

Nonparametric Bounds On The Returns To Language Skills

Libertad González*^{*}

Universidad Pompeu Fabra

Abstract

This paper applies the theoretical literature on nonparametric bounds on treatment effects to the estimation of how limited English proficiency (LEP) affects wages and employment opportunities for Hispanic workers in the United States. I analyze the identifying power of several weak assumptions on treatment response and selection, and stress the interactions between LEP and education, occupation and immigration status. I show that the combination of two weak but credible assumptions provides informative upper bounds on the returns to language skills for certain subgroups of the population. Adding age at arrival as a monotone instrumental variable also provides informative lower bounds.

Keywords: Nonparametric bounds, language skills, Hispanic workers, labor market outcomes.

JEL Codes: C14, J24, J31.

* Contact information: Universitat Pompeu Fabra, Department of Economics and Business, Jaume I building, Ramon Trias Fargas, 25-27, 08005-Barcelona (Spain). Phone (+34) 93 542 2610. Fax (+34) 93 542 1746. Email: libertad.gonzalez@upf.edu.

^{*} I would like to thank Chuck Manski, Chris Taber, Joe Altonji and Bruce Meyer for their guidance and useful comments. I also thank Francesca Molinari and three anonymous referees for their suggestions. All remaining errors are mine.

1. Introduction

Hispanics reached 13.3 percent of the total population in the United States in 2002 (37.4 million).¹ Hispanic workers earn less than Non-Hispanic Whites and are more likely to be unemployed. This has been attributed to assimilation problems for immigrants (2 in 5 Hispanics were foreign born in 2002), low levels of schooling (more than 2 in 5 Hispanics aged 25 and older had not graduated from high school), discrimination, and limited English proficiency (almost 1 in 3 Hispanics admitted speaking English less than “very well”). This paper explores the role of limited language skills in explaining wages and unemployment for Hispanic workers in the US. From a policy perspective, we may be interested in learning whether, and by how much, investing in language acquisition would help labor market outcomes for Hispanics converge to those of Non-Whites.²

Previous studies have estimated the effect of language skills on earnings through OLS regressions.³ Concerns about a possible “ability bias” led some to incorporating IV estimates,⁴ but these results tend to be unstable or implausibly large, and the credibility of the instruments is often questionable. The size of reported OLS and IV estimates varies considerably across studies and methods.

¹ Ramirez and De la Cruz 2003.

² In this paper I focus on how English proficiency translates into labor market outcomes. In terms of policy, we would also be interested in learning how language training translates into the actual mastery of the language. This is outside the scope of this paper. For a study of the determinants of language proficiency see, for example, Chiswick and Miller (1995).

³ For example, see Reimers 1983, McManus, Gould and Welch 1983, Grenier 1984, McManus 1985, Carliner 1996.

⁴ For example, see Chiswick and Miller 1995 or Bleakley and Chin 2002.

I propose a “conservative” approach in the spirit of Manski (Manski 1990, 1995, 1997, Manski and Pepper 2000), according to which a worst-case analysis would precede the study of alternative weak assumptions and their identifying power. Through the introduction of weak but credible assumptions, I provide identification regions for the effects under consideration, so that any point estimates obtained under stronger assumptions should lie within these regions. The results are informative, in the sense that the upper bounds provided are lower than some point estimates reported in previous studies.

The assumptions I explore include monotone treatment response, as introduced by Manski (1997), and monotone instrumental variables, in the sense of Manski and Pepper (2000). I also stress the interactions between LEP and education, occupation and years of residence in the U.S., and study the possibility of positive sorting into treatments.

All the assumptions considered are informative, in the sense that they improve the identification problem with respect to the worst-case situation in which no prior information is available. However, only when we assume positive selection into treatments (MTS) and a monotone response function (MTR) do we get bounds narrow enough to be interesting. The lower bound for the effect of LEP when only MTR and MTS are imposed is always zero, and the upper bound is slightly higher than the estimate obtained when assuming that selection into treatments is random (as in OLS). I also propose the use of immigrants’ age at arrival in the US as a monotone instrumental variable (MIV). This assumption reduces the MTR+MTS bounds by raising the lower bounds above zero.

Once we introduce covariates, MTR+MTS is rejected in some cases, such as low educated workers, allowing us to reject the presence of ability bias or even that treatments are assigned at random.

The bounds also rule out large effects for the later stages of language acquisition and for certain occupations. Improving EP from “well” to “very well” would lead to an average increase in wages no larger than 12.6 percent. Workers in agricultural occupations would experience a wage increase of no more than 6.7 percent when improving their English proficiency from “not at all” to “very well”, and that would reduce their unemployment probability by no more than 0.3 percentage points. For service workers, the overall wage premium would be no higher than 17.4 percent.

The remainder of the paper is organized as follows. In section 2, I briefly review the literature on the labor market effects of LEP. Section 3 analyzes the effect of English proficiency on earnings and unemployment rates. I first introduce the data sources and the variables included in the analysis. Some informative conclusions are then reached by exploring the identifying power of different assumptions. Section 4 summarizes the findings.

2. Literature On The Labor Market Returns to Language Skills

In 2000, 47 million respondents to the U.S. Census of Population reported that they spoke a language other than English at home, and for 60 percent of them this language was Spanish. Moreover, 49 percent of those who spoke Spanish at home admitted to some deficiencies in

their mastery of English. Poor mastery of English has the potential for harming labor market outcomes in an environment where that is the main language in which communication takes place. Several theoretical explanations for this intuitive fact have been attempted. For McManus et al. (1983) and McManus (1985), the extent of the earnings loss caused by English deficiency depends both on the nature of production (integrated vs. non-integrated) and on the nature of the product (personal vs. impersonal).

Language proficiency is viewed in the economics literature as a form of human capital. Grenier (1984), for example, states: *‘people invest in language for the same reasons that they invest in education and other kinds of human capital –that is, in order to maximize the present value of expected net returns’*. For Chiswick and Miller (1995), *“Language skills are an important form of human capital. They satisfy the three basic requirements for human capital: they are embodied in the person; they are productive in the labor market and or in consumption; and they are created at a sacrifice of time and out-of-pocket resources”*.

The standard approach to estimating the effect of language proficiency on earnings is to include some index of English proficiency (EP) in an earnings regression, together with other covariates such as education, experience, and years of residence in the country. The first US survey with detailed questions on EP was the 1976 Survey of Income and Education. Several studies in the early 1980s used this data set to produce a first set of estimates of the effect of EP on earnings. Reimers (1983) used a binary variable indicating English fluency, and found significant effects (18 to 20%) of LEP on wages for Puerto Ricans and other Hispanics. Grenier (1984) found effects of around 15%, concluding that

language completely explains the wage gap between Hispanics and whites. McManus et alia (1983) and McManus (1985) found that the effects of EP on wages increase with schooling and experience, and found higher penalties to LEP in more skilled occupations. McManus (1985) concluded that raising EP to perfect fluency would increase wages by 26% on average. Kossoudji (1988) focused on the simultaneity of the occupation and EP decisions, concluding that LEP reduces earnings in all occupations, with penalties as high as 66% or 30% for certain occupations.

These early studies, however, do not address the possibility that EP and earnings may be correlated simply because more able workers are more likely to speak English and to earn more. If there is a correlation between EP and unmeasured ability, then OLS yields biased estimates. Chiswick and Miller (1995) use IV analysis to correct for this endogeneity, but their results are "unstable" and the validity of their instruments is questionable. Using the 1980 Census, their OLS estimate of the effect of EP on earnings is 17%, which increases to 57% with the IV analysis. Bleakley and Chin (2002) also propose an IV approach.⁵ Their analysis, with data from the 1990 Census, concludes that the IV estimate is greater in magnitude than the OLS one. They provide evidence that the OLS estimate is slightly biased upwards due to the endogeneity of language skills, but substantially biased downward due to measurement error.

Other recent papers have used the pooled 1980 and 1990 US Censuses in their analyses. Mora and Davila (1998) focus on the different penalties imposed by LEP by gender and

⁵ They use immigrants' age at arrival interacted with country of origin as an instrument for English proficiency.

education, and Carliner (1996) estimates an average wage penalty of 21% for men and 9% for women, but the estimates are higher for educated workers. He concludes that the rate of return to moving from no EP to perfect fluency is roughly equal to the rate of return from four years of education.

Berman et alia (2000) provide a different approach to separating language acquisition effects from ability bias. With a data set⁶ that documents the change in immigrants' earnings as well as in language fluency, they are able to introduce individual fixed-effects in the regressions, which they interpret as "ability". They estimate a 26% effect of language fluency on earnings in the cross-section regression, which is reduced to 23% when including the fixed-effects. However, disaggregating by occupation shows that the returns to language proficiency are much higher in high skill occupations, and they are not significantly reduced when introducing fixed effects, while the effects are smaller in low skill occupations, and they turn insignificant in the fixed-effects specifications. This leads them to conclude that improved language fluency only affects wages for high skill workers, and the estimated effects for low skill workers can be entirely attributed to ability bias.

Different studies provide estimates of the effect of language skills on earnings that range from 10 percent to more than 100 percent in some IV estimates.⁷ This paper intends to provide a framework for consensus; an identification region obtained under weak but

⁶ Their data refers to Russian immigrants in Israel.

⁷ Bleakley and Chin (2002) find that improving EP by one unit (for example, from "well" to "very well") increases wages by 33.3 percent (before controlling for education), implying a 100 percent overall effect (from "not at all" to "very well").

credible assumptions, so that any point estimates obtained under stronger assumptions should lie within these regions.

The analysis explicitly addresses the suspected correlation between unobserved ability and language fluency, and the possibility of inter-individual variation in the returns to language fluency. Most previous studies that estimate varying returns to EP across different values of covariates such as education rely on the assumption of exogenous sorting.⁸ Once we relax this assumption, the evidence does not support many of those results.

3. English Proficiency, Wages and Unemployment

3.1 The Data and Variables

I define the population J as “Hispanics 16 to 64 years old in the U.S. in 1989”. Each member j of the population has some covariates $\mathbf{x} \in X$ (education, experience, etc), and they receive treatments $t \in T$ that correspond to their English proficiency. The possible treatments are five:⁹ t_5 (speaks only English at home); t_4 (speaks English very well); t_3 (well); t_2 (not well); t_1 (not at all). The response functions $y_j(\cdot): T \rightarrow Y$ and $u_j(\cdot): T \rightarrow U$ map treatments into outcomes. The two outcomes are earnings¹⁰ and employment status. The realized outcome $y \equiv y_j(z_j)$ is the actual level of earnings (or unemployment rate) for an individual who actually received treatment z_j . The latent outcomes $y_j(t)$, $t \neq z$ tell us what

⁸ See McManus et al. (1983), McManus (1985), Carliner (1996).

⁹ The Census Questionnaire asks everyone whether they speak any language other than English at home. If they do, then they have to answer the question of English proficiency. Most previous studies assumed that Hispanics that reportedly speak only English at home could be considered perfectly fluent in English.

¹⁰ I will refer to y as “earnings”, but I will actually be using the natural logarithm of hourly wages.

individual j would have earned if she had received treatment t . Similarly, the realized outcome $u_j \equiv u_j(z_j)$ is the actual unemployment rate among the individuals receiving a particular treatment, while the latent outcomes $u_j(t)$, $t \neq z$ tell us the probability of being unemployed that an individual would face if she were to receive a different treatment.

We are interested in learning about “average treatment effects”, i.e., we seek bounds on the differences: $E[y(t_i)] - E[y(t_j)]$ and $P[u(t_i)] - P[u(t_j)]$, for $t_i, t_j \in T$, $t_i \neq t_j$. For example, this would answer the question: what would the average increase in earnings be if all the population moved from proficiency level j to i , $i > j$? This is an application of classical treatment analysis, which compares what happens when treating everyone in the population with treatment s versus treating everyone with treatment t . Of course, we are facing a censoring problem, because we don't observe the wage that everyone would earn if all of them were to receive one particular treatment. We only observe the earnings (and unemployment rates) corresponding to actual treatments. To see this, we can decompose $E[y(t)]$ using the law of total probability:

$$E[y(t)] = E[y(t)|z=t] P(z=t) + E[y(t)|z \neq t] P(z \neq t) \quad (1)$$

The sampling process identifies $E[y(t)|z=t]$, $P(z=t)$ and $P(z \neq t)$, but doesn't give us any information about $E[y(t)|z \neq t]$. Here is where the researcher must impose assumptions in order to get identification. To start, we must specify a logical range for $E[y(t)]$.¹¹ In 1990, the minimum (hourly) wage was fixed at \$3.80. This corresponds to a log hourly wage of 1.335. Because there are some observations in the sample that go below this level, I choose

¹¹ Otherwise the bounds will not be informative at all, given that expectations can range from $-\infty$ to $+\infty$.

the value $K_0=1$ as the infimum for y . For an upper limit I choose $K_1=5$, which corresponds to an hourly wage of \$150 (more than 99% of the observed wages are lower than \$55). I will also report the bounds obtained by choosing alternative values of $[K_0, K_1]$.

As for the employment outcomes, we are interested in learning about $P[u(t)]$, i.e., the unemployment rate for a given treatment. This probability can be decomposed as in equation (1):

$$P[u(t)] = P[u(t)|z=t] P(z=t) + P[u(t)|z\neq t] P(z\neq t). \quad (2)$$

Assumptions must be imposed on $P[u(t)|z\neq t]$ in order to get identification. One advantage from dealing with probabilities instead of expectations is that we don't need to impose a range on $P[u(t)]$, since we know it necessarily belongs to $[0,1]$.

I construct my sample from the 1990 1% Public Use Microdata Sample (PUMS) for all 50 states. I consider all persons reporting a Hispanic origin¹², aged 16 to 64. The sample size is 117,998. In the analysis of earnings, I drop from the sample those who report zero earnings, zero usual hours worked per week, or zero weeks worked¹³. The final sample for earnings has 82,250 observations.¹⁴ The outcome variable y , is the natural log of hourly wage, which I calculate with the following information: weeks worked last year, usual hours worked per week last year and wages or salary income in 1989. The employment status variable takes three possible values: employed, unemployed, and out of the labor

¹² I count every individual who reported any "Hispanic" category when asked about his/her first ancestry.

¹³ Keeping females in the sample introduces an additional censoring problem due to lower participation rates, as we do not observe the wages of women out of the labor force. This could potentially bias our results. I do not explicitly model this source of censoring, thus implicitly assuming that, conditioning on the treatment and the covariates, expected wages are the same for women actually working and those out of the labor force. Excluding women from the sample does not however change the main results significantly (see section 3.2.5).

¹⁴ A potential problem with this sample may arise because of illegal immigrants and the way they are counted and report their labor market status. However, there's no obvious way around this problem.

force. The data on treatments received come from the answer to the question on “ability to speak English”. It is important to realize the qualitative character of the EP variable. First, it is a merely ordinal variable; we can’t quantify what speaking English “well” means, so it’s impossible to talk about “distance” between treatments¹⁵. Moreover, there’s some degree of subjectivity implicit in this variable, which opens the door to measurement error problems.¹⁶ I admit this deficiency of the EP measure and the convenience of a better indicator of EP (like the scores on some official test), less subjective and easier to quantify. In absence of such a variable, the best we can do is be aware of these weaknesses, and respect the discrete, ordinal character of our EP variable.

Observed (log) hourly wages are distributed symmetrically around an average of 2 and with a standard deviation of 0.64. More than 99% of the observations lie in the interval [0,4]. Observed wages are clearly increasing in EP, in a merely descriptive sense: people who speak better English earn higher wages, on average (see table 1). Seven percent of the individuals in the sample don't speak English at all, while almost 20 percent speak only English. Treatments t_2 , t_3 and t_4 are received by 15, 18 and 41 percent of the individuals in the sample, respectively. Hispanics with better mastery of English tend to be better educated (average years of education increase with t , from 6 to 12.3). On the other hand, experience shows a very pronounced decreasing trend, declining from 22.9 years for

¹⁵ Although many previous studies do so (see, for example, Mora and Davila 1998 and Berman et. al. 2000).

¹⁶ Dustmann and Van Soest (2000) address the consequences of using self-reported measures of language proficiency, which may contain substantial reporting errors. They decompose misclassification errors into a time-persistent and a time-varying component, using panel data. They find that this type of reporting error can have important consequences for the estimated effect of speaking fluency on earnings. Unfortunately, the cross-sectional nature of my data set does not allow for this type of analysis.

treatment t_1 to 13.3 years for t_5 . Table 1 also shows the proportion of foreign-born and Mexican ancestry for each treatment group.

Observed unemployment rates clearly decrease with EP (see table 2). The unemployment rate is 15.8 percent for Hispanics who don't speak English at all, while only 8.6 percent of those who speak English at home are unemployed. The proportion of people out of the labor force is also decreasing in EP level. The total difference is 12 percentage points (37 percent for t_1 against 25 percent for t_5). Thus, there's reason to suspect that low wages are not the only penalty imposed by the labor market to Hispanics with a low ability to communicate in English. They also seem to suffer higher unemployment rates, and remain out of the labor force with a higher probability.¹⁷

3.2 Exploring the Identifying Power of Different Assumptions

How does English proficiency affect earnings and employment opportunities for Hispanic workers? I start by estimating worst-case bounds, and then explore the identifying power of alternative assumptions. I will focus on the results obtained when combining the assumption that treatment response is monotone and that selection into treatments is positive.¹⁸

If expected log hourly wages belong to $[K_0, K_1]$, then, from equation (1), we must have:

$$E[y(t_i)|z=t_i] P(z=t_i) + K_0 P(z \neq t_i) \leq E[y(t)] \leq E[y(t_i)|z=t_i] P(z=t_i) + K_1 P(z \neq t_i) \quad (3)$$

¹⁷ I do not include the analysis of how LEP affects the probability of staying out of the labor force because the trends there seem to be driven by females.

¹⁸ I will spend more time describing the bounds for earnings, but the ones for unemployment are calculated in a parallel fashion.

The *worst-case bounds* for $E[y(t)]$ for each of the treatments are shown in table 3. A priori, $E[y(t)]$ can range from 1 to 5. We can see how the worst-case bounds already restrict this interval. The worst-case bounds for unemployment rates can be calculated analogously, and are shown in table 5.

3.2.1 Assumptions on Response Functions

These assumptions refer to $y_j(t_i)$ and $u_j(t_i)$, i.e., the individual functions mapping treatments into outcomes. The most usual is the “linear response assumption”, which imposes $y_j(t_i) = a + bt + \varepsilon_j$. This assumption is not applicable in this case, due to the ordinal character of the treatment variable. Instead, I will explore the implications of assuming *monotone treatment response* (MTR). This assumption is specified as follows:

$$t_{i+1} > t_i \rightarrow y_j(t_{i+1}) \geq y_j(t_i) \quad (4)$$

This assumption means that, *ceteris paribus*, response varies monotonically with treatment. In our particular case, it implies that improving someone’s EP cannot lower her wage (or her probability of finding a job), i.e., that each person’s wage (employment) function is weakly increasing in conjectured EP. This assumption is drawn from the human capital approach, which implies that EP is potentially productive in the labor market. Note that assuming MTR does not exclude the possibility of EP being not productive at all.

This weak assumption can narrow the worst-case bounds, since it implies that $E[y(t_j)|z=t_k] \geq E[y(t_i)|z=t_k]$, for any k and any $j > i$. Adding this new restriction to (3), the resulting bounds are calculated according to the following expression (see Minski (1997) for a full derivation of the formula):

$$\begin{aligned}
& K_0 P(z > t_i) + E[y(t_i)|z \leq t_i] P(z \leq t_i) \leq E[y(t_i)] \leq \\
& \leq K_1 P(z < t_i) + E[y(t_i)|z \geq t_i] P(z \geq t_i)
\end{aligned} \tag{5}$$

These bounds can be found in table 3. I also calculate confidence intervals using the percentile bootstrap method. This assumption alone is already informative, in the sense that it narrows the worst-case bounds considerably. Without making additional assumptions about treatment selection, MTR allows us to say that the expected log wage for workers who don't speak English at all lies between 1.05 and 1.98, while for workers who speak English very well this expectation is between 1.77 and 3.22. The MTR bounds for unemployment and the corresponding bootstrap confidence intervals can be found in table 5.

3.2.2 Assumptions on Treatment Selection

These assumptions refer to sorting, i.e., to the process that determines the way treatments are assigned to different individuals in the population.

Exogenous Treatment Selection (ETS)

The usual method when dealing with this question is to assume exogenous treatment selection.¹⁹ This implies that treatments are assigned randomly to the population, and can be expressed formally as:

$$E[y(t)|x,z] = E[y(t)|x] \tag{6}$$

¹⁹ ETS is implicitly assumed in Reimers 1983, McManus, Gould and Welch 1983, Grenier 1984, McManus 1985, Carliner 1996, etc.

This assumption is implied in OLS analyses, but it does not seem appropriate in our case. Individuals receiving different treatments differ substantially from one another, possibly in unobservable dimensions as well as in terms of the observed covariates. However, one advantage of assuming ETS is that it always yields point identification, which probably accounts for its widespread use. The ETS estimates are equivalent to the coefficients obtained by running OLS on earnings with one dummy variable for each EP level and no other covariates²⁰. The results of the OLS analysis can be seen in table 6. The estimated increase in earnings derived from moving from treatment 1 to 4 is 37.8 percent. I also present the results obtained when conditioning on several covariates, with results that replicate those obtained in previous studies.²¹ After including the full set of controls, the estimated effect of moving from treatment 1 to 4 is 12.6 percent. As for the employment outcome, the estimated decrease in the probability of being unemployed derived from moving from treatment 1 to 4 under ETS is 6.3 percentage points (see table 5).²²

²⁰ For examples of this approach, see Carliner (1995) and Lazear (1998). They transform the EP variable into a binary treatment, which is then introduced into a regression as a dummy.

²¹ Education and experience show a significant positive effect on earnings, with coefficients of 0.05 and 0.015, respectively. Gender (a dummy variable for females) is significantly negative. The coefficients on the dummy for US born vs. immigrant and years of residence are consistent with the results suggested by the literature on immigration. The coefficient on the variable “US born” reflects the percent wage differential between natives and immigrants at the time of arrival of the latter. The positive sign indicates that native Hispanics earn more than immigrants at the time of their arrival to this country. The coefficient on the variable “years of residence” is the rate at which earnings of immigrants rise relative to natives. This coefficient is significantly positive, i.e., wages of immigrants appear to rise faster than those of natives. In particular, one more year of residence in the US for an immigrant increases the wage in about 0.4%. I also introduce dummies for ancestry and occupation. After conditioning on all the covariates, moving from t_1 to t_2 increases earnings in an estimated 3%. This percentage is 8% from t_1 to t_3 ; 12.6% from t_1 to t_4 ; and 19% from t_1 to t_5 .

²² The same result could be obtained by estimating a Probit model including observations only for individuals actually receiving treatments 1 and 4, and a dummy variable indicating treatment 4 as the only covariate.

Monotone Treatment Selection (MTS)

If we suspect that sorting into treatment is not exogenous, then a natural alternative to ETS would be to assume that selection into treatments is monotone. This assumption can be defined for earnings as follows:

$$t_j > t_i \rightarrow E[y(t_k)|x, z=t_j] \geq E[y(t_k)|x, z=t_i] \quad (7)$$

This implies that workers with higher EP have weakly higher mean wage functions than people with lower EP. Human capital accumulation theory predicts that persons with higher ability have higher mean wage functions and acquire more human capital than do persons with lower ability. If language proficiency is a form of human capital, then MTS is an implication of this type of model. Many previous studies²³ show concerns that persons with higher ability have higher mean wage functions and speak better English than do persons with lower ability. Some studies have tried to correct for this endogeneity in the framework of a linear model, mainly by using instrumental variables. However, these attempts haven't been convincing. The MTS assumption is a natural way of incorporating this possibility, avoiding the limitations of IV assumptions.²⁴ The new bounds can be derived from inequality (3) by adding the new restrictions implied by MTS, and the final expression is the following (see Manski and Pepper (2000) for the full derivation):

$$K_0 P(z < t) + E[y(t)|z=t] P(z \geq t) \leq E[y(t)] \leq K_1 P(z > t) + E[y(t)|z=t] P(z \leq t). \quad (8)$$

²³ Borjas 1991, Chiswick & Miller 1995, Berman et al. 2000.

²⁴ The validity of an instrument is usually hard to justify and can compromise the credibility of the results. Assuming MTS incorporates the concerns about selection and, although weaker than IV in terms of identification power, does so without compromising credibility.

The resulting bounds and the confidence intervals are reported in table 3. Note that this assumption narrows the expected wage for treatments t_4 and t_5 considerably, but its identifying power is weak for the rest of the treatments.

As for unemployment, MTS implies that more able individuals are more likely to speak English and to be employed. The bounds derived from imposing MTS are shown in table 5.

Monotone Instrumental Variables (MIV)

A traditional way of approaching the selection problem in the literature is to make use of *instrumental variables* (IV). An instrumental variable v is assumed to satisfy mean-independence across specified sub-populations:

$$E[y(t)|x, v=u_2] = E[y(t)|x, v=u_1], \quad u_1 \neq u_2 \quad (9)$$

It is usually a problem to find good instruments. For example, Bleakley and Chin (2002) propose using immigrants' age at arrival as an IV for English proficiency, but admit that "age at arrival probably affects immigrant earnings through channels other than language". For example, younger arrivers may face lower costs of assimilation that are unrelated to language.

An alternative to the IV approach, as proposed by Manski and Pepper (2000), is to use a *monotone instrumental variable* (MIV), which is a weaker, but often more credible assumption.²⁵ An MIV (in the sense of mean-monotonicity) is assumed to satisfy the following:

$$u_2 > u_1 \rightarrow E[y(t)|x, v=u_2] \geq E[y(t)|x, v=u_1], \quad (10)$$

²⁵ Note that MTS is just a special case, where the MIV coincides with the treatment variable. Analogously, ETS can be considered a special case of IV.

I propose using age at arrival in the US as a monotone instrumental variable. Taking age as arrival as an MIV is equivalent to assuming that immigrants who arrived in the US at a younger age have weakly higher wage functions than those who arrived later, which can be attributed to lower costs of assimilation.²⁶ The choice of instrument is based on research on language acquisition that shows that younger children tend to learn languages more easily than adolescents and adults.²⁷

This assumption gives us new bounds on $E[y(t)]$ (see table 4), according to the following expression:²⁸

$$\begin{aligned} \sum_{u \in V} P(v=u) \{ \sup_{u_1 \leq u} (E[y(t)|v=u_1, z=t] P(z=t|v=u_1) + K_0 P(z \neq t|v=u_1)) \} \leq \\ E[y(t)] \leq \\ \sum_{u \in V} P(v=u) \{ \inf_{u_2 \geq u} (E[y(t)|v=u_2, z=t] P(z=t|v=u_2) + K_1 P(z \neq t|v=u_2)) \} \end{aligned} \quad (11)$$

3.2.3 Combining Assumptions on Treatment Response and Selection

It seems reasonable to impose at the same time that the response function is monotone and that selection into treatments is positive, assumptions that are both consistent with conventional theories of human capital accumulation. In addition to shrinking the identification region, once we impose MTR together with MTS, we eliminate K_0, K_1 from the equation, i.e., we do not need to restrict the range of the outcome variables.

Assuming both MTR and MTS yields the following formula (that can be derived starting from (5) and adding the restrictions implied by (7)) for the bounds on earnings:

²⁶ The MIV used is discrete and takes four possible values: US-born, arrived as a child (0 to 11 years old), arrived as a teenager (12 to 17 years old), and arrived as an adult (18 or older).

²⁷ See Bleakley and Chin 2002.

²⁸ Again, see Manski and Pepper (2000) for the original derivation of this formula.

$$\sum_{u < t} E[y(u)|z=u] P(z=u) + E[y(t)|z=t] P(z \geq t) \leq E[y(t)] \leq \sum_{u > t} E[y(u)|z=u] P(z=u) + E[y(t)|z=t] P(z \leq t) \quad (12)$$

The initial worst-case bounds are now considerably reduced. Expected wage lies between 1.67 and 1.98 if we assign treatment t_1 to the whole population, while it lies between 1.97 and 2.05 if treatment t_4 is assigned (see figure 1).

As for unemployment, imposing MTR+MTS yields the result that a worker who doesn't speak English at all faces an unemployment rate between 10.2 and 15.8, while this rate ranges from 9.3 to 10.4 for a worker who speaks English very well (see figure 2).

We can shrink the MTR+MTS bounds for wages further if we add the monotone instrumental variable as defined in the previous section. The new bounds can be seen in table 4 (see also figure 3).

3.2.4 Average Treatment Effects

The “average treatment effect” (henceforth, ATE) is defined simply as $E[y(t_i)] - E[y(t_j)]$. The bounds derived for individual expectations imply bounds on the differences. We are interested in gains in expected wage (and expected probability of being unemployed) when moving from one treatment to the next, because this provides a sense of the gains derived from “investing in language acquisition”. I will pay attention to ATE for t_i to t_{i+1} ; and also for t_1 to t_4 and t_1 to t_5 . These last two quantities give an idea of the global penalty derived from moving from no English proficiency at all to a perfect knowledge of the language.²⁹

²⁹ I consider it more reasonable to focus on t_1 to t_4 . While it is in theory possible for someone to move from treatment t_1 to t_4 in a relatively short period of time, it is not reasonable to assume that someone can easily reach perfect fluency (equivalent to someone who speaks only English at home) starting from zero.

The bounds for the ATEs for wages and the bootstrap confidence intervals are presented in table 3. In order to give an idea of the magnitudes, I graph the bounds for $E[y(t_4)] - E[y(t_1)]$ implied by the assumptions explored in this section (see figure 4). Note that, since y belongs to $[1, 5]$, the ATE could in principle range from -4 to 4 .³⁰ The worst-case bounds imply that the $ATE(t_1-t_4)$ could be anywhere between -3.34 and 2.74 . When we assume MTR+MTS, the worst-case bound is narrowed down to $[0, 0.385]$, i.e., the global wage penalty imposed by LEP is somewhere between 0 and 38.5 percent.³¹ Most of the studies on the subject report estimates of this magnitude (usually ranging from 15 to 35%). Assuming exogenous selection (see table 3), on the other hand, implies that the wage penalty from t_1 to t_4 amounts to 37.8 percent. This estimate is slightly lower than the upper bound when assuming MTR+MTS.³² When we add the MIV to the MTR+MTS assumptions, the upper bound remains unchanged, but the lower bound is now greater than zero (see table 4). Thus the combination of these three weak assumptions allows us to conclude that the overall wage penalty to LEP lies between 3.8 and 38.5 percent. If we use t_5 instead of t_4 as our measure of perfect fluency, then the overall penalty lies between 8.4 and 41.6 percent.

It is also informative to look at the average treatment effects for MTR+MTS from one treatment to the next. We observe that the maximum wage gains vary considerably along the treatments, although the lower bound (LB) is always zero. The upper bounds (UB)

³⁰ See footnote 31 for the bounds obtained by using alternative values of $[K_0, K_1]$.

³¹ Note that assuming MTR immediately implies that the lower bound cannot be lower than zero (although it could be higher).

³² See Appendix for a generalization of this result.

corresponding to moving from treatment t_1 to t_2 (32.2%), and t_2 to t_3 (22.3%) are quite high. However, we can rule out large effects of improving EP at the higher levels of fluency: moving from speaking English well to very well (from very well to perfect fluency) has an effect of at most 12.6% (11.2%) on wages.

The ATEs for unemployment are shown in figure 5. If we assume MTR+MTS, the penalty imposed to lack of EP (ATE(4-1)) amounts to between 0 and 6.46 unemployment points. Alternatively, assuming ETS yields an estimate of 6.3. Unemployment rates can thus be between zero and 6.5 points higher for non-English speakers than for those Hispanics who are fluent in English. Moving from t_1 to t_2 can decrease the probability of unemployment by at most 5.8 points. The upper bounds are 2.3 points for t_2 to t_3 , 1.5 for t_3 to t_4 , and 1.8 for t_4 to t_5 . Again, large effects can be ruled out at the later stages of language acquisition.

Note that the results for wages reported so far have assumed that $E[y(t)]$ is bounded between $[K_0, K_1] = [1, 5]$. This is of course somewhat arbitrary, thus I have also explored how the bounds change with alternative values of $[K_0, K_1]$.³³ This sensitivity analysis is unnecessary for the MTR+MTS bounds since they are not a function of $[K_0, K_1]$. It is also unnecessary for the unemployment outcome since $E[u(t)]$ is naturally bounded between 0 and 1.

³³ I perform a sensitivity analysis by using 0, 0.5 and 1 as possible values for K_0 , and 4, 4.5, 5, 5.5 and 6 for K_1 . The worst-case bounds always reduce the possible range of ATE(4-1) by 26%. The bounds under MTR reduce the worst-case bounds by 54 to 65%, depending on the values of $[K_0, K_1]$, and the MTS bounds reduce the worst-case range by 30 to 43%.

3.2.5 Introducing Covariates

So far I did not condition on any covariates when I calculated the bounds, i.e., I treated the effect of EP as constant across all individuals. In this section I explore how the penalties imposed by LEP may vary across individuals.

Education

Previous research (McManus et al. 1983, Carliner 1996, Mora and Davila 1998) suggests that the wage penalty due to LEP increases with education. For Carliner (1996), education “*is associated with both lower costs and greater benefits to learning English*”. This is the reason why “*the wage premium for speaking English well rises with education*”. In table 7, we can see the values for the bounds on log wages for each of the treatments and the different assumptions considered in the previous subsections, conditioning on three different levels of education. Under ETS, the wage penalty does seem to increase with education: from 9.5 percent for high school dropouts to 45 percent for college attendants. However, once we impose more conservative assumptions, this claim is not sustained. Under MTR+MTS, we cannot reject zero effects for all education levels. Thus, the results reported in the literature may be due simply to higher ability bias in higher educational levels.

Manski and Pepper (2000) show that the joint MTR+MTS assumption is a testable hypothesis, which should be rejected if $E[y(t)|z=t]$ is not weakly increasing in t .³⁴ The MTR+MTS hypothesis is rejected at the 95% confidence level for the lower education

³⁴ See Appendix for a formalization of this *monotonicity condition*.

category (less than a high school diploma), but not for the other two levels.³⁵ Under the human capital approach to language acquisition, we rule out the possibility that improving someone's EP (everything else the same) can lead to decreasing wages. Thus, if the joint MTR+MTS assumption is rejected, we can reasonably conclude that the assumption that is failing to hold is the one about sorting (MTS). This suggests the presence of *negative* ability bias in the group of low educated workers: workers who speak English well have higher mean wage functions (are more skilled) than workers who speak very well, at low levels of education. This possibility has not been suggested in the previous literature.³⁶

Interestingly, the results point to the same direction for the unemployment outcome. Assuming ETS suggests returns to EP that increase with education. However, we can also reject the joint MTR+MTS assumption at the 95% confidence level for workers with less than a high school degree, but not for the rest of the categories.

Country of Birth and Years of Residence in the US

I now consider the returns to EP for US-born Hispanics versus immigrants. ETS suggests higher returns to EP for immigrants, both in terms of wages and unemployment rates. However, zero is a lower bound for both immigrants and US-born once we impose MTR+MTS, and the data are consistent with the actual returns to EP being the same for both groups. This could happen if the ability bias was stronger in the immigrant group. We

³⁵ I compute a 95% confidence band for the estimates of $E[y(t)|z,s1]$ and find that no increasing function can be fit within the band.

³⁶ Of course, without further prior information, we cannot rule out a different explanation for this result: that increasing EP leads to decreasing wages for low educated workers. However, I find the story about selection more compelling.

cannot reject the joint MTR+MTS assumption in any of the two categories, for either wages or unemployment.

The bounds for immigrants are however reduced once we use age at arrival as an MIV. Under the MTR, MTS and MIV assumptions combined, the overall effect of LEP on wages for Hispanic immigrants is bounded between 6 and 43 percent (ATE(1-5)). When I restrict the analysis to workers no older than 45, then the overall effect is between 10.2 and 41.3 percent.

I also analyze the returns to EP for immigrants by years of residence in the US. Again, assuming exogenous selection would lead us to conclude that the return to EP in terms of wages is much higher (60.6 percent) for long-term immigrants (more than 20 years in the US) than for recent ones (24 percent for those who arrived in the US during the previous 10 years). However, assuming MTR+MTS yields bounds that include zero effects for all three categories (see table 8). Moreover, MTR+MTS is rejected at the 95% confidence level for immigrants that arrived in the US during the previous 20 years, but not for the longer-term ones. Thus, without further information about sorting, we cannot support the hypothesis that the returns to EP for immigrants increase with time spent in the US. Previous studies that point to this result are relying on ETS, which seems inadequate in this setting. The patterns are again surprisingly similar for the unemployment outcome. ETS suggests increasing penalties to LEP with time spent in the US. However, this may be due to positive ability bias being present in the category of long-term immigrants, a hypothesis that is rejected for the recent immigrants' category.

Occupation

Different occupations are extremely heterogeneous in their use of language. Thus, the reduction in productivity implied by a lack of EP may vary across different jobs. Here I analyze the interaction between EP, occupations, wages and unemployment. I divide all occupations into six groups (see table 9).³⁷ The occupations with the highest average wage are managerial and repair. The lowest wages correspond to agriculture and services. The goal is to find out how the wage penalty due to LEP varies across occupations. Assuming ETS, the average treatment effect from t_1 to t_4 and t_1 to t_5 varies largely with occupation (see table 9). Managerial and repair occupations show the highest penalties from t_1 to t_4 (62.4% and 40.2%, respectively), while the lowest values are attained in agricultural and service occupations (6.8% and 17.8%). However, this may be due to different degrees of selection across occupations. We can test for the joint MTR+MTS assumption by occupation. Although there are some dips, we cannot reject MTR+MTS for any of the occupational categories at the 95% confidence level. Thus, we can rule out large effects of LEP for certain occupations, such as agriculture and services (MTR+MTS upper bounds of 6.7 and 17.4 percent, respectively). The bounds are wider for the rest.

A similar story can be applied to the analysis of EP and unemployment by occupation (see table 10). The ATE from t_1 to t_4 derived from imposing ETS is very high for managerial occupations with 59 points (!), followed by technical with 7.6 points. The lowest penalties (0.0, 0.0 and 0.3 percentage points) correspond to service, operators and

³⁷ Excluding military occupations and "not classified".

agricultural occupations. Repair occupations seem to be in the middle, with an ETS estimate of 1.5 points.

We can reject the joint MTR+MTS assumption only for service occupations. Thus, we can conclude that the probability of finding a job is barely affected by LEP for agricultural workers and operators (MTR+MTS upper bounds of 0.3 and 0.2 percentage points), while the bounds are wider for the rest of the occupations.

I also study other covariates such as potential experience, ancestry, state of residence and gender (see table 8). The bounds on wages under MTR+MTS are similar for men and women, while in the case of unemployment, large effects can be ruled out for men, but not for women (upper bounds of 3.42 and 11.39 points, respectively). The joint MTR+MTS assumption cannot be rejected for either men or women and both outcomes.³⁸

One conclusion must be highlighted from this section. Assuming exogenous selection into treatments leads to estimates of the labor market penalties imposed by LEP that vary significantly across individuals. The higher returns to EP seem to correspond to highly educated workers and high skill occupations. This coincides with the results obtained by previous studies that use cross-section data, such as Mora and Davila (1998), and leads to viewing language proficiency as a complement of education, or of certain occupational skills. However, my results show that this conclusion is critically dependent on the maintained assumption of random assignment of treatments. If no prior information on sorting is available, we cannot reject the hypothesis that the returns to EP are the same

³⁸ Note that the upper bound under MTR+MTS is 0.386 for the whole sample, and 0.426 when excluding women from the sample.

across individuals, or even that they are all zero. I also show that the presence of positive selection can be tested if we accept the human capital approach to language acquisition. I find that positive selection can be rejected for certain categories, such as low educated workers and recent immigrants.³⁹

3 Summary and Conclusions

This paper studies the effect of limited English proficiency (LEP) on wages and employment opportunities for Hispanic workers in the United States. Imposing assumptions that are weaker but more credible than those implied by OLS or IV estimates, I find that, on average, LEP imposes an overall wage penalty that lies between 3.8 and 38.6 percent, and reduces the probability of finding a job by 0 to 6.5 percentage points. The estimates obtained assuming exogenous sorting (as in OLS) are 37.8 percent and 6.3 points, respectively, but there are strong reasons to believe that this assumption may not hold.

Under exogenous sorting, previous studies (as well as my own analysis) have estimated returns to EP that vary with education, years of residence in the US, and occupation. However, my analysis reveals that strong (and often arbitrary) assumptions about sorting are required in order to support those results. Under MTS+MTR, we cannot reject that the returns to EP are the same for all levels of education or all occupations. We can, however, rule out large effects for low education levels and certain occupations.

³⁹ Note that, if MTR+MTS is rejected and we think MTR always holds, then rejecting MTR+MTS also implies rejecting ETS.

Under MTR+MTS, the increase in wage derived from improving English proficiency from “well” to “very well” is found to be no larger than 12.6 percent on average, while the reduction in the probability of being unemployed would be no larger than 1.5 percentage points. For workers with a high school degree and no college education, that increase in wage would amount to no more than 3.5 percent.

Workers in agricultural occupations would experience a wage increase of no more than 6.7% when improving their English proficiency from “not at all” to “very well”, and that would reduce their unemployment probability by no more than 0.3 percentage points. For service workers, the wage premium would be no higher than 17.4%, and the effect on the probability of not finding a job for operators would be no larger than 0.2 points.

We can also reject the joint MTR+MTS assumption (and thus ETS) in some cases, suggesting the possibility of negative sorting for certain categories, such as recent immigrants and workers who did not graduate from high school.

This study contributes to the literature on the returns to language skills by providing an identification region under weak but credible assumptions, so that any point estimates obtained under stronger assumptions should lie within these regions. I show that the combination of two weak but credible assumptions, monotone treatment response and monotone treatment selection, provides informative upper bounds on the returns to language skills for certain subgroups of the population, in the sense that those upper bounds are smaller than some of the point estimates reported in previous studies. When using age at arrival for immigrants as a monotone instrumental variable, I obtain lower bounds that are higher than zero.

Appendix

Relating The ETS Estimate To The MTR+MTS Bounds

I here formalize the relationship between the ETS point estimate and the MTR+MTS bounds for the case of multiple (discrete) treatments. It is a common practice to use the OLS estimate as an approximation to the true value of an average treatment effect. If the researcher suspects that selection into treatments may not be just random, it is usual to informally use zero as a lower bound for the true ATE, and OLS as an upper bound (if the hypothesized selection is positive), although this is rarely made explicit. I here formalize what this informal practice implies for the case of two and more than two treatments, specifying the assumptions behind it. I will also show that, when the number of treatments is greater than two, the OLS estimate is in general not an upper bound for the ATE, except for the case of comparing the first treatment with the last.⁴⁰

a) Binary treatment ($t \in \{0,1\}$). In this case, there's only one average treatment effect of interest: $ATE_{1-0} = E[y(1)] - E[y(0)]$. If we assume *exogenous treatment selection*,⁴¹ then we obtain what I will call the OLS estimate, directly from equation (6):⁴²

$$ATE_{1-0}^{OLS} = E[y(1)|z=1] - E[y(0)|z=0]. \quad (A1)$$

Claim 1: If we assume instead *monotone treatment response* and *monotone treatment selection* (MTR+MTS),⁴³ then the bounds for the ATE are $[0, OLS]$.

Proof: Straightforward manipulation of equation (12).

Thus, for the case of two (mutually exclusive) treatments, using $[0, OLS]$ as informal bounds for the ATE is equivalent to assuming that response is weakly increasing in t (i.e., $y_j(1) \geq y_j(0)$), and that there is positive selection into treatment 1 (i.e., $E[y(t)|z=1]$

⁴⁰ This section is an extension of the results presented in Manski and Pepper (2000), section 3.4.

⁴¹ See section 3.2.2 for some discussion about what this assumption implies.

⁴² OLS implies not only ETS, but also a linear response function. However, in the case of discrete treatment variables, OLS and ETS can be thought of as equivalent, each treatment being equivalent to a dummy variable in an OLS regression. Thus, the results presented in this section can be used to interpret the coefficients of a multivariate OLS regression with dummies for each of the treatments.

⁴³ See section 3.3 for a discussion of the combined MTR+MTS assumption.

$\geq E[y(t)|z=0]$). When the number of treatments is just two, MTR+MTS is a plausible assumption if and only if $E[y(1)|z=1] \geq E[y(0)|z=0]$, i.e., if $OLS_{1-0} \geq 0$. This is a particular case of the *monotonicity conditions* specified in b). Thus, in the case of only two treatments, using $[0, OLS]$ as informal bounds for the ATE is equivalent to assuming monotone response and monotone selection, instead of exogenous selection.

b) More than two treatments. Suppose we have a treatment variable t , where $t=1,2,\dots,n$, with $n>2$. Then, the number of average treatment effects of interest is $\sum_{t=1}^{n-1} t$. In our particular application, English Proficiency can lie in 5 different categories; therefore there are 10 ATE's of interest.

Monotonicity conditions:

The *weak monotonicity condition* is satisfied when $E[y(t)|z=t]$ is weakly increasing in t .⁴⁴

The *strong monotonicity condition* is satisfied when $E[y(t)|z=t]$ is strictly increasing in t .

Claim 2: MTR+MTS is a plausible assumption if and only if the weak monotonicity condition holds (Manski and Pepper 2000, footnote 9).

Note that, since in our particular application the hypothesized selection into treatments is positive, I will always mean "positive selection" when I say "monotone", i.e., if $i>j$, then MTS will imply that $E[y(i)|z=i] \geq E[y(i)|z=j]$.

Under the assumption of *ETS*, $E[y(t)] = E[y(t)|z=t]$ (see equation (6)). Thus, the ATE from treatment a to b is just:

$$ATE(a-b)_{OLS} = E[y(a)|z=a] - E[y(b)|z=b] \quad (A2)$$

Under *MTR+MTS*, $E[y(t)]$ lies within the bounds shown in equation (3.13). From there we can easily derive the lower and upper bound (LB and UB) for the ATE(a-b) under the MTR+MTS assumption:⁴⁵

$$LB = \text{Max} \{ 0, \sum_{u<a} E[y(u)|z=u] P(z=u) + E[y(a)|z=a] P(z>a) - \quad (A3)$$

⁴⁴ This condition is equivalent to Equation (19) in Manski and Pepper (2000).

⁴⁵ See Manski and Pepper (2000).

$$\begin{aligned}
& \sum_{u>b} E[y(u)|z=u] P(z=u) - E[y(b)|z=b] P(z=b) \} \\
UB = & \sum_{u>a} E[y(u)|z=u] P(z=u) + E[y(a)|z=a] P(z=a) - \\
& \sum_{u<b} E[y(u)|z=u] P(z=u) - E[y(b)|z=b] P(z>b)
\end{aligned} \tag{A4}$$

For the sake of concision, we will from now on use the following notation: $E_i \equiv E[y(i)|z=i]$, and $P_i \equiv P(z=i)$.

Claim 3: If t is a treatment variable, where $t=1,2,\dots,n$, with $n>2$, then the following is true for the average treatment effects $ATE(a-b)$, where a, b are any two treatments, $a>b$:

i) The bounds derived from imposing MTR+MTS take the following values:

If $E_b \sum_{i=1}^b P_i - E_a \sum_{i=a}^n P_i > \sum_{i=1}^b E_i P_i - \sum_{i=a}^n E_i P_i$, then $LB=0$ and $UB>OLS$

If $E_b \sum_{i=1}^b P_i - E_a \sum_{i=a}^n P_i = \sum_{i=1}^b E_i P_i - \sum_{i=a}^n E_i P_i$, then $LB=0$ and $UB=OLS$

If $E_b \sum_{i=1}^b P_i - E_a \sum_{i=a}^n P_i < \sum_{i=1}^b E_i P_i - \sum_{i=a}^n E_i P_i$, then $LB>0$ and $UB<OLS$

ii) If the *weak monotonicity condition* holds, then $LB=0$ and $UB>OLS$.⁴⁶

Proof: Straightforward manipulation of (A2), (A3) and (A4).

Special cases:

ATE(n-k), $k<n$ (from any treatment to treatment n): The bounds derived from imposing MTR+MTS take the following values:

If $E_k \sum_{i=1}^k P_i > \sum_{i=1}^k E_i P_i$, then $LB=0$ and $UB>OLS$

If $E_k \sum_{i=1}^k P_i = \sum_{i=1}^k E_i P_i$, then $LB=0$ and $UB=OLS$

If $E_k \sum_{i=1}^k P_i < \sum_{i=1}^k E_i P_i$, then $LB>0$ and $UB<OLS$

If the strong monotonicity condition holds, then $LB=0$ and $UB>OLS$.

ATE(k-1), $k>1$ (from treatment 1 to any treatment): The bounds derived from imposing MTR+MTS take the following values:

⁴⁶ Claim 3 ii) follows directly from equations (19) and (23) in Manski and Pepper (2000) and is in fact informally stated there (p. 1005).

If $E_k \sum_{i=k}^n P_i < \sum_{i=k}^n E_i P_i$, then $LB=0$ and $UB>OLS$

If $E_k \sum_{i=k}^n P_i = \sum_{i=k}^n E_i P_i$, then $LB=0$ and $UB=OLS$

If $E_k \sum_{i=k}^n P_i > \sum_{i=k}^n E_i P_i$, then $LB>0$ and $UB<OLS$

If the strong monotonicity condition holds, then $LB=0$ and $UB>OLS$.

Corollary 1: For $ATE(n-1)$, it is *always* the case that $LB=0$ and $UB=OLS$.

Corollary 2: For $ATE(n-2)$, $LB=0$ and $UB>(=)OLS$ iff $E_2<(=)E_1$.

Corollary 3: For $ATE((n-1)-1)$, $LB=0$ and $UB>(=)OLS$ iff $E_n>(=)E_{n-1}$.

For example, in the EP application for wages, when assuming MTR+MTS, $ATE(4-1)$ is bounded between zero and an UB higher than the OLS estimate, since $E[y(5)|z=5] > E[y(4)|z=4]$, i.e., average observed wages are higher for workers with assigned treatment 5 than for those with assigned treatment 4.

Thus, we can conclude that:

➤ Under MTR+MTS, $UB=OLS$ only in the following special cases:

- i) Binary treatment.
- ii) More than two treatments, $ATE(n-1)$.
- iii) More than two treatments, $ATE(a-b)$ for any $a>b$, if (and only if):

$$E_b \sum_{i=1}^b P_i - E_a \sum_{i=a}^n P_i = \sum_{i=1}^b E_i P_i - \sum_{i=a}^n E_i P_i$$

➤ If the weak monotonicity condition holds (and this is necessary for MTR+MTS to be a plausible assumption), then for any $ATE(a-b)$, $a>b$, it is the case that $LB=0$, and $UB>OLS$ ($>OLS$ if the strong monotonicity condition holds).

➤ If the monotonicity condition does not hold, then we can have cases where $LB>0$ and $UB<OLS$.

Thus, in cases where the treatment variable is discrete and non-binary, the OLS estimates will in general be interior to the MTR+MTS bounds, except for the specific ATE from the first treatment to the last, and as long as the weak monotonicity condition holds.

Figure 1. Bounds under Monotone Treatment Response and Selection (Wages)

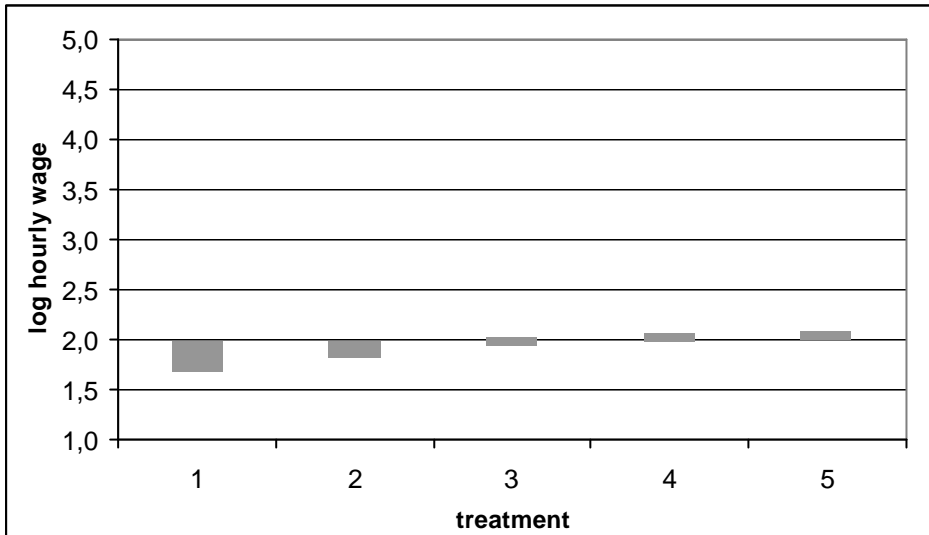


Figure 2. Bounds under Monotone Treatment Response and Selection (Unemployment)

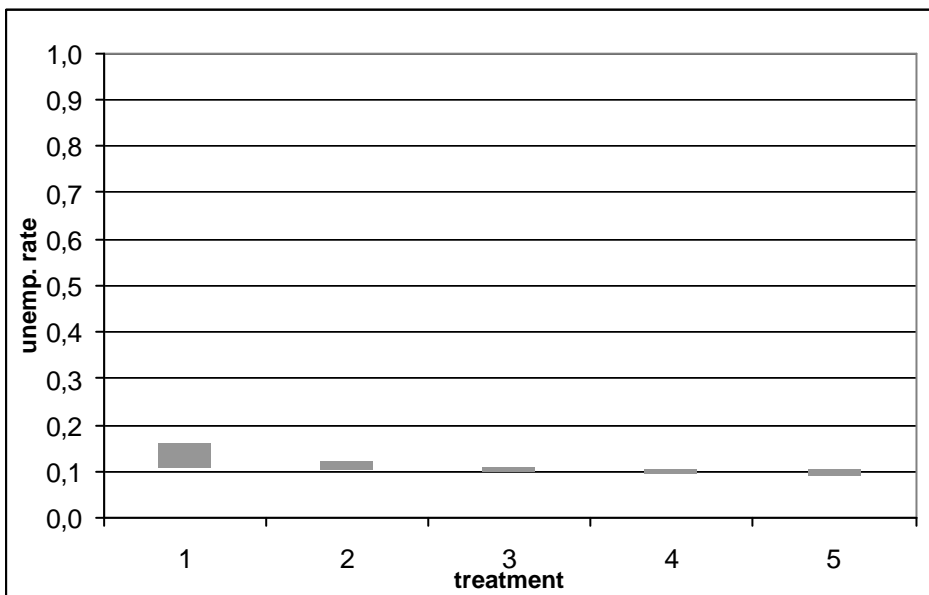


Figure 3. MTR+MTS plus age at arrival as MIV (Wages)

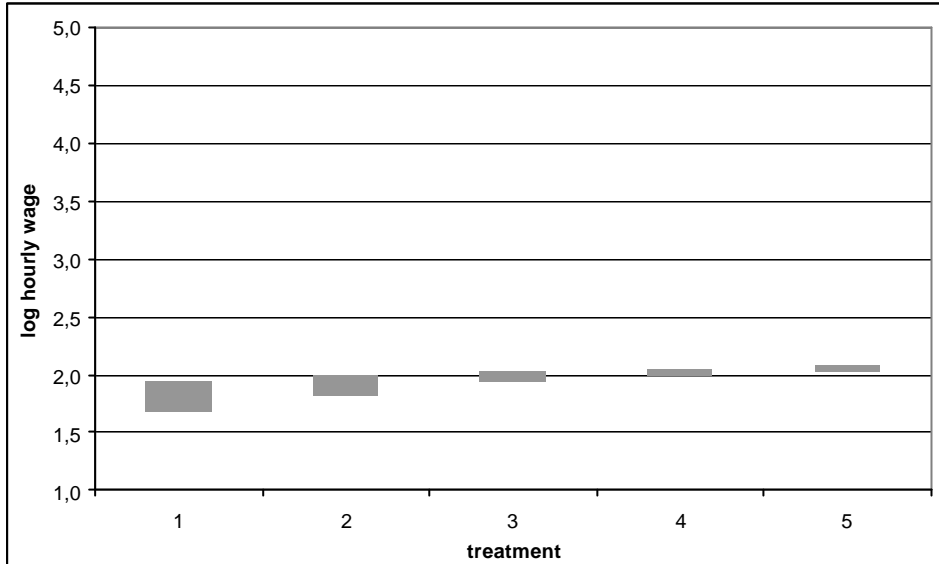


Figure 4. Average Treatment Effects for Wages (t_1-t_4)

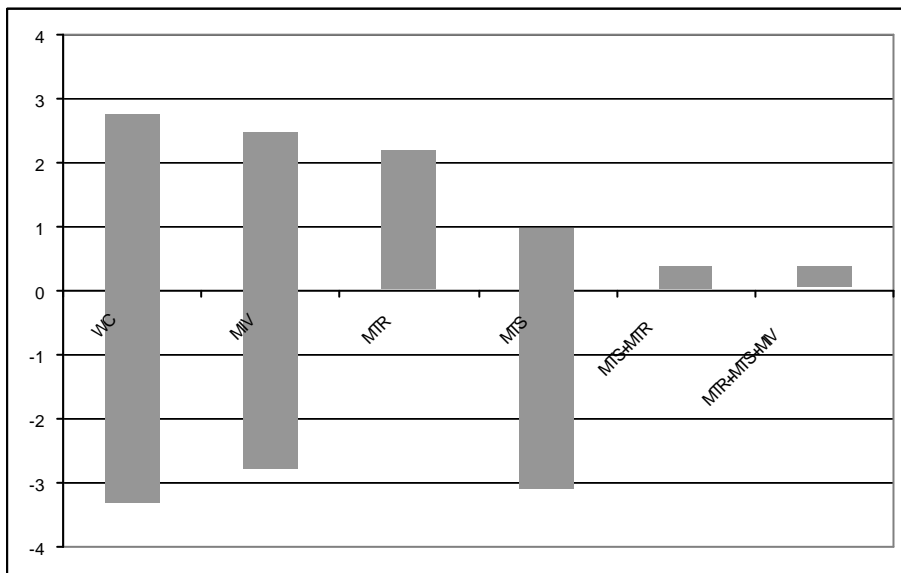


Figure 5. Average Treatment Effects for Unemployment (t_1-t_4)

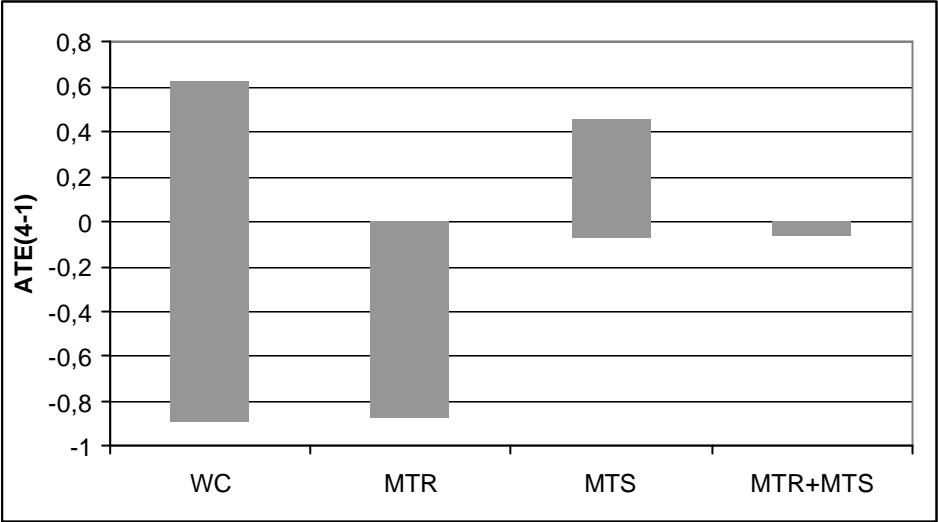


Table 1. English Proficiency and Average Earnings

Treatment	1	2	3	4	5	Totals
N. observations	5701	12175	14674	33725	15975	82250
% total observations	0.0693	0.1480	0.1784	0.4100	0.1942	1.00
Average log wage	1.6658	1.8097	1.9706	2.0440	2.0822	1.9774
Standard deviation	0.6831	0.6625	0.6993	0.7217	0.7505	0.7240
Average education	6.0171	7.9532	10.2220	11.9017	12.3149	10.6899
Average experience	22.9368	21.4180	19.4431	15.2730	13.3314	17.0807
% Born outside U.S.	0.9970	0.9747	0.8947	0.4890	0.6446	0.6987
% Mexican-American	69.5667	63.1458	56.5558	56.1601	55.6808	58.1009

Table 2. English Proficiency and Unemployment

Treatment	1	2	3	4	5
Employed	5379	11551	13649	30867	14472
Unemployed	1006	1511	1547	3227	1355
Out labor force	3785	5406	6064	12780	5399
TOTAL	10170	18468	21260	46874	21226
Unemp. Rate	0.158	0.116	0.102	0.095	0.086
Out l.f. / Total	0.372	0.293	0.285	0.273	0.254

Table 3. Bounds for Wages

Assumptions	ETS	95% confidence int.		Worst-case		95% conf. int.	
		LB	UB	LB	UB	LB	UB
$E[y(t1)]$	1.6658	1.648	1.6835	1.0461	4.7689	1.0444	4.7745
$E[y(t2)]$	1.8097	1.7974	1.821	1.1199	4.5278	1.1171	4.5348
$E[y(t3)]$	1.9706	1.9592	1.9815	1.1732	4.4595	1.1699	4.4677
$E[y(t4)]$	2.044	2.0363	2.0513	1.4281	3.788	1.4237	3.798
$E[y(t5)]$	2.0822	2.0708	2.094	1.2102	4.4333	1.2062	4.4412
$E[y(t5)]-E[y(t4)]$	0.0382	0.0245	0.0523	-2.5778	3.0052	-2.5871	3.0132
$E[y(t4)]-E[y(t3)]$	0.0734	0.0591	0.0874	-3.0315	2.6148	-3.04	2.6246
$E[y(t3)]-E[y(t2)]$	0.1609	0.1434	0.1767	-3.3546	3.3397	-3.3622	3.348
$E[y(t2)]-E[y(t1)]$	0.1439	0.1227	0.165	-3.649	3.4816	-3.6549	3.4887
$E[y(t4)]-E[y(t1)]$	0.3782	0.3584	0.3978	-3.3408	2.7418	-3.3476	2.7519
$E[y(t5)]-E[y(t1)]$	0.4164	0.395	0.4376	-3.5587	3.3871	-3.5653	3.3952

Assumptions	MTR		95% conf. int.		MTS		95% conf. int.	
	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	1.0461	1.9774	1.0444	1.9822	1.6658	4.7689	1.6474	4.7748
$E[y(t2)]$	1.166	2.2085	1.1629	2.2155	1.7536	4.3066	1.7426	4.3154
$E[y(t3)]$	1.3392	2.6808	1.3349	2.6901	1.7597	3.8011	1.7506	3.8117
$E[y(t4)]$	1.7672	3.2212	1.7619	3.2319	1.6308	2.6181	1.6247	2.6279
$E[y(t5)]$	1.9774	4.4333	1.9725	4.4413	1.2102	2.0822	1.2063	2.094
$E[y(t5)]-E[y(t4)]$	0	2.6661	0	2.6733	-1.4079	0.4514	-1.4163	0.4643
$E[y(t4)]-E[y(t3)]$	0	1.8821	0	1.8907	-2.1703	0.8585	-2.1793	0.8716
$E[y(t3)]-E[y(t2)]$	0	1.5148	0	1.5228	-2.547	2.0476	-2.5587	2.0626
$E[y(t2)]-E[y(t1)]$	0	1.1624	0	1.1683	-3.0153	2.6408	-3.0276	2.6608
$E[y(t4)]-E[y(t1)]$	0	2.1751	0	2.1855	-3.1381	0.9523	-3.1455	0.9721
$E[y(t5)]-E[y(t1)]$	0	3.3871	0	3.395	-3.5587	0.4164	-3.5656	0.4377

Table 3. *Continued*

Assumptions	MTR+MTS		95% conf. int.	
	LB	UB	LB	UB
$E[y(t1)]$	1.6658	1.9774	1.6474	1.9822
$E[y(t2)]$	1.7997	1.9874	1.7885	1.9922
$E[y(t3)]$	1.9257	2.0224	1.9165	2.0283
$E[y(t4)]$	1.97	2.0514	1.9644	2.0578
$E[y(t5)]$	1.9774	2.0822	1.9725	2.094
$E[y(t5)]-E[y(t4)]$	0	0.1122	0	0.1251
$E[y(t4)]-E[y(t3)]$	0	0.1258	0	0.1364
$E[y(t3)]-E[y(t2)]$	0	0.2226	0	0.2349
$E[y(t2)]-E[y(t1)]$	0	0.3216	0	0.3392
$E[y(t4)]-E[y(t1)]$	0	0.3856	0	0.4036
$E[y(t5)]-E[y(t1)]$	0	0.4164	0	0.4377

Table 4. Bounds for Wages under Age at Arrival as an MIV

Assumptions	MIV		MTR+MIV		MTR+MTS+MIV	
	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	1,0461	4,4440	1,0461	1,9322	1,6658	1,9322
$E[y(t2)]$	1,1199	3,9857	1,1660	2,2085	1,7997	1,9874
$E[y(t3)]$	1,1732	4,1963	1,3392	2,6808	1,9257	2,0224
$E[y(t4)]$	1,6253	3,5102	1,7672	3,2212	1,9700	2,0514
$E[y(t5)]$	1,3842	4,4333	2,0159	4,4333	2,0159	2,0822
$E[y(t5)]-E[y(t4)]$	0,0000	2,9395	0,0000	1,1624	0,0000	0,3217
$E[y(t4)]-E[y(t3)]$	0,0000	3,0764	0,0000	1,5148	0,0000	0,2226
$E[y(t3)]-E[y(t2)]$	0,0000	2,3370	0,0000	1,8821	0,0000	0,1257
$E[y(t2)]-E[y(t1)]$	0,0000	2,8080	0,0000	2,6661	0,0000	0,1122
$E[y(t4)]-E[y(t1)]$	0,0000	2,4640	0,0000	2,1751	0,0378	0,3857
$E[y(t5)]-E[y(t1)]$	0,0000	3,3871	0,0837	3,3871	0,0837	0,4164

Table 5. Bounds for Unemployment Rates

Assumptions	ETS	95% confidence int.		Worst-case		95% confidence int.	
		LB	UB	LB	UB	LB	UB
$E[y(t1)]$	0.1576	0.1486	0.1665	0.0119	0.9364	0.0112	0.9382
$E[y(t2)]$	0.1157	0.1104	0.1207	0.0179	0.8634	0.017	0.8655
$E[y(t3)]$	0.1018	0.0969	0.1065	0.0183	0.8386	0.0174	0.8409
$E[y(t4)]$	0.0947	0.0914	0.0978	0.0382	0.635	0.0368	0.6379
$E[y(t5)]$	0.0856	0.0814	0.0901	0.016	0.8289	0.0152	0.8313
$E[y(t5)]-E[y(t4)]$	-0.009	-0.0143	-0.0036	-0.619	0.7907	-0.6221	0.7934
$E[y(t4)]-E[y(t3)]$	-0.0072	-0.0125	-0.0016	-0.8004	0.6167	-0.8031	0.6198
$E[y(t3)]-E[y(t2)]$	-0.0139	-0.0212	-0.0067	-0.8451	0.8207	-0.8475	0.8231
$E[y(t2)]-E[y(t1)]$	-0.0419	-0.0525	-0.0318	-0.9185	0.8515	-0.9204	0.8537
$E[y(t4)]-E[y(t1)]$	-0.0629	-0.0725	-0.0534	-0.8982	0.6231	-0.9004	0.6263
$E[y(t5)]-E[y(t1)]$	-0.0719	-0.0816	-0.0619	-0.9204	0.817	-0.9222	0.8197

Assumptions	MTR		95% confidence int.		MTS		95% confidence int.	
	LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t1)]$	0.1022	0.9364	0.1002	0.9382	0.0119	0.1576	0.0112	0.1665
$E[y(t2)]$	0.0903	0.7998	0.0885	0.8024	0.0266	0.1824	0.0253	0.1873
$E[y(t3)]$	0.0725	0.6384	0.0707	0.6415	0.0417	0.3084	0.0397	0.3127
$E[y(t4)]$	0.0542	0.2734	0.0527	0.2763	0.0769	0.4655	0.0743	0.469
$E[y(t5)]$	0.016	0.1022	0.0152	0.1042	0.0856	0.8289	0.0814	0.8313
$E[y(t5)]-E[y(t4)]$	-0.2574	0	-0.2601	0	-0.3799	0.7519	-0.3856	0.7553
$E[y(t4)]-E[y(t3)]$	-0.5842	0	-0.5875	0	-0.2314	0.4238	-0.2362	0.4278
$E[y(t3)]-E[y(t2)]$	-0.7273	0	-0.7301	0	-0.1407	0.2818	-0.1461	0.2858
$E[y(t2)]-E[y(t1)]$	-0.846	0	-0.8485	0	-0.131	0.1706	-0.14	0.1754
$E[y(t4)]-E[y(t1)]$	-0.8822	0	-0.8844	0	-0.0806	0.4536	-0.0901	0.4572
$E[y(t5)]-E[y(t1)]$	-0.9204	0	-0.9222	0	-0.0719	0.817	-0.0816	0.8197

Table 5. *Continued*

Assumptions	MTR+MTS		95% confidence int.	
	LB	UB		
$E[y(t1)]$	0.1022	0.1576	0.1002	0.1665
$E[y(t2)]$	0.0991	0.1188	0.0971	0.1236
$E[y(t3)]$	0.0959	0.1082	0.0933	0.112
$E[y(t4)]$	0.093	0.1039	0.0902	0.1064
$E[y(t5)]$	0.0856	0.1022	0.0814	0.1042
$E[y(t5)]-E[y(t4)]$	-0.0183	0	-0.0231	0
$E[y(t4)]-E[y(t3)]$	-0.0152	0	-0.0197	0
$E[y(t3)]-E[y(t2)]$	-0.023	0	-0.0287	0
$E[y(t2)]-E[y(t1)]$	-0.0585	0	-0.0676	0
$E[y(t4)]-E[y(t1)]$	-0.0646	0	-0.0738	0
$E[y(t5)]-E[y(t1)]$	-0.0719	0	-0.0816	0

Table 6. Summary Linear Regressions for Hourly Wage

	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
Intercept	1.67	[176.24]	1.44	[140.12]	0.98	[77.51]	0.96	[66.66]
Dummy EP=2	0.14	[12.57]	0.07	[6.25]	0.03	[2.71]	0.03	[3.05]
Dummy EP=3	0.30	[27.37]	0.15	[12.87]	0.08	[7.42]	0.08	[6.93]
Dummy EP=4	0.38	[37.01]	0.16	[14.33]	0.15	[13.65]	0.13	[11.33]
Dummy EP=5	0.42	[37.83]	0.18	[15.24]	0.22	[17.81]	0.19	[15.51]
Education			0.04	[52.05]	0.06	[78.01]	0.05	[58.52]
Experience					0.02	[70.58]	0.01	[65.95]
US born					0.01	[1.47]	0.00	[0.34]
Years Res.					0.00	[13.17]	0.00	[11.67]
Gender					-0.23	[-49.15]	-0.23	[-46.00]
CA+TX+NM					0.02	[3.33]	0.01	[1.69]
P.Rican					0.09	[9.23]	0.08	[8.41]
Cuban					0.05	[4.71]	0.02	[2.21]
Spaniard					0.07	[3.97]	0.05	[2.88]
Central Am.					0.01	[1.35]	0.01	[1.01]
South Am.					0.05	[4.39]	0.04	[3.49]
Other non-mex.					0.00	[-0.24]	-0.01	[-0.93]
Managerial							0.47	[37.09]
Technical							0.24	[20.57]
Service							0.03	[2.83]
Repair							0.30	[25.96]
Operators							0.17	[16.01]
Military							-0.01	[-0.16]

Table 7. Bounds on Wages Conditioning on Education

Assumption	ETS	Worst-case		MTR		MTS		MTR+MTS	
Less than a high school degree		LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t5)]-E[y(t4)]$	-0.0667	-2.9871	3.4168	0.000	2.9417	-1.0695	0.4196	0.000	-0.0555
$E[y(t4)]-E[y(t3)]$	-0.0803	-3.0965	2.8754	0.000	2.2446	-1.8078	0.6029	0.000	-0.0279
$E[y(t3)]-E[y(t2)]$	0.1055	-3.0064	3.1376	0.000	1.7686	-2.2110	1.4541	0.000	0.0850
$E[y(t2)]-E[y(t1)]$	0.1363	-3.3451	3.1083	0.000	1.1613	-2.8642	2.1089	0.000	0.1619
$E[y(t4)]-E[y(t1)]$	0.1615	-3.3040	2.9774	0.000	2.6310	-3.2180	0.5008	0.000	0.1544
$E[y(t5)]-E[y(t1)]$	0.0948	-3.4619	3.5649	0.000	3.5649	-3.4619	0.0948	0.000	0.0948
	ETS	Worst-case		MTR		MTS		MTR+MTS	
Just a high school degree		LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t5)]-E[y(t4)]$	0.0522	-2.3248	2.8147	0.000	2.5522	-1.4682	0.3350	0.000	0.0725
$E[y(t4)]-E[y(t3)]$	0.0031	-3.0076	2.4060	0.000	1.5997	-2.4324	0.8416	0.000	0.0352
$E[y(t3)]-E[y(t2)]$	0.1608	-3.5470	3.4037	0.000	1.2487	-2.7677	2.3322	0.000	0.1772
$E[y(t2)]-E[y(t1)]$	0.0702	-3.8535	3.7018	0.000	1.0411	-3.1167	2.8887	0.000	0.2280
$E[y(t4)]-E[y(t1)]$	0.2341	-3.4574	2.5608	0.000	1.8451	-3.2169	0.9625	0.000	0.2467
$E[y(t5)]-E[y(t1)]$	0.2863	-3.6738	3.2671	0.000	3.2671	-3.6738	0.2863	0.000	0.2863
	ETS	Worst-case		MTR		MTS		MTR+MTS	
Some college		LB	UB	LB	UB	LB	UB	LB	UB
$E[y(t5)]-E[y(t4)]$	0.0254	-2.2027	2.5801	0.000	2.3710	-1.6665	0.2719	0.000	0.0629
$E[y(t4)]-E[y(t3)]$	0.1265	-2.9596	2.4120	0.000	1.5941	-2.4257	0.9642	0.000	0.1463
$E[y(t3)]-E[y(t2)]$	0.1786	-3.6908	3.5708	0.000	1.3648	-2.7454	2.4949	0.000	0.2888
$E[y(t2)]-E[y(t1)]$	0.1235	-3.9149	3.8315	0.000	1.2554	-3.0207	2.9757	0.000	0.3996
$E[y(t4)]-E[y(t1)]$	0.4286	-3.3037	2.5527	0.000	1.7939	-2.9516	1.1945	0.000	0.4357
$E[y(t5)]-E[y(t1)]$	0.4540	-3.6056	3.2320	0.000	3.2320	-3.6056	0.4540	0.000	0.4540

Table 8. Bounds on Wages for Some Covariates (ATE 41)

Assumptions		ETS	MTR		MTS		MTR+MTS	
			LB	UB	LB	UB	LB	UB
Years of residence	< 10 years	0.2433	0.00	3.1331	-3.0890	0.3433	0.00	0.2412
	10-20	0.2748	0.00	2.7667	-3.3166	0.4026	0.00	0.2732
	> 20	0.6062	0.00	2.4623	-3.1134	0.8261	0.00	0.6079
Gender	Male	0.4151	0.00	2.2650	-3.1255	0.9785	0.00	0.4262
	Female	0.3557	0.00	1.8927	-3.1763	1.0174	0.00	0.3583
Ancestry	Mexican	0.3419	0.00	2.2018	-3.1494	0.9037	0.00	0.3549
	Puerto Rican	0.2454	0.00	1.9296	-3.1166	0.7146	0.00	0.2399
	Cuban	0.4370	0.00	2.4512	-3.0624	0.7036	0.00	0.4405
	Central Am.	0.3969	0.00	2.8622	-3.1897	0.5822	0.00	0.3987
	South Am.	0.3282	0.00	2.5251	-3.2644	0.6134	0.00	0.3283
	Spain	0.6457	0.00	1.9486	-3.0096	1.5095	0.00	0.6410

Table 9. Wage Penalty Across Occupations under ETS

ATE	$E(w/t2) - E(w/t1)$	$E(w/t3) - E(w/t2)$	$E(w/t4) - E(w/t3)$
1. Managerial	0.3309	0.1183	0.1748
2. Technical	0.1068	0.1083	0.0398
3. Service	0.0764	0.0907	0.0114
4. Agriculture	0.0699	0.0611	-0.063
5. Repair	0.1712	0.191	0.0399
6. Operators	0.1573	0.1501	-0.0135

ATE	$E(w/t5) - E(w/t4)$	$E(w/t4) - E(w/t1)$	$E(w/t5) - E(w/t1)$
1. Managerial	0.048	0.6241	0.672
2. Technical	0.0494	0.2548	0.3043
3. Service	-0.0285	0.1785	0.15
4. Agriculture	-0.0109	0.0679	0.057
5. Repair	0.0764	0.4021	0.4785
6. Operators	0.0096	0.294	0.3036

Note: The occupational groups are defined as follows:

1. Managerial and professional specialty occupations.
2. Technical, sales, and administrative support occupations (technicians of all kinds, sales workers, secretaries, etc).
3. Service occupations.
4. Agricultural and related occupations (farmers, gardeners, fishers, hunters, etc).
5. Precision production, craft, and repair occupations.
6. Operators, fabricators, and laborers.

Table 10. Observed Unemployment Rates by EP and Occupation

Treatments	1	2	3	4	5	Total
<i>1. Managerial</i>	0.686	0.352	0.142	0.092	0.078	0.125
<i>2. Technical</i>	0.148	0.092	0.076	0.072	0.064	0.073
<i>3. Service</i>	0.089	0.077	0.080	0.089	0.102	0.087
<i>4. Agricultural</i>	0.164	0.146	0.158	0.161	0.161	0.157
<i>5. Repair</i>	0.102	0.089	0.082	0.087	0.077	0.085
<i>6. Operators</i>	0.117	0.101	0.107	0.117	0.100	0.109

REFERENCES

- Berman E, Lang K, Siniver E. 2003. Language-skill complementarity: returns to immigrant language acquisition. *Labour Economics* 10(3): 265-290.
- Bleakley H, Chin A. 2002. Language Skills and Earnings: Evidence From Childhood Immigrants. Unpublished draft. University of Chicago and University of Houston.
- Borjas GJ. 1991. Immigrants in the U.S. Labor Market: 1940-1980. *The American Economic Review* 81(2): 287-291.
- Carliner G. 1995. The Language Ability of US Immigrants: Assimilation and Cohort Effects. NBER Working Paper W5222.
- Carliner G. 1996. The Wages and Language Skills of US Immigrants. NBER Working Paper W5763.
- Chiswick BR, Miller PW. 1995. The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics* 13(2): 246-288.
- Dustmann C, Van Soest A. 2001. Language Fluency and Earnings: Estimation with Misclassified Language Indicators. *Review of Economics and Statistics* 83(4): 663-674.
- Grenier G. 1984. The Effects of Language Characteristics on the Wages of Hispanic-American Males. *The Journal of Human Resources* 19: 35-52.
- Kossoudji SA. 1988. English Language Ability and the Labor Market Opportunities of Hispanic and East-Asian Immigrant Men. *Journal of Labor Economics* 6(2): 205-228.
- Lazear EP. 1998. Diversity and Immigration. NBER Working Paper 6535.
- Manski C. 1990. Nonparametric Bounds on Treatment Effects. *American Economic Review Papers and Proceedings* 80: 319-323.
- Manski C. 1995. *Identification Problems in the Social Sciences*. Harvard University Press: Cambridge, MA.
- Manski C. 1997. Monotone Treatment Response. *Econometrica* 65: 1311-1134.
- Manski C, Nagin D. 1998. Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism. *Sociological Methodology* 28: 99-137.

- Manski C, Pepper J. 2000. Monotone Instrumental Variables: With an Application to the Returns to Schooling. *Econometrica* 68: 997-1010.
- Manski C, Sandefur S, McLanahan S, Powers D. 1992. Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation. *Journal of the American Statistical Association*: 25-37.
- McManus WS. 1985. Labor Market Costs of Language Disparity: an Interpretation of Hispanic Earnings Differences. *The American Economic Review* 75: 818-827.
- McManus W, Gould W, Welch F. 1983. Earnings of Hispanic Men: The Role of English Language Proficiency. *Journal of Labor Economics* 1(2): 101-130.
- Mora MT, Davila A. 1998. Gender, Earnings, and the English Skill Acquisition of Hispanic Workers in the U.S. *Economic Inquiry* 26: 631-644.
- Pepper J. 2000. The Intergenerational Transmission of Welfare Receipt: A Nonparametric Bounds Analysis. *Review of Economics and Statistics* 82(3): 472-478.
- Ramirez RR, De la Cruz GP. 2003. The Hispanic Population in the United States: March 2002. Current Population Reports. US Census Bureau.
- Reimers CW. 1983. Labor Market Discrimination Against Hispanic and Black Men. *The Review of Economics and Statistics* 65(4): 570-579.