

# **Who gets left behind? Do Women Disproportionally Benefit from the Jóvenes en Acción Vocational Training Program?**

Nardeen Abdulkareem\*

Student at the Department of Economics, University of Toronto

November 30, 2025

## **Abstract**

This paper revisits the influential work of Attanasio, Kugler, and Meghir (2011) on the Jóvenes en Acción vocational training program in Colombia. Their study provides rigorous causal evidence of the program's effectiveness in improving labor market outcomes among disadvantaged youth, with particularly strong impacts for women. Building on this foundation, I extend the analysis by examining gender-based heterogeneity using modern machine-learning methods designed to capture nuanced, high-dimensional variation in treatment effects. My objective is to determine whether women indeed experience disproportionately larger gains in employment, paid employment, and earnings relative to men following program participation. The broader aim of this work is to contribute robust and externally valid empirical tools for evaluating labor market interventions in low and middle income countries such as Colombia. By documenting gender-differentiated impacts and refining the methods used to estimate them, this paper seeks to inform future normative analyses and policy models that allocate funds to at-risk communities with greater precision and effectiveness.

*Keywords:* Machine Learning, Development Economics, Heterogeneous Effects.

---

\*Thank you to Dr. Nazanin Khazra and our machine learning teaching assistant Adrian Klaus Schroeder for advice and guidance throughout this project. This work is exclusively my own and in no way reflective of the opinions of the University of Toronto. GitHub Repository: <https://github.com/Nardeen-A/Machine-Learning>

# 1 Introduction

Research in development economics outlines that social programs generate substantial gains in employment, earnings, and educational attainment when their design is aligned with constraints faced with the programs implementation. Banerjee and Duflo's (2011) *Poor Economics* book highlights that the success of development programs hinges upon the mechanism of deployment, incentive structure, and established frictions rather than the scale of the program. Intervention in labour markets can exhibit mixed or modest average treatment effects (ATEs) due to behavioural biases, and asymmetric access to information and opportunity. Studies, including Attanasio, Kugler, and Meghir (2011), only find significant positive impacts in outcomes to a segmented portions of the treatment sample. This pattern motivates a shift away from traditional ATEs toward identifying the channels through which differential outcomes emerge.

Attanasio, Kugler, and Meghir (2011), "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial," explores the influence of the Jóvenes en Acción vocational training program for disadvantaged youths on labor market outcomes for men and women. The authors leverage excellent data and empirical analysis by virtue of a randomized controlled trial to explore asymmetric outcomes for men and women. This discontinuity revealed through the large benefits gained by women compared to men compelled me to understand the mechanism by which labor market intervention yields differential effects. Therefore, the main research question I propose is; why do women who participated in the Jóvenes en Acción vocational training program experience significantly larger measured gains than men who received similar treatment? My goal is to illuminate the mechanism behind these differentiated effects, which will be crucial to guiding future programs with a broader goal of provide equitable and efficient outcomes to the vulnerable participants of these programs.

This paper exploits the randomized design of Jóvenes en Acción, which ensures that treatment assignment is exogenous to individual characteristics. Similar to Attanasio, Kugler, and Meghir (2011) paper, rich pre-treatment characteristics serve as controls for labour market outcomes. This analysis provides a novel analysis of how baseline characteristics, such as age, education level, and location, combine with program participation to produce unequal results to participants. I begin by assessing summary statistics and emulating the original paper results in order to validate the data. Descriptive statistics confirm the gender gaps in outcomes, suggesting that treatment effect differences reflect convergence rather than uniformly higher returns to training. We then implement a fixed effects regression to set a benchmark for the results of the machine learning (ML) models which follow. A structural equations model (SEM) allows us to decompose the indirect and direct effects on the dependent variables. A three explanatory variables SEM allows us to analyse how observable and latent factors jointly shape treatment impacts; where baseline human capital, labour market information, and socio-economic constraints are explicitly represented as causal pathways between men and women. Lasso, Ridge, and ensemble methods are used in combinations with the SEM to aid in variables selection to enhance the models predictive power to mitigate the over-fitting of our ML models, and as a visual medium of predictive variables. With this information in hand, we construct directed acyclic graph (DAG) to discipline the specifications of our ML models. Finally, I estimate conditional and heterogeneous treatment effects using matching estimators, doubly robust methods, Double Machine Learning, and causal forests.

The central finding of Attanasio, Kugler, and Meghir (2011) that women benefit dispropor-

tionately from program participation is broadly confirmed. However, richer modelling reveals that women do not outperform men in absolute labour market outcomes when evaluated at the full sample. This is because women enter the program with substantially weaker baseline characteristics, and the large estimated treatment effects reflect convergence toward men’s overall outcomes. While women experience meaningful gains from training, these gains primarily close pre-existing gaps rather than generating persistent advantages over men.

## 2 Literature Review

Banerjee et al. (2007, 2015), Bruhn, Karlan, and Schoar (2018), and McKenzie and Woodruff (2014), indicates that interventions often improve managerial practices, knowledge, and access to credit, but the resulting gains in profits or welfare vary across participants. As noted above, analysis on the variation program design, implementation quality, and baseline characteristics underscores how financial and educational support become necessary to providing lasting economic improvement. Complementary research highlights that development programs often produce indirect or spillover effects that shape aggregate welfare. Miguel and Kremer (2004) and Crépon et al. (2015) show that interventions influences both participants and non-participants by altering market conditions, behaviours, or interactions between peers. Similarly, Duflo, Dupas, and Kremer (2011) and Muralidharan and Sundararaman (2015) demonstrate that peer effects and institutional incentives can amplify or dilute program impacts. This indicates that treatment effects are contextual rather than uniform across individuals. A growing body of work has examined this heterogeneity more directly. McKenzie and Puerto (2021) show that training programs for female entrepreneurs expands market participation and reshape competitive dynamics, revealing that training extend beyond individual outcomes to broader market equilibria. Studies such as McKenzie and Puerto (2021) highlight that training and business education can empower women entrepreneurs, yet the mechanisms behind gender-specific differences remain poorly understood. This paper contributes to this literature by combining experimental identification with structural and machine learning (ML) methods to explicitly model baseline labour market attachment, gender, and local conditions interaction with training to produce heterogeneous outcomes.

The Jóvenes en Acción program constitutes a medium-scale development intervention relative to large national infrastructure or social protection programs. Implemented across seven of Colombia’s largest cities, Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Manizales, and Medellín, The program cost around 60 million dollars, or about 750 dollars per participant. Using experimental evidence from this randomized design, Attanasio, Kugler, and Meghir (2011) document large and statistically significant labour-market gains for women, including a 6 to 7 percentage point increase in employment, a 15 to 17 percentage point increase in paid employment, and roughly a 20 percent increase in earnings. In contrast, estimated effects for men are small and statistically insignificant. A follow-up study by Attanasio, Guarín, Medina, and Meghir (2015) tracks participants nearly a decade later, and finds persistent effects on formal employment outcomes, social security contributions, and wages among women. However, why the same intervention yields divergent effects across demographic groups and localities remains mostly unanswered.

This paper build beyond average effects by explicitly modelling who benefits from vocational training and through which channels. Rather than treating heterogeneity as residual noise, I combine causal identification from random assignment with ML methods to uncover systematic variation

in treatment responses driven by baseline labour market attachment, gender, and location. By integrating structural modelling, causal graphs, and heterogeneous treatment effect estimators, this approach clarifies how pre-treatment constraints shape post-treatment outcomes, thereby improving upon existing work that documents variation but does not explain its sources.

### **3 Data and Summary Statistics**

The dataset, originally compiled by Attanasio, Kugler, and Meghir (2011), was made publicly available through the American Economic Journal: Applied Economics data archive. The data set originates from the Jóvenes en Acción follow up survey and administrative records which provide us detailed information on labour market outcomes for both treated and control individuals. The final sample consists of 3,215 individuals, of whom 1,693 were randomly assigned to receive a vocational training subsidy and 1,759 are women. The randomized controlled trial measured the employment and earning outcomes as a result of the Jovenes en Accion program that ran from 2001 to 2005. The program targeted urban youth aged 18 to 25 years in the two lowest strata of the population in Colombia's seven largest cities. The training intervention consisted of two components, three months of classroom-based instruction provided by accredited training institutions, and three months of on-the-job training with participating firms. Random assignment was conducted once eligible applicants exceeded available program slots, ensuring that selection into treatment was unrelated to baseline characteristics. Within-city randomization is employed to account for systematic differences in local labour market conditions, but treatment effects are estimated at the individual level, not at the district or cluster level. Individual level treatment assignment ensures exogeneity with respect to observable and unobservable characteristics.

Throughout the analysis, treatment is defined as assignment to the program (intention-to-treat) rather than realized training intensity or completion. However, with a 97 percent compliance rate the estimates are nearly identical to that of an Average treatment Effect. Low levels of non-compliance measured through the survey yield a high quality set of controls since there were more candidates than the program has space to accommodate for. Additionally, participants were surveyed at baseline in 2004 and re-interviewed in 2006, approximately one year after program completion. Between baseline and follow-up, approximately 8 to 10 percent of participants dropped out of the survey. Although attrition rates are comparable between the treatment and control groups it nonetheless limits the external validity of the data. Measurement error may arise from self-reported variables such as employment status and monthly earnings but these risks are mitigated by cross checking key variables with administrative labour market records when available. Although treatment assignment was randomized at the individual level, selection bias could still arise through local implementation and peer effects across cities. City, age, marital status, and education dummy variables are included to capture the unobserved regional and characteristic heterogeneity, and matching, doubly robust, and heterogeneous treatment effect estimators can later address the effects of peer based externalities.

Similar to the original study, the summary statistics on table 1 tell us that on average it is difficult to observe significant changes in the overall trend in outcomes for the sample of the population which was studied. Attanasio, Kugler, and Meghir (2011) highlight that in 1999 Columbia experiences its strongest recession in almost 60 years. The country only began recovering to their pre-recession levels of GDP growth in 2003, which happens to be the time this program was rolled out. Thus,

Table 1 suggests that a large amount of the variation in outcomes can be attributed to overall macroeconomic events. However, it is notable that the baseline summary statistics reveal that median pre-treatment salary and paid employment is zero for a large share of the sample. This reflects the high unemployment rates among youth aged 18 to 25 years in the two lowest strata of the population in Colombia's before receiving treatment.

Table 1: Descriptive Statistics by Group (Pre and Post Treatment)

<b>Variable</b>	<b>Period</b>	<b>Group</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Dev.</b>
<b>Age</b>	Post	Control	22.78	23.0	2.05
	Post	Treatment	22.67	22.0	2.12
	Pre	Control	21.23	21.0	2.03
	Pre	Treatment	21.08	21.0	2.05
<b>Education</b>	Post	Control	10.21	11.0	1.77
	Post	Treatment	10.4	11.0	1.54
	Pre	Control	9.99	11.0	1.91
	Pre	Treatment	10.17	11.0	1.67
<b>Employment</b>	Post	Control	0.72	1.0	0.45
	Post	Treatment	0.76	1.0	0.43
	Pre	Control	0.5	1.0	0.5
	Pre	Treatment	0.54	1.0	0.5
<b>Gender</b>	Post	Control	0.55	1.0	0.5
	Post	Treatment	0.53	1.0	0.5
	Pre	Control	0.55	1.0	0.5
	Pre	Treatment	0.53	1.0	0.5
<b>Married</b>	Post	Control	0.27	0.0	0.44
	Post	Treatment	0.25	0.0	0.44
	Pre	Control	0.2	0.0	0.4
	Pre	Treatment	0.19	0.0	0.39
<b>Paid Employment</b>	Post	Control	0.61	1.0	0.49
	Post	Treatment	0.67	1.0	0.47
	Pre	Control	0.35	0.0	0.48
	Pre	Treatment	0.39	0.0	0.49
<b>Salary (1000's) COP</b>	Post	Control	215.79	204.0	203.11
	Post	Treatment	255.49	300.0	221.49
	Pre	Control	99.66	0.0	156.18
	Pre	Treatment	105.67	0.0	155.54

Nonetheless, when the sample is disaggregated by gender, we attain a summary statistic table which reveals pronounced differences in labour market outcomes between men and women. Table 2 shows that women enter the program with substantially weaker baseline labour market conditions, exhibiting lower employment rates, lower participation in paid employment, and massively lower earnings relative to men prior to and following treatment. The table also illustrates that the difference in labour market outcomes for women and men grow in magnitude between the pre-treatment and post-treatment stages. And, although women experience larger proportional gains

than men, men continue to display significantly more favourable employment outcomes. What is especially shocking is our indicator for human capital, education, remains mostly identical across the genders; indicating that former accumulation of schooling is unlikely to be the cause of the observed outcomes. Notably, this table does not isolate the causal effect that the vocational training program had on men and women. The table is simply means for the treated and control individuals within each gender. The observed changes reflect a combination of macroeconomic recovery, general labour market trends, and individual characteristics rather than program participation alone.

Table 2: Descriptive Statistics of Pretreatment and Post-Treatment Variables by Gender

Variable	Women		Men	
	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment
<b>Employment</b>	0.464	0.671	0.581	0.830
<b>Salary (1000's COP)</b>	86.18	196.30	122.171	285.447
<b>Paid employment</b>	0.339	0.581	0.400	0.720
<b>Age</b>	21.324	22.871	20.957	22.543
<b>Education</b>	10.015	10.282	10.168	10.342
<b>Married</b>	0.266	0.324	0.109	0.184

*Notes:* The table reports mean values of demographic and labour market outcomes for women and men in pretreatment and post-treatment periods. This table is a recreation from the original paper.

The outcomes of interest are the employment indicator, paid employment indicator, and real yearly salaries. Represented by the variables `empl_06`, `pempl_06`, and `salary_06`, measured two years after intervention. The treatment variable (`select`) captures random assignment to receive vocational training. To capture gender-differentiated program effects, an interaction term (`select × dwomen`) measures whether the treatment effect differs between men and women. Co-variables include pre-treatment variables such as age, hours worked, marital status, and prior employment outcomes (`empl_04`, `salary_04`). City-level and education level dummy variables are included to control for variation in labour market structures. All continuous or discrete variables were scaled between 0 and 1 to facilitate comparison and to ensure numerical stability in the estimation process.

As a preliminary step, I replicate the original papers regression to confirm the quality and reliability of the dataset.  $\gamma$  denotes the full set of control variables. Fixed effects  $\phi$  are included and they represent an interaction between city, training establishment, and course taken. The data is subset by gender to similar to the original paper. This regression is not used in the main estimation strategy, it is validation exercise to ensure data reproducibility and to establish a benchmark for evaluating subsequent methods. Since the regression separates estimation of the treatment on outcomes by gender, we are unable to observe whether women truly attain outcomes which converge onto the baseline across the entire sample of men and women. This however provides additional robustness to our study, since we can observe the large and statistically significant gains from training present for women reported in the paper. We observe the absence of significant treatment effects for men despite the descriptive statistics pointing towards large gains in employment outcomes. This supports that the gains documented in the summary statistics for men are driven by macroeconomic conditions rather than by the vocational training program itself.

$$Y_i = \beta_0 + \beta_1 \text{Treatment}_i + \gamma' X_i + \phi_{g(i)} + \varepsilon_i$$

Table 3: Replication of Publication

Sex	Obs	Model	Employment	Paid Employment	Salary
Men	1452	Fixed Effects	-0.021 (0.4634)	0.026 (0.4361)	16195.936 (0.3063)
		FE + Control	-0.022 (0.419)	0.015 (0.642)	14938.978 (0.33)
Women	1748	Fixed Effects	0.064 (0.0156) **	0.074 (0.007) ***	40481.508 (4e-04) ***
		FE + Control	0.057 (0.0347) **	0.069 (0.0127) **	36507.172 (0.001) ***

## 4 Regression Results and Machine Learning Models

### 4.1 Ordinary Least Squares

The baseline estimation begins with a simple OLS and logit binary dependent variable model designed to explore the average treatment effects of the vocational training program on key labor market outcomes. Specifically, the model evaluates the impact of program participation on employment, paid employment, and salary outcomes, while accounting for gender differences through an interaction term between treatment status and gender. The inclusion of control variables, denoted by  $\delta$ , captures pre-treatment individual characteristics and city/education dummies that may influence post-treatment outcomes. This specification provides a transparent benchmark for identifying heterogeneous effects by gender before extending the analysis to structural and machine learning models.  $\gamma_{g(i)}$  represents fixed effect for group  $g$  (training center and course taken), with  $g(i)$  showing that individual  $i$  belongs to group  $g$ .

$$Y_i = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 (\text{Treatment}_i \times \text{Woman}_i) + \beta_3 \text{Woman}_i + \delta X'_i + \gamma_{g(i)} + \varepsilon_i,$$

Table 4: OLS &amp; Logit Estimates of Treatment on Employment, Paid Employment, and Salary

	Employment (Logit)	Paid Employment (Logit)	Salary (OLS)
Treatment	-0.071 (0.628)	0.238* (0.051)	30.95** (0.006)
Female	-1.005*** (0.000)	-0.542*** (0.000)	-79.67*** (0.000)
Treatment $\times$ Female	0.311* (0.088)	-0.047 (0.766)	0.26 (0.986)
Constant	0.342** (0.026)	0.245* (0.073)	233.65*** (0.000)
Controls	Yes	Yes	Yes
Observations	3,215	3,215	3,215
Pseudo $R^2$ /Adj. $R^2$	0.111	0.076	0.123
LLR/F Statistic	2.60e <sup>-66</sup> ***	8.22e <sup>-49</sup> ***	27.53***

Robust standard errors. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Pseudo  $R^2$  is reported for logit

The results in table 4 are interpreted as the Average Treatment Effect with a gender treatment interaction. For the employment outcome, coefficients are obtained from a logit specification and should be interpreted in terms of log-odds rather than probability levels. The intercept of 0.342 represents the expected employment outcome for a man in the control group, holding all other covariates constant. The coefficient on Female (-1.005 for employment) indicates that women in the

control group are nearly 100 percentage points less likely to be employed than the men in the control group. This relationship can be observed at a high statistically significant across all the regressions, even after conditioning on pre-treatment characteristics and fixed effects. The treatment effect for men (Treatment =  $-0.071$ ) is negative but small and statistically insignificant with a p value equal to 0.500, suggesting that the vocational training program did not have a meaningful average effect for male participants. However, the interaction term Treatment  $\times$  Female = 0.061 with a p value = 0.311 shows that women in the treatment group experienced an additional employment gain of approximately 31.1 percentage points relative to the male control group in the sample, implying that the programs positive employment effects are concentrated among female participants.

While the Jóvenes en Acción program did not significantly affect men's employment probabilities, it substantially improved women's post-treatment employment outcomes, indicating that the employment effects of the program are concentrated among female participants. This heterogeneity is consistent with the hypothesis that vocational training yields stronger returns for women possibly through mechanisms such as improved access jobs, skill certification, or reduced social barriers to entry in the labour market. It also indicates that women on average have a lower probability of attaining employment than men at the baseline. While the program shows positive effects on overall employment, the story differs for paid employment and salary. Although women begin with much lower baseline levels in both outcomes, the logit and OLS regressions reveal no statistically significant post-treatment gains for treated women when compared to the male control group. This indicates that the women did enjoy positive benefits from participating in the vocational training program, but only upto a point where there is no measurable difference between their labour market outcomes when compares to the male control group. Contrary to the original findings by Attanasio, Kugler, and Meghir (2011), our results suggest that men appear to benefit more from program participation in terms of increased likelihood of paid employment and higher wages. For women, the observed improvements are largely offset by the initial disadvantage in labour market outcomes rather than generating net new gains beyond parity.

## 4.2 Structural Equation Model

The Linear Structural Equation Model (SEM) estimated using three linear models, allows us to interpret our two binary outcomes as average partial effects within a system of simultaneous equations and decompose the total effects into direct and indirect components. Identification of the SEM relies on randomized assignment of treatment, our covariates are pre-treatment variables so they enter the equation exogenously. The SEM extends the OLS framework by allowing us to capture complex causal relationships and simultaneous dependencies among covariates rather than treating them as isolated predictors. In this context, SEM is particularly valuable because it integrates the multiple pathways through which the Jóvenes en Acción program may influence employment, paid employment, and salary outcomes. It informs us on which explanatory variables may affect outcomes both directly and indirectly through intermediate channels. Unlike traditional regressions that estimate one equation at a time, SEM accounts for inter-dependencies such as, how prior employment, education, or marital status indirectly affect post-treatment outcomes through intermediate channels such as formal contract status or hours worked.

$$\begin{bmatrix} empl\_06_i \\ pempl\_06_i \\ salary\_06_i \end{bmatrix} = \begin{bmatrix} \beta_0^{(e)} & \beta_1^{(e)} & \beta_2^{(e)} & \beta_3^{(e)} & \cdots & \beta_{32}^{(e)} \\ \beta_0^{(p)} & \beta_1^{(p)} & \beta_2^{(p)} & \beta_3^{(p)} & \cdots & \beta_{32}^{(p)} \\ \beta_0^{(s)} & \beta_1^{(s)} & \beta_2^{(s)} & \beta_3^{(s)} & \cdots & \beta_{32}^{(s)} \end{bmatrix} \begin{bmatrix} 1 \\ select_i \\ dwomen_i \\ (dwomen \times select)_i \\ \vdots \\ dmarried\_lb_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{e,i} \\ \varepsilon_{p,i} \\ \varepsilon_{s,i} \end{bmatrix}, \quad \varepsilon_i \sim \mathcal{N}(\mathbf{0}, \Sigma_\varepsilon).$$

Unlike the non-linear models which follow, SEM allows us to directly compare our results with the OLS presented earlier. It explicitly highlights which variables to account for through mediation and interdependence. The SEM results reported in Table 5 reveal several noteworthy patterns. Here we can directly map the structure of causal pathways through which vocational training affects labour market outcomes rather than predictive power. The negative and significant coefficients on *dwomen* across all outcomes confirm gender gaps in baseline labour market conditions, consistent with descriptive evidence. The significant contributions of baseline variables such as age, employment, and a measure of formal employment show that a considerable portion of treatment effects flows through prior human capital and formal training channels precisely the type of indirect pathways SEM is designed to capture. The estimated parameters from the SEM provide empirical evidence for the existence, sign, and magnitude of particular causal pathways. The subsequent DAG then acts as a visual and conceptual representation of these pathways, illustrating how the variables relate to one another and indicating where mediation or confounding may occur. The DAG is not merely descriptive, it is a hypothesis of the data generating process that the SEM tests.

Table 5: SEM Regression Results (Significant at 10% level)

	Employment	Paid Employment	Salary
<i>select</i>	-0.0131	0.0418*	42.894***
<i>dwomen</i>	-0.1712***	-0.1193***	-79.530***
<i>select</i> × <i>dwomen</i>	0.0613**	0.0016	-14.644
<i>age_lb</i>	0.1341***		
<i>empl_04</i>	0.1042**		
<i>dformal_04</i>		0.0782*	27.070*
<i>contract_04</i>			24.846*
<i>pempl_04</i>			23.548*
<i>dmarried_lb</i>	-0.0337*	-0.0500**	-18.945***
Residual Variance	0.1679***	0.2078***	22884.29***

Notes: Only coefficients significant at the 10% level are reported. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Standard errors are heteroskedasticity-robust. Dummy variables omitted from table for simplicity.

### 4.3 LASSO and Ridge

We further discipline the construction of the DAG by ensuring that it reflects the appropriate causal structure present in the data by complementing the SEM with Lasso and Ridge variable-selection methods. Lasso and Ridge regressions as regularization based approaches for variable selection and robustness assessment which introduce a penalty term in the loss function, discouraging excessive

model complexity and mitigating multicollinearity among co-variables. We will find the co-variables to include in this model by running LASSO on the outcome variable and the treatment variable and picking variables according to the union of these two LASSO regressions. For the all three models we uncover that the appropriate variables for our DAG are employment, age, education, city, and salary baseline variables as co-variables.

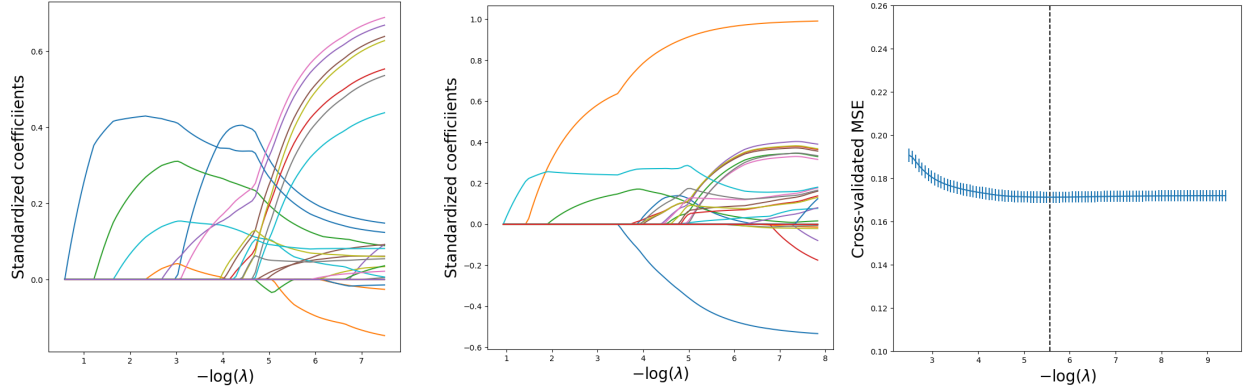


Figure 1: LASSO on Selection, LASSO on Treatment, and LASSO CV

The Ridge regression results reveal which pre-treatment characteristics retain explanatory power for the chosen outcome despite penalization. The variable with the largest absolute coefficient is prior employment, which remains strongly positive even under shrinkage. This indicates that baseline employment status is the most influential predictor of post-treatment outcomes among all co variates: individuals who were employed prior to the program are substantially more likely to display stronger labour market outcomes after treatment. The coefficient on gender (dwomen) is nearly identical in magnitude but negative, suggesting that, even before incorporating treatment interactions, women face a structurally lower predicted outcome relative to men. Other pre-treatment labour market indicators, such as days worked, paid employment, and hours worked also retain relatively large positive coefficients, however these coefficients are strongly correlated so we chose employment, and salary covariates as a proxy for them.

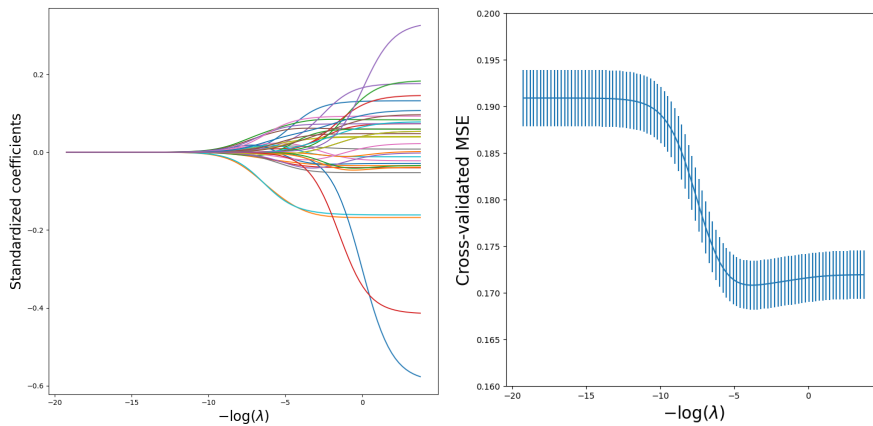


Figure 2: Ridge and Ridge CV

Ridge's preservation of these coefficients confirms that baseline labour market engagement is

a consistent predictor in the data. One of the largest negative coefficients is for city 7, indicating systematically lower predicted outcomes for participants in that locality. Conversely, city 2 shows one of the larger positive effects. In contrast, the coefficients on education dummies are extremely small most near zero indicating that baseline education contributes minimally to predictive power once other labour market indicators are included since ridge shrinkage drives nearly all these coefficients toward zero. However, to get the full story we look at the results in figure 2 point which exhibit pronounced heterogeneity in treatment responses. Individuals with the lowest level of educational attainment (coded as zero) consistently exhibit the largest predicted gains from program participation. Similarities also emerge between ridge and HTE for individuals who were unemployed prior to treatment, as well as for participants residing in city 2 (Bogotá) and city 7 (Medellín). These findings indicate that baseline disadvantage and local labour market conditions are strongly associated with higher potential returns to vocational training.

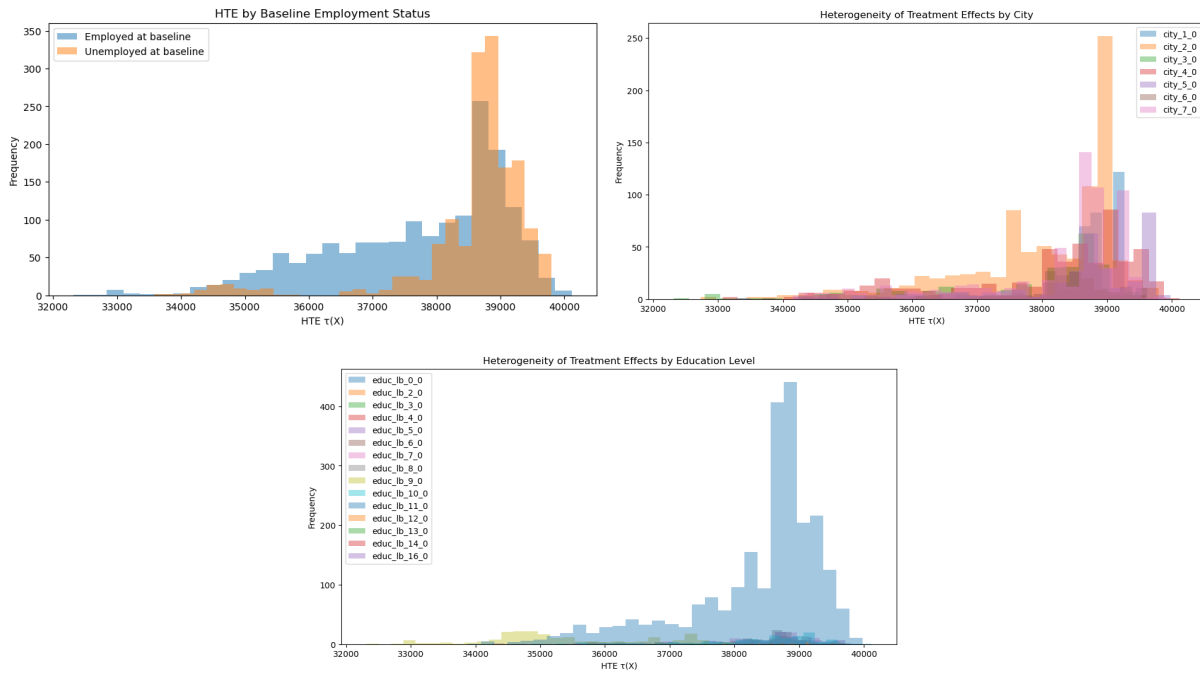


Figure 3: Heterogeneity of Treatment Effect by Covariates

## 4.4 Ensemble Methods

Ensemble methods combine the predictions of multiple models to improve accuracy, reduce variance, and capture non-linear relationships that single estimators might miss. In this analysis, we employ four ensemble approaches; Regression Trees, Bagging, Boosting, and BART. These techniques are employed to explore the variables that provide valuable predictions for labour market outcomes. The regression tree below is pruned and suggests that the variables that are the most important to the prediction of employment in the post period are sex, days worked, and if the student studied in city number 7 (Medellín). The regression tree demonstrates strong predictive accuracy for post-treatment employment outcomes, even when compared to ensemble methods. The tree serves as a visual tool for uncovering sources of heterogeneity and allows us to move past average

treatment effects. We focus on the employment outcome since results for paid employment and salary outcomes are similar for Lasso/Ridge and ensemble methods. The initial split on gender aligns closely with the OLS and SEM findings, reaffirming that gender is a major driver of variation in post-program employment. This division captures the gender gap in labour market outcomes and emphasizes the differential effects between men and women. Additionally, greater work experience prior to intervention indicates stronger networks, job search intensity, and human capital. Meanwhile, spacial heterogeneity for participants trained in Medellín, reflects favourable labour market conditions.

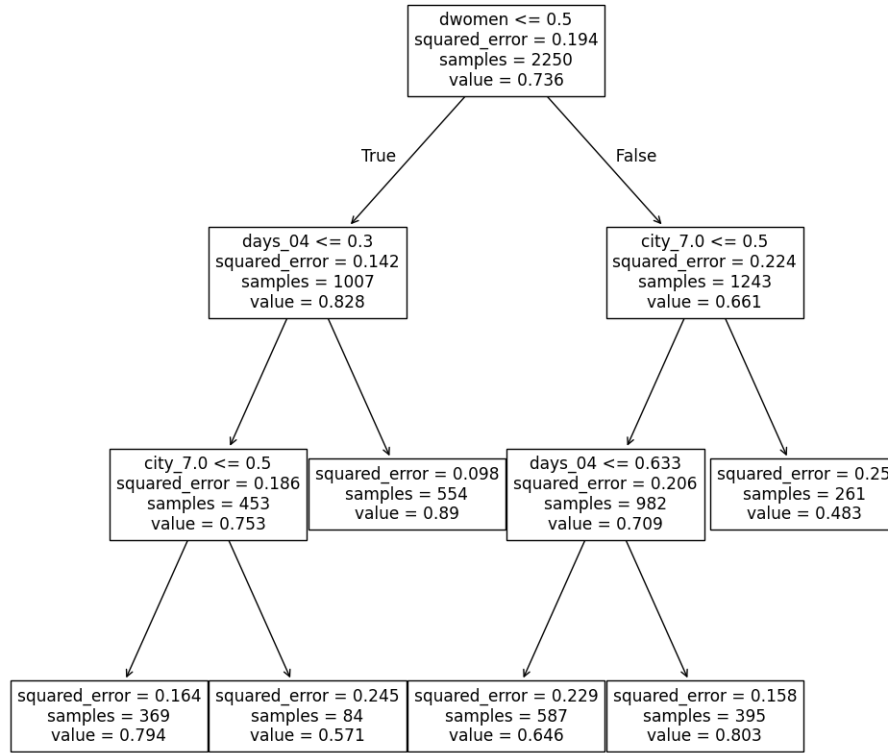


Figure 4: Regression Tree for Employment Outcome.

A Regression Tree recursively partitions the feature space into distinct regions by choosing splits that minimize variance. The Random Forest aggregates hundreds of such trees trained on random subsets of the data and variables, thereby reducing over fitting and providing a stable measure of variable importance. The Random Forest's importance matrix below indicates high predictive power of pre-treatment variables such as age, hours worked in a week, salary, days worked, and whether the student was in city 7 to the prediction of employment outcomes in the post-treatment period. The results display striking heterogeneity in the prediction of employments. While these variables were largely used as controls to the econometric and SEM analyses, the ensemble methods suggest that, the contribution of labour and regional characteristics are vital to the prediction of employment outcomes. The relatively lower importance of the treatment and demographic variables, such as sex, marital status, and pre-treatment employment, offers a revealing contrast. This discrepancy reflects the fundamental difference between causal inference

and prediction. Causal models isolate the treatment’s impact, whereas ensemble methods prioritize variables that best explain variation in observed outcomes.

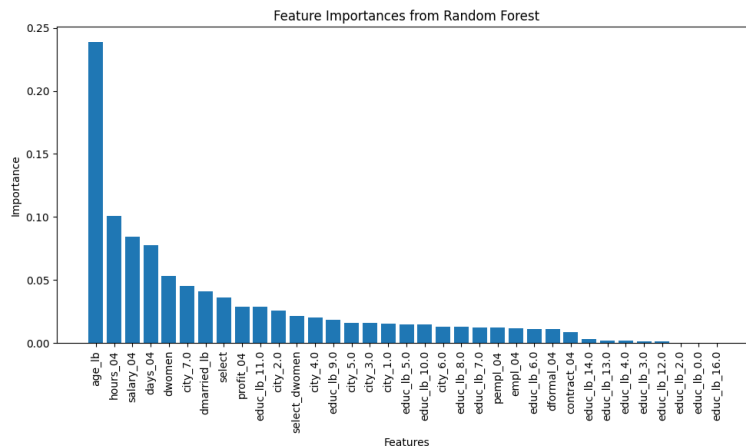


Figure 5: Random Forest’s Importance Matrix.

The boosting algorithm refines predictive accuracy by iteratively building trees that correct the residual errors of preceding ones. This sequential learning process enables boosting to capture subtle non-linearities and treatment heterogeneity that conventional linear or binary models often overlook. In this study, boosting delivers the second-best predictive performance, with test predictions deviating only 0.4077 from the true post-treatment median employment rate.

Table 6: Test Set Prediction Errors Across Models

Model	Test MSE	Test RMSE
Pruned Regression Tree	0.1671	0.4088
Bagging	0.1950	0.4416
Random Forest	0.1948	0.4414
Boosting	0.1648	0.4060
BART	<b>0.1609</b>	<b>0.4011</b>

Bagging and random forest models perform worse than the simple regression tree, indicating that variance reduction alone is insufficient to capture the structure of our data. To further enhance interpret-ability and assess variable relevance, we employ Bayesian Additive Regression Trees (BART). Unlike boosting, which prioritizes pure predictive accuracy, BART offers a framework that balances prediction and inference more effectively. It decomposes complex relationships into an ensemble of shallow trees while naturally quantifying uncertainty around each estimate. This makes BART particularly well-suited for causal exploration in high-dimensional settings such as this one, where multiple correlated pre-treatment and demographic factors interact with treatment assignment. Variables such as gender (dwomen), city dummies, and pre-treatment labor indicators (salary, profit, and days worked) exhibit the highest inclusion rates, highlighting their persistent relevance in explaining post treatment employment outcomes. These results reaffirm the findings from both the structural SEM and penalized regressions as well as the uncovered heterogeneity

from the simple regression tree. Compared to other ensemble methods, BART stands out for its predictive accuracy and provides a data based approach through which to evaluate causal structure across competing specifications.

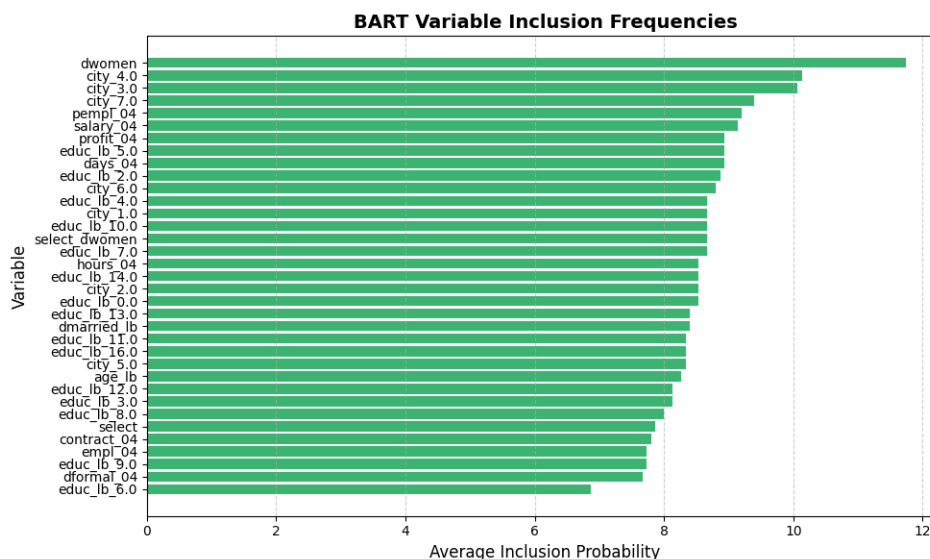


Figure 6: BART's Variable Inclusion Rate

## 4.5 Directed Acyclic Graph

Directed Acyclic Graphs (DAGs) are constructed separately for each outcome; employment, paid employment, and salary. This creates a DAG that is essential for clarifying hypothesized causal pathways and identifying potential confounders or mediation channels that may bias direct regression estimates. By combining empirical evidence from the SEM with theoretical reasoning about the data-generating process, these graphs allow us to explicitly represent conditional dependencies and test whether the assumed causal structure cause any collisions. Together, these inputs discipline which variables enter the causal structure and how they relate to one another. This structure supports the results we attained from the SEM by expressing that the cause of the gap in outcomes between men and women arise through mediated pathways rather than direct effects. The DAGs provide a transparent basis for subsequent causal estimation and guide the interpretation of heterogeneous treatment effects estimated using flexible machine-learning methods.

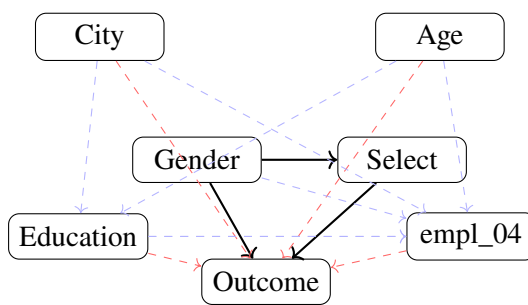


Figure 7: DAG

## 4.6 Propensity Score Matching

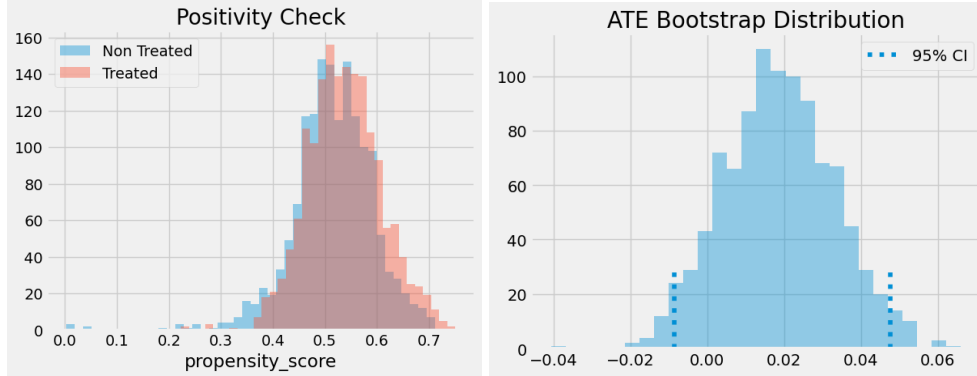


Figure 8: Positivity Check and Propensity Score Bootstrap

I use the propensity score matching model to estimate the causal effect of receiving the vocational train program on employment outcomes. In this application of PSM we use a logistic regression function along side our selected covariates from the DAG. The estimation of these scores across treatment and control groups help us confirm our positivity assumption. The overlap of the distribution of propensity scores for the treatment and control groups indicates on the individual level that those who received the treatment can be matched with one or more people who did not receive the treatment but have a similar propensity score. This creates a synthetic control group that is more comparable to the treatment group. The positivity check graph displays that there is significant overlap between the control and treated groups, particularly within the 0.4 to 0.7 range. This indicates that for most individuals in our data set, there exists a counter-factual group with a similar propensity to be treated. This verifies that we are able to attain an unbiased causal effect estimate for the treatment effect. The matched analysis showed that participants in the treatment group exhibited a higher probability of employment compared to their matched non-participants. After estimating the model, I compared post-treatment employment outcomes between treated and matched non-treated individuals with similar propensity scores. The average treatment effect (ATE) was computed by taking the difference between the mean employment outcomes of the treated.

$$Y_1 = 0.7540 \quad Y_0 = 0.7346 \quad ATE = 0.0195$$

This implies that the average treatment effect increased the probability of being employed by 1.94 percent on average. The results align with my initial hypothesis that treated individuals would exhibit higher employment rates than comparable non-treated individuals. Although the observed effect size is relatively small. This is expected since the value of the average treatment effects is near zero when we do not account for gender differences. This is further proven by bootstrapping the ATE estimate, in this case I obtain a confidence interval that includes zero, indicating that the estimated treatment effect is not statistically significant at 95 percent CI. This means that the sampling uncertainty is large enough that we cannot rule out the possibility that the true average treatment effect is zero. In other words, the observed difference in employment rates between treated and control individuals could have arisen by chance due to sampling variability rather than a true causal effect. This calls for an estimation of the conditional average treatment effect (CATE).

## 4.7 Learner Models

We estimate heterogeneous treatment effects using the meta-learner framework, focusing on post-treatment salary as the outcome variable. Among the three approaches, S-Learner, T-Learner, and X-Learner, the S-Learner is the most appropriate specification in this setting. The S-Learner fits a single outcome model that includes treatment status as a covariate, which stabilizes estimation and mitigates over fitting while still allowing treatment effects to vary flexibly with individual characteristics. Additionally, the randomized treatment assignment gives a balanced dataset with a moderate sample size which allows the S-learner to pool information efficiently across treated and control groups. The S-Learner, while computationally simple and stable, fits a single outcome model for all individuals and therefore tends to smooth out meaningful heterogeneity in treatment effects. The T-Learner allows for separate outcome models for treated and untreated individuals, but becomes statistically inefficient when group sizes differ. The X-Learner combines the advantages of both. It first estimates response functions separately for the treated and control groups, as in the T-Learner, and then constructs imputed individual-level treatment effects. These imputed effects are subsequently re-weighted by estimated propensity scores, making the estimator more efficient and less biased under treatment-control imbalance. This structure makes the X-Learner particularly well-suited for uncovering gender-based heterogeneity in salary outcomes. Figure 10 presents cumulative gain curves for the three meta-learners, along with table 7 that reports the estimated average treatment effect (ATE) and conditional average treatment effects (CATEs) for the first three observations in the sample.

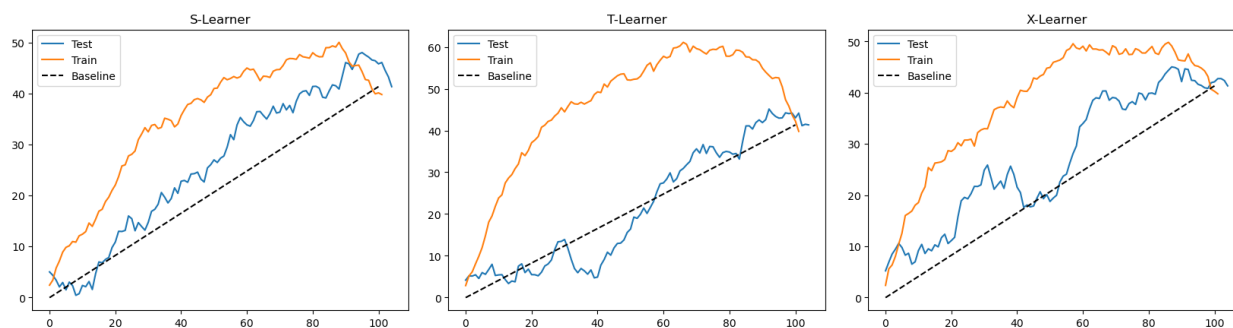


Figure 9: S-Learner, T-Learner, X-learner Gain Curve

The S-learners gain curve exhibits a smooth and relatively steady increasing gain curve. This indicates that the S-learner estimate will more accurately rank the conditional average treatment effects when compared to the T and X learner. The T-learner exhibits more volatility around the middle which reflects sensitivity to noise from fitting separate outcome models. The X-learner exhibits the highest level of volatility across all points on the gain curve, making it the least reliable of the three. Taken together, the results from the learner models provide strong and consistent evidence that participation in Jóvenes en Acción leads to statistically significant increases in post-treatment salary. The fact that all three confidence intervals exclude zero, and that the estimated ATEs closely align across learners, supports the robustness of the program's average impact to a broad set of learner specifications. The CATEs reveal substantial heterogeneity in how participants respond to the intervention. The S-Learner, by construction, smooths over these differences and produces modest, homogeneous estimates. The T-Learner, fitting separate models for treated and untreated

groups, reveals wide dispersion, reflecting genuine variation in how individual characteristics and labour market contexts shape returns to training.

Table 7: Estimated Treatment Effects across Meta-Learners

Learner	ATE	95% CI (Lower)	95% CI (Upper)	CATE <sub>1</sub>	CATE <sub>2</sub>	CATE <sub>3</sub>
S-Learner	34.12	26.13	42.39	55.15	132.48	-13.39
T-Learner	34.63	30.66	40.21	285.98	102.60	-204.56
X-Learner	36.97	29.66	44.16	317.13	59.36	-124.18

Therefore, the program's impact on salary is positive on average but notably heterogeneous across individuals. Some participants experience large earnings gains, while others benefit only modestly, and a small subset may experience negligible or even slightly negative effects. This heterogeneity reflects important differences in baseline labour market attachment, educational attainment, and city-level labour demand factors highlighted in earlier sections of the paper and consistent with the patterns observed in the SEM and DAG analyses. In conclusion, the convergence of evidence across meta-learners suggests that Jóvenes en Acción is effective in raising participants salaries, but its benefits are far from uniform. Identifying the sources of this heterogeneity is crucial for designing more targeted and equitable vocational training programs. Policies should target the groups with high returns, redesign the curriculum for men and women to better align with their skills, and avoid relying solely on the ATE for evaluation. Women-specific constraints (skills, networks, job search) are more binding so training programs can be powerful tools for female economic empowerment. Men may require different forms of support (apprenticeships, and intern ships). In broader development terms, raising women's employment through gender-tailored active labour market policies maximize aggregate welfare.

## 4.8 Doubly Robust Learners

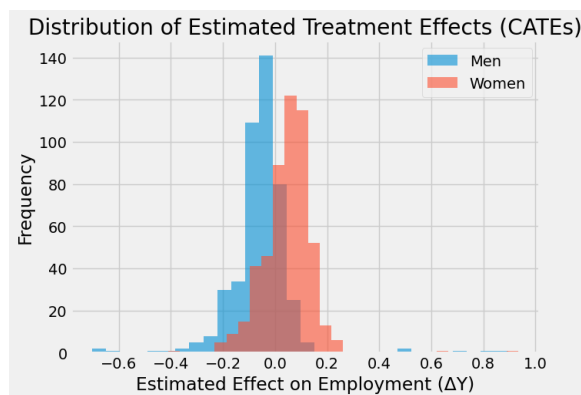


Figure 10: Distribution of CATE through DR learner

Figure 11 is the distribution of conditional average treatment effects (CATE) attained from the Doubly Robust Learners model. This plot shows that women's estimate effect on employment is generally higher than that of men's. A CATE of 0.0338 means that for women, program participation increases the probability of employment by 3.4 percentage points. Although the estimates suggest

Table 8: Gender-Specific CATE Estimates from Doubly Robust Learner

Group	CATE	Std. Dev.	Difference vs. Men
Women	0.0338	0.1064	0.0578
Men	-0.0241	0.1483	

that women experience slightly positive impacts and men experience slightly negative impacts, none of the subgroup treatment effects are statistically distinguishable from zero. The standard deviations substantially exceed the mean effects, indicating that the heterogeneity uncovered by the DR-learner reflects substantial noise rather than subgroup differences. As a result, we cannot assert that the program improves outcomes for women, harms men, or generates a reliable gender differential. These findings suggest either the true effects are small, the data is too noisy, or gender is not the primary moderator of treatment responsiveness. Policy-makers should therefore interpret these subgroup differences cautiously and consider richer sources of heterogeneity.

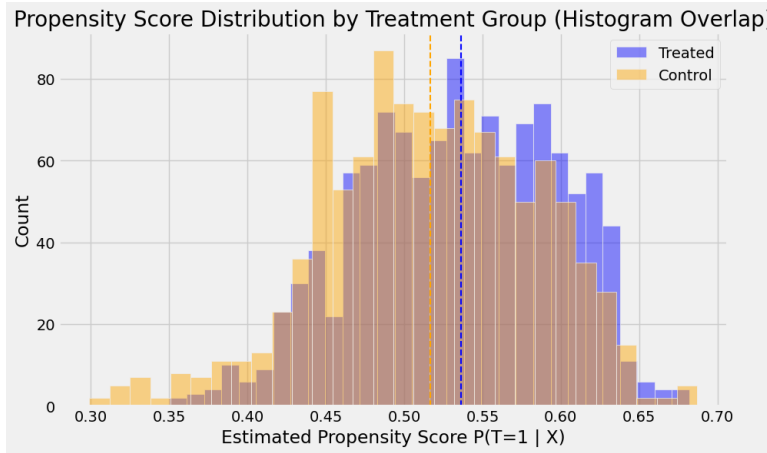


Figure 11: Propensity Score of DR learner

## 4.9 Inverse Probability Weighting (IPW)

Table 9: Average Treatment Effect Estimates (Bootstrap 95% Confidence Intervals)

Estimator	Mean ATE	Lower 95% CI	Upper 95% CI
Inverse Probability Weighting (IPW)	0.0143	-0.0232	0.0526
Doubly Robust (DR)	0.0196	-0.0134	0.0539

IPW is used to check the robustness of previous ATE estimates. It reiterates that the training program does not appear to improve employment outcomes on average. Consequently, while the DR learner provides suggestive evidence of differential treatment responses, these results should be interpreted cautiously. Rather than concluding that the program does not disproportionately benefits women or disadvantages men on average, the findings motivate a deeper exploration of richer sources of heterogeneity. Additionally, if women and men enter the program with different baseline characteristics, the same training could produce different marginal benefits. This motivates

us to explore how estimates for the conditional average treatment effects (CATEs) reveal that women experience notably higher gains from the program, whereas the effects for men are considerably smaller or even slightly negative. Nonetheless, the bootstrapped distribution in figure 13 indicates that on average, the program probably does not strongly help or harm participants, any average true effect is likely very small, and resources might be better targeted toward sub populations.

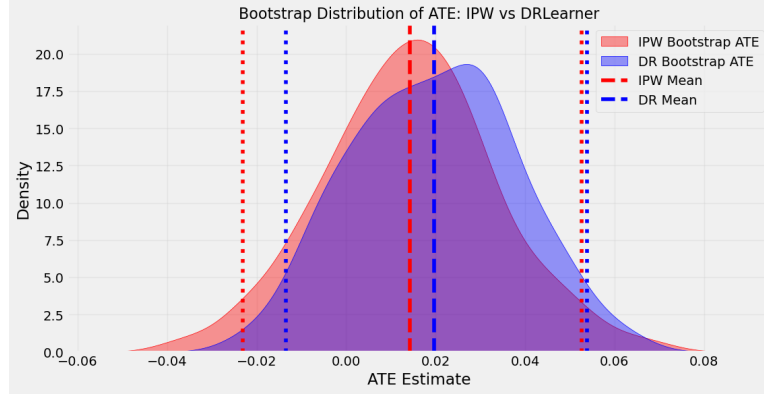


Figure 12: Bootstrapped Distributions

#### 4.10 Causal Forest

Causal forests provide a flexible, non-parametric framework to estimates the treatment effects at the individual level by combining orthogonalized score functions with honest tree construction. We explore the effect on salary outcomes because the effect is larger in magnitude between genders. We implement our previous variable selection standards in order to reduces noise and improves splitting quality in the forest while maintaining the necessary covariates. Using all available variables would introduce unnecessary variance and weaken the heterogeneity signal, whereas restricting the model to data and theory driven predictors improves the stability and interpret ability of the estimated HTE.

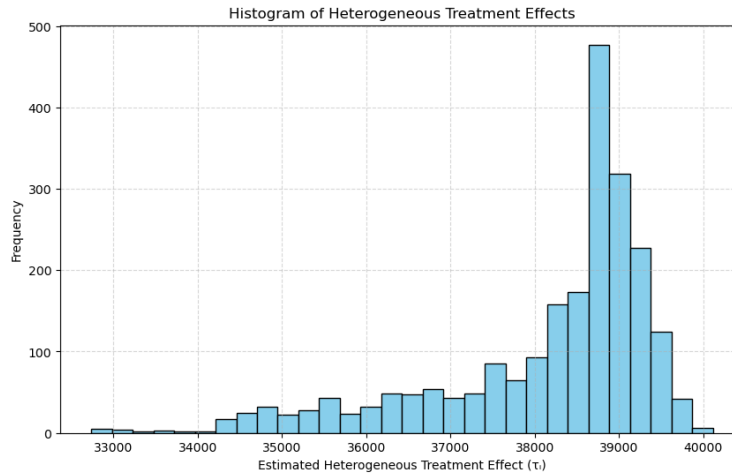


Figure 13: HTE Histogram

The estimated heterogeneous treatment effect (HTE) distribution reveals substantial individual level variation in program impacts. While a small minority of individuals show relatively modest effects, the majority experience large positive effects. A strong mode around 38,000–39,000 indicates the distribution is uni-modal rather than split, suggesting that gender is not the primary source of heterogeneity; instead, baseline characteristics such as prior employment, and earnings appear to drive differences. Overall, the program delivers significantly positive effects for most participants, though the magnitude of these effects varies considerably across individuals. This points to the indirect effect of gender on baseline characteristics on eventual outcomes. Meaning that being a woman influences your outcomes through the attainment of prior employment for instance. Figure 14 presents the distribution of estimated heterogeneous treatment effects (HTEs) for salary outcomes among men and women. The effects are concentrated in the 38,000–39,000 range. Women exhibit a slightly higher mean effect and a longer upper tail, indicating a marginally strong response. However, the distributions overlap extensively, confirming that gender is not the primary source of treatment heterogeneity. Instead, feature importance shows that age and baseline salary are likely the main drivers of heterogeneity. This aligns with the causal forest’s role in adjusting for high-dimensional confounders and explains why the gender gap is much smaller than that obtained through OLS or DR estimators.

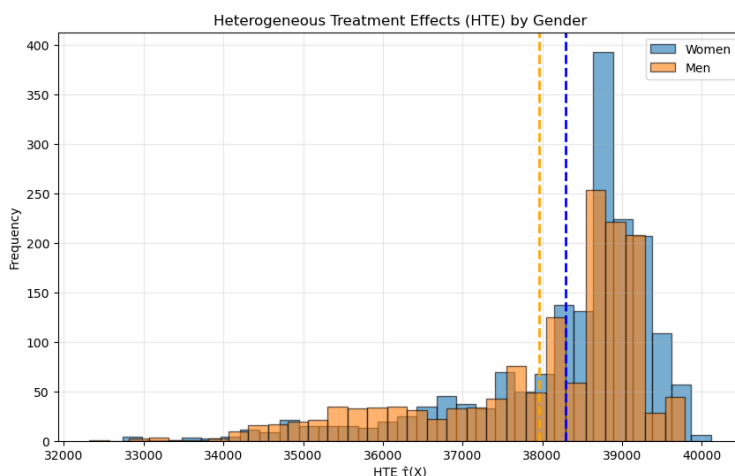


Figure 14: HTE Histogram by Gender

The Quantile comparison of HTE below shows that women consistently have higher predicted treatment effects than men across the entire distribution. The difference is largest at the lower tail, suggesting that the program disproportionately benefits women who are at the lower tail of salary. Taken together, the causal forest results provide a coherent synthesis of the paper’s central findings. Jóvenes en Acción generates large and positive earnings gains for most participants, but the magnitude of these gains varies substantially across individuals. While women tend to benefit slightly particularly among those at the lower end of the earnings distribution—heterogeneity is driven primarily by baseline labour market attachment and local economic conditions. These results reconcile the strong gender effects found in simpler models with the more muted differences observed once rich heterogeneity is accounted for. effects.

Table 10: Quantile Comparison of HTE by Gender

Quantile	Women	Men	Difference
10%	36526.5532	35660.6113	865.9419
25%	37978.8356	37418.3335	560.5021
50% (Median)	38695.6109	38587.0283	108.5826
75%	39055.2535	38915.8548	139.3987
90%	39369.5528	39179.4834	190.0694

### 4.11 Double Machine Learning

The non-linear Double Machine Learning (DML) model using a causal forest as the first stage learner allows us to estimate treatment effects on a flexible model with strong interaction with covariates. This makes this model idea for estimating the differentiated effect of gender and baseline covariates on our outcomes. The causal forest yields an estimated average treatment effect (ATE) for the employment outcome of 0.022 with a 95 percent confidence interval of (0.019, 0.026) which is consistent with our OLS model above. Orthogonalization and cross-fitting are the cause of the narrower confidence interval above.

Gender based DML estimates reveal meaningful heterogeneity and statistically significant outcomes. The estimated ATE for women is 0.040, compared to 0.002 for men. These results closely mirror those obtained from doubly robust learners and causal forests, suggesting that gender-based differences in treatment responsiveness are robust across modelling approaches. Figure 15 illustrates the cumulative gain plot for the linear and non-linear DML models. Both exhibit that they outperform the random assignment, which indicates that predicted treatment effects meaningfully rank individuals by realized gains. The linear model (red dashed line) reaches a peak cumulative gain of roughly 45, outperforming the non-linear model. When individuals are ranked by predicted treatment effect, the linear model identifies the highest-gain participants more sharply.

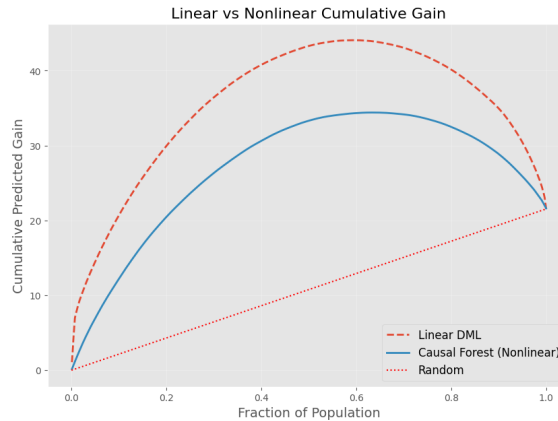


Figure 15: Cumulative Gain Plot

Overall, these findings reinforce the paper's central conclusion. Jóvenes en Acción generates a modest but precisely estimated average employment effect. The DML model does not find insignificant average effects and clarify the structure of heterogeneity by confirming that gender-related differences persist even under flexible specifications. The consistency of findings across

OLS, SEM, doubly robust estimators, causal forests, and DML provides a coherent and robust assessment of the program’s impact.

## 5 Discussion of Results

The paper examines whether participating in the Jóvenes en Acción vocational training program yields disproportionately larger labour market outcomes for women when compared to men. By combining variable selection methods with structural equation modelling, we were able to determine the appropriate variables and DAG structure to use for our subsequent machine learning models. This step ensures that we do not include variables that may yield spurious relationships within the ML models above. We deviated from the original paper by comparing treated women to a control group consisting of both men and women, rather than subsetting the data based on individual characteristics.

Across all models, we obtain similar positive average treatment effects relative to our OLS benchmark. The primary difference between the OLS benchmark and the machine learning models is the ability of the latter to uncover statistically meaningful relationships driven by individual characteristics that generate differentiated outcomes among program participants. The near-identical ATE estimates across linear and non-linear specifications indicate that the program’s average impact is not driven by functional form assumptions, strengthening confidence in the core finding that Jóvenes en Acción modestly improves labour market outcomes on average.

Many of the machine learning methods yield large differences between the estimated effects for men and women. However, the vast majority of these models do not exhibit statistically significant conditional average treatment effects when comparing men and women directly. This pattern is evident from the outset, as the structural equation modelling and causal forest results show that these differences primarily reflect women’s weaker baseline labour market attachment rather than intrinsically higher returns to training. Women experience larger gains because they face more binding pre-treatment constraints, and training facilitates convergence toward men’s outcomes.

Several results that appear surprising in isolation become intuitive within this broader framework. First, much of the gain experienced by female participants only lifts them to the level of men in the control group. Second, increasingly complex estimations of the average treatment effect yield nearly identical and robust results, indicating that the effect is well captured by linear specifications. Third, large positive heterogeneous effects are more clearly observed among participants with weaker baseline characteristics, such as lower levels of education, residence in Bogotá or Medellín, or no employment prior to the program. Finally, the imprecision of gender-specific CATE estimates in doubly robust learners highlights that gender alone is an insufficient dimension along which to partition treatment responsiveness, with baseline employment, earnings, and local labour market conditions playing a more central role.

The policy implications are as follows. Jóvenes en Acción is effective, but its benefits are unevenly distributed. Women with weak baseline labour market attachment experience the largest gains, suggesting that vocational training programs should be explicitly targeted toward participants facing the strongest entry barriers. Overall, this analysis shows that Jóvenes en Acción works not by uniformly raising outcomes, but by alleviating binding constraints for the most disadvantaged participants. Understanding and exploiting this heterogeneity is essential for designing more equitable and effective labour market policies in developing economies.

## 6 Conclusion

This paper revisits the Jóvenes en Acción vocational training program with renewed attention to who benefits most and why. By extending the original analysis of Attanasio, Kugler, and Meghir (2011), I compare outcomes across the full sample rather than within gender subgroups. This approach reveals that, consistent with the original findings, women experience substantially larger and more robust post-treatment employment gains relative to men. In fact, the magnitude of these gains appears even stronger than the results reflected in the initial study. However, the pattern reverses when examining paid employment and salary outcomes. While men experience notable improvements in both dimensions, women’s gains are more modest and largely serve to narrow their initial disadvantage rather than generate net new parity-adjusted advances.

Structural Equations Model (SEM) and Lasso/Ridge regularization validates the robustness of the aforementioned results and strengthens the internal consistency of the model. The use of machine-learning approaches underscores a crucial policy insight; geographic context and prior labor attachment strongly predict post-treatment outcomes. City-level effects and pre-treatment labor indicators—such as hours worked, prior salary, age, and days of employment emerge as the most powerful predictors across regression trees, random forests, boosting, and BART models. Ignoring these features risks miss-estimating the program impact and obscuring the heterogeneous pathways through which training influences employment trajectories.

These results move the discussion beyond average treatment effects to a richer understanding of heterogeneous treatment mechanisms. The machine-learning evidence reveals that geographically based labor characteristics and prior employment outcomes play decisive roles in shaping outcomes under Jóvenes en Acción. Participants from cities with stronger labor demand and higher training quality are more likely to translate newly acquired skills into employment. Similarly, individuals stronger prior work experience convert vocational training into paid employment and higher salaries more efficiently. This is consistent with the regression constant, which reflects men’s higher baseline labor attention.

Moving forward we look to prove economically intuitive mechanisms that are yet deeply relevant for policy. In regions where employment demand is stronger, women benefit disproportionately from training because vocational certification helps overcome entry barriers in the labor market. Conversely, men, who often enter the program with higher baseline experience, convert those skills more readily into paid work. These findings suggest that enhancing the Jóvenes en Acción program requires both gender-sensitive and location-tailored design. Incorporating an internship or job-placement component targeted toward women could bridge the remaining gap in paid employment and salary outcomes, improving the program’s overall efficiency and equity.

Ultimately, this study reaffirms that Jóvenes en Acción remains an effective intervention to the labor-market outcomes for men and women. But it highlights that the models success depends on understanding heterogeneity rather than the average benefit across and between groups. The program works, and it can work better. Policymakers should therefore view gender not simply as a demographic category but as a structural determinant of program impact, interacting with local demand, prior labor experience, and institutional quality. Vocational training initiatives in developing economies should build upon this insight—combining causal identification, and predictive analytics to create programs that are not only effective, but also equitable and context-responsive in promoting sustainable labor-market outcomes.

## 7 Appendix

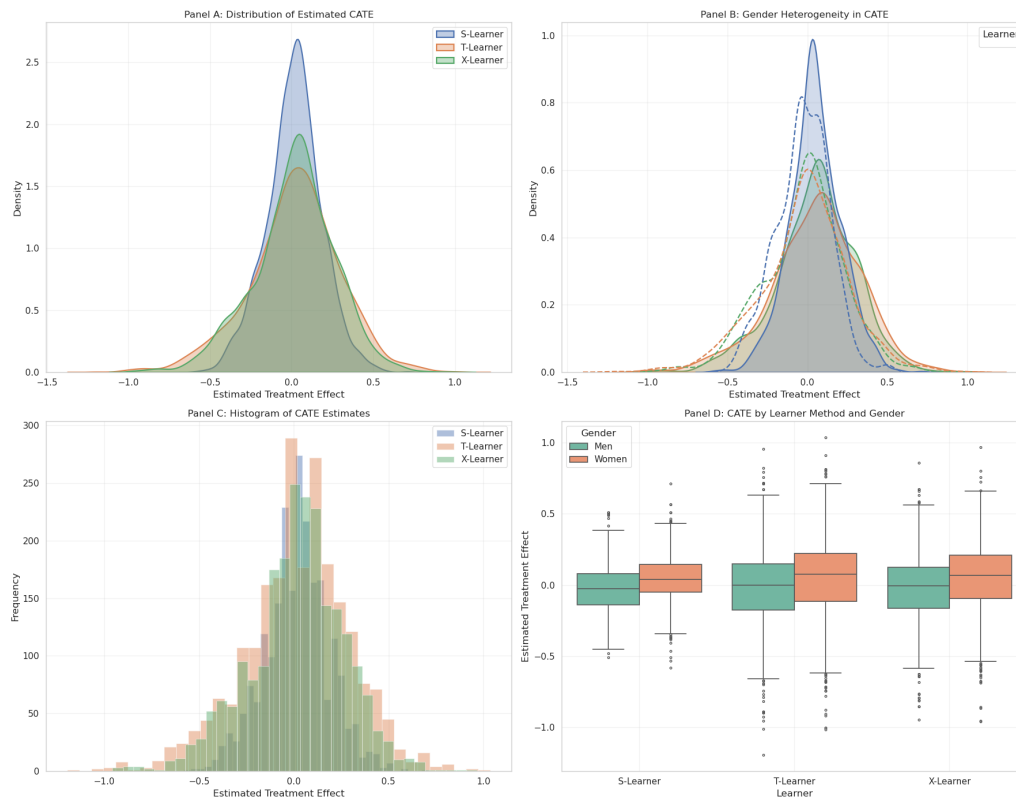


Figure 16: Various DR Learner Plots

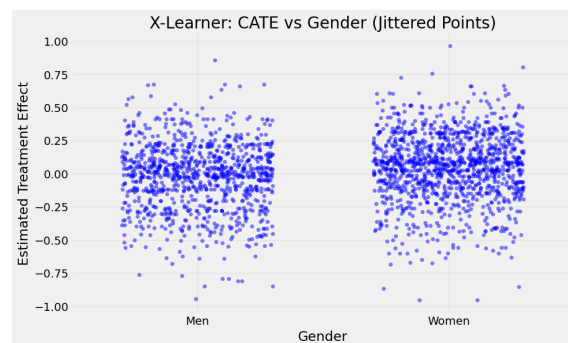


Figure 17: DR Learner Cate Jitter plot

The downward sloping graph illustrates that Lower baseline wages cause a higher treatment effects.



Figure 18: Heterogeneity of Treatment Effect by Baseline Salary

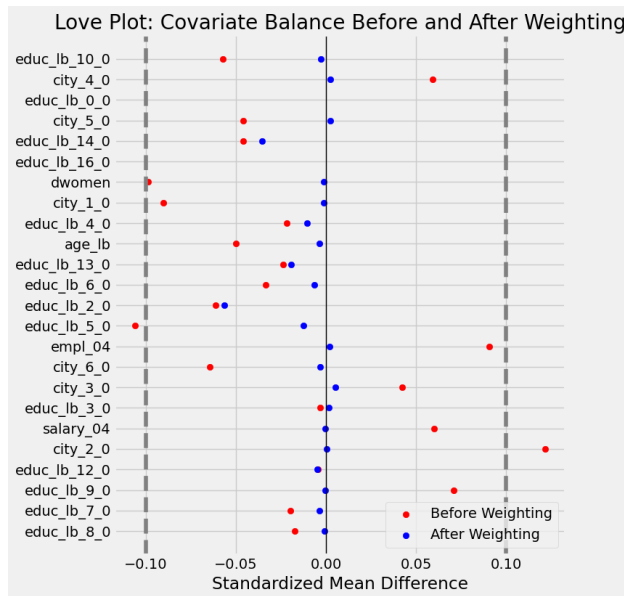


Figure 19: Love Plot for Matching

## References

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. “Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial.” *American Economic Journal: Applied Economics*, 3(3): 188–220.

Banerjee, Abhijit V., and Esther Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: PublicAffairs.

Banerjee, Abhijit V., Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. “The Miracle of Microfinance? Evidence from a Randomized Evaluation.” *American Economic Journal: Applied Economics*, 7(1): 22–53.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. “Remedying Education: Evidence from Two Randomized Experiments in India.” *The Quarterly Journal of Economics*, 122(3): 1235–1264.

Baird, Sarah, J. Aislinn Bohren, Craig McIntosh, and Berk Özler. 2014. “Designing Experiments to Measure Spillover Effects.” *World Bank Policy Research Working Paper No. 6824*.

Bouguen, Adrien, Yue Huang, Michael Kremer, and Edward Miguel. 2019. “Using Randomized Controlled Trials to Estimate Long-Run Impacts in Development Economics.” *Annual Review of Economics*, 11: 523–561.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2018. “The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico.” *Journal of Political Economy*, 126(2): 635–687.

Crépon, Bruno, Florencia Devoto, Esther Duflo, and William Parienté. 2015. “Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco.” *American Economic Journal: Applied Economics*, 7(1): 123–150.

Dhaliwal, Iqbal, Esther Duflo, Rachel Glennerster, and Caitlin Tulloch. 2013. “Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education.” In *Education Policy in Developing Countries*, ed. Paul Glewwe, 285–338. Chicago: University of Chicago Press.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.” *American Economic Review*, 101(5): 1739–1774.

Gertler, Paul. 2004. “Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA’s Control Randomized Experiment.” *American Economic Review*, 94(2): 336–341.

Karlan, Dean, and Jonathan Morduch. 2010. “Access to Finance.” In *Handbook of Development Economics*, Vol. 5, eds. Dani Rodrik and Mark Rosenzweig, 4703–4784. Amsterdam: Elsevier.

McKenzie, David. 2017. “How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence.” *World Bank Policy Research Working Paper No. 8011*.

McKenzie, David, and Christopher Woodruff. 2014. “What Are We Learning from Business Training and Entrepreneurship Evaluations Around the Developing World?” *The World Bank Research Observer*, 29(1): 48–82.

McKenzie, David, and Susana Puerto. 2021. “Growing Markets through Business Training for Female Entrepreneurs: A Market-Level Randomized Experiment in Kenya.” *American Economic Journal: Applied Economics*, 13(2): 297–332.

Miguel, Edward, and Michael Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica*, 72(1): 159–217.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. “The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India.” *The Quarterly Journal of Economics*, 130(3): 1011–1066.