census into cohorts he shows that whites attained higher levels of education

[1]The racial hierarchy arises most likely because the proportion of Africans with high levels of education—a legacy of the apartheid education system—is extremely low. This result is despite the restricted employment opportunities for highly educated Africans.

than non-whites, and that levels of education have been increasing over time for all race groups. Indians and Coloureds have the second highest levels of educational attainment, with Africans acquiring the lowest amount of education. Following the Soweto school riots of 1976, the government increased expenditure on African education within South Africa (but not the homelands which were supposed to be funding their own students). Despite the increased expenditure, by 1991 African education continued to lag behind that of the other races.

Mwabu and Schultz (1996) use the Project for Standards of Living Survey Data (PSLSD) set of 1993 to measure the returns to education. Using the working age population, an OLS regression finds that the return to education is 16% for secondary education and 27% for higher education for Africans while for whites the comparative returns are 8% and 15%. The quantile regression results find that the return to primary education for Africans is between 10% and zero, the return to some secondary education is between 10% and 18% (although the difference is not significant) and the return to higher education ranges from around 23% to around 30%. For whites, the return to primary education is zero, the return to secondary ranges from zero to 20% and the return to higher education ranges from around 7% to around 15%. They also apply the quantile regression approach to determine the direction of correlation between education and ability. The negative correlation between the two suggests that people substitute education for ability.

In a later paper, Mwabu and Schultz (2000) look again at returns to education and find that the returns are greater at higher levels of educational attainment for Africans than for whites. This is most likely a result of both the low quantity of Africans who have attained higher education as well as job reservation which means that Africans and whites do not compete for the same jobs and therefore do not compete for wages. Using the PSLSD, they spline education into three groups (primary, secondary and higher), apply OLS and find that the return to African education (for men) is 8.4% at the primary level, 15.8% at the secondary level and 29.4% at the tertiary level. For white men they find 0% for primary, 8.4% for secondary and 15.1% for tertiary. They claim that the results are similar to those found using a Heckman two stage procedure. The authors are concerned with how the returns might change in the future and they note that as more people acquire higher levels of education the return is likely to drop for those levels.

Chamberlain and van der Berg (2002) try to account for differences in the quality of education across race groups in determining the return to education. They proxy for quality using test scores from the PSLSD survey of 1993 and weight the years of schooling an individual has attained in the October Household Survey 1995 by using a predicted test score. Using a two stage selection procedure, they find that the return to education is around 5% before accounting for quality and that it increases to around 6% after accounting for quality.

Serumage-Zake and Naudé (2003) use a double hurdle model where they predict simultaneously whether a person will enter the labor market as well as whether they will find a job. They find returns around 12% using the 1995 October Household Survey.

Hertz (2003) shows that errors in the reporting of educational attainment can bias the estimated return

to education upward. Using the PSLSD and KIDS panel data set in an OLS regression he shows that failing to correct for reporting error results in a return to education of 11 to 13%. Whereas, correcting for the error and using a within-family fixed effects approach reduces the return to between 5 and 6%. He shows that errors in the schooling variable are strongly correlated within families.

Keswell (2004), in a paper examining differences in the return to similar levels of schooling across race groups finds that the rate of return for Africans was around 11% at the end of Apartheid. This finding is from the PSLSD data set. He finds, using the Labour Force Surveys (LFS) of 2001 and 2002, that this return declined to 7%. These private returns to education are measured in OLS and Tobit regressions.

Keswell and Poswell (2004) use several data sets (PSLSD of 1993, October Household Survey (OHS) of 1995 and 1997 and LFS of September 2000) to show that returns to higher levels of education in South Africa are convex. The estimation procedure they use is OLS allowing for non-linear returns to education in the form of polynomials in the second and third degree on the education variables. They find the return to primary school in 1993 is 2%, secondary school 28% and tertiary is between 68 and 72%. In 1995 the return to primary school decreases to zero, secondary school remains around 28% and tertiary education increases to between 71 and 86%. The return to secondary school decreases to 21% in 1997, and that to tertiary education to between 54 and 61%. Finally, in 2000, the return to primary education is negative, secondary drops to between 15 and 16% and tertiary remains constant.

Leibbrandt, Levinsohn and McCrary (2005) use the October Household Survey and Income and Expenditure Survey of 1995 to compare South Africa's income distribution to that found using the Labour Force Surveys and Income and Expenditure Survey in 2000. Using both descriptive methods and non-parametric techniques they find that the income distribution has shifted to the left over the five year period. In determining causes of the shift they find that the return to attributes has declined from 1995 to 2000. Specifically with respect to education, they find that the return to additional years of education decreased for African men and increased for white men. They claim that this result is expected due to continuing labor market rigidities. In 1995 for African men under 60 years of age, the return to education is between 11% and 14%. By 2000 the return for the youngest cohorts has declined by 4 percentage points. It remains constant for older cohorts.

Maitra and Vahid (2005) examine the effect of household characteristics on living standards between 1993 and 1998 in KwaZulu Natal. They use the KIDS panel data set. They account for non-random sample attrition since it appears that wealthier households were more likely to attrite. Using quantile regression techniques, they find that the return to education on log wage ranges from zero at the highest quantile and lowest level of education to 108% for the lowest quantile at the highest level of education. They find a negative correlation between education and ability, possibly a result of limited African access to occupations during apartheid. They find that by 1998 there is no longer any difference in the return across quantiles which they suggest is due to the openness of the labor market after the end of apartheid.

3. DATA

The data we use are from four surveys, two for 1995 and two for 2000 as there is no single data set that contains adequate information on both demographics and incomes for either year. We are able to match households from the two surveys within each year but not across the years, resulting in two cross-sectional data sets.

We combine the 1995 October Household Survey (OHS95) and the 1995 Income and Expenditure Survey (IES95), both collected by Statistics South Africa. The OHS95 is one of a series of general household surveys run from 1994 intermittently until 1999 during October of each year. The surveys contain information on household members' demographic characteristics, education and health levels, employment status, access to infrastructure, and dwelling structures. In 1995 96,261 individuals from 29,700 households were interviewed.

We supplement the OHS95 data with income data from the IES95 which contains detailed information on household and personal income and expenditure. The survey consists of 29,582 households and 128,917 individuals.[2] The merged OHS95 and IES95 data set consists of 114,568 individuals from 27,135 households.[3] Finally, we inflate the data by 39%, the increase in the consumer price index between 1995 and 2000.

In 2000, the Labour Force Survey (LFS) replaced the OHS. It is implemented biannually once in March and once in September. We use the September 2000 survey as it largely contains the same households as those in the IES 2000 (IES00). The LFS00 consists of 105,370 individuals from 26,571 households.The IES00 consists of 104,153 individuals from 26,263 households. Upon merging the IES00 with the LFS00, we have 103,732 individuals from 26,150 households.[4]

We restrict our analysis to African males between the ages of 18 and 65 who are employed and for whom we have education information.[5] We have 9,527 individuals in 1995 and 8,837 individuals in 2000. We ignore females to avoid potential complications with labor force and child rearing decisions.

The basic Manski and Pepper non-parametric analysis uses only income and education data, we provide summary statistics for these two variables in Table I and Table II. Table I shows that the mean real income of African males has declined between 1995 and 2000, consistent with the findings in Hoogeveen and Özler (2004) and Leibbrandt at el. (2005). Table II summarizes educational attainment in 1995 and 2000 by highest educational qualification obtained. A higher percentage of people completed high school in 2000 than in 1995, while a lower percentage obtained no education at all. Yet, overall, the aggregate educational attainment of employed African males hardly changed over the five year period.

One problem that Manski and Pepper do not deal with is sample selection bias. In a developed country environment, focusing the analysis on the working sample does not lead to bias since such a

[2]Leibbrandt, Levinsohn and McCrary (2005) point out that the reason the OHS95 income data is so poor might be because Statistics South Africa anticipated re–interviewing the same households for the IES95 later in the year.

[3]We drop people whose gender or race changes and those who age more than a year between surveys.

[4]Again we delete any observations that do not match up such as people whose gender or race changed between the surveys or those who age more than a year.

[5]The African population is the group most negatively affected by apartheid as well as the most numerous.

TABLE I

—Data: Summary Statistics on Income, OHS/LFS and IES

		1995	2000
Mean		23,432	22,470
	s.d.	22,289	62,937
	Skewness	4.01	51.53
	Kurtosis	42.10	3,448
Log Mean		9.68	9.49
	s.d.	0.94	1.08
	Skewness	-0.46	0.75
	Kurtosis	2.98	5.49
Income percentile:	1	1,668	550
	5	2,836	2,000
	10	4,170	3,480
	25	8,674	7,200
	50	18,354	14,560
	75	30,024	25,200
	90	48,372	43,200
	95	63,384	62,400
	99	100,080	132,000
Observations		9,527	8,837

Note.—In 2000 Rand. Employed African males.

large portion of the adult population is actually in the labor force. However, in developing countries this is often not the case. In particular, the labor force participation rates for African males in the post apartheid period in South Africa are comparatively low. Table III shows that for both 1995 and 2000, African male labor force participation is remarkably low and is declining from 1995 to 2000.

While such a low labor force participation is likely to have substantial social and economic implications, for our purposes the main concern is the resulting sample selection bias - why do some work and others not? And what is the wage of a non-worker likely to be if he had a job? The simplest way to answer this question would be to assume that a working man's wage is the same as the wage a non-worker would earn if he had a job and then to proceed with the estimation. However, since we do not know how a person selects himself into work, we are reluctant to make that assumption. Instead we would like to use a segment of the working sample to place bounds on the wage a non-worker could earn.

In Section 4 we show that the upper bound on the ATE is given by the difference between the upper bound to the return on a more advanced year of education less the lower bound on a less advanced year of education. It follows that we need one segment of the working sample to provide an upper bound on the return a non-worker could earn if he had received a more advanced level of education. In addition, we need a different segment of the working sample to provide a lower bound on the return a non-worker with few years of education could earn. This will give us the unconditional upper bound on the return to education.

TABLE II

—Data: Distribution of Education, OHS/LFS and IES

	1995	2000
No schooling	14.74	11.03
Sub A/Sub B/Standard 1	5.88	7.11
Grade4/Standard 2	6.46	5.24
Grade5/Standard 3	6.08	5.57
Grade6/Standard 4	7.43	6.79
Grade7/Standard 5	8.60	8.98
Grade8/Standard 6	10.35	9.27
Grade9/Standard 7	6.46	6.87
Grade10/Standard 8/NTCI	8.54	8.78
Grade11/Standard 9/NTCII	4.59	6.53
Grade12/Standard 10/NTCIII	12.42	16.17
Diploma	6.42	5.38
Degree	2.05	2.29

Note.—In percent. Employed African males.

As we show in the estimation section, to deal with sample selection bias we need two additional variables to assist us in bounding the ATE a non-worker could earn - one to restrict the upper bound on several years of education, and one to restrict the lower bound on few years of education. These need to be variables that we know affect the choice of whether to work or not. Table IV provides information for two possible variables: the number of migrant workers and the number of pensioners. We use migrant worker wages to establish an upper bound on the return to education for a given year of schooling since we believe that migrant workers are at least as productive as non-migrant workers (references). Due to historical restrictions on population movement in South Africa before and during apartheid a large migrant labor force continues to exist today. We assume that those people who do not leave, regardless of whether they work or not, are unlikely to be as productive as those who have left. This provides us with an upper bound on the return a non-worker with some education might earn. At most, he can earn what a migrant worker earns, not more, since if he was as productive as a person who is already a migrant worker, he would be one too.

For the lower bound, we look at the members of the household who stayed behind and are not migrant workers. Since we do not know from which household a migrant worker comes, we cannot look at his remaining relatives to calculate a lower bound on their wages. However, we do know whether a household has a pensioner or not. Pensioners are valuable members of poor households in South Africa. The South African old age pension is the primary system of redistribution in the economy. It is around twice the average per capita income of African households and is an important source of income for around 1/3 of African households. Pension eligibility is dependent on age (60 years for women and 65 years for men) and is not means tested, this means that any poor household with an elderly person receives the pension. Ardington et al. (2007), Posel et al. (2004), Bertrand

TABLE III

—Data: Labor Force Participation Rates, OHS/LFS

	1995	2000
Age 15 - 19	0.05	0.04
Age 20 - 24	0.20	0.17
Age 25 - 29	0.49	0.43
Age 30 - 34	0.66	0.53
Age 35 - 39	0.68	0.59
Age 40 - 44	0.72	0.57
Age 45 - 49	0.68	0.56
Age 50 - 54	0.65	0.51
Age 55 - 59	0.54	0.53
Age 60 - 64	0.34	0.30
Age 65 - 69	0.18	0.22
Total	0.48	0.40
Observations	18096	20215

Note.—African males aged between 18 and 65.

et al. (2003), Case and Deaton (1998, 1999) have shown that the receipt of a pension affects other household members' labor force participation decisions. They find a lower labor force participation of prime-aged males. Furthermore, they show that households with pensioners are more able to overcome credit constraints. This implies that households with pensioners are more likely to be able to send off their most productive males (as migrant workers) while the less productive males remain at home. We therefore assume that the labor force participation decisions of those males who remain at home with the pensioner are somewhat random, as the more productive members of the household have already left. Therefore, there should be no unobservable differences in the characteristics of workers and non-workers who live with pensioners. Table IV shows that, indeed, migrant workers earn more tan non-migrants on average, suggesting this variable is suitable as a restriction on the upper bound. The table also shows that workers in households without pensioners earn more and have more potential experience, suggesting that the variable measuring whether or not a person lives with a pensioner is suitable for restricting the lower bound. Section 4 explains in further detail how we use these two variables.

TABLE IV

—Data: Migrant Workers and Pensioners

Migrant Workers						
	1995			2000		
Migrant workers	12.44%			29.89%		
Migrant (characteristics of workers with $z_i = 8$)	Yes	No	Stat. diff.	Yes	No	Stat. diff.
Log Average wage	9.69	9.61	No	9.83	9.34	Yes
Average experience (E=1)	24.22	25.39	Yes	24.57	25.62	No
Average experience all	24.15	22.08	Yes	22.62	19.70	Yes

Pensioners						
	1995			2000		
Households with pensioners	8.77%			12.79%		
Pensioner (characteristics of workers with $z_i = 0$)	Yes	No	Stat. diff.	Yes	No	Stat. diff.
Log Average wage	8.94	8.99	No	8.29	9.05	Yes
Average experience (E=1)	35.52	37.50	Yes	40.61	39.51	No
Average experience all	33.72	38.37	Yes	38.11	39.01	Yes

Note.—African males aged between 18 and 65.

4. DERIVING MEANINGFUL AND FEASIBLE BOUNDS ON ATE

This section states the estimation objective, sets up the model and estimation framework and summarizes Manski and Pepper's (2000) results.

4.1. *Estimation Objective*

We need to introduce some non–conventional notation first. A person $i \in I$ has a realized treatment $z_i \in T$ and a realized outcome $y_i := y_i(z_i) \in Y$, both of which are observable. The function $y_i(\cdot) : T \to Y$ is called *response function*. The latent or conjectural outcomes $y_i(t)$ with $t \neq z_i$ are not observed. Specifically, the treatment z_i is years of education and the response $y_i(\cdot)$ is log–earnings.

Our goal is to derive an upper bound on the average treatment effect of one year of education on earnings defined as

$$\Gamma(s,t) := \frac{\Delta(s,t)}{t-s},$$
$$\text{where:} \quad \Delta(s,t) := E[y(t)] - E[y(s)],$$

for $s < t$. In words, we compute the difference between the expected value of log–earnings after t years of education and the expected value of log–earnings after s years of education and average it out on a per–year basis. This will tell us by how much, on average, one year of education affects log–earnings.

Why is this difficult? A naive idea for computing, for example, $E[y(t)]$ would be to simply average out the log–earnings for those individuals in the sample who have t years of education. The problem, of course, is that we are dealing with a select sample; people have already sorted themselves into their preferred level of schooling and thus causing selection bias.

To circumvent the selection problem we will use a nonparametric framework suggested by Manski and Pepper (2000). Compared to parametric identification we will impose relatively mild assumptions to obtain the ATE. This, of course, comes with a disadvantage: The object of interest, $\Gamma(s,t)$, is not point identified. Instead, we will only be able to bound it below a certain threshold. Nevertheless, a bounded estimate can be informative, which we will see below.

4.2. *Correcting for Omitted Variables (Ability) Bias*

The selectivity of the sample is caused by ability bias. We ignore bias due to sample selection into work for now. For developed countries, conditioning on white males, ability bias is the main threat to estimating the ATE of education on earnings. The typical justification for omitted variables bias is that people with higher ability tend to select themselves into higher levels of education, and vice versa. This seems reasonable ex–ante in economies in which there are almost no frictions: no discrimination, easy access to educational institutions, no borrowing constraints (or free education), no constraints on individuals due to household bargaining (free choice). Developed countries are most likely to offer such conditions.

In that light, the framework of Manski and Pepper (2000) for identifying the ATE of education on earnings is promising. Manski and Pepper solve the problem of ability bias by imposing the following two mild monotonicity assumption:

Assumption 1 (Monotone Treatment Response)
Let T be an ordered set. For each $i \in I$,

$$t_2 \geq t_1 \Rightarrow y_i(t_2) \geq y_i(t_1).$$

Assumption 2 (Monotone Treatment Selection)
Let T be an ordered set. For each $t \in T$,

$$u_2 \geq u_1 \Rightarrow E\left[y(t)|z = u_2\right] \geq E\left[y(t)|z = u_1\right].$$

What do these assumptions mean and how do they differ? As Manski and Pepper (2000) write, both assumptions are distinct versions of the statement "wages increase with schooling." The MTR Assumption concerns the functional form of the income equation, it does not address the stochastic selection process that makes people choose different levels of education. All it says is that more education will weakly increase a person's income, *holding ability constant*. The MTR Assumption deals with the *direct* or *pure* effect that education has on earnings. The assumption does not deal with the indirect effect that education could have through its correlation with ability (or any other covariates). This is a statement regarding the (human capital production) functional form.

The MTS Assumption, in contrast, is concerned with the stochastic selection process that runs in the background of the model. Schooling is an *endogenous treatment*, different people tend to select themselves into different education levels for a multitude of (often unobserved) reasons. Consider the following social experiment: You can force people to attend school for exactly t years. How would this

affect the subset of people who in the absence of your intervention would have chosen to attend school for u_2 years? How would this affect the subset of people who in the absence of your intervention would have chosen to attend school for $u_1 < u_2$ years? The MTS Assumption claims that you would expect a higher income for the former group (i.e., the subset of people who chose u_2 years of education).

The MTR Assumption is consistent with the human capital accumulation model. The MTS Assumption is weaker than a standard instrumental variable assumption and is consistent with the idea of people with high ability selecting themselves into higher paying jobs. The joint validity of the MTR and MTS Assumptions can be tested statistically.

Manski and Pepper (2000) establish the following two important results (a comprehensive derivation is in Appendix B):

Proposition 1 (Sharp MTS–MTR Bounds on Unconditional Expectation)
The bounds on the expected value of earnings conditional on being employed but unconditional on education, $E[y(t)|E=1]$, under Assumptions 1 and 2 are

$$\sum_{u<t} E[y|z=u,E=1]\cdot Pr(z=u|E=1) + E[y|z=t,E=1]\cdot Pr(z\geq t|E=1)$$
$$\leq E[y(t)|E=1] \leq$$
$$\sum_{u>t} E[y|z=u,E=1]\cdot Pr(z=u|E=1) + E[y|z=t,E=1]\cdot Pr(z\leq t|E=1),$$

where E is an employment dummy variable that takes on the value one if a person worked and zero otherwise.

Last, to put an upper bound on the average treatment effect $\Delta(s,t|E=1) := E[y(t)|E=1] - E[y(s)|E=1]$, for $s<t$, under both the MTR and MTS Assumptions, we subtract the lower bound on $E[y(s)|E=1]$ from the upper bound on $E[y(t)|E=1]$:

Corollary 2 (Sharp Upper Bound on ATE)
The upper bound on the average treatment effect of education on earnings, $\Delta(s,t|E=1)$, under Assumptions 1 and 2 is

$$\begin{aligned}\Delta(s,t|E=1) \leq \overline{\Delta}(s,t|E=1) := & \\ & \sum_{u<s}\big(E[y|z=t,E=1] - E[y|z=u,E=1]\big)\cdot Pr(z=u|E=1)\ldots \\ & + \big(E[y|z=t,E=1] - E[y|z=s,E=1]\big)\cdot Pr(s\leq z\leq t|E=1)]\ldots \\ & + \sum_{u>t}\big(E[y|z=u,E=1] - E[y|z=s,E=1]\big)\cdot Pr(z=u|E=1).\end{aligned}$$

Last, defining the ATE per year as $\Gamma(s,t|E=1) = \frac{\Delta(s,t|E=1)}{t-s}$ we obtain the upper bound on the ATE per year as

$$\Gamma(s,t|E=1) \leq \frac{\overline{\Delta}(s,t|E=1)}{t-s}. \tag{4.1}$$

4.3. *Correcting for Selection into Employment*

The Manski and Pepper (2000) bounds from Proposition 1 and Corollary 2 take care of the omitted variables bias due to unobserved ability. As discussed earlier, ability bias presents the biggest challenge in the estimation of the returns to education for developed countries and conditioning on white males. Manski and Pepper apply their estimator to the US National Longitudinal Survey of Youth data and find informative upper bounds on the ATE that lie below parametric returns to education estimates by Card (1993) and Card (1994).

We are interested in bounding the ATE of education on earnings for African males in South Africa for the years 1995 and 2000. In addition to omitted variables bias we will also have to deal with sample selection into employment. As discussed in the data section, labor force participation rates for that group are low compared to data for white males in the US (see Table III). For the subpopulation of white males in the US data it is reasonable to ignore the sample selection problem because almost any white male above the age of 26 will work. The MTR and MTS Assumptions 1 and 2 will likely not break down if a small degree of sample selection is present. However, in the case of South Africa, this will no longer be true.

We therefore adjust the bounds to deal with the sample selection problem. To fix ideas, every expectation and probability in Proposition 1 is taken conditional on being employed, i.e., the employment indicator variable E is equal one. For the US data it is implicitly assumed that $\Pr(E = 1) \approx 1$ in which case suppressing the conditioning on $E = 1$ is mostly a matter of notational convenience.

In the case of South Africa the employment probability is closer to 50% and we therefore derive a bound estimator that accounts for such a degree of sample selection. Instead of bounding $E[y(t)|E = 1]$ we would like to bound $E[y(t)]$. To outline our extension of Manski and Pepper's framework, break up the unconditional expectation into two parts:

$$E[y(t)] = \Pr(E = 0) \cdot E[y(t)|E = 0] + \Pr(E = 1) \cdot E[y(t)|E = 1].$$

Both probabilities in the last equation are estimable from the data. For $E[y(t)|E = 1]$ we can, in principal, apply Manski and Pepper's method. The main complication stems from the term $E[y(t)|E = 0]$ because the earnings of non–employed people are not observed. Even imposing monotonicity–type assumptions will not solve that problem. To deal with those counterfactual earnings we will impose mild relations between certain sub–populations for which $E = 0$ and $E = 1$.

To derive an upper bound on $E[y(t)]$ we introduce a monotone instrumental variable w_1 which is a dummy variable that takes on value one if a person migrated away from home in order to work and send remittances back to his household and zero otherwise. We impose two assumptions on w_1:

Assumption 3 (Complementarity)

$\Pr(E = 0, w_1 = 1) = 0$.

Assumption 4 (Monotone Instrumental Variable 1)

$E[y|v_1 = u_1, w_1 = 0, z \geq t, E = 0] \leq E[y|v_1 = u_1, w_1 = 1, z \geq t, E = 1]$.

Assumption 3 states that a person cannot send remittances back home if he is not working. Keeping true to that assumption is merely a matter of defining the variable w_1 appropriately. The MIV Assumption 4 plays a more important role in establishing the upper bound on $E[y(t)]$. It creates a link between non–workers and workers. We require that the expected earnings of non–employed people who do not send remittances be weakly less than the expected earnings of those persons who work and send remittances back home. The idea behind this assumption is that households internally negotiate and decide do send their relatively productive members away from home in order to find employment and send remittance money back home. An example would be a household member who moves away from home to work in a diamond mine.

To derive a lower bound on $E[y(s)]$, with $s < t$, we introduce a monotone instrumental variable w_2 which is a dummy variable that takes on value one if a person lives in a household with a pensioner who receives state pension benefits and zero otherwise. We impose two assumptions on w_2:

Assumption 5 (Monotone Instrumental Variable 2)

$E[y|v_1 = u_1, w_2 = 0, z \leq s, E = 0] \geq E[y|v_1 = u_1, w_2 = 1, z \leq s, E = 0]$.

Assumption 6 (Conditional Mean Independence)

$E[y|v_1 = u_1, w_2 = 1, z \leq s, E = 0] = E[y|v_1 = u_1, w_2 = 1, z \leq s, E = 1]$.

The MIV Assumption 5 says that the expected earnings of people who live in households without pensioners weakly exceed those of people who live together with pensioners. This assumption is imposed only on the non–employed sub–population, in that sense we are referring to counterfactual earnings. The justification for this assumption comes from research which finds that households with pensioners are more likely to send individual members away from home to find work and send back remittances. If those migrating individuals are on average more productive than their non–migrating family members then Assumption 5 holds. The conditional mean independence Assumption 6 links non–workers and workers. For people who live at home together with a pensioner household member we assume that the expected earnings for non–employed and employed people are the same. This assumption could be restrictive for general levels of s. However in the computation of the lower bound, to find the ATE, we will restrict ourselves to values of $s = 0$. In that case, Assumption 6 says that for the zero education group conditional on having a pensioner in the household that the expected earnings of workers and non-workers are the same. This implies that selection into work is not driven by unobserved ability for that sub–population.

We get the following result.

Proposition 3 (Bounds Adjusting for Sample Selection)

The bounds on the unconditional expectation of earnings, $E[y(t)]$, under Assumptions 1, 2, 3, 4, 5,

and 6 are:

$$\begin{aligned}
&\sum_{u<t} \Pr(z=u)\cdot\Big(E[y|z=u,w_2=0,E=1]\cdot Pr(w_2=0,E=1)\ldots\\
&\quad + E[y|z=u,w_2=1,E=1]\cdot\big(Pr(w_2=1,E=1)+Pr(E=0)\big)\Big)\ldots\\
&\quad + \Pr(z\geq t)\cdot\Big(E[y|z=t,w_2=0,E=1]\cdot Pr(w_2=0,E=1)\ldots\\
&\quad + E[y|z=t,w_2=1,E=1]\cdot\big(Pr(w_2=1,E=1)+Pr(E=0)\big)\Big)\\
&\qquad\qquad \leq E[y(t)] \leq\\
&\sum_{u>t} \Pr(z=u)\cdot\Big(E[y|z=u,w_1=0,E=1]\cdot Pr(w_1=0,E=1)\ldots\\
&\quad + E[y|z=u,w_1=1,E=1]\cdot\big(Pr(w_1=0,E=0)+Pr(w_1=1,E=1)\big)\Big)\ldots\\
&\quad + \Pr(z\leq t)\cdot\Big(E[y|z=t,w_1=0,E=1]\cdot Pr(w_1=0,E=1)\ldots\\
&\quad + E[y|z=t,w_1=1,E=1]\cdot\big(Pr(w_1=0,E=0)+Pr(w_1=1,E=1)\big)\Big).
\end{aligned}$$

Proof. See Appendix C. ■

Last, to put an upper bound on the average treatment effect $\Delta(s,t) := E[y(t)] - E[y(s)]$, for $s < t$, under Assumptions 1 through 6, we merely add subtract the lower bound on $E[y(s)]$ from the upper bound on $E[y(t)]$:

Corollary 4 (Sharp Upper Bound on ATE)

The sample selection adjusted upper bound on the average treatment effect of education on earnings, $\Delta(s,t)$, under Assumptions 1, 2, 3, 4, 5, and 6 are

$$\begin{aligned}
\Delta(s,t) \leq \overline{\Delta}(s,t) :=& \sum_{u<s} \Pr(z=u)\cdot\Big(E[y|z=u,w_2=0,E=1]\cdot Pr(w_2=0,E=1)\ldots\\
&\quad + E[y|z=u,w_2=1,E=1]\cdot\big(Pr(w_2=1,E=1)+Pr(E=0)\big)\Big)\ldots\\
&\quad + \Pr(z\geq s)\cdot\Big(E[y|z=s,w_2=0,E=1]\cdot Pr(w_2=0,E=1)\ldots\\
&\quad + E[y|z=s,w_2=1,E=1]\cdot\big(Pr(w_2=1,E=1)+Pr(E=0)\big)\Big)\ldots\\
&- \sum_{u>t} \Pr(z=u)\cdot\Big(E[y|z=u,w_1=0,E=1]\cdot Pr(w_1=0,E=1)\ldots\\
&\quad + E[y|z=u,w_1=1,E=1]\cdot\big(Pr(w_1=0,E=0)+Pr(w_1=1,E=1)\big)\Big)\ldots\\
&\quad - \Pr(z\leq t)\cdot\Big(E[y|z=t,w_1=0,E=1]\cdot Pr(w_1=0,E=1)\ldots\\
&\quad - E[y|z=t,w_1=1,E=1]\cdot\big(Pr(w_1=0,E=0)+Pr(w_1=1,E=1)\big)\Big).
\end{aligned}$$

Next, defining the ATE per year as $\Gamma(s,t) = \frac{\Delta(s,t)}{t-s}$ *we obtain the sample selection adjusted upper bound on the ATE per year as*

$$\Gamma(s,t) \leq \frac{\overline{\Delta}(s,t)}{t-s}. \tag{4.2}$$

Proof. See Appendix C. ■

5. RESULTS

5.1. *Main Findings*

Table V presents, for the years 1995 and 2000, the empirical mean of log-income given t years of education, $\widehat{E}[y|z=t]$, the distribution of years of education in the data, $\widehat{\Pr}(z=t)$, and the number of observations in each education category. We restrict our sample to employed African males between the ages of 18 and 65. The columns for $\widehat{E}[y|z=t]$ and $\widehat{\Pr}(z=t)$ are the main ingredients for the estimation of the MTR–MTS bounds. What is striking in the South African case is the high number of uneducated ($t=0$) people: 1,489 out of 11,702 persons in 1995 and 947 out of 9,143 persons in 2000 do not have any formal education. The low education level in South Africa is in stark contrast to developed countries. Manski and Pepper (2000), using the United States National Longitudinal Survey of Youth data set, have a minimum level of schooling of 8 years and a maximum level of schooling of 20 years. In the South African data, for African males, the maximum level of schooling is 16 years (for a very small fraction of the sample).

We consider it advantageous, for bound estimation, that the minimum schooling level is zero. We want to estimate the bounds in equations (4.1) and (4.2) holding s fixed at zero and letting $t \geq 8$. Like that, we compare, for example, $E[y(8)]$ to $E[y(0)]$, i.e., we calculate the average treatment effect of eight years of education. Equations (4.1) and (4.2) break the ATE down to a per–year basis. For each ordered pair $(s=0,t)$ with $t \geq 8$ we compute two alternative bounds on the ATE. First, we get Manski and Pepper's (2000) upper bound, equation (4.1), that corrects for unobserved ability bias. Second, we compute the adjusted bound, equation (4.2), that corrects for unobserved ability bias and sample selection into employment.

Table VI contains the results for African males aged 18–65. The first row in column (1) reports the result for $\Gamma(0,8|E=1)$ of equation (4.1), i.e., the bounds without sample selection correction. The upper bound estimate on the return to education is equal to 11.74%. This means that people with eight years of education have received an additional return of *at most* 11.74% per year on each of their eight years of education. In column (2) we report the 95% quantile of the empirical distribution of the bound estimate. For $\Gamma(0,8|E=1)$ the 95% quantile equals 13.57% which provides a more conservative estimate of the upper bound.

Continuing downwards in column (1), the bound estimate is more or less stable up to $t=12$ which represents the final year of high school. The final high–school year is in fact the most frequent level of education and we estimate an upper bound on the ATE of a high–school education on earnings of 11.85% per year. For $t=14,15$ the bound estimates increase markedly which we attribute to a

TABLE V

—Empirical Mean of Log-Income and Distribution of Years of Schooling

	1995			2000		
s	$\widehat{E}[y\|z=t]$	$\widehat{\Pr}(z=t)$	N	$\widehat{E}[y\|z=t]$	$\widehat{\Pr}(z=t)$	N
0	9.92	0.127	1489	8.93	0.104	947
1				8.76	0.008	77
2	8.90	0.057	662	8.88	0.019	169
3				9.12	0.040	362
4	9.04	0.058	682	9.13	0.047	433
5	9.01	0.056	650	9.16	0.052	472
6	9.18	0.070	823	9.24	0.067	613
7	9.39	0.087	1021	9.30	0.090	825
8	9.61	0.110	1288	9.41	0.094	863
9	9.69	0.070	813	9.43	0.072	659
10	9.94	0.094	1102	9.72	0.098	896
11	10.04	0.049	571	9.56	0.063	580
12	10.30	0.145	1702	9.87	0.172	1571
14	10.69	0.056	653	10.64	0.050	453
15	11.13	0.021	246	11.08	0.017	152
16				11.16	0.008	71
Total		1	11,702		1	9,143

Note.—In 1995 years of schooling $t = 1, 2, 3$ are comprised in one category. Everything conditional on employed African males, age 18–65.

combination of additional returns due to post–high school education and low sample size.

In all cases, the MTR–MTS bound lies below the OLS estimator of 15.59%. To obtain this estimate we ran a typical Mincerian wage regression of log-income on a constant, education, potential experience (linear and quadratic), and province (there are nine provinces in the data). The OLS estimator is, of course, biased. Card (2001) comprehensively discusses different sources of bias in the OLS estimator of the return to schooling. Summarizing research findings in the area, Card argues that, if anything, the OLS estimator is downwards biased (for example, comparing to instrumental variables estimation of discrete dynamic choice programming model estimation). We therefore regard the OLS estimator as a lower bound for parametric estimation. In this sense then the bound estimates of column (1) in Table VI for the year 1995 are informative because they fall below estimates that result from parametric estimation. Even if we base our conclusions on only the conservative 95% quantile, the bound estimates for $t = 8, 9, 10, 11, 12, 14$ all fall below the parametric estimates.

The results for 1995 are qualitatively similar if instead we use our adjusted upper bounds that account for sample selection. The first row in column (3) reports the result for $\Gamma(0, 8)$ of equation (4.2), i.e., the bounds with sample selection correction. The upper bound on the return to education is equal to 12.89% which is higher than the non–adjusted bound in column (1). In column (4) the corresponding conservative 95% quantile of the empirical distribution equals 15.43%.

TABLE VI

—Estimation Results: Upper Bounds on Return to Schooling, Age 18–65

		1995				2000			
		UB w/o correction	95% Quantile	UB with correction	95% Quantile	UB w/o correction	95% Quantile	UB with correction	95% Quantile
s	t	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	8	11.74	13.57	12.89	15.43	8.74	10.65	12.90	14.52
0	9	10.94	12.91	12.15	14.74	7.91	9.95	11.16	12.99
0	10	11.47	13.27	12.10	14.61	8.83	11.13	11.60	13.61
0	11	11.05	12.61	11.80	14.54	7.02	9.79	9.47	12.13
0	12	11.85	13.75	12.64	14.91	8.39	10.78	10.68	13.13
0	14	12.69	14.40	12.66	15.19	12.26	14.72	14.19	16.65
0	15	14.74	16.74	15.90	17.36	14.34	17.20	16.70	19.60
0	16					13.91	15.80	15.03	11.50
	N			11,702				9,143	
	OLS			15.59	(0.15)			15.31	(0.20)

Note.—Numbers in percent. Columns (1) and (5) contain upper bound estimates on ATE without correcting for sample selection, based on equation (4.1). Columns (3) and (7) contain upper bound estimates adjusting for sample selection, based on equation (4.2). Corresponding quantiles via bootstrap with 2,000 repetitions. Bottom row presents OLS regression of log-income on constant, education, potential experience (linear and quadratic), province. Standard errors in parentheses.

Continuing downwards in column (3), the bound estimate again is relatively stable up to $t = 14$. The ATE of high–school education over zero education is at most 12.64% per year of education with a 95% quantile of 14.91%. Comparing columns (1) and (3), the adjusted upper bounds that correct for sample selection (column (3)) are (by construction) at least as large as the Manski and Pepper (2000) bounds (column (1)) that only correct for ability bias. The corrected bounds are however still informative vis-á-vis the OLS estimates. For all values of t between 8 and 14 do the bounds fall to the left of the OLS point estimates.

For the year 2000 a qualitatively similar picture emerges and the bound results are generally stronger. For example, the estimate for the ATE of high–school education on earnings is at most 8.39% (without sample selection correction) and 10.68% (with correction). Both numbers are considerably lower than the parametric estimate of 15.31%. And again it is true that all nonparametric results lie to the left of the OLS estimate. For education levels beyond high–school ($t = 14, 15, 16$) the ATE jump up to 14.19%, 16.70%, and 15.03% which indicates that the added value of attending higher education for African males is increasing. However, the increase in ATE for those levels of education could also be the result of small sample size. We also caution that by the nature of the upper bounds it is not logically correct to claim that, for example, the ATE of education on earnings is higher for $t = 14$ (upper bound of 14.19%) than for $t = 12$ (upper bound of 10.68%).

By the same reasoning we cannot claim that on average the ATE of education on earnings has decreased between 1995 and 2000. The ATE of a high–school degree on earnings was bounded above by 12.64% in

TABLE VII

—Estimation Results: Upper Bounds on Return to Schooling, Age 18–39

		1995				2000			
		UB w/o correction	95% Quantile	UB with correction	95% Quantile	UB w/o correction	95% Quantile	UB with correction	95% Quantile
s	t	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	8	12.84	14.74	13.86	16.07	10.12	13.44	15.37	17.36
0	9	12.33	14.37	13.01	16.02	9.58	12.81	13.83	16.20
0	10	12.54	14.67	12.64	15.23	9.86	13.13	13.48	15.55
0	11	12.25	14.07	12.58	13.20	8.33	11.74	11.72	14.52
0	12	12.84	14.99	13.06	15.08	9.56	12.87	12.62	15.53
0	14	13.52	15.73	12.84	16.17	13.36	16.43	15.80	18.08
0	15	14.98	17.51	15.95	17.54	14.77	19.35	17.50	20.78
0	16					14.72	17.64	15.75	12.77
	N			6,958				5,507	
	OLS			18.48	(0.20)			18.61	(0.27)

Note.—Numbers in percent. Columns (1) and (5) contain upper bound estimates on ATE without correcting for sample selection, based on equation (4.1). Columns (3) and (7) contain upper bound estimates adjusting for sample selection, based on equation (4.2). Corresponding quantiles via bootstrap with 2,000 repetitions. Bottom row presents OLS regression of log-income on constant, education, potential experience (linear and quadratic), province. Standard errors in parentheses.

1995 and by 10.68% in 2000 (bounds adjusting for sample selection). Because we cannot point identify the ATE it is conceivable that the actual treatment effect of education was lower in 1995 than in 2000. It is curious that even the OLS estimator—which as a point identified estimation method should be able to establish whether the ATE increased or decreased over time—is not able to distinguish significantly the 1995 from the 2000 estimates.

5.2. *Sensitivity to Cohort Effects*

We could be concerned that the MTS Assumption 2 breaks if older cohorts (age groups) had more restricted access to education than younger ones. Recall that the justification for the MTS assumption was that people select themselves into education according to their abilities. Higher ability people tend to have more education. The cohort effect could theoretically reverse this causation between education and ability. For example, if high ability people of older cohorts could on average only go to school for six years they will be observed in the data with little education and high income (because their high ability is rewarded in the market and also because they have high levels of work experience). On the other hand, if low ability people of younger cohorts go to school for an average of eight years (due to improved access to education) they will be observed in the data with relatively high education levels but low incomes (because of lower ability and less experience). This example illustrates how the the MTS assumption might be violated due to cohort effects.

We therefore drop people older than 39 years from the sample and repeat the calculations from Table

VI. The new results are collected in Table VII. Qualitatively, everything that was true for the 18–65 age group still holds true for the 18–39 age group. The bounds for the younger cohort are generally higher. For example, for the year 1995, the ATE of high–school education is at most 12.84% (without correction for sample selection) and 13.06% (with correction) while the OLS point estimate is 18.48%. For the year 2000 the upper bound on the high–school ATE is 9.56% (without correction) and 12.62% (with correction). Again, the conservative 95% quantiles for education levels less than 15 fall below the OLS estimates.

We take the results from this subsections as evidence that cohort effects do not violate Assumption 2. We have repeated this exercise for even younger cohorts with similar findings.

6. CONCLUSION

Our research objective is to estimate the average treatment effect on earnings of one year of education for African males in South Africa. We construct a cross–sectional data set combining the OHS95 and IES95 in 1995 and LFS00 and IES00 in 2000. We cannot use OLS to point identify the ATE because of the well–known problem of unobserved ability bias. The parametric alternative of instrumental variables estimation is not feasible because there are no strong instrumental variables in the data.

Instead we use Manski and Pepper's (2000) nonparametric bound estimator which allows us to estimate an upper bound on the ATE. There is one complication with Manski and Pepper's estimator: It is only useful for populations with high labor force participation rates. Labor force participation of Africans in South Africa, however, is very low. We therefore modify Manski and Pepper's estimator to account for such degrees of sample selection.

Using the modified nonparametric estimator, we find an upper bound on the return to a high–school degree of 12.64% (per year of high–school) in 1995 and 10.68% in 2000. Compared to standard parametric estimates, our bound is informative. The OLS estimates of the return to education for 1995 and 2000 are 15.59% and 15.31%. Our bound estimates for almost all levels of education fall below the OLS estimates.

APPENDIX A: PARAMETRIC IDENTIFICATION OF ATE

A.1. *The Human Capital Earnings Function: Schooling and Ability*

Our goal is to estimate the causal effect *ceteris paribus* of education on earnings. To do that, assume that there exists a human capital production function for individual i

$$y_i(\cdot,\cdot) : T \times A \to Y \tag{A.1}$$

that maps years of schooling $s \in T$ and ability $a \in A$ into individual earnings outcomes $y_i(s,a) \in Y$. This function is a smaller version of Mincer's (1974) human capital earnings function. Manski (1997) calls equation (A.1) a *response* function. The function simply illustrates that schooling s and ability a have a *direct* or *pure* effect on earnings.
For practical purposes the presence of ability in the response function creates at least three problems:

(i) Ability is not well defined.

(ii) Ability is not measured with sufficient accuracy.

(iii) Ability is not part of most data sets, and hence unobserved.

Standard cross–sectional data sets typically collect years of education, s, along with income y. Estimation has to rely on these variables only. Going back to equation (A.1) we define

$$y_i(\cdot) := y_i(\cdot,\bar{a}) : T \to Y,$$

which is a univariate function (holding ability constant) mapping schooling into earnings outcomes. Our goal is to measure the pure effect of schooling on earnings *ceteris paribus*:

$$\begin{aligned}\Delta(s,t) :=& E\left[y(t,\bar{a})\right] - E\left[y(s,\bar{a})\right] \\ =& E\left[y(t)\right] - E\left[y(s)\right],\end{aligned} \tag{A.2}$$

where $s \in T$ and $t \in T$ are years of education and $s < t$. The object $\Delta(s,t)$ is the average treatment effect of schooling on earnings, the estimation of which is our research objective.

A.2. *Digression: Naive Nonparametric Estimation*

In the sample data we observe, for a every person i, her level of schooling, $z_i \in T$, and her income $y_i := y_i(z_i) \in Y$. Note that we distinguish between the *latent* levels of schooling, denoted s and t, and the *observed* level of schooling for person i, denoted z_i. The latent outcomes $y_i(s)$ or $y_i(t)$ are not observed whenever $s \neq z_i$ or $t \neq z_i$. To estimate the effect of schooling on earnings we use a random sample of people for which we observe data pairs on schooling and income ($\{z_i, y_i\} \in T \times Y$). A naive nonparametric solution for estimating $\Delta(s,t)$ is to simply average those income observations y_i for which $z_i = s$, and compare them to the average of the income observations for which $z_i = t$. The problem with this approach is that ability and schooling are correlated. The observed data are realizations of peoples' optimization decisions in which ability can be seen as a state variable and schooling as a choice variable.[6] High ability people are more likely to choose higher levels of schooling (and vice versa). Schooling is hence endogenous, the resulting estimator for $\Delta(s,t)$ is biased upward.

A.3. *Problems of Parametric Estimation*

There are at least two problems with parametric estimation: functional form and selection. Writing log–earnings as a linear function of schooling looks simplistic but it is convention in labor economics. Card (2001) argues for an additional quadratic schooling term so that the marginal effect of schooling is declining in schooling (assuming that the coefficient of the quadratic schooling term is negative). But even this assumptions seems arbitrary. Because the relationship between earnings and schooling is not governed by a deterministic law, there will always remain different opinions about functional form.

[6] Keane and Wolpin (1997) develop a dynamic choice programming model along those lines.

Regarding selection, the OLS model simply disregards the problem of correlation between schooling and ability. Consider two persons with different levels of ability. If in a social experiment we could force both of them to obtain the same amount of schooling t then disregarding any selection we would expect both of them to have the same income:

$$E[y_i(t)|a_i = a_1] = E[y_i(t)|a_i = a_2] \qquad \text{for } a_1 \neq a_2. \tag{A.3}$$

It is more realistic, however, to think that the person with higher ability would receive a higher income, in which case the equality in equation (A.3) turns into an inequality. The availability of an instrumental variable changes this interpretation a bit. Equation (A.3) is replaced by

$$E[y_i(t)|q_i = q_1] = E[y_i(t)|q_i = q_2],$$

for $q_1 \neq q_2$. This claim holds by the definition of an instrumental variable.

A.4. *Ordinary Least Squares Estimation*

The first assumption of any parametric analysis always is linearity. Countless papers in labor economics run versions of the following regression:

$$y_i = z_i\beta + a_i + \varepsilon_i, \tag{A.4}$$

where y_i is the logarithm of earnings, a_i is unobserved ability, ε_i is a random error, and β is some coefficient. Equation (A.4) is a simplification of Mincer's (1974) human capital earnings function, with two essential features: the linear link between schooling and log–earnings and the effect of unobserved ability on earnings. Using a random sample $\{z_i, y_i\}$, the classical linear regression model simply estimates β as the slope coefficient using the assumption

$$E[a_i + \varepsilon_i|z_i] = 0, \tag{A.5}$$

Under equations (A.4) and (A.5) the average treatment effect $\Delta(s, s+1)$ equals

$$\begin{aligned}
\Delta(s, s+1) &= E\left[y_i(s+1)\right] - E\left[y_i(s)\right] \\
&= E\left[(z_i+1)\beta + a_i + \varepsilon_i|(z_i+1)\right] - E\left[z_i\beta + a_i + \varepsilon_i|z_i\right] \\
&= E\left[(z_i+1)\beta|z_i\right] - E\left[z_i\beta|z_i\right] \\
&= (z_i+1)\beta - z_i\beta \\
&= \beta.
\end{aligned}$$

To estimate the average treatment effect we therefore only need to run an OLS regression and obtain the slope coefficient.

A.5. *Instrumental Variables Estimation*

Parametric estimation seems like a convenient way to estimate $\Delta(s, s+1)$. And given the assumptions so far, it is also the best linear unbiased estimator. An obvious drawback is assumption (A.5). A better set of assumptions would be

$$\begin{aligned}
E[a_i|z_i] &\neq 0 \\
E[\varepsilon_i|z_i] &= 0.
\end{aligned}$$

Ability is unobserved in the data. Running an ordinary least squares regression of log–earnings on education yields an inconsistent estimate for β. The average treatment effect is not identified. The way around this problem is instrumental variables estimation. In order to identify β we need an instrumental variable, $v \in V$, which satisfies:

(i) Constant treatment response:[7] $y(s, v_1) = y(s, v_2) = y(s)$ for all $v_1 \neq v_2$

(ii) Correlation: $E[s_i|v_i] = \pi v_i$ with $\pi \neq 0$

(iii) Exogeneity: $E[a_i|v_i] = 0$.

[7]The term 'constant treatment response' was first defined by Manski and Pepper (2008).

The average treatment effect equals

$$\begin{aligned}\Delta(s, s+1) &= E\left[y_i(s+1)\right] - E\left[y_i(s)\right] \\ &= E\left[(s_i+1)\beta + a_i + \varepsilon_i | s_i\right] - E\left[s_i\beta + a_i + \varepsilon_i | s_i\right] \\ &= E\left[(s_i+1)\beta + a_i + \varepsilon_i | v_i\right] - E\left[s_i\beta + a_i + \varepsilon_i | v_i\right] \\ &= \beta.\end{aligned}$$

APPENDIX B: DERIVING MANSKI AND PEPPER'S (2000) BOUNDS

B.1. *No Assumptions Bound*

Although the subsection title promises a bound without any assumptions, we will need to assume that the range of the dependent variable is bounded. The range of Y is $[K_0, K_1]$. The misnomer "no assumptions bound" is due to Manski (1989) and we do not intend to deviate from his convention. In any case, the assumption of bounded support will become redundant later; it only serves as an auxiliary assumption to derive the main results.
We start with the following decomposition of the conditional expectation[8]:

$$E[y(t)|v=u] = E[y(t)|v=u, z=t] \cdot \Pr(z=t|v=u) + E[y(t)|v=u, z\neq t] \cdot \Pr(z\neq t|v=u).$$

Whenever the conjectural treatment t equals the actual treatment z we rewrite $E[y(t)|v=u, z=t] = E[y|v=u, z=t]$ and thus

$$E[y(t)|v=u] = E[y|v=u, z=t] \cdot \Pr(z=t|v=u) + E[y(t)|v=u, z\neq t] \cdot \Pr(z\neq t|v=u).$$

This bit of rewriting helps see that the only term on the right hand side of the last equation that is not identifiable by the data is $E[y(t)|v=u, z\neq t]$. What is the interpretation of $E[y(t)|v=u, z\neq t]$? It is the expected value of income after exactly t years of education for the subset of people who would choose to attend school for more or less than t years of schooling. It is the expected value of a counterfactual event. We do not observe this latent value in the data. The only thing we know about $E[y(t)|v=u, z\neq t]$ is that it lies between K_0 and K_1. Therefore we use those extremes to bound the expectation:

$$\begin{aligned}E[y(t)|v=u] &\geq E[y|v=u, z=t] \cdot \Pr(z=t|v=u) + K_0 \cdot \Pr(z\neq t|v=u) \\ E[y(t)|v=u] &\leq E[y|v=u, z=t] \cdot \Pr(z=t|v=u) + K_1 \cdot \Pr(z\neq t|v=u).\end{aligned} \tag{B.1}$$

B.2. *Monotone Instrumental Variable*

The bounds in inequalities (B.1) are sharp in the absence of more information. But for practical purposes the no assumptions bounds have at least two drawbacks: They are likely not informative (if the range of Y is wide) and researchers usually do not know the values of K_0 and K_1. To ameliorate the first problem, Manski and Pepper (2000) introduce their monotone instrumental variable assumption.

Assumption 7 (Monotone Instrumental Variable)
Let V be an ordered set. The variable v is a monotone instrumental variable in the sense of mean–monotonicity if, for each $t \in T$, and all $(u_1, u_2) \in (V \times V)$ such that $u_2 \geq u_1$

$$E[y(t)|v=u_2] \geq E[y(t)|v=u_1].$$

How does this assumption help us? So far, the variable v did not play any active role in the derivation of the bounds. Now, if we observe different individuals with different values for v we can refine the bounds. By Assumption 7 we get, for all $u_1 \leq u \leq u_2$

$$\begin{aligned}E[y(t)|v=u] &\geq E[y(t)|v=u_1] \\ E[y(t)|v=u] &\leq E[y(t)|v=u_2].\end{aligned} \tag{B.2}$$

[8] All expressions within Appendix B should be understood as conditional on being employed, i.e., the employment indicator variable E is equal to 1.

For each of the expectations on the right hand side of the two inequalities (B.2), equation (B.1) provides bounds. The inequality $E[y(t)|v = u] \geq E[y(t)|v = u_1]$ holds for all $u_1 \leq u$. To tighten the bound on $E[y(t)|v = u]$ we can pick the maximum value of the lower bound on $E[y(t)|v = u_1]$, hence

$$E[y(t)|v = u] \geq \sup_{u_1 \leq u} \{E[y|v = u_1, z = t] \cdot \Pr(z = t|v = u_1) + K_0 \cdot \Pr(z \neq t|v = u_1)\}. \tag{B.3}$$

Likewise, to refine the upper bound on $E[y(t)|v = u]$ we can pick the minimum value of the upper bound on $E[y(t)|v = u_2]$, hence

$$E[y(t)|v = u] \leq \inf_{u_2 \geq u} \{E[y|v = u_2, z = t] \cdot \Pr(z = t|v = u_2) + K_1 \cdot \Pr(z \neq t|v = u_2)\}. \tag{B.4}$$

B.3. *Monotone Treatment Selection*

The assumption of monotone treatment selection is a modification of the MIV Assumption 7. When instrument and treatment coincide, $v = z$, then the MIV Assumption collapses to

Assumption 8 (Monotone Treatment Selection)
Let T be an ordered set. For each $t \in T$,

$$u_2 \geq u_1 \Rightarrow E[y(t)|z = u_2] \geq E[y(t)|z = u_1].$$

In order to derive the bounds on $E[y(t)|v = u] = E[y(t)|z = u]$ we will use equations (B.3) and (B.4), the next proposition states the result.

Proposition 5 (MTS Bounds)
For a given latent treatment level $t \in T$ and for different levels u of actual treatment z, the bounds on $E[y(t)|z = u]$ under Assumption 8 are

$$u < t \Rightarrow K_0 \leq E[y(t)|z = u] \leq E[y|z = t]$$
$$u = t \Rightarrow E[y(t)|z = u] = E[y|z = t]$$
$$u > t \Rightarrow E[y|z = t] \leq E[y(t)|z = u] \leq K_1.$$

Proof. See section B.7. ■

B.4. *Monotone Treatment Response*

The next step towards tightening the bounds on $E[y(t)|v = u]$ involves an assumption on the response function $y(\cdot) : T \to Y$.

Assumption 9 (Monotone Treatment Response)
Let T be an ordered set. For each $i \in I$,

$$t_2 \geq t_1 \Rightarrow y_i(t_2) \geq y_i(t_1).$$

Under the MTR Assumption Manski (1997, Corollary M1.2) derives bounds on $E[y(t)|v = u]$ that make more use of the information contained in the sample data than the MIV bounds (B.3) and (B.4).

Proposition 6 (Monotone Treatment Response Bounds)
The bounds on $E[y(t)|v = u]$ under Assumption 9 are

$$E[y(t)|v = u] \geq E[y|v = u, z \leq t] \cdot Pr(z \leq t|v = u) + K_0 \cdot Pr(z > t|v = u) \tag{B.5}$$
$$E[y(t)|v = u] \leq E[y|v = u, z \geq t] \cdot Pr(z \geq t|v = u) + K_1 \cdot Pr(z < t|v = u). \tag{B.6}$$

Proof. See section B.7. ■

What is the advantage of the MTR bounds (B.5) and (B.6) over the MIV bound (B.3) and (B.4)? The MTR bounds use more information from the sample data. For example, the MTR lower bound uses the relatively uninformative range restriction K_0 on Y only for treatment values $z > t$, for all values $z \leq t$ it exploits the more informative sample data. Compare this to the MIV lower bound which uses K_0 for all values of $z \neq t$ and exploits the sample data only for $z = t$. The MIV upper bound also makes productive use of the sample data only for $z = t$ while the MTR upper bound exploits all $z \geq t$. Altogether the MTR bounds thus use the entire range for z while the MIV bounds only exploits the sample data for $z = t$.

B.5. *Combining MIV with MTR*

Our goal is to derive sharp bounds on $E[y(t)]$ without restricting the support Y of $y(\cdot)$ to the interval $[K_0, K_1]$. Combining the MIV Assumption 7 with the MTR Assumptions 9 will yield bounds that still depend on K_0 and K_1, however the MIV-MTR bounds are the last crucial step that we need before we are ready to derive unrestricted bounds on $E[y(t)]$.

Proposition 7 (MIV–MTR Bounds)
The bounds on $E[y(t)|v = u]$ under Assumptions 7 and 9 are

$$E[y(t)|v = u] \geq sup_{u_1 \leq u} \left\{ E[y|v = u_1, z \leq t] \cdot Pr(z \leq t|v = u_1) + K_0 \cdot Pr(z > t|v = u_1) \right\}. \tag{B.7}$$

$$E[y(t)|v = u] \leq inf_{u_2 \geq u} \left\{ E[y|v = u_2, z \geq t] \cdot Pr(z \geq t|v = u_2) + K_1 \cdot Pr(z < t|v = u_2) \right\}. \tag{B.8}$$

Proof. See section B.7. ■

How do the MIV-MTR bounds (B.7) and (B.8) compare to the MIV bounds (B.3) and (B.4)? The MIV-MTR bounds make better use of the sample data than the MIV bounds. This is a direct consequence of the properties of the MTR bounds as discussed above in section B.4.

B.6. *Combining MTS with MTR*

Now we are ready to derive bounds on $E[y(t)]$ that do not rely on the artificial restriction that the support Y of $y(\cdot)$ is bounded below by K_0 and above by K_1. Starting point is the observation that under MTS the instrument v and the treatment z coincide. In order to derive the bounds on $E[y(t)|v = u] = E[y(t)|z = u]$ we will adapt equations (B.7) and (B.8) to the new environment in which $v = z$. We need to be careful about how the actual treatment value u relates to the latent value t. The discussion here is similar to section B.3. The next proposition states the MTS–MTR bounds.

Proposition 8 (Sharp MTR–MTS Bounds on Conditional Expectation)
The bounds on $E[y(t)|v = u] = E[y(t)|z = u]$ under Assumptions 8 and 9 are

$$\begin{aligned}
&u < t \Rightarrow E[y|z = u] \leq E[y(t)|z = u] \leq E[y|z = t] \\
&u = t \Rightarrow E[y|z = t] \leq E[y(t)|z = u] \leq E[y|z = t] \\
&\qquad \Leftrightarrow E[y|z = t] = E[y(t)|z = u] \\
&u > t \Rightarrow E[y|z = t] \leq E[y(t)|z = u] \leq E[y|z = u].
\end{aligned}$$

Proof. See section B.7. ■

To obtain bounds on the unconditional expectation, 'integrate' out over all values of z:

Corollary 9 (Sharp MTR–MTS Bounds on Unconditional Expectation)
The bounds on $E[y(t)]$ under Assumptions 8 and 9 are

$$\begin{aligned}
&\sum_{u<t} E[y|z = u] \cdot Pr(z = u) + E[y|z = t] \cdot Pr(z \geq t) \\
&\qquad \leq E[y(t)] \leq \sum_{u>t} E[y|z = u] \cdot Pr(z = u) + E[y|z = t] \cdot Pr(z \leq t).
\end{aligned}$$

Last, to put an upper bound on the average treatment effect $\Delta(s,t) := E[y(t)] - E[y(s)]$, for $s < t$, under both the MTR and MTS Assumptions, we merely add subtract the lower bound on $E[y(s)]$ from the upper bound on $E[y(t)]$:

$$\begin{aligned}\Delta(s,t) \leq & \sum_{u<s} \big(E\left[y|z=t\right] - E\left[y|z=u\right]\big) \cdot \Pr(z=u) \\ & + \big(E\left[y|z=t\right] - E\left[y|z=s\right]\big) \cdot \Pr(s \leq z \leq t) \\ & + \sum_{u>t} \big(E\left[y|z=u\right] - E\left[y|z=s\right]\big) \cdot \Pr(z=u).\end{aligned} \tag{B.9}$$

B.7. *Proofs*

Proof of Proposition 5. We need to be careful about how the actual treatment value u relates to the latent treatment value t. There are three cases.

(i) If $u < t$ then the term $E[y|v = u_1, z = t]$ in equation (B.3) equals $E[y|z = u_1, z = t]$ which is undefined for $u_1 \leq u$. The probability $\Pr(z = t|v = u_1) = \Pr(z = t|z = u_1) = 0$ for $u_1 \leq u$. The only thing remaining for the lower bound is K_0, noting that $\Pr(z \neq t|v = u_1) = \Pr(z \neq t|z = u_1) = 1$ for $u_1 \leq u$. Hence, if $u < t$ then $K_0 \leq E[y(t)|z = u]$. Next, to derive the upper bound when $u < t$ consider equation (B.4). How do we find the infimum? We have the degree of freedom to set $u_2 = t$ which gives an upper bound of $E[y(t)|z = t] =: E[y|z = t]$. For all other values of $u_2 \geq u$ the upper bound will just be K_1 which exceeds $E[y|z = t]$. Thus, if $u < t$ then $E[y(t)|z = t] =: E[y|z = t] \geq E[y(t)|z = u]$.

(ii) If $u = t$ then the supremum of the lower bound in equation (B.3) is $E[y(t)|z = t]$ which equals $E[y|z = t]$. The infimum of the upper bound is also equal to $E[y|z = t]$.

(iii) If $u > t$ in the derivation of the lower bound from equation (B.3) we have the degree of freedom to set $u_1 = t$ which gives a lower bound of $E[y(t)|z = t] =: E[y|z = t]$. For all other values of $u_1 \leq u$ the upper bound will just be K_0 which is less than $E[y|z = t]$. Thus, if $u > t$ then $E[y(t)|z = t] =: E[y|z = t] \leq E[y(t)|z = u]$. Next, to derive the upper bound when $u > t$ consider equation (B.4). How do we find the infimum? If $u > t$ then the term $E[y|v = u_2, z = t]$ in equation (B.4) equals $E[y|z = u_2, z = t]$ which is undefined for $u_2 \geq u$. The probability $\Pr(z = t|v = u_2) = \Pr(z = t|z = u_2) = 0$ for $u_2 \geq u$. The only thing remaining for the upper bound is K_1, noting that $\Pr(z \neq t|v = u_2) = \Pr(z \neq t|z = u_2) = 1$ for $u_2 \geq u$. Hence, if $u > t$ then $K_1 \geq E[y(t)|z = u]$.

In summary, we have

$$\begin{aligned}& u < t \Rightarrow K_0 \leq E[y(t)|z = u] \leq E[y(t)|z = t] =: E[y|z = t] \\ & u = t \Rightarrow E[y(t)|z = u] = E[y|z = t] \\ & u > t \Rightarrow E[y(t)|z = t] =: E[y|z = t] \leq E[y(t)|z = u] \leq K_1.\end{aligned}$$

■

Proof of Proposition 6. By the MTR Assumption 9 we have

$$\begin{aligned}& t < z \Rightarrow K_0 \leq y(t) \leq y(z) \\ & t = z \Rightarrow y(t) = y(z) \\ & t > z \Rightarrow y(z) \leq y(t) \leq K_1.\end{aligned}$$

This carries over immediately to the conditional expectations:

$$t < z \Rightarrow K_0 \leq E[y(t)|v = u] \leq E[y(z)|v = u, z > t] =: E[y|v = u, z > t] \tag{B.10}$$

$$t = z \Rightarrow E[y(t)|v = u] = E[y(z)|v = u, z = t] =: E[y|v = u, z = t] \tag{B.11}$$

$$t > z \Rightarrow E[y(z)|v = u, z < t] =: E[y|v = u, z < t] \leq E[y(t)|v = u] \leq K_1. \tag{B.12}$$

Now we combine the two lower bounds from equations (B.10) and (B.12) by weighing them with their probabilities, $P(t < z|v = u)$ and $P(t > z|v = u)$, and we also involve equation (B.11) to obtain the MTR lower bound

$$E[y(t)|v = u] \geq K_0 \cdot \Pr(z > t|v = u) + E[y(z)|v = u, z \leq t] \cdot \Pr(z \leq t|v = u).$$

Likewise, we combine the two upper bounds from equations (B.10) and (B.12) by weighing them with their probabilities, $P(t < z|v = u)$ and $P(t > z|v = u)$, and we also involve equation (B.11) to obtain the MTR upper bound

$$E[y(t)|v = u] \leq E[y(z)|v = u, z \geq t] \cdot \Pr(z \geq t|v = u) + K_1 \cdot \Pr(z < t|v = u).$$

■

Proof of Proposition 7. The MIV-MTR bounds follow immediately by adapting the derivation of the MIV bounds in section B.2. Starting point is Assumption 7. We get, for all $u_1 \leq u \leq u_2$

$$\begin{aligned} &E[y(t)|v = u] \geq E[y(t)|v = u_1] \\ &E[y(t)|v = u] \leq E[y(t)|v = u_2]. \end{aligned} \tag{B.13}$$

For each of the expectations on the right hand side of the two inequalities (B.13), equations (B.5) and (B.6) provide bounds. The inequality $E[y(t)|v = u] \geq E[y(t)|v = u_1]$ holds for all $u_1 \leq u$. To tighten the bound on $E[y(t)|v = u]$ we can pick the maximum value of the lower bound (B.5), hence

$$E[y(t)|v = u] \geq \sup_{u_1 \leq u} \left\{E[y|v = u_1, z \leq t] \cdot \Pr(z \leq t|v = u_1) + K_0 \cdot \Pr(z > t|v = u_1)\right\}. \tag{B.14}$$

Likewise, to refine the upper bound on $E[y(t)|v = u]$ we can pick the minimum value of the upper bound (B.6), hence

$$E[y(t)|v = u] \leq \inf_{u_2 \geq u} \left\{E[y|v = u_2, z \geq t] \cdot \Pr(z \geq t|v = u_2) + K_1 \cdot \Pr(z < t|v = u_2)\right\}. \tag{B.15}$$

■

Proof of Proposition 8. There are three cases.

(i) If $u < t$ the lower bound will be the supremum (across u_1) of $E[y|z = u_1, z \leq t] \cdot \Pr(z \leq t|z = u_1) + K_0 \cdot \Pr(z > t|z = u_1)$. The term $\Pr(z > t|z = u_1) = 0$ while $\Pr(z \leq t|z = u_1) = 1$ so the supremum has to be based on $E[y|z = u_1, z \leq t]$. Therefore, the lower bound is $\sup_{u_1 \leq u} \{E[y|z = u_1, z \leq t]\}$ which is equal to $\sup_{u_1 \leq u} \{E[y|z = u_1]\}$ when $u < t$. Contrast the lower bound here to the one in section B.3 for $u < t$. Imposing the MTS assumption only, the lower bound was K_0. Already it emerges that the lower bound under MTR–MTS is more informative than under MTS only. As for the upper bound, the infimum of $E[y|z = u_2, z \geq t] \cdot \Pr(z \geq t|z = u_2) + K_1 \cdot \Pr(z < t|z = u_2)$ will be based on $E[y|z = u_2, z \geq t]$. Therefore, the upper bound for the case $u < t$ equals $\inf_{u_2 \geq t} \{E[y|z = u_2, z \geq t]\}$ which is equal to $\inf_{u_2 \geq t} \{E[y|z = u_2]\}$. (Note that we use $\inf_{u_2 \geq t}$ and not $\inf_{u_2 \geq u}$. This holds because the intersection of $\{z = u_2\}$ and $\{z \geq t\}$ when $u < t$ and for all $u_2 \geq u$ equals $\{z \geq t\}$.)

(ii) If $u = t$ then the supremum of the lower bound in equation (B.7) is $\sup_{u_1 \leq t} \{E[y|z = u_1]\}$. The infimum of the upper bound is also equal to $\inf_{u_2 \geq t} \{E[y|z = u_2]\}$ which is the same as for the case $u < t$.

(iii) If $u > t$ the lower bound will be the supremum (across u_1) of $E[y|z = u_1, z \leq t] \cdot \Pr(z \leq t|z = u_1) + K_0 \cdot \Pr(z > t|z = u_1)$. The term $\Pr(z > t|z = u_1) = 0$ while $\Pr(z \leq t|z = u_1) = 1$ so the supremum has to be based on $E[y|z = u_1, z \leq t]$. Therefore, the lower bound is $\sup_{u_1 \leq t} \{E[y|z = u_1, z \leq t]\}$ which is equal to $\sup_{u_1 \leq t} \{E[y|z = u_1]\}$ when $u < t$. (Note that we use $\sup_{u_1 \leq t}$ and not $\sup_{u_1 \leq u}$. This holds because the intersection of $\{z = u_1\}$ and $\{z \leq t\}$ when $u > t$ and for all $u_1 \leq u$ equals $\{z \leq t\}$.) As for the upper bound, the infimum of $E[y|z = u_2, z \geq t] \cdot \Pr(z \geq t|z = u_2) + K_1 \cdot \Pr(z < t|z = u_2)$ will be based on $E[y|z = u_2, z \geq t]$. Therefore, the upper bound for the case $u > t$ equals $\inf_{u_2 \geq u} \{E[y|z = u_2, z \geq t]\}$ which is equal to $\inf_{u_2 \geq u} \{E[y|z = u_2]\}$. Contrast the upper bound here to the one in section B.3 for $u > t$. Imposing the MTS assumption only, the upper bound was K_1. Again it is true that the bound under MTR–MTS is more informative than under MTS only.

In summary, we have

$$\begin{aligned} &u < t \Rightarrow \sup_{u_1 \leq u} \{E[y|z = u_1]\} \leq E[y(t)|z = u] \leq \inf_{u_2 \geq t} \{E[y|z = u_2]\} \\ &u = t \Rightarrow \sup_{u_1 \leq t} \{E[y|z = u_1]\} \leq E[y(t)|z = u] \leq \inf_{u_2 \geq t} \{E[y|z = u_2]\} \\ &u > t \Rightarrow \sup_{u_1 \leq t} \{E[y|z = u_1]\} \leq E[y(t)|z = u] \leq \inf_{u_2 \geq u} \{E[y|z = u_2]\}. \end{aligned} \tag{B.16}$$

Next, note that

$$
\begin{aligned}
u_1 \leq u_2 \Rightarrow E[y|z=u_1] :=& E[y(u_1)|z=u_1] && \text{(definition)}\\
\leq& E[y(u_2)|z=u_1] && \text{(by MTR)}\\
\leq& E[y(u_2)|z=u_2] && \text{(by MTS)}\\
\leq& E[y|z=u_2] && \text{(definition).}
\end{aligned} \tag{B.17}
$$

The observation that $E[y|z=u_1] \leq E[y|z=u_2]$, which may appear obvious, does not follow from the MTR Assumption or from the MTS Assumption alone, but only from combining the two. The result that $E[y|z=u_1] \leq E[y|z=u_2]$ implies that

$$
\begin{aligned}
\sup_{u_1 \leq u} \{E[y|z=u_1]\} &= E[y|z=u]\\
\inf_{u_2 \geq t} \{E[y|z=u_2]\} &= E[y|z=t]\\
\sup_{u_1 \leq t} \{E[y|z=u_1]\} &= E[y|z=t]\\
\inf_{u_2 \geq u} \{E[y|z=u_2]\} &= E[y|z=u],
\end{aligned}
$$

and therefore by combining the results from equations (B.16) and equations (B.17) we obtain informative bounds that do not rely on bounded support Y of $y(\cdot)$:

$$
\begin{aligned}
u < t &\Rightarrow E[y|z=u] \leq E[y(t)|z=u] \leq E[y|z=t]\\
u = t &\Rightarrow E[y|z=t] \leq E[y(t)|z=u] \leq E[y|z=t]\\
&\Leftrightarrow E[y|z=t] = E[y(t)|z=u]\\
u > t &\Rightarrow E[y|z=t] \leq E[y(t)|z=u] \leq E[y|z=u].
\end{aligned}
$$

■

APPENDIX C: SELECTION CORRECTED BOUNDS

This section contains the proofs of Proposition 3 and Corollary 4. To get the upper bound on the ATE we first derive the upper bound on $E[y(t)]$ and then we derive the lower bound on $E[y(s)]$ for $s < t$. The upper bound on the ATE results as the upper bound on $E[y(t)]$ minus the lower bound on $E[y(s)]$ which finally we will average out over the years to obtain equation (4.2).

C.1. *Upper Bound on $E[y(t)]$*

Starting point for deriving nonparametric bounds that deal with endogenous sample selection is the MTR upper bound from equation (B.6). Recall that we obtained as the upper bound

$$
E[y(t)|v=u] \leq E[y|v=u, z \geq t] \cdot \Pr(z \geq t|v=u) + K_1 \cdot \Pr(z < t|v=u).
$$

The conditioning variable v now has two dimensions and we make this explicit by defining the vectors $v := (v \quad w_1)'$ and $u := (u \quad \omega_1)'$ and rewriting

$$
\begin{aligned}
E[y(t)|v=u, w_1=\omega_1] \leq E[y|v=u, w_1=\omega_1, z \geq t] \cdot \Pr(z \geq t|v=u, w_1=\omega_1) \ldots\\
+ K_1 \cdot \Pr(z < t|v=u, w_1=\omega_1).
\end{aligned}
$$

The last inequality holds for all ordered pairs $(u, \omega_1) \in (U, W_1)$. By the law of total probability we can now expand

$$
\begin{aligned}
&E[y(t)|v=u, w_1=\omega_1] \leq\\
&\quad E[y|v=u, w_1=\omega_1, z \geq t, E=1] \cdot \Pr(z \geq t, E=1|v=u, w_1=\omega_1) \ldots\\
&\quad + E[y|v=u, w_1=\omega_1, z \geq t, E=0] \cdot \Pr(z \geq t, E=0|v=u, w_1=\omega_1) \ldots\\
&\quad + K_1 \cdot \Pr(z < t|v=u, w_1=\omega_1).
\end{aligned}
$$

The newly introduced conditioning variable E is a labor force participation dummy which takes on value 1 if the person is working and 0 if the person is not working (a rigorous definition of what counts as working is part of the data section). The main identification problem is created by the term $E[y|v = u, w_1 = \omega_1, z \geq t, E = 0]$ which cannot be estimated from the data. There are two crude ways of circumventing this counterfactual expectation. A bounded support assumption in the spirit of Manski's worst–case bounds would solve this problem via $E[y|v = u, w_1 = \omega_1, z \geq t, E = 0] \leq K_1$. In many applications this is likely to be uninformative. Alternatively, we could assume positive selection into work, for example: $E[y|v = u, w_1 = \omega_1, z \geq t, E = 0] \leq E[y|v = u, w_1 = \omega_1, z \geq t, E = 1]$. However, as Blundel et al. (2007) point out there is no good reason to assume this.

A weaker set of assumptions is centered around the variable w_1 which we restrict to be binary. In the case of binary w_1 we can integrate out to obtain

$$
\begin{aligned}
&E[y(t)|v = u] \leq \\
&\quad \Pr(w_1 = 0|v)\Big(E[y|v = u, w_1 = 0, z \geq t, E = 1] \cdot \Pr(z \geq t, E = 1|v = u, w_1 = 0) \dots \\
&\quad + E[y|v = u, w_1 = 0, z \geq t, E = 0] \cdot \Pr(z \geq t, E = 0|v = u, w_1 = 0) \dots \\
&\quad + K_1 \cdot \Pr(z < t|v = u, w_1 = 0)\Big) \dots \\
&\quad + \Pr(w_1 = 1|v)\Big(E[y|v = u, w_1 = 1, z \geq t, E = 1] \cdot \Pr(z \geq t, E = 1|v = u, w_1 = 1) \dots \\
&\quad + E[y|v = u, w_1 = 1, z \geq t, E = 0] \cdot \Pr(z \geq t, E = 0|v = u, w_1 = 1) \dots \\
&\quad + K_1 \cdot \Pr(z < t|v = u, w_1 = 1)\Big) =: RHS.
\end{aligned}
$$

We get

$$
\begin{aligned}
RHS &= E[y|v = u, w_1 = 0, z \geq t, E = 1] \cdot \Pr(z \geq t, w_1 = 0, E = 1|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 0, z \geq t, E = 0] \cdot \Pr(z \geq t, w_1 = 0, E = 0|v = u) \dots \\
&\quad + K_1 \cdot \Pr(z < t, w_1 = 0|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 1, z \geq t, E = 1] \cdot \Pr(z \geq t, w_1 = 1, E = 1|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 1, z \geq t, E = 0] \cdot \Pr(z \geq t, w_1 = 1, E = 0|v = u) \dots \\
&\quad + K_1 \cdot \Pr(z < t, w_1 = 1|v = u) \\
&= E[y|v = u, w_1 = 0, z \geq t, E = 1] \cdot \Pr(z \geq t, w_1 = 0, E = 1|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 0, z \geq t, E = 0] \cdot \Pr(z \geq t, w_1 = 0, E = 0|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 1, z \geq t, E = 1] \cdot \Pr(z \geq t, w_1 = 1, E = 1|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 1, z \geq t, E = 0] \cdot \Pr(z \geq t, w_1 = 1, E = 0|v = u) \dots \\
&\quad + K_1 \cdot \Pr(z < t|v = u). \qquad \text{(C.1)}
\end{aligned}
$$

Plugging Assumption 3 into equation (C.1) gives

$$
\begin{aligned}
RHS &= E[y|v = u, w_1 = 0, z \geq t, E = 1] \cdot \Pr(z \geq t, w_1 = 0, E = 1|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 0, z \geq t, E = 0] \cdot \Pr(z \geq t, w_1 = 0, E = 0|v = u) \dots \\
&\quad + E[y|v = u, w_1 = 1, z \geq t, E = 1] \cdot \Pr(z \geq t, w_1 = 1, E = 1|v = u) \dots \\
&\quad + K_1 \cdot \Pr(z < t|v = u).
\end{aligned}
$$

Next, using Assumption 4 in the last equation yields

$$
\begin{aligned}
&E[y(t)|v=u] \leq \\
&\quad E[y|v=u, w_1=0, z\geq t, E=1]\cdot \Pr(z\geq t, w_1=0, E=1|v=u)\ldots \\
&\quad + E[y|v=u, w_1=1, z\geq t, E=1]\ldots \\
&\quad \times \Big(\Pr(z\geq t, w_1=0, E=0|v=u) + \Pr(z\geq t, w_1=1, E=1|v=u)\Big)\ldots \\
&\quad + K_1\cdot \Pr(z<t|v=u).
\end{aligned}
$$

Introducing the MIV Assumption 7 the upper bound becomes

$$
\begin{aligned}
&E[y(t)|v=u] \leq \\
&\quad \inf_{\tilde{u}\geq u}\Big\{E[y|v=\tilde{u}, w_1=0, z\geq t, E=1]\cdot \Pr(z\geq t, w_1=0, E=1|v=\tilde{u})\ldots \\
&\quad + E[y|v=\tilde{u}, w_1=1, z\geq t, E=1]\ldots \\
&\quad \times \Big(\Pr(z\geq t, w_1=0, E=0|v=\tilde{u}) + \Pr(z\geq t, w_1=1, E=1|v=\tilde{u})\Big)\ldots \\
&\quad + K_1\cdot \Pr(z<t|v_1=\tilde{u})\Big\}.
\end{aligned}
$$

Next, applying the MTS Assumption 2 and realizing that $v=z$ we can define three cases around the latent variable t as follows:

$$
\begin{aligned}
u<t \Rightarrow E[y(t)|z=u] \leq \inf_{\tilde{u}\geq t}\Big\{&E[y|z=\tilde{u}, w_1=0, E=1]\cdot \Pr(w_1=0, E=1)\ldots \\
&+ E[y|z=\tilde{u}, w_1=1, E=1]\cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big)\Big\} \\
u=t \Rightarrow E[y(t)|z=u] \leq \inf_{\tilde{u}\geq t}\Big\{&E[y|z=\tilde{u}, w_1=0, E=1]\cdot \Pr(w_1=0, E=1)\ldots \\
&+ E[y|z=\tilde{u}, w_1=1, E=1]\cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big)\Big\} \\
u>t \Rightarrow E[y(t)|z=u] \leq \inf_{\tilde{u}\geq u}\Big\{&E[y|z=\tilde{u}, w_1=0, E=1]\cdot \Pr(w_1=0, E=1)\ldots \\
&+ E[y|z=\tilde{u}, w_1=1, E=1]\cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big)\Big\}.
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
u_1\leq u_2 \Rightarrow E[y|z=u_1, w_1=1, E=1] :=& E[y(u_1)|z=u_1, w_1=1, E=1] && \text{(definition)} \\
\leq& E[y(u_2)|z=u_1, w_1=1, E=1] && \text{(by MTR)} \\
\leq& E[y(u_2)|z=u_2, w_1=1, E=1] && \text{(by MTS)} \\
\leq& E[y|z=u_2, w_1=1, E=1] && \text{(definition).}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
u<t \Rightarrow E[y(t)|z=u] \leq{}& E[y|z=t, w_1=0, E=1]\cdot \Pr(w_1=0, E=1)\ldots \\
&+ E[y|z=t, w_1=1, E=1]\cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big) \\
u=t \Rightarrow E[y(t)|z=u] \leq{}& E[y|z=t, w_1=0, E=1]\cdot \Pr(w_1=0, E=1)\ldots \\
&+ E[y|z=t, w_1=1, E=1]\cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big) \\
u>t \Rightarrow E[y(t)|z=u] \leq{}& E[y|z=u, w_1=0, E=1]\cdot \Pr(w_1=0, E=1)\ldots \\
&+ E[y|z=u, w_1=1, E=1]\cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big).
\end{aligned}
$$

Integrating out over U results in

$$\begin{aligned} E[y(t)] \leq \sum_{u>t} \Pr(z=u) \cdot \Big(& E[y|z=u, w_1=0, E=1] \cdot \Pr(w_1=0, E=1) \dots \\ & + E[y|z=u, w_1=1, E=1] \cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big)\Big) \dots \\ & + \Pr(z \leq t) \cdot \Big(E[y|z=t, w_1=0, E=1] \cdot \Pr(w_1=0, E=1) \dots \\ & + E[y|z=t, w_1=1, E=1] \cdot \big(\Pr(w_1=0, E=0) + \Pr(w_1=1, E=1)\big)\Big). \end{aligned} \tag{C.2}$$

This establishes the upper bound on $E[y(t)]$ in Proposition 3.

C.2. *Lower Bound on $E[y(s)]$*

The next step in deriving nonparametric bounds that deal with endogenous sample selection is the MTR lower bound from equation (B.5). Recall that we obtained as the lower bound

$$E[y(s)|v=u] \geq E[y|v=u, z \leq s] \cdot \Pr(z \leq s|v=u) + K_0 \cdot \Pr(z > s|v=u).$$

The conditioning variable v now has two dimensions and we make this explicit by by defining the vectors $v := (v \quad w_2)'$ and $u := (u \quad \omega_2)'$ and rewriting

$$\begin{aligned} E[y(s)|v=u, w_2=\omega_2] \geq & E[y|v=u, w_2=\omega_2, z \leq s] \cdot \Pr(z \leq s|v=u, w_2=\omega_2) \dots \\ & + K_0 \cdot \Pr(z > s|v=u, w_2=\omega_2). \end{aligned}$$

The last inequality holds for all ordered pairs $(u, \omega_2) \in (U, W_2)$. By the law of total probability we can now expand

$$\begin{aligned} & E[y(s)|v=u, w_2=\omega_2] \geq \\ & \quad E[y|v=u, w_2=\omega_2, z \leq s, E=1] \cdot \Pr(z \leq s, E=1|v=u, w_2=\omega_2) \dots \\ & \quad + E[y|v=u, w_2=\omega_2, z \leq s, E=0] \cdot \Pr(z \leq s, E=0|v=u, w_2=\omega_2) \dots \\ & \quad + K_0 \cdot \Pr(z > s|v=u, w_2=\omega_2). \end{aligned}$$

Analogous to the previous subsection, the main identification problem is created by the term $E[y|v=u, w_2=\omega_2, z \leq s, E=0]$ which cannot be estimated from the data. There are two crude ways of circumventing this counterfactual expectation. A bounded support assumption in the spirit of Manski's worst–case bounds would solve this problem via $E[y|v=u, w_2=\omega_2, z \leq s, E=0] \geq K_0$. In many applications this is likely to be uninformative. Alternatively, we could assume negative selection into work, for example: $E[y|v=u, w_2=\omega_2, z \leq s, E=0] \geq E[y|v=u, w_2=\omega_2, z \leq s, E=1]$. This assumption is very counter–intuitive.

A weaker set of assumptions is centered around the variable w_2 which we restrict to be binary. In the case of binary w_2 we can integrate out to obtain

$$\begin{aligned} & E[y(s)|v=u] \geq \\ & \quad \Pr(w_2=0|v)\Big(E[y|v=u, w_2=0, z \leq s, E=1] \cdot \Pr(z \leq s, E=1|v=u, w_2=0) \dots \\ & \quad + E[y|v=u, w_2=0, z \leq s, E=0] \cdot \Pr(z \leq s, E=0|v=u, w_2=0) \dots \\ & \quad + K_0 \cdot \Pr(z > s|v=u, w_2=0)\Big) \dots \\ & \quad + \Pr(w_2=1|v)\Big(E[y|v=u, w_2=1, z \leq s, E=1] \cdot \Pr(z \leq s, E=1|v=u, w_2=1) \dots \\ & \quad + E[y|v=u, w_2=1, z \leq s, E=0] \cdot \Pr(z \leq s, E=0|v=u, w_2=1) \dots \\ & \quad + K_0 \cdot \Pr(z > s|v=u, w_2=1)\Big) =: RHS. \end{aligned}$$

We get

$$
\begin{aligned}
RHS = {} & E[y|v=u, w_2=0, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=0, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=0, z\leq s, E=0]\cdot \Pr(z\leq s, w_2=0, E=0|v=u)\ldots \\
& + K_0\cdot \Pr(z>s, w_2=0|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=1, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=0]\cdot \Pr(z\leq s, w_2=1, E=0|v=u)\ldots \\
& + K_0\cdot \Pr(z>s, w_2=1|v=u) \\
= {} & E[y|v=u, w_2=0, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=0, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=0, z\leq s, E=0]\cdot \Pr(z\leq s, w_2=0, E=0|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=1, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=0]\cdot \Pr(z\leq s, w_2=1, E=0|v=u)\ldots \\
& + K_0\cdot \Pr(z>s|v=u). \qquad \text{(C.3)}
\end{aligned}
$$

Plugging Assumption 5 into equation (C.3) gives

$$
\begin{aligned}
E[y(s)|v=u] \geq {} & E[y|v=u, w_2=0, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=0, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=1, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=0]\cdot \Pr(z\leq s, E=0|v=u)\ldots \\
& + K_0\cdot \Pr(z>s|v=u).
\end{aligned}
$$

Next, using Assumption 6 in the last equation yields

$$
\begin{aligned}
E[y(s)|v=u] \geq {} & E[y|v=u, w_2=0, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=0, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=1, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=1]\cdot \Pr(z\leq s, E=0|v=u)\ldots \\
& + K_0\cdot \Pr(z>s|v=u) \\
= {} & E[y|v=u, w_2=0, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=0, E=1|v=u)\ldots \\
& + E[y|v=u, w_2=1, z\leq s, E=1]\ldots \\
& \times \Big(\Pr(z\leq s, w_2=1, E=1|v=u) + \Pr(z\leq s, E=0|v=u)\Big)\ldots \\
& + K_0\cdot \Pr(z>s|v=u).
\end{aligned}
$$

Introducing the MIV Assumption 7 the lower bound becomes

$$
\begin{aligned}
& E[y(s)|v=u] \geq \\
& \sup_{\tilde{u}\leq u}\Big\{ E[y|v=\tilde{u}, w_2=0, z\leq s, E=1]\cdot \Pr(z\leq s, w_2=0, E=1|v=\tilde{u})\ldots \\
& + E[y|v=\tilde{u}, w_2=1, z\leq s, E=1]\ldots \\
& \times \Big(\Pr(z\leq s, w_2=1, E=1|v=\tilde{u}) + \Pr(z\leq s, E=0|v=\tilde{u})\Big)\ldots \\
& + K_0\cdot \Pr(z>s|v=\tilde{u})\Big\}.
\end{aligned}
$$

Next, applying the MTS Assumption 2 and realizing that $v=z$ we can define three cases around the latent variable s

as follows:

$$\begin{aligned}
u < s \Rightarrow E[y(s)|z=u] \geq \sup_{\tilde{u}\leq u}\Big\{ & E[y|z=\tilde{u},w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=\tilde{u},w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big)\Big\} \\
u = s \Rightarrow E[y(s)|z=u] \geq \sup_{\tilde{u}\leq s}\Big\{ & E[y|z=\tilde{u},w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=\tilde{u},w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big)\Big\} \\
u > s \Rightarrow E[y(s)|z=u] \geq \sup_{\tilde{u}\leq s}\Big\{ & E[y|z=\tilde{u},w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=\tilde{u},w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big)\Big\}.
\end{aligned}$$

Next, note that

$$\begin{aligned}
u_1 \leq u_2 \Rightarrow E[y|z=u_1,w_2=1,E=1] :=& E[y(u_1)|z=u_1,w_2=1,E=1] && \text{(definition)} \\
\leq& E[y(u_2)|z=u_1,w_2=1,E=1] && \text{(by MTR)} \\
\leq& E[y(u_2)|z=u_2,w_2=1,E=1] && \text{(by MTS)} \\
\leq& E[y|z=u_2,w_2=1,E=1] && \text{(definition).}
\end{aligned}$$

Therefore,

$$\begin{aligned}
u < s \Rightarrow E[y(s)|z=u] \geq & E[y|z=u,w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=u,w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big) \\
u = s \Rightarrow E[y(s)|z=u] \geq & E[y|z=s,w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=s,w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big) \\
u > s \Rightarrow E[y(s)|z=u] \geq & E[y|z=s,w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=s,w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big).
\end{aligned}$$

Integrating out over U results in

$$\begin{aligned}
E[y(s)] \geq \sum_{u<s}\Pr(z=u)\cdot\Big(& E[y|z=u,w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=u,w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big)\Big)\ldots \\
& + \Pr(z\geq s)\cdot\Big(E[y|z=s,w_2=0,E=1]\cdot \Pr(w_2=0,E=1)\ldots \\
& + E[y|z=s,w_2=1,E=1]\cdot\big(\Pr(w_2=1,E=1)+\Pr(E=0)\big)\Big). \qquad \text{(C.4)}
\end{aligned}$$

This establishes the lower bound on $E[y(t)]$ in Proposition 3.

C.3. *Combining Upper Bound on $E[y(t)]$ with Lower Bound on $E[y(s)]$*

The last step in obtaining the average treatment effect from equation (4.2) in Corollary 4 is to subtract the lower bound on $E[y(s)]$ in equation (C.4) from the upper bound on $E[y(t)]$ in equation (C.2).

REFERENCES

Ardington, C., A. Case, and V. Hosegood (2007): "Labor Supply Responses to Large Social Transfers: Longitudinal Evidence from South Africa," *University of Cape Town Working Paper*.

Bertrand, M., S. Mullainathan, and D. Miller (2003): "Public Policy and Extended Families: Evidence from Pensions in South Africa," *World Bank Economic Review*, 17, 27–50.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323–363.

Card, D. (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450.

Case, A. and A. Deaton (1998): "Large Cash Transfers to the Elderly in South Africa," *The Economic Journal*, 108, 1330–1361.

——— (1999): "School Inputs and Educational Outcomes in South Africa," *The Quarterly Journal of Economics*, 114, 1047–1084.

Chamberlain, D. and S. van der Berg (2002): "Earnings Functions, Labor Market Discrimination and Quality of Education in South Africa," *BER Working Paper*.

Fedderke, J., J.Manga, and F. Pirouz (2004): "Challenging Cassandra: Household and per capita household income in the October Household Survey 1995-1999, income and expenditure surveys 1994 and 2000 and the Labour Force Survey 2000," *ERSA Working Paper*.

Hoogeveen, J. and B. Özler (2005): "Not Separate, Not Equal: Poverty and Inequality in Post–Apartheid South Africa," *William Davidson Institute Working Paper*.

Keane, M. and K. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

Keswell, M. (2004): "Education and Racial Inequality in Post Apartheid South Africa," *Santa Fe Institute Working Paper*.

Keswell, M. and L. Poswell (2004): "Returns to Education in South Africa: A Retrospective Sensitivity Analysis of the Available Evidence," *South African Journal of Economics*, 72, 834 – 860.

Leibbrandt, M., J. Levinsohn, and J. McCrary (2005): "Incomes in South Africa since the Fall of Apartheid," *NBER Working Paper*.

Maitra, P. and F. Vahid (2006): "The Effect of Household Characteristics on Living Standards in South Africa 1993-1998: A Quantile Regression Analysis with Sample Attrition," *Journal of Applied Econometrics*, 21, 999–1018.

Manski, C. (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334.

Manski, C. and J. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica: Notes and Comments*, 68, 997–1010.

——— (Forthcoming): "More on Monotone Instrumental Variables," *The Econometrics Journal*.

Mariotti, M. (2009): "Desegregating Labour Markets During Apartheid in South Africa," *Working Paper*.

May, J., M. Carter, and J. Maluccio (2000): "Kwazulu-Natal Income Dynamics Study (KIDS) 1993-1998: A Longitudinal Household Database for South African Policy Analysis," in *Working Paper Series*, Centre for Social and Development Studies, University of Natal, Durban, 21.

Mincer, J. (1974): *Schooling, Experience and Earnings*, New York: Columbia University Press.

Mwabu, G. and T. Schultz (1996): "Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa," *American Economic Review Papers and Proceedings*, 86, 330 – 334.

——— (2000): "Wage Premiums for Education and Location of South African Workers, by Gender and Race," *Economic Development and Cultural Change*, 48, 307 – 334.

Ntuli, M. (2007): "Exploring Gender Wage "Discrimination in South Africa, 1995- 2004: A Quantile Regression Approach," *IPC Working Paper Series*.

Posel, D., J. Fairburn, and F. Lund (2004): "Labour Migration and Households: A Reconsideration of the Effects of the Social Pension on Labour Supply in South Africa," *Working Paper*.

Serumage-Zake, P. and W. Naudé (2003): "Private Rates of Return to Education of Africans in South Africa: A Double Hurdle Model," *Development Southern Africa*, 20, 515 – 528.

Thomas, D. (1996): "Education Across Generations in South Africa," *American Economic Review Papers and Proceedings*, 86, 330 – 334.