Estimating Marginal Treatment Effects of High School Mathematics Credits on Income

Alfredo Sosa*

February 11, 2018

Abstract

In this paper I estimate marginal treatment effects of advanced high school mathematics credits on total labor income at age 28. Based on the work of Carneiro, Heckman and Vytlacil (2011) and Heckman, Tobias and Vytlacil (2001), I estimate the marginal returns to advanced math during high school. The results indicate that, the average gain for a randomly selected individual is larger than the gain for an individual who is already selected into the treatment. Also, most importantly, individuals who are not selected into the treatment are the ones who could potentially benefit the most by selecting themselves into the treatment.

I Introduction

Several studies have attempted to measure the impact of high school mathematics credits on labor market outcomes (Altonji, 1995; Levine & Zimmerman, 1995; Rose & Betts, 2004; Joensen & Nielsen, 2009; Goodman, 2012; Gaertner, Kim, DesJardins & McClarty, 2014; Kim, Kim, DesJardins & McCall, 2015). In all cases the starting point is the following equation:

$$Y_i = \alpha_0 + \beta Math_i + \mathbf{X}_i\gamma + \epsilon_i \tag{1}$$

If equation (1) is estimated via Ordinary Least Squares (OLS), the estimate of β is inconsistent since $Math_i$ is endogenous, i.e., it is correlated with ϵ_i. This literature has addressed this problem by using instrumental variables approaches. Therefore, these studies have provided estimates of the Local Average Treatment Effects ($LATE$) that are, by construction, instrument-specific (Angrist and Pischke, 2009). The problem with this approach as stated in Carneiro, Heckman and Vytlacil (2011) is that people induced by the instrument to change their choice might not be the same people that change their choice in response to a policy change; the returns might be different for these two groups.

*Alfredo Sosa, Ph.D. Higher Education and Economics - alfredosos@gmail.com

There is another problem that this literature has overlooked: β is assumed to be constant in the population, and, consequently, implicitly these studies assume that individuals act as if they don't know their idiosyncratic returns or, if they know, they do not use this information when choosing the optimal level of math credits. As pointed out by Carneiro, Heckman and Vytlacil (2011), selection on gains complicates the estimation of the impacts of education on earnings.

The idea of heterogeneous returns to education is not new to the literature. Card (2001) used heterogeneous returns to education to explain why the 2SLS estimates are generally larger than the corresponding OLS estimates. He concluded that the marginal returns to education among low-education groups tend to be relatively high, reflecting their high marginal cost of schooling.

A concept that is suited to measure heterogeneous effects is the Marginal Treatment Effect (MTE) which was introduced in the literature by Björklund and Moffitt (1987) and extended in Heckman & Vytlacil (1999, 2001, 2005, 2007). Moffitt (2008) estimated marginal treatment effects of higher education in the UK by using power series or splines. Carneiro, Heckman & Vytlacil (2011), provided two methods for estimating marginal treatment effects of college attendance in the US: the first is based on a normal selection model, and the second utilizes Local Instrumental Variables (LIV). In addition, Heckman and Li (2004) utilized marginal treatment effects to identify heterogeneous returns to college in China. For a comprehensive treatment of marginal treatment effects theory and applications please refer to Heckman and Vytlacil (2007).

By borrowing the methods in Carneiro, Heckman and Vytlacil (2011), and Heckman, Tobias and Vytlacil (2001), in this paper I estimate marginal treatment effects of obtaining advanced mathematics credits during high school on income. The treatment is a dummy variable that indicates whether or not individuals earned credits in advanced mathematics during high school. Following the definitions of NLSY 97, advanced mathematics include Algebra 2 through Pre-Calculus, Calculus, AP/IB and Advanced Mathematics-Other.

These methods allow me to estimate not only the marginal treatment effects (MTE), but also other relevant parameters commonly presented in the literature of returns to education such as the Average Treatment Effect (ATE), Treatment on the Treated (TT), and, Treatment on the Untreated (TUT). The results indicate that for each Carnegie unit of advanced mathematics earned during high school, the Average Treatment Effect (ATE) is about 4.4%, the Treatment on the Treated (TT) is about 2.45% and the Treatment on the Untreated (TUT) is about 7.39%. The Marginal

Treatment Effect (MTE) varies between -0.01% and 10%.

Interestingly, the results indicate that individuals who are already enrolled in advanced math courses would benefit the least from taking an additional year of advanced math courses. Also, individuals who are not enrolled in advanced math courses, would benefit the most from taking advanced math during high school. Finally, the results also indicate that obtaining an additional year of advanced math courses is beneficial for all individuals.

These results are consistent with Heckman, Tobias and Vytlacil (2001) who found that, for a randomly chosen individual, the average gain from obtaining some form of college is about 9%, whereas the average gain for those who actually select into college is about 4%. Even when they do not report TUT, implicitly these estimates need to be higher than 9%.

The paper is organized as follows. In section (II), I briefly describe the literature that examines the returns to high school math credits. In section (III), I describe the methods for estimating marginal treatment effects. Section (IV) is devoted to describing the data, the first stage and reduced form relationships and the parameter estimates of ATE, TT, TUT and MTE from the normal selection model. Conclusions are presented in section (V).

II Brief Review of the Literature on the Returns to High School Math Credits

All the studies that have attempted to provide a causal impact of high school mathematics credits on earnings utilize instrumental variables. Although these studies do not mention which parameter they are estimating, by using instrumental variables, they provide an estimate of the Local Average Treatment Effect or LATE (Angrist & Pischke, 2009). In other words, they gauge the impact of an additional unit of mathematics credits during high school on earnings *only* for the subgroup of individuals who are induced by the instrument to change their course-taking behavior.

In this study, I depart from the traditional IV framework approach utilized in the returns to math credits literature and, based on the work of Carneiro, Heckman and Vytlacil (2011) and Heckman, Tobias and Vytlacil (2001), I present four parameters of interest in the treatment evaluation literature: Average Treatment Effect (ATE), Treatment on the Treated (TT), Treatment on the Untreated (TUT) and Marginal Treatment Effect (MTE). The rationale of presenting the results of previous studies is to compare the magnitudes and place this study's estimates in context.

An important difference of this study with respect to other studies in the returns to high school math credits literature is that, instead of using a continuous treatment, I use a binary treatment. In this way, I can use the methods presented in Carneiro, Heckman and Vytlacil (2011) and Heckman, Tobias and Vytlacil (2000).

In this section, I briefly describe the results of the studies that examine the causal impact of high school mathematics credits on labor market outcomes[1]. For a complete review of the literature please refer to Sosa (2017a).

Altonji (1995) concluded that one more year of the combination of science, math, English, social studies and foreign language leads to an increase of wages of about 0.3%. Levine & Zimmerman (1995) found that the number of high school math courses does not impact individuals' weekly wages ten years after high school graduation. The magnitudes of the estimates was -0.017 for men and -0.060 for women; both statistically insignificant. In addition, Rose & Betts (2004) estimated that credits earned in algebra/geometry increased earnings by 8%. No statistically significant effects were found for intermediate algebra (-0.107), advanced algebra (-0.77) and calculus (-0.132).

Goodman (2012) calculated that each additional year of math increases black males' earnings between 5-9 %. The impact for white males is about the same magnitude but statistically insignificant. The results for women are small and statistically insignificant at about 0.035 for black women and 0.005 for white women. Finally, departing from the US context, Joensen & Nielsen (2009) concluded that taking advanced math credits coupled with advanced chemistry, increases earnings by about 20%.

In Sosa (2017a), I estimate that each additional Carnegie unit (year) of high school advanced mathematics increases total labor income by about 3%. For comparison purposes, I use Goodman (2012) and Sosa (2017a) as the references to benchmark the estimates found in this study. The reasons behind choosing these two studies are the following: first, they both provide convincing instruments; second, they depart from the traditional instrument proposed by Altonji (1992,1995) - the per-high school average number of math credits - and followed by Levine & Zimmerman (1995) and Rose & Betts (2004). Finally, Joensen & Nielsen's (2009) study was conducted in Denmark, and, thus, might not be relevant to the US context.

[1] Earnings or income

III Methods for Estimating Marginal Treatment Effects

In this section, I present the methods to identify and estimate four parameters of interest in the program evaluation literature: Average Treatment Effect (ATE), Average Treatment Effect on the Treated (TT), Average Treatment Effect on the Untreated (TUT), and, Marginal Treatment Effect (MTE). Specifically, I estimate these parameters in the context of the effects of high school math credits. To my knowledge, this is the first study of marginal treatment effects of high school mathematics credits on labor market outcomes.

Mostly, the methods presented in this paper are based on the study by Heckman, Tobias and Vytlacil (2001) who presented the aforementioned four parameters - ATE, TT, TUT and MTE in the context of returns to college education. The main takeaway on their methods is that these parameters can be estimated using the normal selection model. For the case of the Marginal Treatment Effect (MTE), I combine the methods of Heckman et al (2001), with those of Carneiro, Heckman and Vytlacil (2011). The rationale behind this choice is due to the fact that, in Carneiro et al. (2011), it is easier to interpret the MTE as a function of the propensity to receive the treatment in a support bounded between 0 and 1.

Since the treatment is binary, the starting point in Heckman et al. (2001) is the Generalized Roy Model. If individual i obtains a positive (> 0) number of advanced math credits during high school, then $D = 1$; $D = 0$ otherwise. The selection equation is:

$$D_* = Z\theta + U_D \quad where \quad D = 1[Z\theta + U_D \geq 0] \tag{2}$$

In equation (2), D_* represents a latent index of the propensity to obtain advanced math credits during high school. The support of D_* is $(-\infty, +\infty)$. The potential outcome equations are:

$$Y_1 = X\beta_1 + U_1 \quad and \quad Y_0 = X\beta_0 + U_0 \tag{3}$$

It is worth noting that equations (3) are *potential* outcome equations instead of outcome equations because we observe individuals only in one of the two states, $D = 1$ or $D = 0$ but not in both. For example, if individual i earned advanced math credits during high school, and hence, $D = 1$, then, Y_1 represents her measure of log income, and, Y_0 represents what would her log income be if she did not earn advanced math credits during high school ($D = 0$). Combining the outcome and treatment in the same equation we have $Y = DY_1 + (1 - D)Y_0$ for all individuals.

In equations (3) X is a matrix of variables that influence income assumed to be uncorrelated with (U_1, U_0). Also, in equation (2), Z includes some or all elements of X but also, it includes at least one variable that induces individuals into the treatment but is not included in any of the potential outcome equations. U_D is an unobserved scalar component that induces individuals into the treatment.

Define $\Delta \equiv Y_1 - Y_0$. The ATE Average Treatment Effect is defined as the expected gain from the treatment for a randomly chosen individual. The ATE conditional on x is:

$$ATE(x) = E[\Delta | X = x] = x(\beta_1 - \beta_0)$$

Assuming that the joint distribution of X is $F_X(X)$ then, the unconditional ATE is:

$$ATE = E[\Delta] = \int ATE(X)dF(X) \approx \frac{1}{n}\sum_{i=1}^{n} ATE(x_i) = \bar{x}(\beta_1 - \beta_0) \tag{4}$$

The Treatment on the Treated (TT) is the average gain from treatment for those who actually select into the treatment.

$$TT(x, z, D[z] = 1) = E[\Delta | X = x, Z = z, D[z] = 1]$$

In the previous equation $D[z] = 1$ indicates that the treatment is equal to 1 and also that it depends on the vector of instruments z.

$$TT(x, z, D[z] = 1) = x(\beta_1 - \beta_0) + E[U_1 - U_0 | U_D \geq -z\theta, X = x, Z = z]$$

Assume that (U_D, U_1, U_0) is independent from (X, Z) then

$$TT(x, z, D[z] = 1) = x(\beta_1 - \beta_0) + E[U_1 - U_0 | U_D \geq -z\theta]$$

Integrating over the joint distribution of (X,Z) conditional on $D = 1$ we have:

$$TT = E[\Delta|D[Z] = 1] = \int TT[X, Z, D[Z] = 1]dF(X, Z|D[Z] = 1)$$

Then, define n_t as the number of observations for which $D = 1$ then

$$TT \approx \frac{1}{n_t}\sum_{i=1}^{n} D_i TT[x_i, z_i, D[z_i] = 1] \tag{5}$$

Finally, the Marginal Treatment Effect, MTE is defined as the treatment effect for individuals with a given value of U_D.

$$MTE(x, u_D) = E[\Delta|X = x, U_D = u_D]$$

$$MTE(x, u_D) = x(\beta_1 - \beta_0) + E[U_1 - U_0|X = x, U_D = u_D]$$

$$MTE(x, u_D) = x(\beta_1 - \beta_0) + E[U_1 - U_0|U_D = u_D]$$

The unconditional (of X) version is

$$MTE(u_D) = \int MTE(X, u_D)dF(X) \approx \frac{1}{n}\sum_{i=1}^{n} MTE(x_i, u_D) \tag{6}$$

$$MTE(u_D) = \bar{x}(\beta_1 - \beta_0) + E[U_1 - U_0|U_D = u_D]$$

Treatment parameters assuming joint normality

Similar to Heckman, Tobias and Vytlacil (2001) assume joint normality of (U_D, U_1, U_0)

$$\begin{bmatrix} U_D \\ U_1 \\ U_0 \end{bmatrix} \sim N\left[0, \begin{bmatrix} 1 & \sigma_{1D} & \sigma_{0D} \\ \sigma_{1D} & \sigma_1^2 & \sigma_{10} \\ \sigma_{0D} & \sigma_{10} & \sigma_0^2 \end{bmatrix}\right]$$

The variance parameter in the selection equation has been normalized to unity. The estimation procedure, according to Heckman, Tobias and Vytlacil (2001) is as follows:

1. Obtain $\hat{\theta}$ from a probit model on the decision to take the treatment, i.e.,estimate equation 2 using a probit model.

$$D = 1[Z\theta + U_D \geq 0]$$

2. Compute the appropriate selection correction term evaluated at $\hat{\theta}$ (i.e., $\widehat{\lambda_{1i}} = \phi[Z_i\hat{\theta}]/\Phi[Z_i\hat{\theta}]$ when $D_i = 1$ and $\widehat{\lambda_{0i}} = \phi[Z_i\hat{\theta}]/(1 - \Phi[Z_i\hat{\theta}])$ when $D_i = 0$). For each individual in the sample calculate either $\widehat{\lambda_{1i}}$ or $\widehat{\lambda_{0i}}$.

3. Run treatment-outcome specific regressions for the groups $[i : D_i = 1]$ and $[i : D_i = 0]$ with the inclusion of the appropriate selection-correction term obtained from the previous step.

$$Y_1 = X\beta_1 + \sigma_{1D}\widehat{\lambda_1} + U_1$$

$$Y_0 = X\beta_0 + \sigma_{0D}\widehat{\lambda_0} + U_0$$

4. Given $\hat{\beta}_1$, $\hat{\beta}_0$, $\widehat{\sigma_{1D}}$ and $\widehat{\sigma_{0D}}$ obtained from step 3, and $\hat{\theta}$ from step 1, calculate the following equations:

Conditional ATE

$$ATE(x_i) = x_i(\hat{\beta}_1 - \hat{\beta}_0)$$

Unconditional ATE

$$ATE = \bar{x}(\hat{\beta}_1 - \hat{\beta}_0) \tag{7}$$

$$ATE = \frac{1}{n}\sum_{i=1}^{n} ATE(x_i) \tag{8}$$

Conditional TT

$$TT[x_i, z_i, D[z_i] = 1] = x_i(\widehat{\beta}_1 - \widehat{\beta}_0) + (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})\frac{\phi(z_i\widehat{\theta})}{\Phi(z_i\widehat{\theta})}$$

Unconditional TT

$$TT = \frac{1}{n_t}\sum_{i=1}^{n} D_i TT[x_i, z_i, D[z_i] = 1] \tag{9}$$

Conditional TUT

$$TUT[x_i, z_i, D[z_i] = 0] = x_i(\widehat{\beta}_1 - \widehat{\beta}_0) - (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})\frac{\phi(z_i\widehat{\theta})}{(1 - \Phi(z_i\widehat{\theta}))}$$

Unconditional TUT

$$TUT = \frac{1}{n - n_t}\sum_{i=1}^{n}(1 - D_i) TUT[x_i, z_i, D[z_i] = 0] \tag{10}$$

Conditional MTE

$$MTE[x_i, U_D = u_D] = x_i(\widehat{\beta}_1 - \widehat{\beta}_0) + (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})u_D \tag{11}$$

Unconditional MTE

$$MTE[U_D = u_D] = \bar{x}(\widehat{\beta}_1 - \widehat{\beta}_0) + (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})u_D \tag{12}$$

Since the Heckman, Tobias and Vytlacil (2001) study does not present charts of marginal treatment effects, for comparison purposes I use the same unobserved component of the propensity to participate in the treatment (U_S) as in Carneiro, Heckman and Vytlacil (2011). Transforming (11) to correspond to Carneiro, Heckman and Vytlacil (2011). Under Carneiro et al (2011), in the choice model I_S is defined as the net benefit from choosing $S = 1$[2]. The latent variable I_S is a function of a vector of observable characteristics ($\mathbf{Z}$) and an unobserved scalar component V in the following way:

$$I_S = \mu_S(\mathbf{Z}) - V \tag{13}$$

[2] In Carneiro et al. (2011) S represents a binary treatment of whether or not individuals were ever enrolled in college.

Thus, we observe

$$S = 1 \quad if \quad I_S \geq 0; \quad S = 0 \quad otherwise \tag{14}$$

Similarly to Heckman, Tobias and Vytlacil (2001) each choice is associated with a potential outcome equation:

$$Y_1 = \mu_1(\mathbf{X}) + U_1 \quad and \quad Y_0 = \mu_0(\mathbf{X}) + U_0 \tag{15}$$

Again, in equations (15), $\mathbf{X}$ is a vector of observed characteristics that influence income. In equation (13), $\mathbf{Z}$ is a vector of observable characteristics which might include some (or all), elements of $\mathbf{X}$ but also includes at least one instrument, i.e., a variable that induces individuals to change their advanced math credits but does not belong to any of the potential outcome equations (15).

In the net benefit equation, (13), V is assumed to be a continuous random variable with a cumulative distribution function F_V. The rationale behind this transformation is the following: whereas the support of V is $(-\infty, +\infty)$, the support of $F_V(V)$ is $[0, 1]$. This transformation allows working with a more mathematically tractable variable $F_V(V)$ instead of V. Define $U_S = F_V(V)$, then, the MTE is defined as:

$$MTE(x, u_s) \equiv E[\beta|\mathbf{X{=}x}, U_s = u_s] \tag{16}$$

According to Carneiro et al. (2011), in their context "the marginal treatment effect is the mean return to schooling for individuals with characteristics $\mathbf{X}$ and $U_S = u_S$." (pp. 2727). In this paper, the marginal treatment effect is interpreted as the average return of having earned advanced math credits during high school for persons who are indifferent between obtaining or not advanced math credits.

By comparing equations (2) and (13) we know that $U_D = -V$. Thus, equation (12) becomes

$$MTE[V = v] = \bar{x}(\widehat{\beta}_1 - \widehat{\beta}_0) - (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})V \tag{17}$$

Similar to Carneiro et al (2011), I use the transformation $U_S = F_V(V)$ to be able to work with a more tractable variable U_S instead of V. Finally, the marginal treatment effect formula that I use in this study is:

$$MTE[U_S = u_S] = \bar{x}(\widehat{\beta}_1 - \widehat{\beta}_0) - (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})F^{-1}(u_S) \tag{18}$$

Implications of the joint normality assumption

The estimates generated in this study depend on whether or not the joint normality assumption holds in the population. Since the error terms are by construction unobserved, no test can tell us whether or not the assumption of joint normality holds. Even when this is a strong assumption, it is also true that this is the first paper that explores marginal treatment effects in the context of returns to high school math credits. In further versions of the paper, I am planning to estimate marginal treatment effects by the semiparametric method of Local Instrumental Variables as presented in Carneiro, Heckman and Vytlacil (2011). For now, this paper is the first step into the integration of the Marginal Treatment Effect framework and the returns to high school math credits literature.

IV Estimated Marginal Treatment Effects

A Data

Treatment, Outcome and Controls

The individual-level data for this paper were drawn from the National Longitudinal Survey of Youth 1997 (NLSY 97) which is a nationally representative sample of around 9,000 individuals who were 12 to 16 years old as of 12/31/1996. This survey allows researchers to link individual choices such as high school mathematics course taking behavior to labor market outcomes later in life. NLSY 97 includes transcript information for about 70 percent of all individuals in the sample. Transcript information has been made homogeneous across all schools and years via the Carnegie units system.

According to NCES, one Carnegie unit is defined as the number of credits a student receives for a course taken every day, one period per day, for a full school year[3]. In addition to high school Carnegie units, NLSY 97 also includes information about labor market outcomes such as employment, income, earnings and wages. The dependent variable or outcome in this paper is

[3] https://nces.ed.gov/nationsreportcard/glossary.aspx?nav=y

the total labor income, which is obtained by asking "During last year, how much income did you receive from wages, salary, commissions, or tips from all jobs, before deductions for taxes or anything else?". Since this question is asked repeatedly, I constructed a measure of income, in 2011 real USD, centered around age 28. I chose 28 because all individuals in the sample turn 28 within the observation period.

In this paper I define advanced mathematics (*advanced_math*) as a dummy variable equal to 1 if individual i obtained credits in advanced mathematics as defined by NLSY 97; that is, in any of the following courses: Algebra 2 through Pre-Calculus, Calculus, AP/IB and Advanced Mathematics-Other. As controls, in all the models, I included: a dummy for female, a dummy for white, age in years as of 12-31-1996 (age at the beginning of the study), the average of non-missing values of household gross income between 1996 and 1999 (in 1997 real USD), the average of non-missing values of household income to poverty ratio between 1996 and 1999, household size in 1997, a dummy that indicates whether the household had both biological parents in 1997, state-level number of Carnegie units (years) of math required to obtain a high school diploma in 1997, number of years of exposure to a high school math reform[4]. State and cohort[5] fixed effects were also included.

The sample was restricted to individuals who ever enrolled in high school and who had non-missing values of advanced mathematics, total labor income at age 28 as well as in all the aforementioned controls. A total of 4,219 individuals were included in the analysis sample.

Instrument / Exclusion Restrictions

In the formulation presented in section (III), the matrix Z includes some or all elements in X but also includes at least one instrument, i.e., a variable that induces individuals to modify their course taking choices but does not otherwise affect outcomes. The instruments in this paper are state-level measures of shocks to teacher labor supply. Specifically, as instruments, I use the implementation of state-level financial incentive programs that induce current or future teachers

[4]Reform indicates whether a state changed the number of years of mathematics required for high school graduation between 1995 and 2005. Exposure to math reforms was based on the first year of high school enrollment using the following rules: if the reform year occurred before first year of enrollment, exposure to math reforms is equal to 4; if the reform year occurred after the last year of high school enrollment, exposure to math reforms is equal to zero; finally, if the reform year occurred in between the first and last years of enrollment, exposure to math reforms equals the difference between the last year of high school enrollment and reform year.

[5]Cohort is defined as the year individuals entered high school.

and/or recent college graduates into teaching in shortage subject areas such as math and science, and/or geographic critical shortage areas (e.g., schools located in high-poverty neighborhoods). Throughout this paper I call these programs STEM Programs. For a comprehensive description of the STEM programs please refer to Sosa (2017b).

For clarity, I call *instruments* to individual-level measures of exposure to STEM programs. When the instruments are *interacted* with state dummies, I call these variables *exclusion restrictions*. The main instrument utilized in this paper is the number of years of potential exposure to STEM programs while individuals were enrolled in high school; this variable is called **expo**.

To construct expo, first, for each individual, I calculated the first and last years of potential enrollment in high school; the first year of enrollment is equal to the birth year plus 17 and the last year of enrollment is the first year of enrollment plus 3. In this way, the instrument does not depend on *actual* enrollment which is endogenous but only on potential enrollment which depends on the year individuals were born. For each individual, I created a row vector - $enrollment_{is}$ - which is a 1X34 vector with 0's in the years of no enrollment and ones in the years of potential enrollment.

Also, for each state, I constructed a row vector A_s with ones on the years state s had at least one active STEM program, and 0's on the years state s did not have an active STEM program. The 34 columns refer to all the years between 1983 to 2016 utilized in the STEM programs data. Thus, expo was calculated by the following formula: $expo = enrollment_{is} * A'_s$.

In addition to *expo*, I also used other STEM program characteristics such as number of recipients, expenditures, recipients per 1,000 teachers, expenditures per teacher, recipients per 1,000 secondary teachers, expenditures per secondary teacher, recipients per 1,000 students, expenditures per student, recipients per 1,000 high school students, and, expenditures per high school student.

Let $stem_s$ be any of the mentioned STEM program characteristics. For the sake of explanation, let $stem_s$ be $expo\_recipients_s$ which is a 1X34 row vector with zeros in the years state s did not have any STEM program, and, the average number of recipients on the years state s did have at least one active STEM program. Thus, to calculate potential exposure to all STEM program characteristics I used the following formula: $expo\_stem_s = enrollment_{is} * stem'_s$.

To abbreviate, for the remaining of the paper, I call exposure to STEM program characteristics in1,...,in11. Please refer to table (1) for the definition of all the instruments in1,...,in11. Please refer to Sosa (2017b) to learn more about the different STEM Program characteristics like program type (e.g., loan forgiveness, scholarship, etc), program focus (e.g., math and science, shortage geographic areas, minorities) as well as the magnitude of the programs in terms of the number of participants and expenditures. I also provide maps that capture the variation across states regarding the aforementioned STEM program variables. Finally, in table (1), I define all the variables included in this paper.

Table 1: Definitions of the Variables Used in the Empirical Analysis

Variable	**Definition**
Y	Log of total labor income at age 28
D=1	if respondent obtained high school advanced math credits; 0 otherwise.
X	Female, white, age in years as of 12-31-1996, average household gross income between 1996 and 1999, average household income to poverty ratio between 1996 and 1999, household size in 1997, household both biological parents in 1997, state-level years of math required to obtain a high school diploma in 1997, years of exposure math reforms.
Instruments	
in1	Years of potential exposure to STEM programs during high school.
in2	Years of potential exposure to STEM program recipients.
in3	Years of potential exposure to STEM program expenditures.
in4	Years of potential exposure to STEM program recipients per 1,000 teachers.
in5	Years of potential exposure to STEM program expenditures per teacher.
in6	Years of potential exposure to STEM program recipients per 1,000 secondary teachers.
in7	Years of potential exposure to STEM program expenditures per secondary teacher.
in8	Years of potential exposure to STEM program recipients per 1,000 students.
in9	Years of potential exposure to STEM program expenditures per K-12 student.
in10	Years of potential exposure to STEM program recipients per 1,000 high school students.
in11	Years of potential exposure to STEM program expenditures per high school student.

All the analyses utilize the same controls. The exclusion restrictions are the interactions of the instruments and state dummies. State and cohort fixed effects are also included.

B Sample Characteristics

The NLSY 97 sample includes 8,984 individuals of which 6,120 have transcript information. After dropping records with missing values of income the resulting sample included 4,841 individuals. By dropping missing observations on the following variables: average household gross income between 1996 and 1999 (545 observations), average household income to poverty ratio between 1996 and 1999 (17 observations), and the number of years of math required to obtain a high school diploma (60 observations), the final sample size for all the analyses is 4,219. None of the remaining controls have missing values.

Table (2) includes descriptive statistics of the variables employed in the empirical analyses. The analysis sample consists of N=4,219 individuals of whom 50% are women; about 73% are white with an average age at the beginning of the study of about 14.68 years. The average household gross income between 1996 and 1999 is about $56,141, and, the household income to poverty ratio, also between 1996 and 1999, is about 3.56. The average household size in 1997 was about 4.36 and, 55 percent of the households had both biological parents. All the summary statistics are weighted.

On average, the number of Carnegie units of advanced math credits is about 1.02, and 59% of the sample obtained advanced math credits. The mean income (total labor income) at age 28 is about $26,894. Moreover, the average years of potential exposure during high school is 2.14. It is noteworthy that potential exposure to STEM programs varies greatly from 0 to 4 with a standard deviation of 1.82 years.

C First Stage and Reduced Form

One condition for identification of marginal treatment effects is the existence of at least one variable that influences individuals choices without being included in the potential outcome equations. The exclusion restrictions $(Z|X)$ are measures of potential exposure to STEM program characteristics (e.g., *expo*) interacted with state dummies.

In table (3), I estimated separate OLS regressions in which the dependent variables are, in column (1), advanced math, and, in column (2), the log of total labor income. I only present estimates of the first stage and reduced form relationships for the first instrument (*expo*). For more information about the $F - statistics$ regarding the association between $Z|X$ and advanced math,

Table 2: Summary Statistics of Analysis Sample N= 4,219

Variable	**Mean**	**Sd**	**Min**	**Max**
Treatment, Outcome, Instrument				
Advanced math	1.02	1.11	0	7.50
Advanced math (binary)	0.59	0.49	0	1
income	$26,894	$19,194	$2.70	$128,535
expo	2.14	1.82	0	4
Demographics				
female	0.50	0.50	0	1
white	0.73	0.44	0	1
age as of 12-31-1996	14.68	1.1	13	16
Family				
household gross income 1996-1999	$56,141	$46,039	$233	$417,074
household income to poverty ratio 1996-1999	3.56	3.05	0.01	32.27
household size 1997	4.36	1.42	2	16
both bio parents in household	0.55	0.50	0	1

All the means are calculated using the 1997 weight. The analysis sample includes individuals who ever enrolled in high school who also have transcript information and whose income at age 28 is non-missing. State and cohort fixed effects are also included in the analyses but not shown in this table.

please refer to Sosa (2017a). In all cases (instruments), the $F - statistics$ are much bigger than the rule of thumb of 10.

As table (3) indicates, the impact of *expo* on treatment and outcome vary greatly by state and it is positive and statistically significant in most cases. For space limitations, I present a trimmed version of the first stage and reduced form table; the complete table is located in the appendix.

Table 3: **First Stage and Reduced Form**. Impact of interactions of exposure to STEM programs and state dummies on advanced math and log income. Please refer to the appendix for the complete table.

	(1) advanced math	(2) ln(income)
expo*state 4	0.115***	0.707***
	(0.0202)	(0.0531)
expo*state 5	0.293***	1.502***
	(0.0335)	(0.0747)
⋮	⋮	⋮
expo*state 50	0.0836***	0.353***
	(0.0193)	(0.0391)
expo*state 51	0.304***	1.159***
	(0.0333)	(0.0766)
N	4,219	4,219

State-level cluster robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. OLS regressions in which the dependent variable are for column 1, a dummy variable that indicates whether students earned a positive number of credits in advanced math, and 0 otherwise. In column 2, the dependent variable is the natural logarithm of income at age 28. State and cohort fixed effects were also included. All regressions use the 1997 weight.

D Parameter estimates from the normal selection model

As mentioned in section (III), in order to estimate ATE, TT, TUT and MTE, under the assumption of joint normality of the error terms, I estimated an OLS regression for each group $D = 1$ and $D = 0$ with the corresponding sample correction term. Table (4) includes the parameter estimates of such regressions.

The important contribution of table (4) is that, the parameter estimates of β_1, β_0, σ_{1D} and σ_{0D} along with θ are the basis for calculating ATE, TT, TUT and MTE using the formulas presented in section (III). For instance, to calculate ATE, I first calculated $ATE(x_i) = x_i(\hat{\beta}_1 - \hat{\beta}_0)$ for all the members in the sample. Next, I calculated the average across all individuals.

Similarly, to calculate TT, first, I calculated $TT[x_i, z_i, D[z_i] = 1] = x_i(\widehat{\beta_1} - \widehat{\beta_0}) + (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})\frac{\phi(z_i\widehat{\theta})}{\Phi(z_i\widehat{\theta})}$, or equivalently, $TT[x_i, z_i, D[z_i] = 1] = x_i(\widehat{\beta_1} - \widehat{\beta_0}) + (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})\lambda_{1i}$ for each individual with $D = 1$, and, next, I calculated the average across all treated individuals. For the TUT, I calculated $TUT[x_i, z_i, D[z_i] = 0] = x_i(\widehat{\beta_1} - \widehat{\beta_0}) - (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})\frac{\phi(z_i\widehat{\theta})}{(1-\Phi(z_i\widehat{\theta}))}$ for all individuals with $D = 0$, and, next, I averaged it across all the untreated.

In order to calculate MTE, first, I generated a grid of U_S ranging from 0.05 to 0.95 in intervals of 0.05. Next, for each i and U_S, I calculated $MTE[x_i, U_s = u_s] = x_i(\widehat{\beta_1} - \widehat{\beta_0}) - (\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})\Phi^{-1}(u_s)$.

Table 4: **Parameter Estimates - Normal Sample Selection Model**

	(1) $E[Y|D=1, X, lamda1]$	(2) $E[Y|D=0, X, lamda0]$
female	-0.297***	-0.603***
	(0.0545)	(0.0565)
white	-0.0295	0.135
	(0.0682)	(0.0881)
age as of 12-31-1996	-0.0296	0.0239
	(0.130)	(0.0916)
hh gross inccome 1996-1999	1.56e-07	-1.74e-06
	(2.34e-06)	(2.06e-06)
hh income to poverty ratio	0.0181	0.0483
1996-1999	(0.0261)	(0.0299)
household size 1997	0.0101	0.0295
	(0.0285)	(0.0216)
both biological parents	-0.0238	-0.0600
	(0.102)	(0.0992)
expo reforms	-0.0708	-0.0294
	(0.0755)	(0.0725)
high school math grad req 1997	-0.00887	-0.119
	(0.209)	(0.199)
λ_1	-0.236	
	(0.404)	
λ_0		0.492*
		(0.292)
Constant	10.79***	9.717***
	(1.497)	(1.270)
N	2,354	1,859
R^2	0.060	0.165

Column 1 includes the parameter estimates of the OLS regression of log income on X and λ_1 conditional on $D = 1$. Column 2 includes the parameter estimates of the OLS regression of log income on X and λ_0 conditional on $D = 0$.

This step resulted in 19 columns: MTE_{05}, MTE_{10},...,MTE_{95}. Finally, the graph (figure (1)) showing $MTE(U_S)$ is generated averaging each of the 19 columns across all members of the sample. The 90% standard errors were obtained by bootstrapping this process 50 repetitions[6].

Similar to Carneiro, Heckman and Vytlacil (2011) who divided by their estimates by four in order to measure the average gain of one year of college, I also divided the estimates in this paper by 20 since most states require 20 units to obtain a high school diploma. By doing so, I assume that all of the credits are concentrated in advanced mathematics.

It is important mentioning the parameter estimates of λ_1 and λ_0 in table (4). These parameters are estimates of the covariances between the error terms in the outcome equations (Equation 3) and the error term in the selection equation (Equation 2) in Heckman, Tobias and Vytlacil (2001). In particular $\sigma_{1D} = Cov(U_1, U_D)$ and $\sigma_{0D} = Cov(U_0, U_D)$. Since U_D is positive in equation (2)), a higher value of U_D leads to a higher propensity to select into the treatment. Now, the fact that $\widehat{\sigma}_{1D} < 0$ indicates that, for individuals who select into the treatment, an increase in U_D is associated with a decrease in U_1, an unobserved component of income. Furthermore, $\widehat{\sigma}_{0D} > 0$ also indicates that, for individuals who do not select themselves into the treatment, an increase in U_D is associated with an increase in U_0; in other words, for individuals who are not taking advanced math courses, increasing the propensity to obtain advanced math credits is associated with a higher unobserved component of income.

E Estimates of the Returns to Advanced Mathematics Credits: ATE, TT, TUT, MTE.

The main results of the paper are presented in this section. These results were obtained using the first instrument: *expo*. Table (5) includes the estimates of the returns to advanced math credits for different groups. For each Carnegie unit of advanced math, for a random individual in the population, the Average Treatment Effect (ATE) is about 4.47%. For individuals who actually obtained advanced math credits the effect (TT) is about 2.45%. Finally, for the group who did not earn advanced math credits, the average effect (TUT) is about 7.39%.

The results are consistent with those in Heckman, Tobias and Vytlacil (2011) who examined

[6] According to the Stata Bootstrap Manual, N=50 is an adequate number of repetitions for normal approximation confidence intervals.

the impact of college education on log wages. They found that for a randomly chosen person, receiving some form of higher education leads to an increase in hourly wages of about 9%. For individuals who select into college, the average gain is about 4%. They conclude that "individuals with unobservables making them most likely to enroll in college receive the smallest return to a college education", Heckman, Tobias and Vytlacil (2001, p.221).

In addition, the results presented in this study are consistent with other studies in the literature of returns to high school math credits. For example, Goodman (2012) found that the returns to math credits for males varied between 5% and 9%. In Sosa (2017a), I found that one Carnegie unit of advanced math increases total labor income by about 3%. Let's recall that both Goodman (2012) and Sosa (2017a) present estimates of weighted Local Average Treatment Effects (LATE's). Unlike all other studies in the literature of returns to high school math credits, this is the first study that provides measures of different estimands other than weighted LATE's.

Table 5: Returns to one credit of advanced math

$ATE = E(\beta)$	0.0447
	(0.0002345)
$TT = E(\beta \mid D = 1)$	0.0245
	(0.000387)
$TUT = E(\beta \mid D = 0)$	0.0739
	(0.0196)

The parameters ATE, TT and TUT are calculated according to the formulas presented in this paper which are based on Heckman, Tobias and Vytlacil (2001).

An important concern is whether the average gain to advanced math credits is constant in the population. If this is the case, then individuals either do not know their idiosyncratic returns to math credits, or if they do know, this information does not play any role when choosing whether or not to take advanced math during high school. When individuals act on the information about their idiosyncratic returns to math credits, selection on gains complicates the estimation of returns to math credits. On top of the endogeneity problem characterized by the correlation between $Math_i$ and ϵ_i in equation (1), we also have a selection on gains problem because of the correlation between $Math_i$ and β.

The framework presented in this study allows testing whether or not β is constant in the population. This is accomplished by testing the null hypothesis of equality of covariances: H_0 : $\sigma_{1D} = \sigma_{0D}$. Table (6) includes the results of this test. By bootstrapping the difference in covariances $(\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}})$ 250 times, I calculated the standard error. The conclusion is that I cannot reject H_0.

Even when we cannot reject the null hypothesis of equality of covariances, and consequently, constancy of MTE, the p-value (0.133) is very close to 0.1, therefore we cannot conclude that MTE is constant. Heckman and Vytlacil (2007) summarize the evidence regarding the constancy of MTE in different contexts. In their analysis, studies such as Lee (1978), Farber (1983) and Duncan and Leigh (1985) do not reject the null hypothesis of equality of covariances in the normal selection model. Most studies, however, reject the null hypothesis, and conclude that MTE is not constant.

Table 6: Test of Constancy of MTE. $H_0 : \sigma_{1D} = \sigma_{0D}$

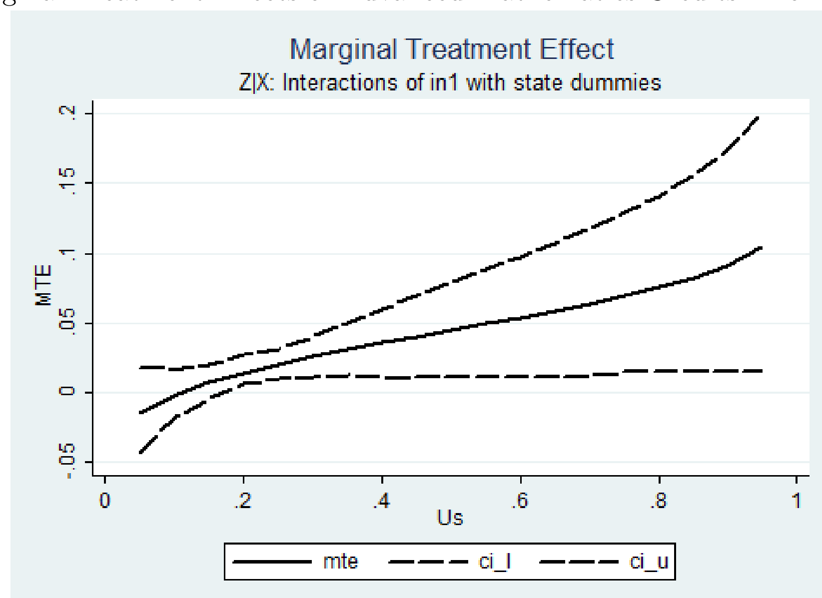
$\widehat{\sigma_{1D}} - \widehat{\sigma_{0D}}$	-0.7273
Bootstrap standard error	(0.485)
$P - value$	0.133
Decision	Cannot Reject H_0

The standard error was calculated bootstrapping 250 repetitions.

Figure (1) shows the mapping of the marginal treatment effect onto the unobserved propensity to receive the treatment, U_S. The 90% confidence intervals were estimated by bootstrapping 50 repetitions. Since I transformed the unobserved component of the propensity to receive the treatment from U_D to U_S, higher values of U_S decrease the propensity to receive the treatment, and viceversa, lower values of U_S increase the propensity to receive the treatment.

Figure (1) reflects some important features. First, the average gain varies across the distribution of the propensity to select into the treatment. On the one hand, individuals with high values of U_S, those who are less likely to participate in the treatment, are the ones who gain more. On the other hand, individuals with low values of U_S; i.e., those with high propensity to obtain advanced math credits are the ones who benefit the least. Second, the average gain for individuals who are indifferent between $D = 1$ and $D = 0$ varies between -0.05% and 10%. For individuals at the margin the effects can be substantial, especially for those with high values of U_S.

Figure 1: Marginal Treatment Effects of Advanced Mathematics Credits - Normal Selection Model



Notes: The 90% confidence intervals were calculated using bootstrap. The exclusion restrictions $Z|X$ are the interactions of in1 (years of exposure to STEM programs) and state dummies. State and cohort fixed effects were also included in the outcome and selection equations.

F Sensitivity Analysis

In this section I estimate three parameters of interest, ATE, TT, TUT and MTE using the entire set of instruments (one by one) defined at the beginning of this section. As table (7) indicates, the ATE ranges from 3.35% to 4.77%. Also, the TT varies between 1.95% and 2.67%, and, finally, the TUT varies between 5.37% and 7.90%.

Table 7: Returns to one credit of advanced math. All instruments.

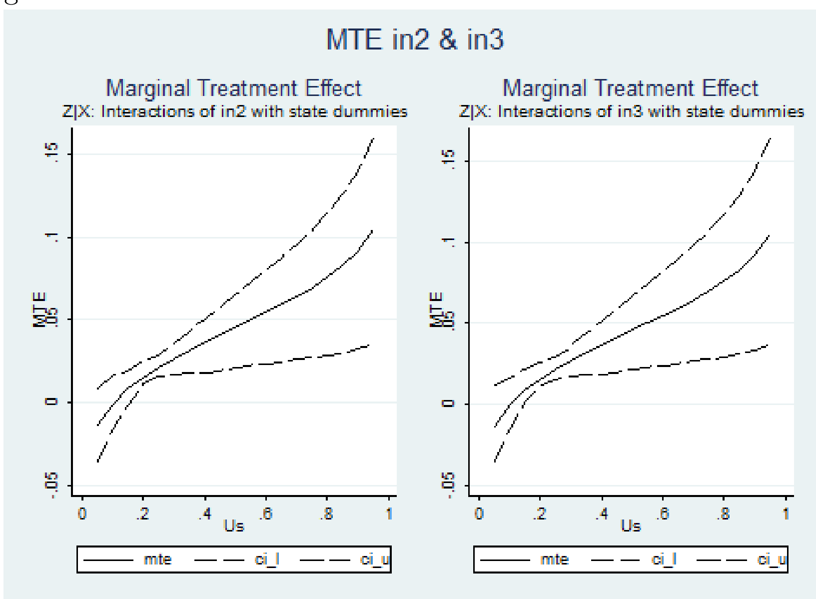
Instrument	**Definition: exposure to**	**ATE**	**TT**	**TUT**
in1	years of STEM programs	0.0447	0.0245	0.0739
		(0.000235)	(0.000387)	(0.000456)
in2	Recipients	0.0456	0.0266	0.0729
		(0.000239)	(0.000414)	(0.0215)
in3	Expenditures	0.0456	0.0267	0.0727
		(0.000240)	(0.000416)	(0.000502)
in4	Recipients/1,000 teachers	0.0335	0.0195	0.0537
		(0.000237)	(0.00034)	(0.000373)
in5	Expenditure/teacher	0.0378	0.0218	0.0609
		(0.000235)	(0.000359)	(0.000407)
in6	Recipients/1,000 sec teachers	0.0470	0.0254	0.0781
		(0.000235)	(0.0004004)	(0.000481)
in7	Expenditures/sec teacher	0.0428	0.0241	0.0696
		(0.000235)	(0.000384)	(0.000452)
in8	Recipients/ 1,000 students	0.0460	0.0260	0.0748
		(0.000237)	(0.000407)	(0.000490)
in9	Expenditure/student	0.0436	0.0244	0.0712
		(0.000235)	(0.000387)	(0.000456)
in10	Recipients/1,000 HS students	0.0477	0.0260	0.0790
		(0.000236)	(0.000408)	(0.00050)
in11	Expenditure/HS student	0.0448	0.0258	0.0721
		(0.000237)	(0.000404)	(0.000483)

The parameters ATE, TT and TUT are calculated according to the formulas presented in this paper which are based on Heckman, Tobias and Vytlacil (2001).

The small variation in ATE, TT and TUT across the different instruments is remarkable. In contrast to Sosa's (2017a) study in which the variation in weighted LATE's across the first five instruments is very large, in this study, the parameters estimates are more robust to the selection of the instrument. This is not surprising since the weighted LATE estimates are instrument dependent (Angrist & Pischke, 2009) whereas the Marginal Treatment Effects framework presented here helps to dissociate the instruments from the parameter estimates. Thus, it is a more robust framework to examine the gains from advanced math courses.

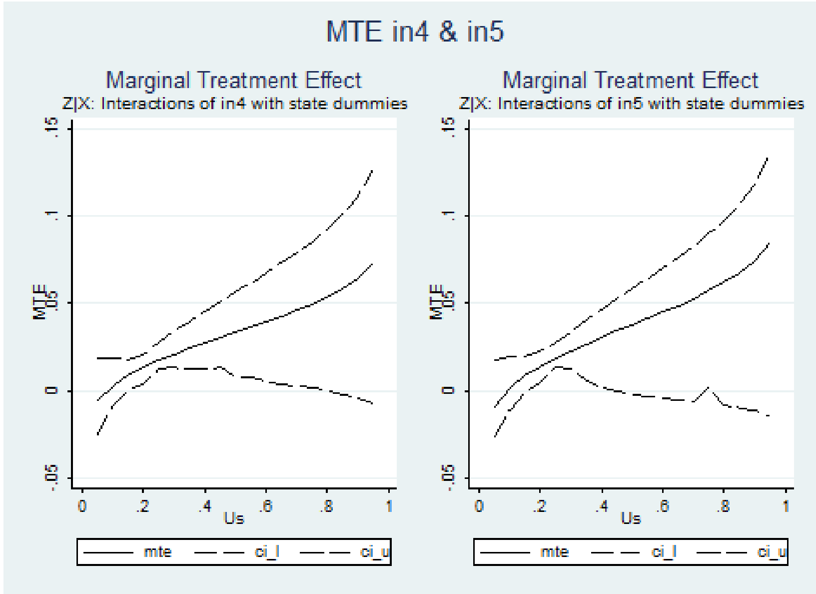
In figures (2) to (6), I estimated the marginal treatment effects using instruments 2 through 11. All the graphs follow the same pattern as shown in figure (1). Regardless of the instrument utilized, the average gain of advanced math credits is not constant in the population. In fact, it is a increasing function of U_S which implies that, for those with high values of U_S, and, hence, those who are less likely to earn advanced math credits, the gains are high whereas for those with low values of U_S, i.e., those more likely to earn advanced math credits, the returns are close to zero.

Figure 2: Marginal Treatment Effects of Advanced Mathematics Credits: Instruments n2 and n3



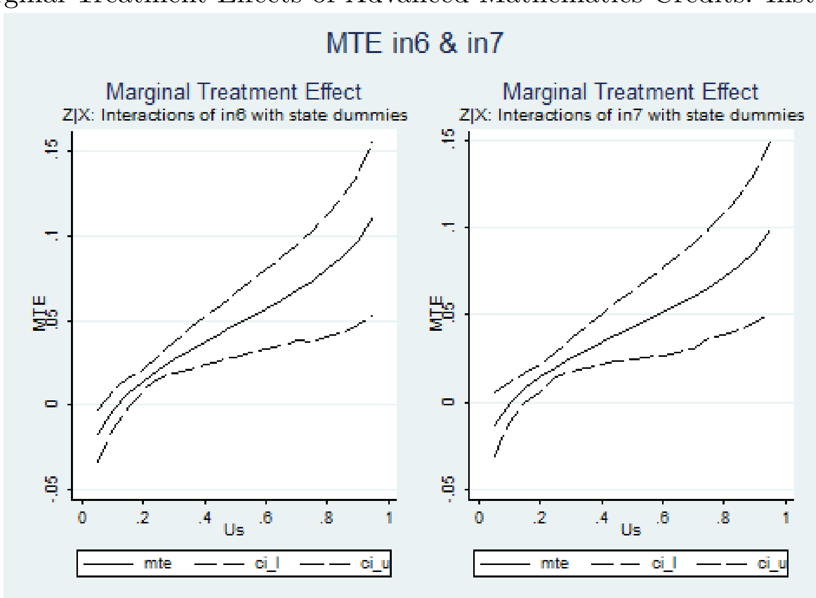
Notes: The 90% confidence intervals were calculated using bootstrap. The exclusion restrictions $Z|X$ are the interactions of in2 & in3 respectively and state dummies. State and cohort fixed effects were also included in the outcome equations and in the selection equation.

Figure 3: Marginal Treatment Effects of Advanced Mathematics Credits: Instruments n4 and n5



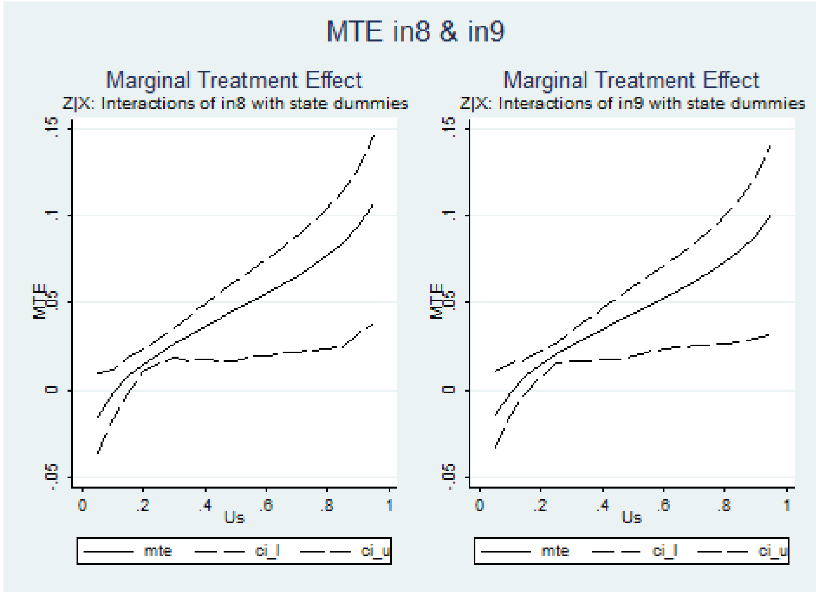
Notes: The 90% confidence intervals were calculated using bootstrap. The exclusion restrictions $Z|X$ are the interactions of in4 & in5 respectively and state dummies. State and cohort fixed effects were also included in the outcome equations and in the selection equation.

Figure 4: Marginal Treatment Effects of Advanced Mathematics Credits: Instruments n6 and n7



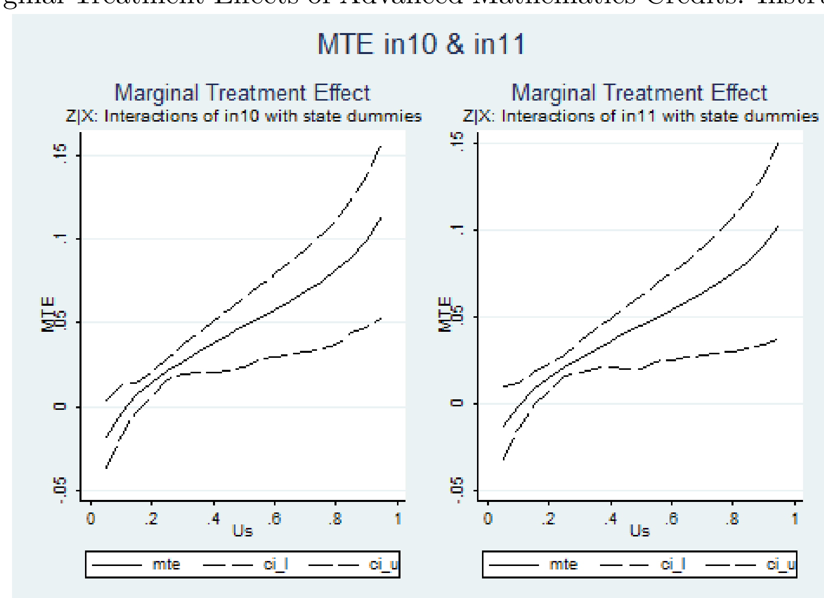
Notes: The 90% confidence intervals were calculated using bootstrap. The exclusion restrictions $Z|X$ are the interactions of in6 & in7 respectively and state dummies. State and cohort fixed effects were also included in the outcome equations and in the selection equation.

Figure 5: Marginal Treatment Effects of Advanced Mathematics Credits: Instruments n8 and n9



Notes: The 90% confidence intervals were calculated using bootstrap. The exclusion restrictions $Z|X$ are the interactions of in8 & in9 respectively and state dummies. State and cohort fixed effects were also included in the outcome equations and in the selection equation.

Figure 6: Marginal Treatment Effects of Advanced Mathematics Credits: Instruments n10 and n11



Notes: The 90% confidence intervals were calculated using bootstrap. The exclusion restrictions $Z|X$ are the interactions of in10 & in11 respectively and state dummies. State and cohort fixed effects were also included in the outcome equations and in the selection equation.

V Conclusions

This study is the first to estimate marginal treatment effects in the context of the returns to high school math credits. Several conclusions and contributions are worth mentioning.

First, the framework presented in this paper allows estimating different parameters relevant to the program evaluation literature: ATE, TT, TUT and MTE. The average gain in total labor income from obtaining advanced math credits during high school varies between 3.35% and 4.77%. These results are consistent with the parameter estimates obtained in Sosa (2017a) of about 3%.

Second, for individuals who are already enrolled in advanced mathematics credits during high school, increasing one more year of advanced math yields a return of about 2.45%. In other words, even when this group benefits the least, there is still some gain to increase the amount of advanced math credits.

Third, for individuals who are not enrolled in advanced math credits, by adding one year of advanced math credits to their transcripts, they have the potential to increase earnings by about 7.39%. From the public policy perspective this result is the most important because, the potential to benefit individuals from low-income backgrounds who do not have access to advanced mathematics due to lack of resources, teachers in particular, is very large.

Fourth, unlike Carneiro, Heckman and Vytlacil (2011), this study suggests that: (1) both treated and untreated individuals could benefit from one additional year of high school advanced mathematics. (2) There is no selection on gains, in other words, individuals with the highest potential gains do not select themselves into the treatment. This might be associated with the fact that, whereas in Carneiro, Heckman and Vytlacil (2011), individuals choose whether or not to attend college, and therefore, most of them are high school graduates, in this study, individuals are choosing whether or not to enroll in advanced math, thus, high school dropouts are included.

Also, because marginal individuals in Carneiro, Heckman and Vytlacil (2011) are older than those in this study, they might be more prone to include future earnings into their cost-benefits analyses of the decisions to attend college, when compared to the cost-benefits analyses done by high school students when choosing which courses to take. It is likely that high school students are not thinking on future earnings when deciding whether or not to take algebra II.

Finally, unlike the previous literature on returns to math credits, this study advances our understanding of the distribution of average benefits across different groups of the population. No other study in the returns to high school math credits literature discusses the parameters of interest in the program evaluation literature: ATE, TT, TUT and MTE.

References

[1] Altonji, J. (1995). The Effects of High School Curriculum on Education and Labour Outcomes. *The Journal of Human Resources*, Vol. 30, No. 3 (Summer, 1995), pp. 409-438

[2] Angrist, J. & Pischke, J. (2009). Mostly Harmless Econometrics, An Empiricist's Companion. Princeton University Press.

[3] Bjorklund, A. & Moffitt, R. (1987). The Estimation of Wage Gains and Welfare Gains in Self-Selection. *Review of Economics and Statistics*, 69(1): 49-49.

[4] Card, D. (2001). Estimating the Returns to Schooling: Progress on some Persistent Econometric Problems. *Econometrica*, Vol. 69, No. 5 September, 2001., 1127 - 1160

[5] Carneiro, P., Heckman, J. & Vytlacil, E. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101 October 2011, pp.2754-2781.

[6] Duncan, G.M., & Leigh, D.E. (1985). "The endogeneity of union status: An empirical test". *Journal of Labor Economics*, 3(3), 385–402 (July).

[7] Farber, H.S. (1983). "Worker preferences for union representation". *Research in Labor Economics*, Volume Supplement 2: New Approaches to Labor Unions. JAI Press, Greenwich, CT.

[8] Gaertner, M., Kim, J., DesJardins, S. & McClarty, K.(2014). Preparing Students for College and Careers: The Causal Role of Algebra II. *Research in Higher Education*, (2014) 55:143 - 165.

[9] Goodman, J. (2012). The Labor of Division: Returns to Compulsory Math Coursework. Faculty Research Working Paper Series. Harvard Kennedy School. August 2012.

[10] Heckman, J. & Li, X. (2004). Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000. *Pacific Economic Review*, 9:3 (2004) pp. 155-171.

[11] Heckman, J., Tobias, J. & Vytlacil, E. (2001). Four Parameters of Interest in the Evaluation of Social Programs. *Southern Economic Journal*, 68:2 (Oct 2001) pp. 210-223.

[12] Heckman, J., Urzua, S. & Vytlacil, E. (2006). Understanding Instrumental Variables with Essential Heterogeneity. *The Review of Economics and Statistics*, August 2006, 88(3): pp.389-432.

[13] Heckman, J. & Vytlacil, E. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proc. Natl. Acad. Sci. USA*, Vol. 96, pp. 4730 - 4734, April 1999, Economic Sciences.

[14] Heckman, J. & Vytlacil, E. (2001). Policy-Relevant Treatment Effects. *The American Economic Review*, Vol. 91, No. 2, Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association (May, 2001),pp. 107-111

[15] Heckman, J. & Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, Vol. 73, No. 3 (May, 2005), 669 - 738.

[16] Heckman, J. & Vytlacil, E. (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Handbook of Econometrics* Volume 6B

[17] Joensen, J. & Nielsen, H. (2009). Is there a Causal Effect of High School Math on Labor Market Outcomes? *The Journal of Human Resources*, 44 (2009).

[18] Kim, J., Kim, J., DesJardins, S. & McCall, B. (2015). Completing Algebra II in High School: Does It Increase College Access and Success? *The Journal of Higher Education*, Vol. 86, Number 4, July/August 2015 pp.628-662.

[19] Lee, L.-F.(1978)."Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables". *International Economic Review*, 19(2), 415-433 June

[20] Levine, P. & Zimmerman, D. (1995). The Benefit of Additional High-School Math and Science Classes for Young Men and Women. *Journal of Business & Economic Statistics*. Vol. 13, No. 2, JBES Symposium on Program and Policy Evaluation (Apr., 1995), pp. 137-149.

[21] Moffitt, R. (2008). Estimating Marginal Treatment Effects in Heterogeneous Populations. *Annales d'Economie et de Statistique*, No.91/92, Econometric Evaluation of Public Policies: Methods and Applications (JULY-DECEMBER 2008), pp. 239-261.

[22] Rose, H. & Betts, J. (2004). The Effect of High School Courses on Earnings. *The Review of Economics and Statistics*, May 2004, 86(2): 497-513

[23] Sosa, A. (2017a). The Impact of High School Mathematics Credits on Earnings: Evidence from Shocks to Teachers' Labor Supply. *Job Market Paper*, University of Michigan May 2017.

[24] Sosa, A. (2017b). Financial Incentives for Teachers in STEM fields: A National Data Set. *Dissertation Chapter*, University of Michigan May 2017.

Appendix

Table 8: **First Stage / Reduced Form** Impact of interactions of potential years of exposure to STEM programs on advanced math and natural logarithm of income.

	(1) advanced math	(2) ln(income)
expo*state 4	0.115***	0.707***
	(0.0202)	(0.0531)
expo*state 5	0.293***	1.502***
	(0.0335)	(0.0747)
expo*state 6	-0.0943***	0.0297
	(0.00973)	(0.0187)
expo*state 7	0.170***	-0.104***
	(0.0110)	(0.0250)
expo*state 8	0.346***	1.568***
	(0.0318)	(0.0719)
expo*state 10	0.0984***	0.732***
	(0.0200)	(0.0521)
expo*state 11	0.141***	0.705***
	(0.0185)	(0.0455)
expo*state 14	0.335***	1.473***
	(0.0339)	(0.0753)
expo*state 15	0.351***	1.499***
	(0.0347)	(0.0758)
expo*state 18	0.0906***	0.561***
	(0.0207)	(0.0540)
expo*state 19	0.171***	0.314***
	(0.0217)	(0.0518)
expo*state 20	0.164***	1.526***
	(0.0329)	(0.0691)
expo*state 21	0.119***	0.754***
	(0.0196)	(0.0517)
expo*state 25	0.425***	1.542***
	(0.0321)	(0.0796)
expo*state 26	-0.137***	-0.152***
	(0.0238)	(0.0432)
expo*state 32	0.128***	0.784***
	(0.0196)	(0.0539)
expo*state 33	0.0332***	0.00890
	(0.0108)	(0.0279)
expo*state 34	0.0425***	0.128***
	(0.0101)	(0.0226)
expo*state 35	0.00435	0.190***
	(0.00826)	(0.0233)
expo*state 37	0.220***	0.825***
	(0.0169)	(0.0416)
expo*state 39	0.118***	0.687***

	(0.0205)	(0.0522)
expo*state 41	-0.0361***	-0.0365***
	(0.000873)	(0.00184)
expo*state 42	-0.320***	0.0604
	(0.0263)	(0.0486)
expo*state 43	0.192***	0.786***
	(0.0163)	(0.0411)
expo*state 44	0.0451***	-0.0375
	(0.0117)	(0.0243)
expo*state 45	0.284***	1.540***
	(0.0338)	(0.0734)
expo*state 47	0.0687***	0.154***
	(0.0136)	(0.0371)
expo*state 48	0.314***	1.543***
	(0.0344)	(0.0765)
expo*state 49	0.258***	1.598***
	(0.0360)	(0.0762)
expo*state 50	0.0836***	0.353***
	(0.0193)	(0.0391)
expo*state 51	0.304***	1.159***
	(0.0333)	(0.0766)
female	0.0603***	-0.370***
	(0.0182)	(0.0329)
white	0.0761***	0.105**
	(0.0257)	(0.0422)
age as of 12-31-1996	-0.159***	-0.159***
	(0.0153)	(0.0365)
hh gross income 1996-1999	2.29e-06***	1.32e-06
	(8.17e-07)	(1.35e-06)
hh income poverty ratio 1996-1999	-0.0124	0.0193
	(0.0115)	(0.0190)
hh size 1997	-0.0280***	-0.00636
	(0.00750)	(0.0150)
both bio parents	0.152***	0.0903***
	(0.0157)	(0.0315)
expo reforms	-0.0857***	-0.118**
	(0.0114)	(0.0554)
high school grad reg 1997	0.880***	3.236***
	(0.0630)	(0.162)
N	4,219	4,219

State-level clustered robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. OLS regressions in which the dependent variable are for column 1, a dummy variable that indicates whether students earned a positive number of credits in advanced math, and 0 otherwise. In column 2, the dependent variable is the natural logarithm of income at age 28. State and cohort fixed effects were also included. All regressions use the 1997 weight.