Effects of vocational training on formality: Evidence of a payment-for-success program in Colombia

Michael Cardona* Juan Camilo Chaparro† Gustavo A. García‡

April 2022

Preliminary version, please do not quote

Abstract

Social impact bonds are a novel mechanism for financing social programs and operate under a payment-for-success scheme. We studied the effects on the probability of obtaining formal employment of the *Empleando Futuro* program, the first social impact bond in a developing country. The program offered hard and social skills training, job search assistance, and socioemotional counseling to its participants. Using staggered difference-in-differences design, we found that participants' probability of obtaining formal employment increased significantly in the short term (for both genders), and in the medium term (for women only). The effects on treated females are larger and more persistent than the effects on treated males. Finally, we find that high-intensity social skills training has larger effects on participants.

Keywords: Vocational training, payment-for-success, formality, impact evaluation, staggered difference-in-differences, Colombia

JEL classification: C21, C22, I25, J24, O17

*Department of Economics, Universidad EAFIT, Medellín, Colombia. Email: macardonar@eafit.edu.co

†Department of Economics, Universidad EAFIT, Medellín, Colombia. Email: jcchaparrc@eafit.edu.co

‡Department of Economics, Universidad EAFIT, Medellín, Colombia. Email: ggarci24@eafit.edu.co

1 Introduction

To lead with the unemployment persistency, governments have implemented both Passive Labor Market Policies (PLMP) and Active Labor Market Policies (ALMP). PLMPs aim to hold a consumption level through cash transfers, while ALMPs aim to train and assist the jobless to increase their likelihood of getting and staying in formal employment. Job search assistance programs and training programs are the most used ALMPs. Job search assistance programs are cheaper programs that seek to help the unemployed find employment quickly and their approach is on-the-job learning. Training programs are programs that seek to improve the participant's hard or social skills (or both) and their approach is the human capital increase. Nevertheless, there is an ongoing debate due to the mixed evidence about the ALMP's effectiveness (Card, Kluve, and Weber (2018)).

This paper aims to provide some evidence about the ALMP effects on the low-income individuals' probability to get a formal job. To the best of our knowledge, our paper provides the first evidence of treatment effect heterogeneity of training programs when these are implemented across payment-for-success schemes in Colombia. Our results could incentivize developing country governments to implement training programs through the payment-for-success mechanism to improve labor market outcomes and spread social investment risk. The payment-for-success system involves three main actors: investors, executors, and governments. First, the investors provide the working capital needed to run the training programs. Then, the executors conduct the training programs. When the results have been verified by an independent entity, the government pays to the investors the initial capital plus a return for the risk assumed (Instiglio (2019)).

Between 2017 and 2018, was implemented the first Social Impact Bond in a developing country. This program was designed under a payment-for-success scheme, and was called *Empleando Futuro*. This program aimed to improve job attachment and stability for low-income individuals in Bogota, Cali, and Pereira, three Colombian cities. The program offered training in both social and hard skills, job search assistance, and socio-emotional accompaniment. Furthermore, each executor could offer more intensive training in social or hard skills. The intensive focus on hard skills aimed to improve computer or customer service skills, for example. While the intensive focus on social skills aimed to improve social and communication skills, among others. The defined goals were the attachment in formal employment and permanence of three and six months in formal employment.

The structure of paper is as follow. After this introduction, Section 2 gives an overview of the state of the art on training programs. Section 3 describes the "*Empleando Futuro*" design, accounting for the payment-for-success scheme. Section 4 details the data sources. Section 5 presents the evaluation framework and estimation procedure. Section 6 discusses the obtained results. Finally, Section 7 concludes.

2 Related literature

To improve the employment level in an economy, governments may adopt Active Labor Market Policies (ALMP), for instance: job search assistance, training programs, or subsidized jobs (Crépon and Van Den Berg (2016)). Job search assistance programs and training programs are the most common types of implemented interventions to help people out of joblessness. The former is relatively cheaper and highly successful in the short term, but its effect may fade out through a few periods. The latter es is the most expensive and might not have short-term effects but have long-term effects (Osikominu (2016)).

Schultz (1961) and Becker (1964) were the first to comment on the positive effects of human capital accumulation on economic growth and employment. On the other hand, authors such as Spence (1973) and Arrow (1973) have refuted the effects of human capital accumulation, arguing that it does not affect real economic variables. Training programs are a type of ALMP that aim to increase the human capital of less-skilled workers. There have been multiple evaluations of these programs and there is still no consensus among economists on their effectiveness in improving labor market outcomes. The vast majority of these studies have been conducted in developed countries, evaluating programs with different characteristics: classroom training, hard and social skills training, job search assistance, on-the-job training, and the inclusion of internships. The programs evaluated differ not only in their characteristics but also in the intensity of the training they offer.

While one part of the literature reports that training programs have positive effects on participants' labor market outcomes, another part of the literature reports only short-term or even null effects on employment and wages. Following Card et al. (2018), the labor market outcomes of treated individuals may vary depending on the type of program in which they participate. For instance, training programs that emphasize human capital accumulation suggest positive and sustained effects over time (Barrera-Osorio, Kugler, and Silliman (2020); Chakravarty, Lundberg, Nikolov, and Zenker (2019); Attanasio, Guarín, Medina, and Meghir (2017); Diaz and Rosas (2016); Attanasio, Kugler, and Meghir (2011)), and training programs that emphasize job search assistance may have positive effects only in the short term (Osikominu (2013); Biewen, Fitzenberger, Osikominu, and Paul (2014)). Other authors have found that training programs have neither short- nor long-term effects (Hirshleifer, McKenzie, Almeida, and Ridao-Cano (2016); Groh, Krishnan, McKenzie, and Vishwanath (2016)).

Program objectives matter. If the program designers' objective is that the unemployed overcome their joblessness spells immediately, then the literature suggests that a program with a greater emphasis on job search assistance is more appropriate. Conversely, if the goal is to counteract the depreciation of the human capital of unemployed people, the literature suggests implementing programs with an emphasis on hard and social skills training. Osikominu (2013) and Biewen et al. (2014) highlight that in both the short and long term, job search assistance programs and training programs have different outcomes. Programs with greater emphasis on job search assistance are short-term programs that consider that job skills are best acquired on the job. These programs have positive effects in the short term but no effects in the medium or long term.

In contrast, programs with greater emphasis on hard and social skills training have negative short-term effects ("lock-in-effects") that are offset by increases in employment rates and higher wages in the medium and long term (Osikominu, 2013; Biewen et al., 2014; Card et al., 2018). These training programs have greater effects in countries with low levels of human capital (Hirshleifer et al, 2014). In short, there is a propensity to combine training programs with job search assistance programs to obtain both short-term benefits and medium- and long-term benefits of each approach (Osikominu, 2013).

We highlight four investigations of the effects of training programs conducted in high-income countries. These investigations found time-varying effects. First, Osikominu (2013) studied several training programs in Germany and found that they increase participants' human capital. When human capital increases, the effects on job stability and labor income of training programs are increasing and persistent in the medium and long term. Biewen et al. (2014) found that publicly funded training programs increased participants' employment and labor income in the medium term. In addition, Biewen et al. (2014) provide evidence that training programs are more effective for people with long periods of unemployment. Brunner, Dougherty, and Ross (2021) study the effects of training schools on specific employability skills. The authors find that in treated males there are short-term effects on income and labor participation, and warn that because the training is very specific, the effects may fade in the medium term. On the other hand, they find no statistically significant effects for females. Silliman and Virtanen (2022) find that vocational training programs persistently increase annual earnings and do not increase the probability of working in jobs that are at risk of being replaced by automation.

We highlight four investigations of the effects of training programs in upper-middle-income countries. Three research find positive and persistent effects. Hirshleifer et al. (2016) studied the training programs effects in Turkey. The authors found that the effects on job quality tend to be larger when training is offered by private implementers. However, the effects disappear over time. Reis (2015) found that vocational training programs increased labor income and the likelihood of getting a job for participants in Brazilian metropolitan areas. However, he did not find statistically significant effects on formality. Diaz and Rosas (2016) found that the *Projoven* program in Peru increased the formality rate of the treatment group by about 20%. Acevedo, Cruces, Gertler, and Martinez (2017) studied the effects of hard and social skills training in the Dominican Republic. The authors find that the training had larger effects on labor market outcomes for women, particularly labor participation.

Some research has studied the effects of training programs in Colombia. These programs had an experimental design, and provide relevant evidence for our research. Attanasio et al. (2017) analyzed the long-term effects of *Jovenes en Acción (JeA)* program. The authors found an increase of 4% in males' probability to be a formal employee, while the females' probability increase was around 3%. Respecting formal earnings, Attanasio et al. (2017) found an average increase of COP 35.000 to both genders. Nevertheless, this increase represents an increase of 17.5% in women's earnings and 10.7% in men's earnings. Barrera-Osorio et al. (2020) analyze an experiment with three random assignments. The authors highlight two main findings: on the one hand, the program focused on hard skills has higher short-term effects on overall employment. On the other hand, the program focused on social skills has increased employment and earnings in the long run.

Finally, a first approximation to the evaluation of *Empleando Futuro* effects was made by Chaparro-Cardona, García-Cruz, and Cardona-Botero (2020). Nonetheless, there are methodological aspects that should be taken with caution. We contribute to an application of a novel Difference-in-differences method that allows estimating the ATT in a staggered design, which allows us to obtain an unbiased estimation and a robust parallel trends test, even if covariates are included.

3 Program description

Between June 2017 and December 2018 was executed the first payment-for-success program in a developing country, it was called *Empleando Futuro*, which was carried out in Colombia and targeted skills training and employment support to vulnerable and unemployed individuals. The payment-for-success design is a novel approach that seeks to increase both the effectiveness and the quality of social programs, where the payment of all or part of the funding depends on the achievement of previously agreed results. This scheme involves three main actors: the public sector (results payer), private investors, and executors.

The payment-for-success design works as follows. First, the public sector and the private investor agreed on concrete goals. Second, the private investor financing the executors working capital: inputs, activities, and operative costs of intervention. Third, the executors design and implement the program to targeted people, from the call for registration to the end of the intervention. Fourth, an independent organization verifies if the agreed goals were achieved. Finally, whether agreed results were achieved, then the public sector returns the investment to the private investor, plus a reward for the assumed risk. It means that the payment is contingent upon the achievement of desired social outcomes.

Empleando Futuro supported a range of employment measures, including skills training, psychosocial support, and intermediation services for job placement and retention (at least three months in a formal employment) for vulnerable and unemployed individuals in Bogotá, Cali, and Pereira.[1] The program results payer were the Administrative Department of Social Prosperity of Colombia (DPS), and the IADB Lab. These organizations represent the public sector. On the other hand, *Fundación Corona*, *Fundación Mario Santo Domingo*, and *Fundación Bolivar Davivienda* were the private investors which financed the working capital and operative cost of four program executors: *Fundación Carvajal*, *Corporación Volver a la Gente*, *Kuepa*, and *Fundación Colombia Incluyente*, which both trained the participants in social-and-hard skills and assisted trained in job-search process. These executors were chosen across closed bidding. Finally, *"Deloitte"* was the independent auditor of reached results.

[1] Poor people, unskilled people, youth, women, violence victims, disabled population, LGBT population, ethnic minorities, among others. To more details, please see "Resultados de la Agenda de Aprendizajes". Instiglio (2019).

6,717 individuals registered voluntarily in the *Sistema de Gestión de Desempeño(GdD)*[2] as a response to the four executors' calls for registration. However, only 4,411 people met the eligibility criteria. The first step of eligible criteria filtered participants who had a *SISBEN*[3] score between 0 and 41.74, belong to *Red Unidos*[4], and/or belong to the Colombian registry of victims of the arm conflict. The second eligibility criterion filtered participants between 18 and 40 years, those who had secondary school education, those who did not participate in other social programs [5], and those who did not have formal employment at the start of the program (Instiglio (2019)).

Table 1 reports the registrants' path in the program. We identified 1,352 individuals who did not provide the complete information or didn't meet the eligibility criteria (rows 1 and 2). Next, 2,795 registrants provided complete information and meet the eligibility criteria but did not complete their enrollment process[6], conforming to the control group[7] (row 3). Rows 4 and 5 show the individuals with incomplete treatment: those who did not complete their training process (row 4), and those who complete their training process but did not participate in the job-search assistance module (row 5). Finally, the treatment group[8] is conformed by 1616 eligible participants who completed both the training and job search assistance module. Rows 6, 7, 8, and 9 report the employability information of treated individuals.

[2] *Sistema de Gestión de Desempeño (GdD)* was a system for collecting information from the participants. The GdD platform allows identifying the pathway of the eligible participant through the program phases (application, document sending). Their objective was to generate lessons learned and improve future implementations of this type of program. See *Reporte Final de la Agenda de Aprendizajes, Instiglio (2019).*

[3] It is a classification system that provides living conditions and income information of potential beneficiaries of social programs.

[4] It is a Colombian initiative that seeks to improve the living conditions of the poorest households. It provides family accompaniment and preferential access to relevant public and private social services in health, education, labor, and housing.

[5] Programs like "Inclusión Productiva", "Empleo para la prosperidad", "Mi Negocio", among others.

[6] The reason was not recorded by the program

[7] 1,982 in Bogotá, 760 in Cali, and 53 in Pereira

[8] 1,292 in Bogotá, 235 in Cali, and 89 in Pereira

Table 1: Composition of treatment and control groups

Program flow	Individuals	%	Group
1. Incomplete pre-registrations	621	9.2	
2. Non-eligible pre-registrants	732	10.9	
3. Eligible but non-enrolled registrants	2,795	41.6	Control
4. Enrolled but not certified	705	10.5	
5. Certified but not intermediated	248	3.7	
6. Intermediated but not employed	733	10.9	Treatment
7. Employed at less than 3 months	215	3.2	
8. Employed between 3 and 6 months	363	5.4	
9. Employed more than 6 months	305	4.6	
Total	**6,717**	**100**	

From June 2017 to August 2018, eleven treated groups started their formation programs and intermediation processes staggeringly. The vast majority of participants started their treatment process between July-October 2017 (915) and June-July 2018 (579). Each executor trained a different number of participants. Executor 1 treated 589 participants, Executor 2 treated 522 participants, Executor 3 trained 235 participants, and Executor 4 trained 270 participants. Table 2 shows the treated participants by each executor by gender.

Table 2: Treated individuals by gender and executor

	Females	**Males**	**Total**
Executor 1	409	180	589
Executor 2	402	120	522
Executor 3	214	21	235
Executor 4	208	62	270
Total	**1233**	**383**	**1616**

The executors offered between 100 hours and 300 hours of social-and-hard skills formation, and independently they did choose whether they offered greater emphasis in teaching social skills or hard skills. The hard skills courses targeted developing specific skills to increase job performance. Executors offered basic systems courses, customer service, and administrative assistant modules, among others. The social skills courses targeted improving labor relationships. Executors offered communication skills courses, relational skills, and conflict management, among others.

Regarding the training emphasis, we define a cross-section through formation hours median. The median of hours of hard-skills (h) training was 70 hours, and the median of hours of social-skills (s) training was 50 hours. Hence, we can build four training emphasis: high-intensity both hard and social skills formation ($H_h - H_s$), high-intensity in hard skills but low-intensity in social skills ($H_h - L_s$), low-intensity in hard skills but high-intensity in social skills formation ($L_h - H_s$), and low-intensity both hard-and-social skills training ($L_h - L_s$). For instance, in $H_h - H_s$ classification are those who received 70 hours or more of hard-skills training and 50 hours or more of social-skills training. Table 3 reports our treatment classifications.

Table 3: Training intensity approaches

	$H_h - H_s$	$H_h - L_s$	$L_h - H_s$	$L_h - L_s$	**Total**
Executor 1	18	307	0	264	589
Executor 2	247	1	272	2	522
Executor 3	235	0	0	0	235
Executor 4	92	178	0	0	270
Total	**592**	**486**	**272**	**266**	**1616**

4 Data and descriptive statistics

4.1 Data sources

We track both treatment and control groups matching the *Gestión de Desempeño (GdD)* records, *SISBEN*, and *PILA* database, through participant's ID. This merge allows us to re-build labor histories and socio-demographic characteristics of treated (1,616) and untreated individuals (2,795), and to compare their observable characteristics before and after the treatment. Description of data sources is as follows:

- ***Empleando Futuro* administrative records:** the program registration records collected demographic characteristics from those interested in participating in the program before they knew whether met the eligible criteria. Demographic characteristics included information like participant id, participant age, gender, ethnicity, living city, born city, civil status, education level, their home composition, among others. Furthermore, for treated participants was possible to record information at each stage of the program, like training hours that they received, the start and end date of training courses, and the occupation of each treated after the intermediation process, among others. Survey information and the additional treated information at each stage of the program allowed built the *Gestión de Desempeño (GdD)* platform.

- ***Planilla Integrada de Liquidación de Aportes (PILA):*** It is the national information system that monthly records mandatory contributions to social security (health and pensions). Formal employees are reported in *PILA*. We can re-built the monthly labor history of the treatment and control groups since January 2015. In addition, we can follow up on the short-and-medium-term labor market outcomes of both groups through March 2020.

- ***SISBEN 3:*** It is the Identification System for Potential Beneficiaries of Social Programs in Colombia. From this data set, we gather socioeconomic information of 4,411 eligible people (treatment group and control group). *SISBEN* contains information as household characteristics, communication system access (internet, telephone), household size (number of persons), household income, score, among others. The *SISBEN* information is available before the start of the treatment (January 2016). The cutoff-date guarantees that we can control by the observable socioeconomic characteristics, which weren't affected by treatment status (treatment started in June 2017).

4.2 Descriptive statistics

Table 4 reports the observable differences between treatment and control groups. The first column contains the descriptive statistics of the treatment group, the fourth column contains the descriptive statistics of the control group, and the seventh column shows the difference in characteristics between the treatment and control groups. In both groups, females represent 76% of overall individuals. This fact shows that the vast majority of low-income individuals interested in participating in the programs were females. At the start of the treatment, the treatment group average age was 26.8 years old, while the control group average age was 26.4 years old. Although there is significative difference between the mean age of the groups, it is too small to be taking account.

Table 1' rows 3, 4, and 5 show the distribution of treatment and control groups by residence city. Except for Pereira, we find significant differences between this distribution between groups. Nevertheless, we aren't concerned about this, due to the vast majority of eligible individuals being placed in Bogotá (80% and 70% respectively). Regarding the *SISBEN* score, we observed that the treatment group has a higher score than the control group. Despite this, both groups, on average, are classified as *SISBEN* level 1. The latest row reports the missing registrants in the *SISBEN* database, in which we did not find significant differences between both groups.

Table 4: Treatment and control group differences (Covariate balance check)

	Treatment Group (N = 1616)			Control Group (N = 2795)			Differences			
	Mean (Percentage)	Std. dev.	Observations	Mean (Percentage)	Std. dev.	Observations	Mean	Std. err.	95% conf. interval	
Females	0.7629	0.4253	1233	0.7645	0.4243	2137	-0.0015	0.01327	-0.0276	0.0244
Age	26.8842	5.8435	1616	26.4411	6.1521	2795	0.443	0.1887	0.0729	0.8131
Bogotá	0.7995	0.4004	1292	0.7091	0.0102	1982	0.0903	0.0155	0.0599	0.1207
Cali	0.1454	0.3526	235	0.2719	0.445	760	-0.1264	0.0317	-0.1887	-0.0642
Pereira	0.055	0.2281	89	0.0189	0.1364	53	0.0361	0.0345	-0.0321	0.1044
SISBEN score	22.5664	0.3717	1242	16.8249	0.3273	2136	5.7415	0.5144	4.7329	6.7501
SISBEN missing	0.2314	0.4218	374	0.2357	0.4245	659	-0.0043	0.0274	-0.0581	0.0494

Table 5 shows the labor history of the treatment and control groups. In March 2015, 287 treatment group individuals (18%) were reported in *PILA*, while 523 control group individuals (19%) were reported in *PILA*. Two years later, in March 2017, 360 treatment group individuals (22%) were reported in *PILA*, while 603 control group individuals (22%) were reported in *PILA*. In conclusion, we find that there aren't significant differences in mean formality rates between the groups during the 24 months before treatment starts, implying that in the labor spectrum, both groups are, on average, identical.

Table 5: Treatment and control group formality rates

	Treatment Group (N = 1616)			Control Group (N = 2795)			Differences			
	Mean (Percentage)	Std. dev.	Observations	Mean (Percentage)	Std. dev.	Observations	Mean	Std. err.	95% conf. interval	
March 2015	0.1775	0.3822	287	0.1871	0.39	523	-0.0095	0.0284	-0.0653	0.0463
September 2015	0.2116	0.4085	342	0.2014	0.4011	563	0.0102	0.0276	-0.0441	0.0645
March 2016	0.2184	0.4133	353	0.2125	0.4091	594	0.0059	0.0276	-0.0482	0.06
September 2016	0.258	0.4376	417	0.2372	0.4254	663	0.0208	0.0268	-0.0319	0.0735
March 2017	0.2227	0.4162	360	0.2157	0.4114	603	0.007	0.0275	-0.0469	0.061

5 Econometric framework

The Two-Way Fixed Effects (TWFE) specification has been broadly used to estimate the effects of a program in the Difference-in-Differences research designs. This estimation method captures both the cross-section and temporal variation. Using the canonical TWFE specification represented by Equation 1, we estimate the Average Treatment Effect on the Treated (ATT) on the probability to get a formal job after participating in *Empleando Futuro*:

$$f_{i,t} = \delta_i + \lambda_t + \{\mathbb{1}[T = t] \cdot D_i\}\beta + \gamma \cdot X + \mu_{i,t}, \tag{1}$$

where $f_{i,t}$ is a dummy equal to 1 if individual i was reported in *PILA* at month t, δ_i is individuals fixed effects, λ_t is months fixed effects, $\mathbb{1}[T = t]$ is an indicator function that turns on the periods after treatment, D_i is equal to 1 to every treated individual and equal to 0 to every untreated individual, and X is a covariates vector. Our covariates vector allows us to control by observable characteristics such as residency city, age, gender, and the duo formed by being a *SISBEN* beneficiary and the *SISBEN* score. β is our interest parameter such as it captures the differences between participants and not participants of *Empleando Futuro*, conditional on controls. Under canonical TWFE specification, β is the causal effect of the program over the probability to get a formal job, so that month-to-month control group changes build the counterfactual for treated individuals, had they not participated in the program.

From a dynamic context, Equation 2 describes the dynamic TWFE specification to *Empleando Futuro*:

$$f_{i,t} = D_i \cdot \sum_{t=-q}^{-2} \theta_t \mathbb{1}[\tau - \tau_i^* = t] + D_i \cdot \sum_{t=0}^{T} \theta_t \mathbb{1}[\tau - \tau_i^* = t] + \delta_i + \lambda_t + \gamma \cdot X + \mu_{i,t} \tag{2}$$

The variable $D_{i,t}$ is equal to 1 if the individual i participated in the program at month t, and takes the value of zero otherwise. Indicator variables $\mathbb{1}[\tau - \tau_i^* = t]$ capture the period relative to the training month (τ_i^*) and are zero in all months for non-training individuals. Each θ_t estimation captures the ATT of participating in *Empleando Futuro* relative to the untreated individuals in month t, normalized relative to the last pre-treatment period ($t = -1$). If formality rates for treated and untreated were similar before the treatment started, then we expect that θ_t coefficients between $-q$ and -2 aren't statistically significant. Thus, the dynamic TWFE specification allows us to test the Parallel Trends Assumption (PTA).

Nevertheless, *Empleando Futuro* was a staggering program executed between June 2017 and December 2018. There is novel literature that shows shortcomings of TWFE specifications on more than two-period Difference-in-Differences designs. In this case, we have twelve treatment groups, created by their respective training month of start: June 2017, July 2017, August 2017, September 2017, October 2017, November 2017, December 2017, January 2018, March 2018, May 2018, June 2018, and July 2018. Regarding canonical specification described by Equation (1), Callaway and Sant'Anna (2021) and Goodman-Bacon (2021) show that TWFE with regressors can be severely biased. Goodman-Bacon (2021) proved that the TWFE estimator is a weighted sum of all 2x2 ATT estimators: earlier treated versus untreated groups, later treated versus untreated groups, earlier treated versus later treated groups, and later treated versus earlier treated groups.

Following Goodman-Bacon (2021), we have $g = 1, .., 12$ groups of treated individuals ordered by treatment time t_g^*, and one never-treated group (nt). Further, suppose that l represents the individuals whose was "treated later", e.g., the periods after k ($l > g$). We can denote as n_g the share of individuals in group g, and $\bar{D_k}$ as the share of periods that group g spends under treatment. On the other hand, n_l is the share of individuals in group l, and $\bar{D_l}$ is the share of periods that group l spends under treatment. Equation 3 shows the Goodman-Bacon decomposition of the weighted sum of all n-groups' Difference-in-Differences estimators.

$$\widehat{\beta} = \sum_{g \neq nt} (W_{g,nt} \cdot \widehat{\beta}_{g,nt}) + \sum_{g \neq nt} \sum_{l > g} (W_{g,l} \cdot \widehat{\beta}_{g,l} + W_{l,g} \cdot \widehat{\beta}_{l,g}), \tag{3}$$

where, W represents the weight of each 2x2 comparison. These weight are given by:

$$W_{g,nt} = \frac{(n_g + n_{nt})^2 \cdot \widehat{V}_{g,nt}}{\widehat{V}(\tilde{D}_{it})}; \quad W_{g,l} = \frac{((n_g + n_l)(1 - \bar{D}_l))^2 \cdot \widehat{V}_{g,l}}{\widehat{V}(\tilde{D}_{it})}; \quad W_{l,g} = \frac{((n_g + n_l)\bar{D}_g)^2 \cdot \widehat{V}_{l,g}}{\widehat{V}(\tilde{D}_{it})}$$

Such that $\sum_{g \neq nt} W_{g,nt} + \sum_{g \neq nt} \sum_{l>g} (W_{g,l} + W_{l,g}) = 1$. All the above shows that TWFE estimates a variance-weighted average of ATT, and when already-treated units act as controls, changes in their outcomes are subtracted and these changes may include time-varying treatment effects. We must take into account this heterogeneity and seek up which treatment groups matter most. Also, the post-treatment variable is undefined for the covariates vector (Goodman-Bacon (2021)). Regarding the dynamic TWFE (Equation (2)), Sun and Abraham (2021) show that any parameter θ_t might be contaminated by other period's effects.

To address TWFE shortcomings, our paper estimates the effects of *Empleando Futuro* over the probability to get a formal job up to March 2020, using Callaway and Sant'Anna's (2021) ATT (g, t) causal parameter. The Callaway and Sant'Anna's (2021) estimand uses the Goodman-Bacon decomposition to exploit the heterogeneity of the treatment and returns the Average Treatment Effects on the Treated (ATT) for each group (g) at each period (t). When covariates are included, Callaway and Sant'Anna (2021) suggest using *Doubly-Robust* (D-R) estimator which allows covariate-specific trends across groups (Callaway and Sant'Anna (2021)).

There are two additional reasons to use the D-R estimator. First, it estimator combines linear regression and Inverse Probability Weighting (IPW) to remove the bias in the causal inference estimations, modeling either treatment mechanism (logistic regression) or outcome mechanism (linear regression). It means that it only needs either of both mechanisms to be correct to work, and also that it returns a more efficient estimation. Second, it addresses the possible effect of self-selection to receive the treatment: the D-R estimator measures differences in the distribution of covariates between treated and untreated individuals, assigning more weight to observations that are similar to each other on the covariates.

Following Callaway and SantAnna (2021), we can use two control groups: never-treated units or not-yet-treated units. The former is the traditional control group, that builds a month-to-month counterfactual to treated individuals if had they not participated in the program. The latest is a novel control group that includes never-treated units and not-yet-treated units to build the counterfactual. We use not-yet-treated as the control group for two reasons. First, we can get a larger and more dynamic control group, using Goodman-Bacon decomposition. Second, we mitigate possible self-selection by comparing earlier treated units with later treated units plus never treated units, building a cleaner counterfactual to trained individuals. Under No-Anticipation and Conditional Parallel Trends assumptions, we propose Equation 4 to estimate the effects of *Empleando Futuro* on the probability to get a formal job. This equation is built under the D-R estimator and the not-yet-treated control group (ny).

$$F = \alpha_{g,t} + \eta_{g,t} \cdot G_g + \rho_{g,t} \cdot \mathbb{1}\{T_g = t_g^*\} + \tilde{\beta}_{g,t}(G_g \cdot \mathbb{1}\{T_g = t_g^*\}) + \omega X' + \epsilon_{g,t} \tag{4}$$

Where,

- F is a binary variable which takes the value one if the individual "i" of group "g" at period "t" was employed in a formal job, and it is equal to zero otherwise.
- The G_g variable identifies the treatment group. This variable contains the start month of the treatment for every twelve groups. It is particularly important to both test the PTA and to measure effects estimation of the *Empleando Futuro* program.
- $\mathbb{1}\{T_g = t_g^*\}$ is an indicator function that turns on the periods after treatment starting for group g (t_g^*).
- $\tilde{\beta}_{g,t}$ is our parameter of interest, as it captures the effect of *Empleando Futuro* on the probability to get a formal job when covariates are included. It will be interpreted as a causal effect when $t \geq g$, and PTA pre-test otherwise.

$$\tilde{\beta}_{g,t} = ATT_{dr}^{ny} = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)(1-D_t)(1-G_g)}{1-p_{g,t}(X)}}{\mathbb{E}\left[\frac{p_{g,t}(X)(1-D_t)(1-G_g)}{1-p_{g,t}(X)}\right]}\right)(f_t - f_{g-1} - \mathbb{E}[f_t - f_{g-1}|X, C=1])\right]$$

- ω is the covariates parameter and X is the vector of covariates.
- $\epsilon_{g,t}$ is the error term.

6 Results

In this section, we study the effect of *Empleando Futuro* on formality. Also, we compare our results when the TWFE static and dynamic estimation is used, to our results coming from Callaway & Sant'Anna method, and we provide the Goodman-Bacon decomposition for the program. We think that allowing for treatment effect heterogeneity measures accuracy the effects on the formality of staggered training programs. Table 6 reports our estimations. Panel A shows the results for the overall database, conducting the canonical TWFE estimation (Equation 1), and the simple, the dynamic, the group, and the calendar Callaway & Sant'Anna's aggregations (as of Equation 4). Panel B shows the results by gender, conducting the Callaway & Sant'Anna dynamic aggregations (as of Equation 4) for both females and males. Panel C shows the results by training intensity, conducting the Callaway & Sant'Anna dynamic aggregations (as of Equation 4) for $High_{hard}$ and $High_{social}$ training emphasis, $High_{hard}$ and Low_{social} training emphasis, Low_{hard} and $High_{social}$ training emphasis, and Low_{hard} and Low_{social} training emphasis.

When we conduct the canonical TWFE estimation (first row of Panel A), we observed that *Empleando Futuro* increased the probability to get a formal job by 6.8 percentage points for the treated group, nevertheless this effect is not significant at a 95% confidence level. This canonical estimation just uses a binary variable to indicate the post-treatment periods but doesn't take into account the staggering of the program.

Table 6: *Empleando Futuro.* TWFE and Callaway & Sant'Anna aggregations

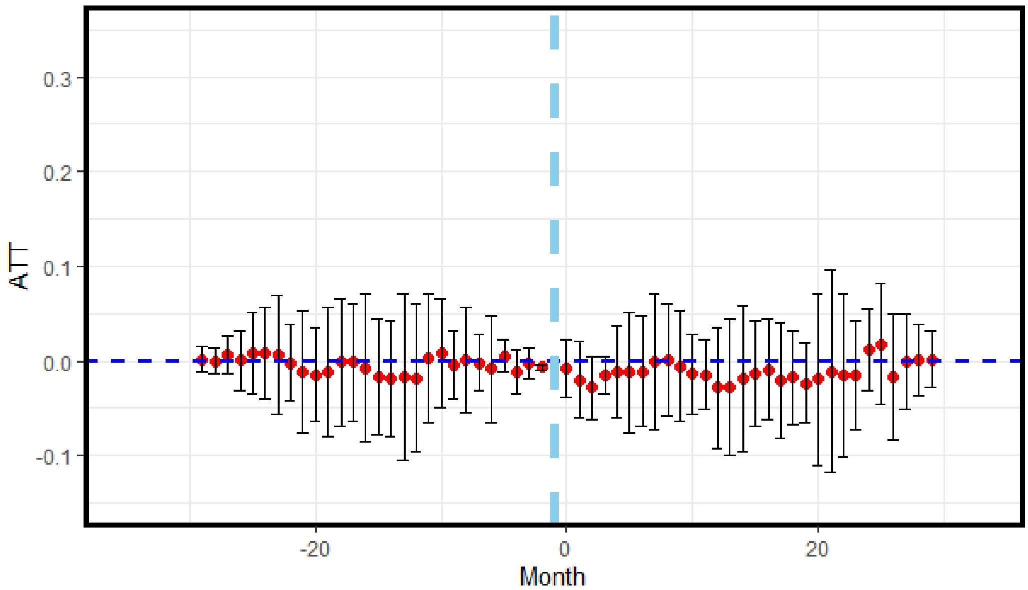
	ATT	Std.Error	95% Conf. Inter.	
Panel A. Overall results				
TWFE estimation	0.0678	0.0192	-0.0150	0.1506
Simple aggregation	0.2162*	0.0118	0.1931	0.2394
Dynamic aggregation	0.2122*	0.0114	0.1898	0.2346
Group aggregation	0.2182*	0.011	0.1966	0.2398
Calendar aggregation	0.2048*	0.0108	0.1837	0.2258
Panel B. Results by gender				
Females dynamic aggregation	0.2382*	0.0159	0.207	0.2695
Males dynamic aggregation	0.1703*	0.0244	0.1226	0.2181
Panel C. Results by training intensity				
$H_{hard} - H_{social}$ dynamic aggregation	0.2068*	0.0182	0.1712	0.2424
$H_{hard} - L_{social}$ dynamic aggregation	0.1967*	0.0185	0.1605	0.2330
$L_{hard} - H_{social}$ dynamic aggregation	0.2452*	0.0245	0.1971	0.2932
$L_{hard} - L_{social}$ dynamic aggregation	0.2045*	0.0257	0.1542	0.2548

The star (*) reports the significance at the 95% level. Column 1 reports the β coefficients, interpreted as the ATT for the canonical TWFE estimation and each Callaway & Sant'Anna's aggregation. Panel A reports the TWFE estimation (Equation 1) and the Callaway & Sant'Anna aggregations (as of Equation 4) for the overall database. The TWFE standard errors were clustered by city. The Callaway & Sant'Anna's standard errors were bootstrapped with 1000 interactions but did not cluster at the city level, due to the small variation across the cities (see Callaway & Sant'Anna (2021) and Table 4). In addition, Callaway & Sant'Anna (2021) propose deleting small groups that have five or fewer treated individuals, hence was delete the following groups to conduct the estimations with the overall database: groups that started at June 2017, March 2018, and May 2015. Panel B reports Callaway & Sant'Anna's dynamic aggregation by gender, and its standard errors were bootstrapped with 1000 interactions but not cluster at the city level. Panel C reports Callaway & Sant'Anna's dynamic aggregation by the four emphases of training, and its standard errors were bootstrapped with 1000 interactions but not cluster at the city level.

Figure 1 shows event-study estimation for *Empleando Futuro*, calculated by Equation 2. On the abscissas axis is the time relative to the treatment's start. On the ordinate axis are the ATT effects by the time of exposure to treatment. The vertical dashed line indicates the program's start. To the left of this dashed line, we observe the formality rates difference between the treatment and the control group before the program's start. On average, there aren't significant differences between both formality rates. It indicates that the control group represents a counterfactual to the treatment group because their formality rates trends are parallel in the pre-treatment period.

To the right of the vertical dashed line, we observe the formality rates difference between the treatment and the control group after the program's start. During the training period (t=0), the treated individuals experienced a negative effect on their formality rate, it was called in literature as "Lock-in Effects" (see Spinnewijn (2013), Crépon and Van Den Berg (2016), Diaz and Rosas (2016), Osikominu (2016), Card et. al. (2018)). Consistently to the canonical TWFE estimation, on average there aren't significant effects on the probability to obtain a formal job after participating in *Empleando Futuro*.

Figure 1: *Empleando Futuro*. Dynamic TWFE



The dynamic TWFE's standard errors were bootstrapped (400 interactions) and clustered by city.

The TWFE specification estimates a variance-weighted mean of the ATT parameters. As demonstrated with the Goodman-Bacon decomposition (Equation 3) in section 5, the TWFE specifications for staggered difference-in-differences designs could be biased. The measure to which the differential trend of a given time group biases the overall estimate is equal to the difference between the total weight in the 2x2 comparison where it is the treatment group and the total weight in the 2x2 comparison where it is the control group. Since treated units near the beginning or end of the database have the lowest treatment variance, they may have more weight as controls than as treatments. (Goodman-Bacon, 2021). Figure 2 shows the Goodman-Bacon decomposition to *Empleando Futuro*. The abscissa axis corresponds to the weight that this comparison between the treatment and control groups should have. The ordinate axis corresponds to the estimation proposed by Goodman-Bacon Decomposition.

Figure 2: *Empleando Futuro.* Goodman-Bacon decomposition

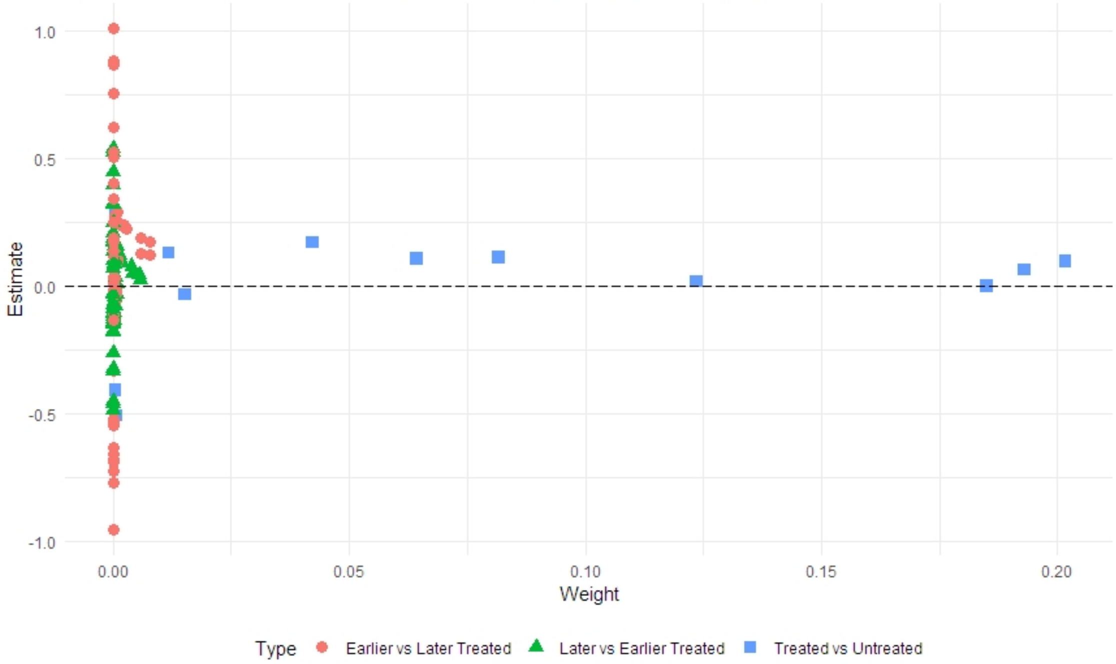


Figure 2 shows that the Treated vs Untreated individuals comparisons must have the highest weights. These weights sum the 92% of overall decomposition. Nevertheless, there is heterogeneity within this comparison. The highest weights were assigned to the groups that start their treatment in July 2017 (0.19) August 2017 (0.20), June 2018 (0.18), and July 2018 (0.12). This result suggests that we must be taking into account the staggering of *Empleando Futuro* to conduct our study on the program's effects on the probability to obtain a formal job.

The program trained individuals at different periods, between June 2017 and July 2018. We calculate the Average Treatment Effects on the Treated for each group at each period following Equation 4. These results were aggregated and are reported by Panel A in Table 6, from the second row to the last one. The second row of Table 6 shows the "simple aggregation" of ATT (g, t) parameters that can be compared to canonical TWFE results. Following Callaway and Sant'Anna (2021), this "simple aggregation" calculates the average effect that *Empleando Futuro* had on each group, and then averages these effects across groups to summarize the overall effect of participating in the program. We observe that the likelihood to obtain a formal job for individuals who participated in the program increased by 21.6% percentage points, respecting the control group. This effect is statistically significant at a 95% confidence level.

From the estimation of Equation 4, we also obtain a dynamic aggregation. The dynamic aggregation proposed by Callaway and Sant'Anna (2021) calculates an average of the mean program effects for each length of treatment exposure. Therefore, the dynamic aggregation's results can be compared with the dynamic TWFE's results. Figure 3 shows the dynamics effects of *Empleando Futuro* on the probability of obtaining formal employment for treated individuals. On the abscissas axis is the time relative to the treatment's start. On the ordinate axis are the ATT effects by the time of exposure to treatment. We observe the pre-treatment period to the left of the dashed vertical line. On average, there are no statistically significant differences between the formality rates of the treatment and the control groups. This implies the PTA meeting and that the counterfactual for the treatment group is constructed with the control group.

To the right of the vertical dashed line in Figure 3, we observe the short-and-medium term differences between the formality likelihood of the treatment and the control group after the program's start. When the staggering of the program is taken into account in the estimation, we observed that participation in the program increased the probability of work in formal employment. The first two months correspond to the training's duration, so the absence of significant differences is an expected result. From the third month on, treated individuals were more likely to be employed in the formal sector (7 pp). After seven months, the probability of being employed in the formal sector was 30 pp higher in the case of individuals who participated in the program. After one year, this likelihood falls to 20 pp but still was significant with 95% simultaneous confidence intervals. On average, *Empleando Futuro* increased the treatment group's likelihood of being employed in the formal sector by 21 pp.

Figure 3: *Empleando Futuro.* Dynamic Aggregation

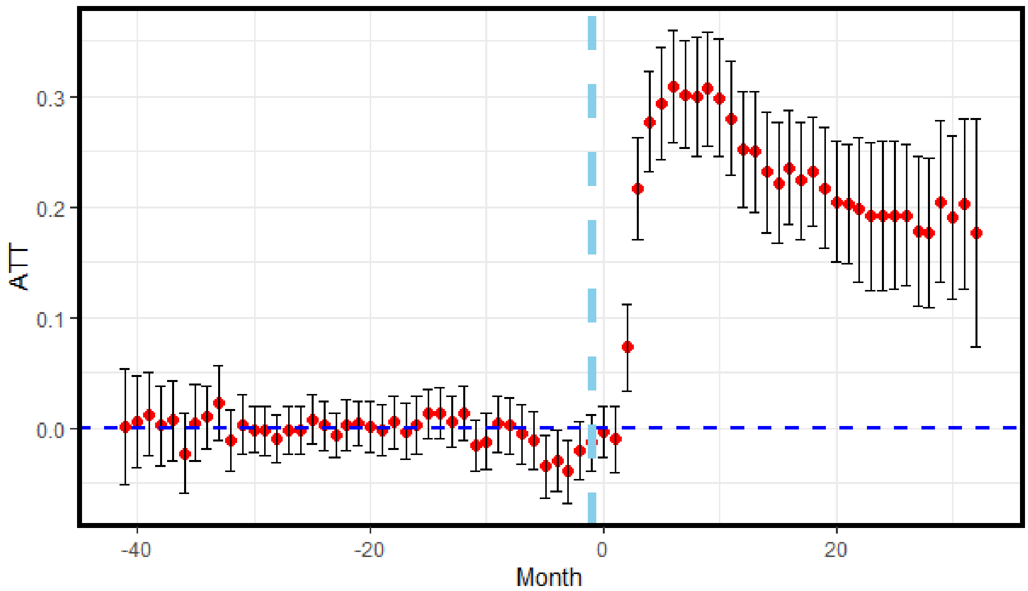


Figure 4 shows the treatment effect heterogeneity of *Empleando Futuro*, plotting the average effect experienced by each group. On the ordinate axis are the different groups, represented by the start date of their training. On the abscissas axis are the effects experienced by each group. This result is interesting because it highlights that the likelihood of being employed in a formal job is higher for some groups. Overall, people who received training in 2017 experienced an average increase in their probability of being employed in the formal sector of 17 pp relative to the control group. On the other hand, the average increase experienced by people who received training in 2018 was 26 pp concerning the control group. In particular, Figure 4 shows that the group that started their training period in January 2018 experienced the largest effect (34 pp). The group that started its training period in November 2017 did not experience statistically significant effects.

Figure 4: *Empleando Futuro.* Group Aggregation

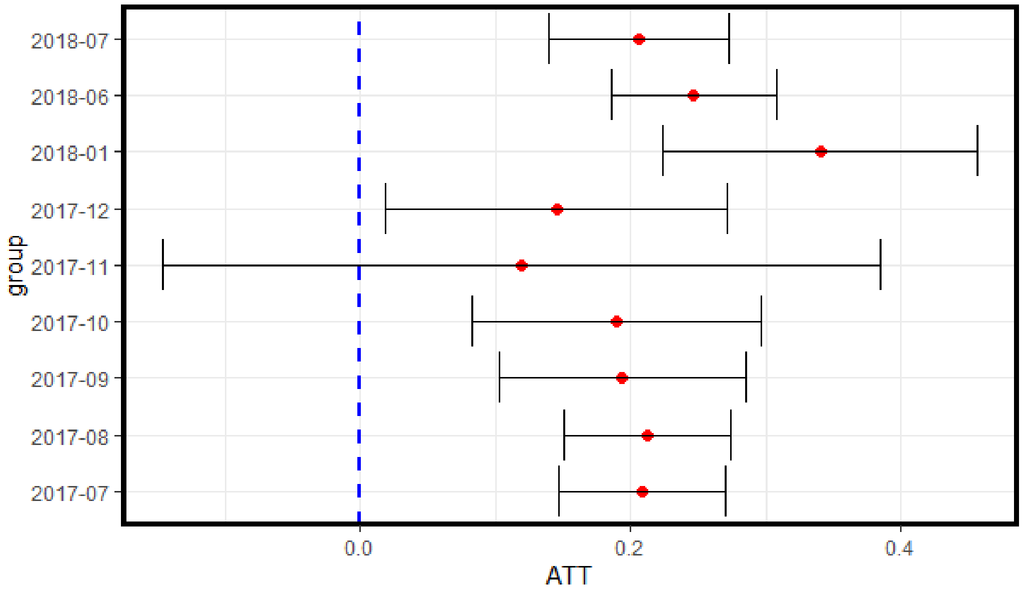
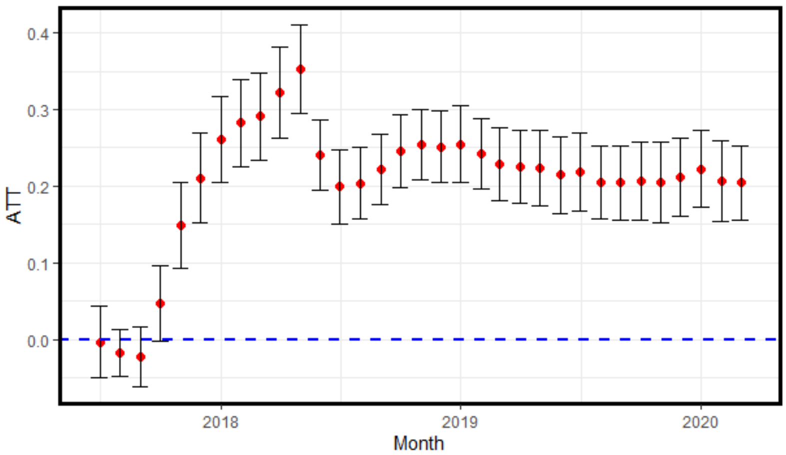


Figure 5 accumulates the experienced effect of all groups at a determinate month. We observe that the accumulative effect increases rapidly, reaching 35 pp after the finish of the training of the group that started in January 2018. Next, this accumulative falls almost 10 pp. The last row of Table 6' Panel A shows that the average accumulative effect on the likelihood of getting a formal job was 20 for the treatment group's individuals.

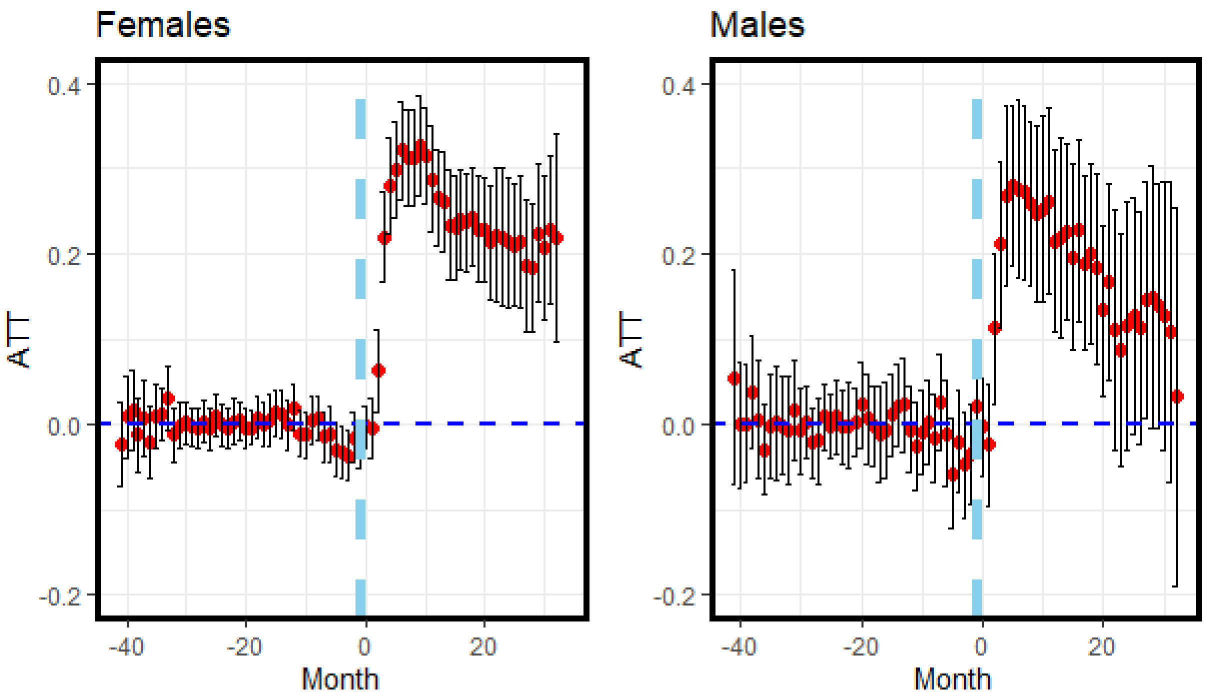
Figure 5: *Empleando Futuro.* Calendar Aggregation



Panel B of Table 6 and Figure 6 present the results by gender. We observe that both genders experienced an increase in their probability of obtaining formal employment after participating in *Empleando Futuro*. However, some differences are worth noting. The first is that the program had a greater effect on women. After overcoming the "lock-in-effect", the probability of women who participated in the program obtaining formal employment increased by more than 30 pp concerning women in the control group. After several months, this probability decreases and stabilizes at about 20 pp. These results are similar to those found by Acevedo et al. (2017), which show that social skills have long-term effects on treated women. After overcoming the "lock-in effect," the probability of men who participated in the program obtaining formal employment increased by less than 30 pp concerning men in the control group. After several months, this probability decreases and stabilizes at about 10 pp.

The second is that the effects experienced by women are more persistent over time. While the medium-term effects of treated men are not statistically significant, the medium-term effects of the program on women remain statistically significant. Although these results should be interpreted with caution due to possible self-selection into the program (76% of the total observations are women), Figure 6 suggests that the human capital of treated women may depreciate more slowly than that of treated men. Another possibility is that the combination of hard and soft skills training is more effective for low-income women.

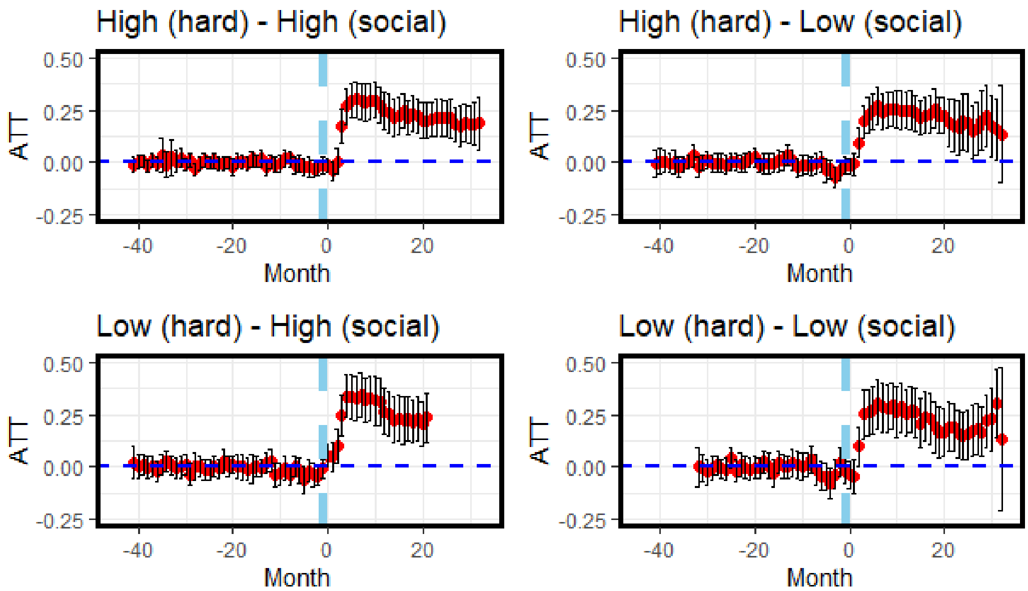
Figure 6: *Empleando Futuro.* Dynamic Aggregation by gender



Panel C of Table 6 examines the impacts separately for those receiving more emphasis on social skills training or hard skills training. To run these pooled models, we extract subsets of each training approach and compare them to the full sample of individuals in the control group. Although all treated participants experienced a positive effect, our results show that those who received greater emphasis on social skills experienced a greater likelihood of being formal workers. For example, individuals who received low-intensity hard skills training and high-intensity social skills training experienced a 24 pp increase over the control group.

Figure 7 shows the dynamics of the treatment effects of each training approach. We observe that regardless of the intensity of hard skills training, programs with greater emphasis on social skills training had larger effects (above 25 pp). In addition, programs with greater emphasis on social skills training show statistically significant effects at all months post-treatment. On the other hand, the effects of programs that offered lower intensity in social skills training lost statistical significance in the medium term. This may be because social skills are more valuable for the low-wage jobs in which low-skilled people tend to be placed. Typically, these low-paying jobs are human relations and customer service-intensive and do not require computer or accounting skills.

Figure 7: *Empleando Futuro.* Dynamic Aggregation by training intensity



7 Concluding remarks

Empleando Futuro was the first Social Impact Bond implemented in a developing country. Between 2017 and 2018, this program provided hard and social skills training, job search assistance, and socio-emotional accompaniment to low-income people in Bogota, Cali, and Pereira, three cities in Colombia. We study the effects of this program on the probability of obtaining formal employment in treated individuals using a unique database that combines administrative records (*SISBEN* and *PILA*) with program records to reconstruct the labor and socioeconomic history of the treatment and control groups. Our research is novel because it uses a recent staggered difference-in-differences technique, that allows us to measure more precisely the *Empleando Futuro* effects, for each treated group at each point in time. In addition, we analyze the effects of the program by gender and by the intensity of training in hard and social skills.

Overall, we find that participation in the program increased the probability of obtaining formal employment by 21 percentage points. Importantly, we find that the effects are sustained in the medium term, two years after participating in the program. Treated women experienced a 7 pp higher probability of obtaining formal employment than treated men and 23 pp higher than untreated women. In addition, we find that the effects for women are positive and statistically significant in both the short and long term. For males, the effects are positive but cease to be significant in the medium term. We also find that training approaches that offered a greater emphasis on social skills produced higher effects on treated individuals.

The next steps in our research are to measure the effects on the intensive margin: quality of work, labor income, and labor stability. In addition, we will conduct a cost-benefit analysis of *Empleando Futuro.*

References

David Card, Jochen Kluve, and Andrea Weber. What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16 (3):894–931, jun 2018. ISSN 15424774. 10.1093/jeea/jvx028.

Instiglio. Agenda de Aprendizajes - BIS Empleando Futuro. 2019.

Bruno Crépon and Gerard J. Van Den Berg. Active Labor Market Policies. *Annual Review of Economics*, 8:521–546, 2016. ISSN 19411391. 10.1146/annurev-economics-080614-115738.

Aderonke Osikominu. The dynamics of training programs for the unemployed. *IZA World of Labor*, 2016. 10.15185/izawol.277.

Theodore W. Schultz. Investment in Human Capital. *American Economic Association*, 51 (5):1035–1039, 1961.

Gary Becker. *Human Capital.* The University of Chicago Press, Chicago, 1964.

Micheal Spence. Job Market Signaling. *The Quarterly Journal of Economics*, 87(3):355–374, 1973.

Kenneth J. Arrow. Higher education as a filter. *Journal of Public Economics*, 2(3):193–216, 1973. ISSN 00472727. 10.1016/0047-2727(73)90013-3.

Felipe Barrera-Osorio, Adriana D Kugler, and Mikko I Silliman. Hard and Soft Skills in Vocational Training: Experimental Evidence from Colombia. Technical report, 2020. URL `http://www.nber.org/papers/w27548`.

Shubha Chakravarty, Mattias Lundberg, Plamen Nikolov, and Juliane Zenker. Vocational training programs and youth labor market outcomes: Evidence from Nepal. *Journal of Development Economics*, 136:71–110, jan 2019. ISSN 03043878. 10.1016/j.jdeveco.2018.09.002.

Orazio Attanasio, Arlen Guarín, Carlos Medina, and Costas Meghir. Vocational training for disadvantaged youth in Colombia: A long-term follow-up. *American Economic Journal: Applied Economics*, 9(2):131–143, apr 2017. ISSN 19457790. 10.1257/app.20150554.

Juan Jose Diaz and David Rosas. Impact Evaluation of the Job Youth Training Program Projoven Labor Markets and Social Security Division Inter-American Development Bank. Technical report, 2016. URL http://www.iadb.org.

Orazio Attanasio, Adriana Kugler, and Costas Meghir. Subsidizing vocational training for disadvantaged youth in Colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 3(3):188–220, jul 2011. ISSN 19457782. 10.1257/app.3.3.188.

Aderonke Osikominu. Quick job entry or long-term human capital development? The dynamic effects of alternative training schemes. *Review of Economic Studies*, 80(1): 313–342, jan 2013. ISSN 00346527. 10.1093/restud/rds022.

Martin Biewen, Bernd Fitzenberger, Aderonke Osikominu, and Marie Paul. The effectiveness of public-sponsored training revisited: The importance of data and methodological choices. *Journal of Labor Economics*, 32(4):837–897, oct 2014. ISSN 0734306X. 10.1086/677233.

Sarojini Hirshleifer, David McKenzie, Rita Almeida, and Cristobal Ridao-Cano. The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey. *Economic Journal*, 126(597):2115–2146, nov 2016. ISSN 14680297. 10.1111/ecoj.12211.

Matthew Groh, Nandini Krishnan, David McKenzie, and Tara Vishwanath. The impact of soft skills training on female youth employment: evidence from a randomized experiment in Jordan. *IZA Journal of Labor and Development*, 5(1), dec 2016. ISSN 21939020. 10.1186/s40175-016-0055-9.

Eric Brunner, Shaun Dougherty, and Stephen L Ross. The Effects of Career and Technical Education: Evidence from the Connecticut Technical High School System. Technical report, 2021. URL http://www.nber.org/papers/w28790.

Mikko Silliman and Hanna Virtanen. Labor Market Returns to Vocational Secondary Education. *American Economic Journal: Applied Economics*, 14(1):197–224, jan 2022. ISSN 1945-7782. 10.1257/app.20190782.

Mauricio Reis. Vocational training and labor market outcomes in Brazil. *B.E. Journal of Economic Analysis and Policy*, 15(1):377–405, jan 2015. ISSN 19351682. 10.1515/bejeap-2013-0023.

Paloma Acevedo, Guillermo Cruces, Paul Gertler, and Sebastian Martinez. LIVING UP TO EXPECTATIONS: HOW JOB TRAINING MADE WOMEN BETTER OFF AND MEN WORSE OFF. Technical report, 2017. URL http://www.nber.org/papers/w23264.

Juan Camilo Chaparro-Cardona, Gustavo Adolfo García-Cruz, and Marlen Yamilet Cardona-Botero. Evaluación de Impacto de "EMPLEANDO FUTURO". Technical report, 2020.

Brantly Callaway and Pedro H.C. Sant'Anna. Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, dec 2021. ISSN 18726895. 10.1016/j.jeconom.2020.12.001.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021. ISSN 18726895. 10.1016/j.jeconom.2021.03.014. URL https://doi.org/10.1016/j.jeconom.2021.03.014.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, dec 2021.

ISSN 18726895. 10.1016/j.jeconom.2020.09.006.
Johannes Spinnewijn. Training and search during unemployment. *Journal of Public Economics*, 99:49–65, 2013. ISSN 00472727. 10.1016/j.jpubeco.2013.01.002. URL `http://dx.doi.org/10.1016/j.jpubeco.2013.01.002`.