Effects of job training on labor formality: evidence of a payment-for-success program in Colombia

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

July 2022

Preliminary version, please do not quote

Abstract

This study evaluates the effects of a job training program in Colombia (*Empleando Futuro*), financed trough a payment-for-success scheme. We measure the effects of *Empleando Futuro* on the probability of obtaining formal employment on low-income individuals. *Empleando Futuro* was the first job training program in a developing country financed through a Social Impact Bond (SIB), a novel mechanism for covering the costs of social programs. The program offered training in hard and social skills, job search assistance, and socio-emotional counseling to all participants. Using a staggered Difference-in-Differences design, we found that participants' probability of obtaining formal employment increased significantly in the short term for both females and males, and in the medium term only for females. We also found that high-intensity training in social skills has a larger effect on participants.

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

Training programs 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 (Card et al., 2018; González-Velosa et al., 2012; Osikominu, 2013). 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.

This paper aims to provide additional evidence regarding the training programs effects on 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 job training programs in developing countries when these are implemented using a payment-for-success scheme. Our results could incentivize governments in developing countries to implement training programs through a payment-for-success mechanism to improve labor market outcomes and spread social investment risk. The payment-for-success mechanism involves three main actors: investors, executors, and governments. First, the investors provide the working capital needed to run the social program. Then, the executors carry out the social program. 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).[1]

Between 2017 and 2018, the first Social Impact Bond (SIB) in a developing country was implemented. 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 support in different months of 2017 and 2018. Furthermore, each executor could offer more intensive training in social skills or hard skills. The intensive focus on hard skills aimed at improving computer or customer service skills, for example. While the intensive focus on social skills aimed to improve social and communication skills, among others. Payments to the social investors were conditioned on the number of participants who obtained a formal job and were able to keep it for at least three months. Additional payments were made to the social investor if the participant kept its formal job for at least six months (Instiglio, 2019).

[1]To more details of payment-for-success mechanism see Gustafsson-Wright and Gardiner (2016), Galitopoulou and Noya (2016), Intiglio and GPOBA (2018)

We built a unique data panel with information from three administrative databases: program records, the *Planilla Integrada de Liquidación de Aportes (PILA)* and *SISBEN*. This information is available between January 2015 and March 2020. *Empleando Futuro* was a staggered program. Therefore, we estimate the effects of the program on the probability of obtaining formal employment both by the Two-Way Fixed Effects (TWFE) method and by the method proposed by Callaway and Sant'Anna (2021). On the one hand, when we estimated with the TWFE method we found that the program had no statistically significant effect on the probability of obtaining formal employment. On the other hand, when we estimated with the method proposed by Callaway and Sant'Anna (2021) we found that *Empleando Futuro* increased the probability of obtaining formal employment of treated individuals by 21 percentage points. These results are statistically significant at a 95% confidence level.

Our findings contribute evidence to the ongoing debate on ALMP effects. 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 achieve positive and sustained effects over time (Attanasio et al., 2017,1; Barrera-Osorio et al., 2020; Chakravarty et al., 2019; Diaz and Rosas, 2016), and training programs that emphasize job search assistance may have positive effects only in the short term (Biewen et al., 2014; Osikominu, 2013). Other authors have found that training programs have neither short- nor long-term effects (Groh et al., 2016; Hirshleifer et al., 2014).

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 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-effect") that are offset by increases in employment rates and higher wages in the medium and long term (Biewen et al., 2014; Card et al., 2018; Osikominu, 2013). These training programs have greater effects in countries with overall 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 and long-term benefits (Osikominu, 2013).

We highlight four researches on 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 positive 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 et al. (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.

In upper-middle-income countries there are four impact evaluations of the effects of training programs. Hirshleifer et al. (2014) studied the training programs effects in Turkey, finding that the effects on job quality tend to be larger when training is offered by private implementers. However, the authors showed that 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 et al. (2017) studied the effects of hard and social skills training in the Dominican Republic, and found that the training had larger effects on labor market outcomes for women, particularly on labor participation.

Some researches have 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 *Jóvenes en Acción (JeA)*. The authors found an increase of 4% in males' probability to be a formal employee, while the females' probability increase was around 3%. Regarding formal earnings, the authors found an average increase of COP 35,000 for 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) carried out a randomized experiment to study the effects of vocational training programs with three random assignments. The first step of randomization chooses treated individuals. The second step of randomization allocates treated individuals in hard-skills-intensive vocational training or social-skills-intensive vocational training. The last step of randomization provides a stipend for some participants. The authors highlight two main findings: first, the program focused on hard skills has higher short-term effects on overall employment. In addition, the program focused on social skills has increased employment and earnings in the long run.

Finally, a first approximation to the impact evaluation of *Empleando Futuro* was made by Chaparro et al. (2020). Nonetheless, there are different methodological aspects that should be taken with caution. We contribute to an application of a novel Difference-in-Differences method that allows the estimation of average treatment effects on the treated (ATT) in a staggered design, which allows us to obtain an unbiased estimation and a robust parallel trends test, even when covariates are included.

The structure of paper is as follow. After this introduction, Section 2 describes the *Empleando Futuro* design, accounting for the payment-for-success scheme. Section 3 details the data sources and reports the descriptive statistics. Section 4 presents the evaluation framework and estimation procedure. Section 5 discusses the obtained results. Finally, Section 6 concludes.

2 Program description

Empleando Futuro was the first payment-for-sucess job training program carried out in a developing country. The program targeted skills training and job searching support of vulnerable and unemployed individual.The program took place between June 2017 and December 2018. The payment-for-success design is a novel approach that seeks to increase both the effectiveness and the quality of social programs, where total or partial payment depends on the achievement of previously agreed employment results. This scheme involves three main actors: the public sector, 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 finances the executors working capital: inputs, activities, and operative costs of intervention. Third, the executors design the program and implement it with targeted individuals. Fourth, an independent organization verifies if the agreed goals were achieved. Finally, if the 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 included a range of employment support policies, such as skills training, psychosocial support, and intermediation services for job placement and retention (for at least three months in a formal employment) for vulnerable and unemployed individuals in Bogotá, Cali, and Pereira.[2] The outcome funder 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 trained the participants in social-and-hard skills and assisted trained in job-search process. These executors were chosen through closed bidding. Finally, *"Deloitte"* was the independent auditor of results.

[2]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*[3] 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*[4] score between 0 and 41.74 points, belong to *Red Unidos*[5], and/or belong to the Colombian registry of victims of the armed 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 [6], and those who did not have formal employment at the start of the program (Instiglio, 2019).

The program's intervention scheme defined seven steps in the treatment process. Each executor had to meet this pathway mandatorily. The steps are as follows: socio-demographic characterization; socio-occupational orientation; pre-training and post-training evaluation; training in hard and soft skills; psychosocial support; labor intermediation; and accompaniment after employment attachment. Psychosocial support was very important to avoid the desertion of the program. First, socio-occupational orientation seeks to help participants to build their life project plan. It is relevant because participants can identify their strengths and weakness and be targeted to improve them. Second, the program assigned a psychologist to the individuals treated to identify needs and motivate them to complete training courses preventing them from quitting the program. We highlight the program's effort to recognize that the targeted people were low-income individuals that need more motivation than less vulnerable individuals (Instiglio, 2019).

Table 1 reports the registrants' path along the program. We identified 1,352 individuals who did not provide the complete information or didn't meet the eligibility criteria (see rows 1 and 2). Next, 2,795 registrants provided complete information and met the eligibility criteria but did not complete their enrollment process[7], which will be used as the control group[8] (see row 3). Rows 4 and 5 show the individuals with incomplete treatment: those who did not complete their training process (see row 4), and those who complete their training process but did not participate in the job-search assistance module (see row 5). Finally, the treatment group[9] consists of 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.

[3] *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).

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

[5] 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.

[6] Programs such as *Inclusión Productiva*, *Empleo para la prosperidad*, *Mi Negocio*, among others.

[7] The reason was not recorded by the program

[8] 1,982 in Bogotá, 760 in Cali, and 53 in Pereira

[9] 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**	

Eleven treatment groups started their staggering job training programs and intermediation processes between June 2017 and August 2018. The vast majority of participants started their treatment process in the third quarter of 2017: 327 individuals in July, 341 individuals in August, and 138 individuals in September[10]. 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 executor and 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**

As part of their operating contracts, executors had to offer 100 hours of training at least, and 300 hours of training at most. Independently, executors chose whether to offer greater emphasis on soft skills or hard skills training. The hard skills courses targeted specific skills with the goal of increasing increase job performance. Executors offered basic computer systems courses, customer service, and administrative assistant modules, among others. The social skills courses targeted improving work environment. Executors offered communication skills courses, relational skills, and conflict management, among others.

[10] 962 participants started their treatment process in 2017: 1 individual in June, 327 individuals in July, 341 individuals in August, 138 individuals in September, 138 individuals in October, 20 individuals in November, and 26 individuals in December. 654 participants started their treatment process in 2018: 73 individuals in January, 1 individual in March, 1 individual in May, 344 individuals in June, and 235 individuals in July

We are interested in analyzing the differences between the training approaches. 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 groups: high-intensity both in hard and social skills formation ($High_{hard} - High_{social}$), high-intensity in hard skills but low-intensity in social skills ($High_{hard} - Low_{social}$), low-intensity in hard skills but high-intensity in social skills formation ($Low_{hard} - High_{social}$), and low-intensity both in hard-and-social skills training ($Low_{hard} - Low_{social}$). For instance, in $High_{hard} - High_{social}$ 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 the number of individuals in each treatment category.

Table 3: Training intensity approaches

	$High_{hard}-$ $High_{social}$	$High_{hard}-$ Low_{social}	$Low_{hard}-$ $High_{social}$	$Low_{high}-$ Low_{social}	**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**

3 Data and descriptive statistics

3.1 Data sources

We track both treatment and control groups matching the *Sistema de Gestión de Desempeño* records, *SISBEN*, and *PILA* database, through participant's national 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 job training and job search assistant program. 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 they met the eligible criteria. Demographic characteristics included information like participant id, age, gender, ethnicity, city of residence, city of birth, civil status, education level, their home composition, among others. Furthermore, for treated participants it 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.

- ***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. Access to the *PILA* database was granted by the Colombian Ministry of Labor. Wage data was not available due to individual data disclosure restrictions.

- ***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).

3.2 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 into account.

Rows 3, 4, and 5 in Table 1 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, the treatment group has a higher score than the control group. Despite this, both groups, on average, are classified as *SISBEN* level 1. The last 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 a summary of 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.

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

4 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-sectional 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*:

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

where $f_{i,t}$ is a dummy equal to 1 if individual i was reported in *PILA* at month t, δ_i is an individuals fixed effect, λ_t is a monthly fixed effect, $\mathbb{1}[T = t]$ is an indicator function that turns on in the periods after treatment, D_i is equal to 1 for every treated individual and equal to 0 for every untreated individual, and X_i 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 parameter of interest, as it captures the differences in labor formality rates between participants and not participants of *Empleando Futuro*, conditional on controls. Under the canonical TWFE specification, β measures the causal effect of the program on the probability of getting a formal job. The monthly changes in the formality rate of the control group build a counterfactual for treated individuals, had they not participated in the program.

From a dynamic context, Equation 2 describes the dynamic TWFE specification:

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

where $D_{i,t}$ 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 month when training began (τ_i^*) and are zero in all months for non-training individuals. Each θ_t parameter 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 individuals 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 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 the 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 the TWFE shortcomings, our paper estimates the effects of *Empleando Futuro* on the probability of obtaining a formal job up to March 2020, using the ATT (g, t) causal parameter proposed by Callaway and Sant'Anna (2021). The Callaway and Sant'Anna's 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, is appropriate to use Doubly-Robust (D-R) estimator which allows covariate-specific trends across groups (Callaway and Sant'Anna, 2021; Sant'Anna and Zhao, 2020).

There are two additional reasons to use the D-R estimator. First, it 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 Sant'Anna (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 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 the 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 of getting a formal job. This equation is built under the D-R estimator and the not-yet-treated control group (ny).

$$F_{i,g,t} = \alpha_{g,t} + \eta_{g,t} \cdot G_{i,g} + \rho_{g,t} \cdot \mathbb{1}\{T_g = t_g^*\} + \tilde{\beta}_{g,t}(G_g \cdot \mathbb{1}\{T_g = t_g^*\}) + \omega X_i' + \epsilon_{g,t}, \quad (4)$$

where , $F_{i,g,t}$ is a binary variable which takes the value of one if the individual i of group g was reported in *PILA* at period t, and it is equal to zero otherwise. The $G_{i,g}$ variable identifies each of twelve treatment groups. This variable takes the value of 1 if the individual i belongs to the group g, and takes zero value otherwise. $\mathbb{1}\{T_g = t_g^*\}$ is an indicator function that turns on the periods after treatment starting for group g (t_g^*). ω is the covariates parameter, X_i is the vector of covariates, and $\epsilon_{g,t}$ is the error term. $\tilde{\beta}_{g,t}$ is our parameter of interest, as it captures the effect of *Empleando Futuro* on the probability of getting a formal job. It will be interpreted as a causal effect when $t \geq g$, and must be interpreted as pre-test of Parallel Trends Assumption (PTA) otherwise. Using "Not-Yet-Treated" individuals as control group, the "Doubly-Robust" estimation of $\tilde{\beta}_{g,t}$ is:

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

5 Results

In this section, we study the effect of *Empleando Futuro* on labor formality. We compare our results when the TWFE static and dynamic estimation is used, to our results coming from Callaway and Sant'Anna's method. Also we provide the Goodman-Bacon decomposition for the program. We think that allowing for treatment effect heterogeneity measures accurately 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 and Sant'Anna's aggregations (as of Equation 4). Panel B shows the results by gender, conducting the Callaway and Sant'Anna dynamic aggregations (as of Equation 4) for both females and males. Panel C shows the results by training intensity, conducting the Callaway and 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 of getting a formal job by 6.8 percentage points (pp) 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. We also estimate the effect of *Empleando Futuro* with the method proposed by Callaway and Sant'Anna (2021). We find that when treatment staggering is taken into account, the effect on the probability of getting formal employment is statistically significant at 95% confidence level and substantially higher than TWFE results. Row 2 of Panel B in Table 6 shows the "Simple aggregation" of equation 4. Overall, we found that the *Empleando Futuro* participants had a probability of 21 pp higher of getting a formal job, after participating in the program.

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

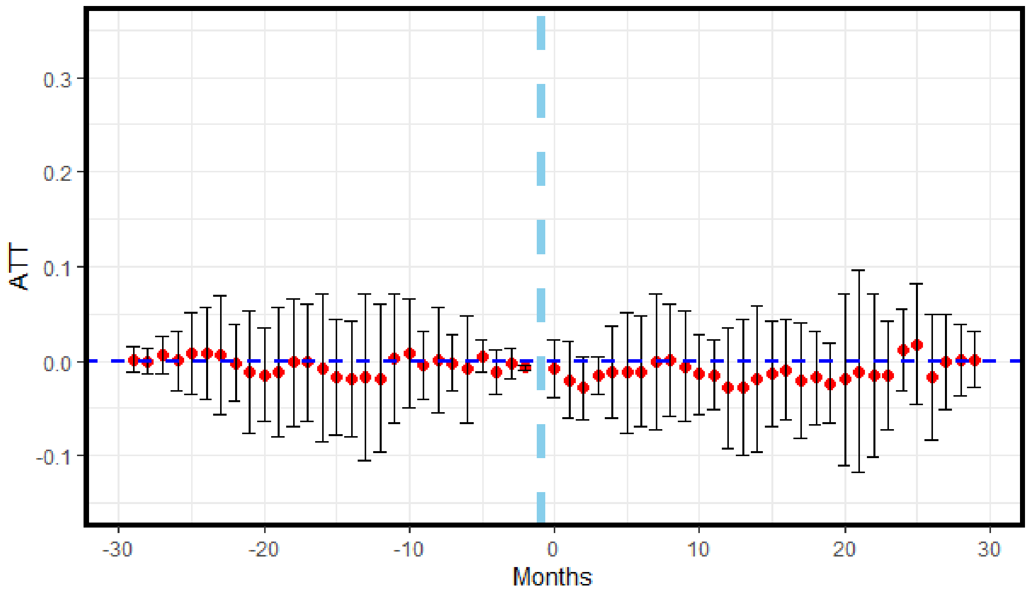
	ATT	Std.Error	95% Conf. Inter.	
Panel A. Overall results				
TWFE estimation	0.0678	0.0192	-0.0150	0.1506
Simple aggregation	0.2162*	0.0118	0.1931	0.2394
Dynamic aggregation	0.2122*	0.0114	0.1898	0.2346
Group aggregation	0.2182*	0.011	0.1966	0.2398
Calendar aggregation	0.2048*	0.0108	0.1837	0.2258
Panel 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				
$High_{hard} - High_{social}$ dynamic aggregation	0.2068*	0.0182	0.1712	0.2424
$High_{hard} - Low_{social}$ dynamic aggregation	0.1967*	0.0185	0.1605	0.2330
$Low_{hard} - High_{social}$ dynamic aggregation	0.2452*	0.0245	0.1971	0.2932
$Low_{hard} - Low_{social}$ dynamic aggregation	0.2045*	0.0257	0.1542	0.2548

Note: The star (*) reports the significance at the 95% level. All Panel A and Panel C regressions include the following controls: residency city, age, gender, and the duo formed by being a SISBEN beneficiary and the SISBEN score. All Panel B regressions include the following controls: residency city, age, and the duo formed by being a SISBEN beneficiary and the SISBEN score. Column 1 reports the β coefficients, interpreted as the ATT for the canonical TWFE estimation and each Callaway and Sant'Anna's aggregation. Panel A reports the TWFE estimation (Equation 1) and the Callaway and Sant'Anna aggregations (as of Equation 4) for the overall database. The TWFE standard errors were clustered by city. The Callaway and 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 and Sant'Anna (2021) and Table 4). In addition, Callaway and 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 and 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 and 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 the event-study estimation for *Empleando Futuro*, calculated by Equation 2. On the horizontal axis is the time relative to the treatment's start. On the vertical axis are the ATT effects by the time of exposure to treatment. The vertical dashed line indicates the start of the program. 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 good 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-effect" (Card et al., 2018; Crépon and Van Den Berg, 2016; Diaz and Rosas, 2016; Osikominu, 2016; Spinnewijn, 2013). 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: Dynamic TWFE

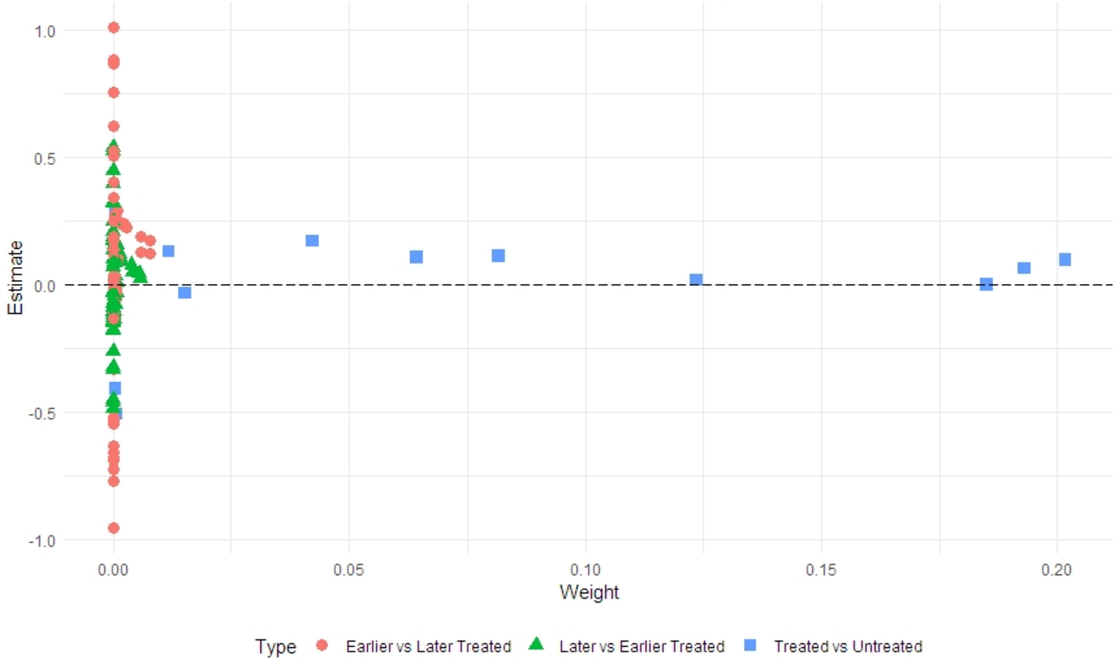


Note: this figure shows the event-study estimation for *Empleando Futuro*, calculated by Equation 2. The dynamic TWFE's standard errors were bootstrapped (400 interactions) and clustered by city.

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 of *Empleando Futuro*. The horizontal axis corresponds to the weight that this comparison between the treatment and control groups should have. The vertical axis corresponds to the estimation proposed by Goodman-Bacon Decomposition.

Figure 2: Goodman-Bacon decomposition



Note: this figure shows the Goodman-Bacon decomposition for *Empleando Futuro*, calculated by Equation 3.

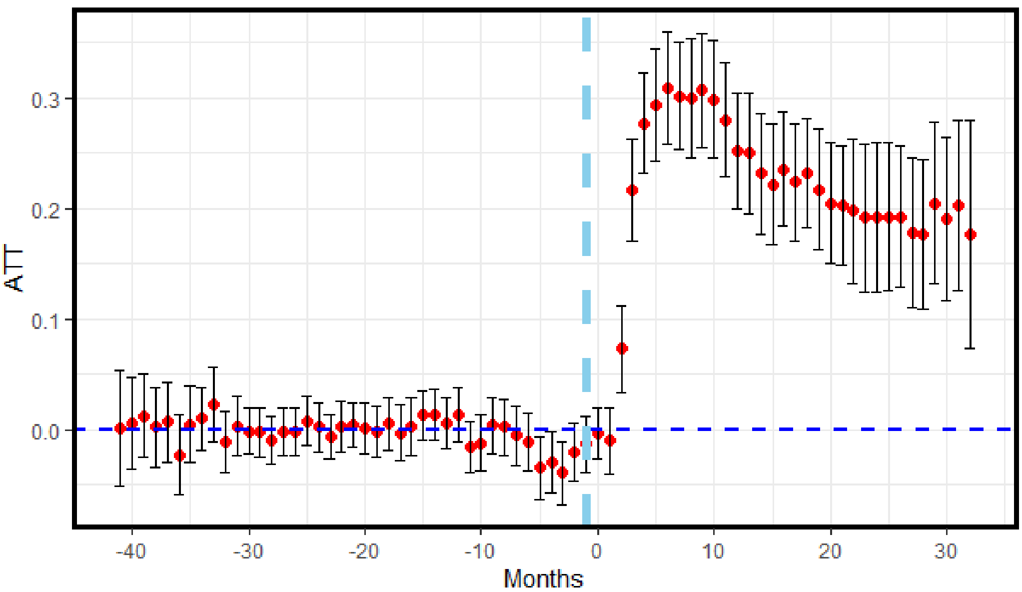
Figure 2 shows that the Treated vs Untreated individuals comparisons must have the highest weights. These weights sum the 92% of overall decomposition. Nevertheless, there is heterogeneity within this comparison. The highest weights were assigned to the groups that start their treatment in July 2017 (0.19) August 2017 (0.20), June 2018 (0.18), and July 2018 (0.12). This result suggests that we must take into account the staggering start period of *Empleando Futuro* to conduct our study on the program's effects on the probability of obtaining 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, when compared to 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 dynamic effects of *Empleando Futuro* on the probability of obtaining formal employment for treated individuals. On the horizontal axis is the time relative to the treatment's start. On the vertical 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 before treatment. It implies the met of the parallel trends assumption. Therefore, the control group is a good counterfactual for the treatment 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 onwards, 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.

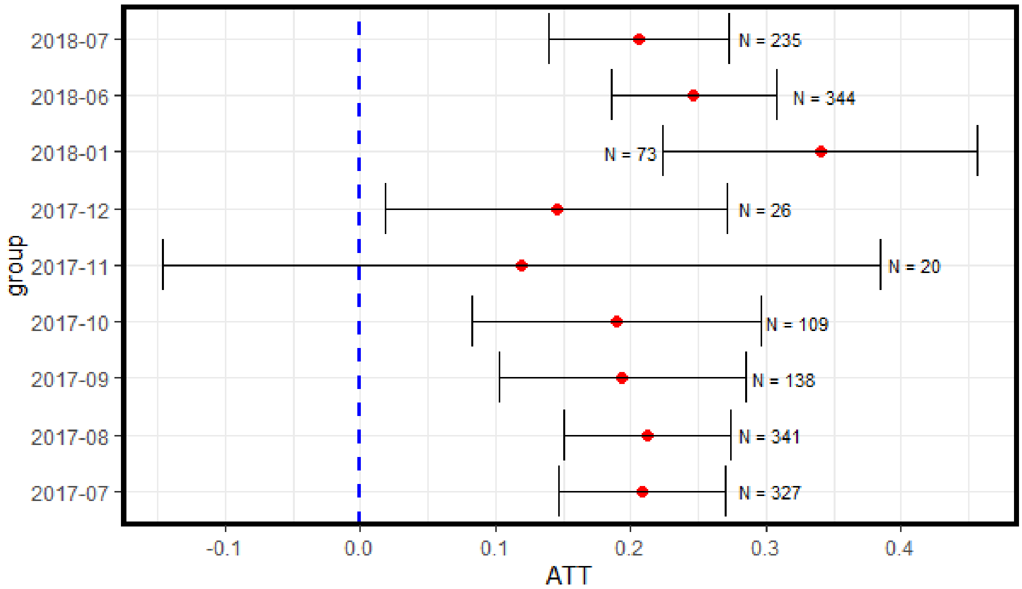
Figure 3: Dynamic Aggregation



Note: this figure shows the Dynamic Aggregation, calculated from Equation 4.

Figure 4 shows the treatment effect heterogeneity of *Empleando Futuro*, plotting the average effect experienced by each group. On the vertical axis are the different groups, represented by the start date of their training. On the horizontal 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.

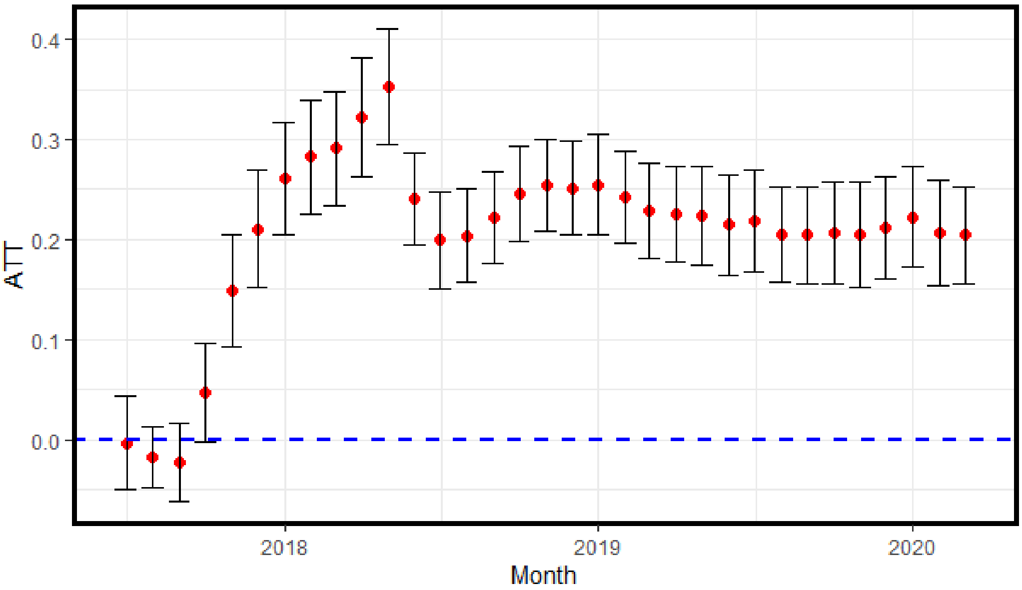
Figure 4: Group Aggregation



Note: this figure shows the Group Aggregation, calculated from Equation 4.

Figure 5 accumulates the experienced effect of all groups at a specific months. 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 Panel A in Table 6 shows that the average accumulative effect on the likelihood of getting a formal job was 20 pp for the treatment group's individuals.

Figure 5: Calendar Aggregation

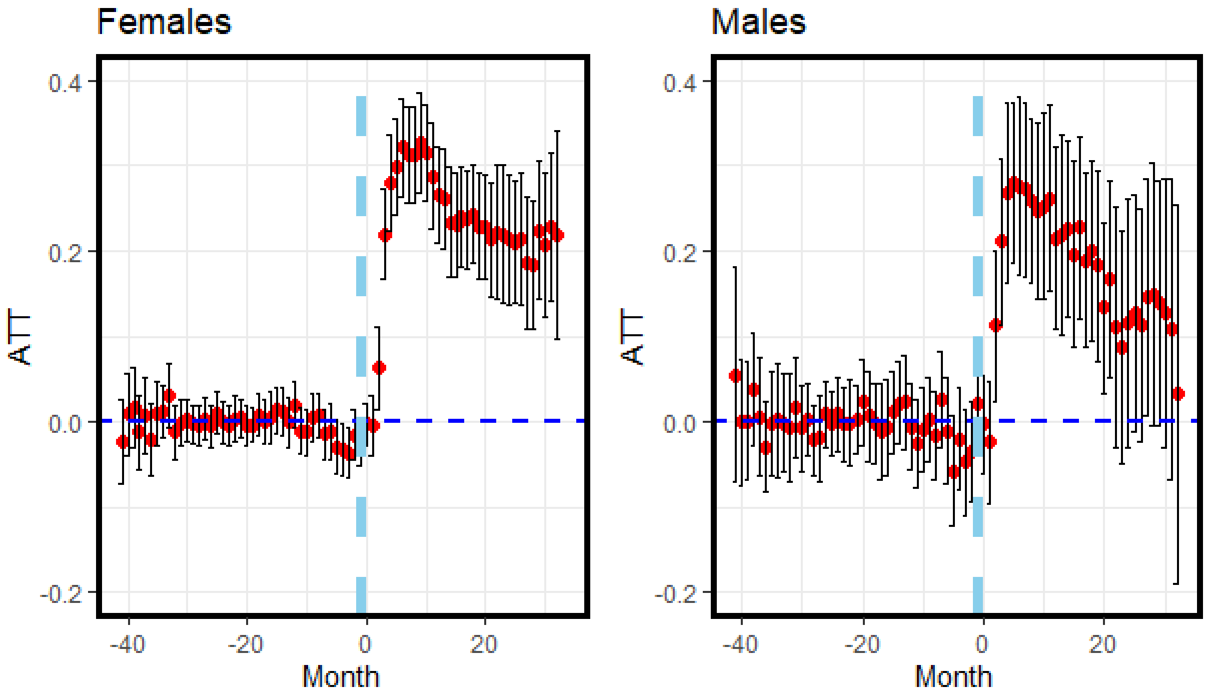


Note: this figure shows the Calendar Aggregation, calculated from Equation 4.

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 of 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: Dynamic Aggregation by gender

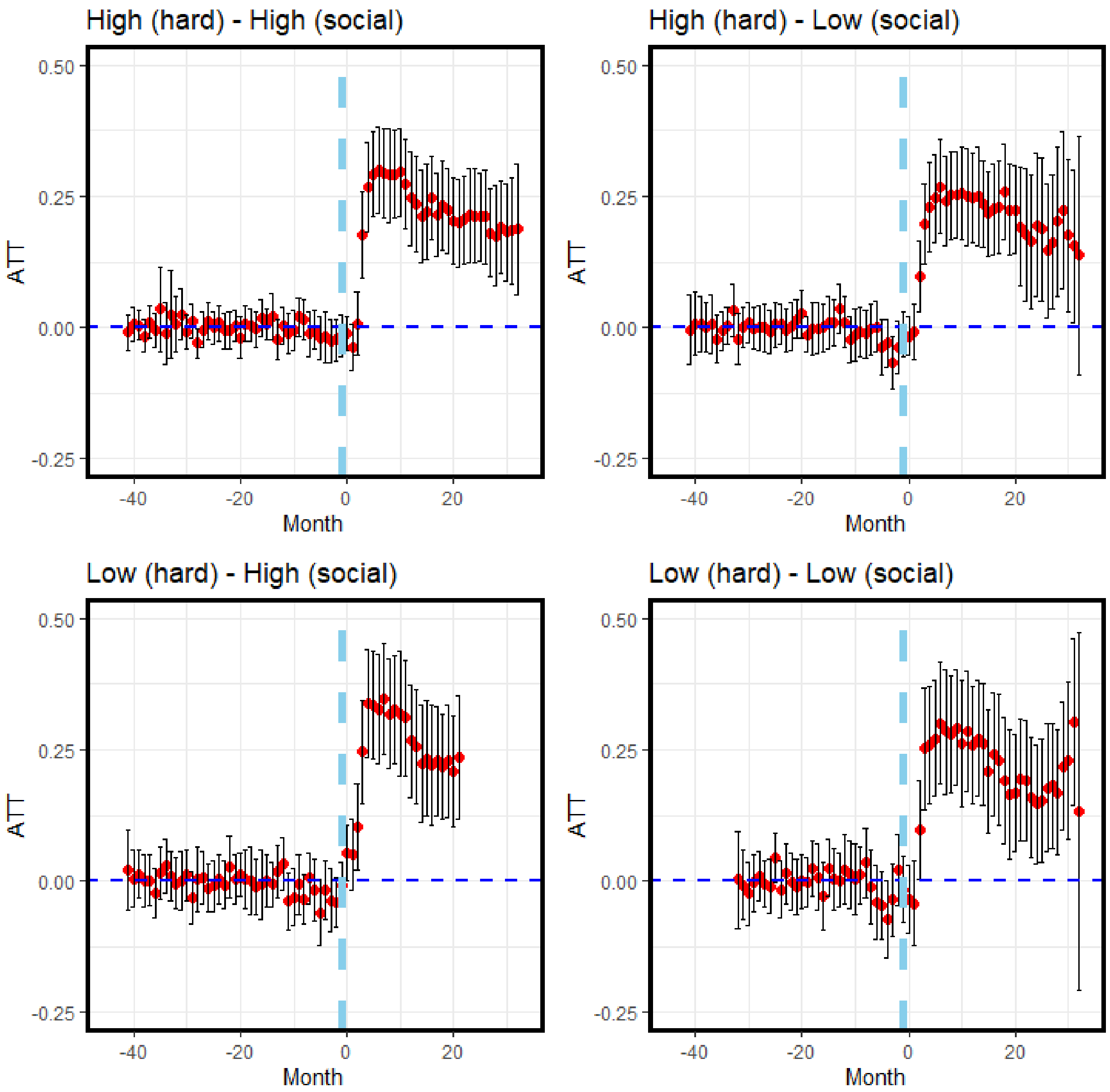


Note: this figure shows the Dynamic Aggregation calculated from Equation 4 distinguished by gender.

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



Note: this figure shows the Dynamic Aggregation calculated from Equation 4 distinguished by training intensity.

6 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.

From the previous related literature, the programs evaluated by Kugler et al. (2022) and Barrera-Osorio et al. (2020) are the most similar to *Empleando Futuro*. Kugler et al. (2022) found that participating in *Jóvenes en Acción* increased the probability of getting a formal job in 7,1 pp in short-term and 5.7 pp in long-term. Barrera-Osorio et al. (2020) found that participating in Inclusive Employment Program increased participants' probability of employment. Those who were assigned to higher social skills training emphasis increased their probability of getting a job in 10 pp. Those who were assigned to higher technical skills training emphasis increased their probability of getting a job in 13 pp.

Nonetheless, our results differ largely from the outcomes found in Kugler et al. (2022) and Barrera-Osorio et al. (2020). It could be due to two features of the program. First, *Empleando Futuro* was a Social Impact Bond (SIB) designed in a payment-for-success scheme. It implied that executors must be synchronized to the labor demand, in order to offer pertinent training and achieve the employment attachment of treated individuals. Second, psycho-social support offered by *Empleando Futuro* was relevant to avoid the desertion in the program. Motivation and socio-occupational orientation might enhance the effects of training programs, strengthening aspects such as self-esteem, confidence, and resilience.

The next steps in our research are to estimate the *Empleando Futuro* effects using other novel techniques in staggering Difference-in-Differences setups. Proposals made by de Chaisemartin and D'Haultfoeuille (2022), Borusyak et al. (2021), and Sun and Abraham (2021) are relevant for us. Further, we will measure the effect of *Empleando Futuro* on other outcome variables, from an intensive margin perspective. Our database allows us to estimate the effects on quality of work, labor income, and labor stability of treated individuals. It is important to analyze if a payment-for-success program improves both the extensive (formality rates) and the intensive (quality of job) margins. Finally, we will conduct a cost-benefit analysis of the program.

References

Acevedo, P., Cruces, G., Gertler, P., and Martinez, S. (2017). LIVING UP TO EXPECTATIONS: HOW JOB TRAINING MADE WOMEN BETTER OFF AND MEN WORSE OFF. Technical report.

Attanasio, O., Guarín, A., Medina, C., and Meghir, C. (2017). Vocational training for disadvantaged youth in Colombia: A long-term follow-up. *American Economic Journal: Applied Economics*, 9(2):131–143.

Attanasio, O., Kugler, A., and Meghir, C. (2011). Subsidizing vocational training for disadvantaged youth in Colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 3(3):188–220.

Barrera-Osorio, F., Kugler, A. D., and Silliman, M. I. (2020). Hard and Soft Skills in Vocational Training: Experimental Evidence from Colombia. Technical report.

Biewen, M., Fitzenberger, B., Osikominu, A., and Paul, M. (2014). The effectiveness of public-sponsored training revisited: The importance of data and methodological choices. *Journal of Labor Economics*, 32(4):837–897.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. pages 1–54.

Brunner, E., Dougherty, S., and Ross, S. L. (2021). The Effects of Career and Technical Education: Evidence from the Connecticut Technical High School System. Technical report.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Card, D., Kluve, J., and Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3):894–931.

Chakravarty, S., Lundberg, M., Nikolov, P., and Zenker, J. (2019). Vocational training programs and youth labor market outcomes: Evidence from Nepal. *Journal of Development Economics*, 136:71–110.

Chaparro, J. C., García, G. A., and Cardona, M. Y. (2020). Evaluación de Impacto de "EMPLEANDO FUTURO". Technical report.

Crépon, B. and Van Den Berg, G. J. (2016). Active Labor Market Policies. *Annual Review of Economics*, 8:521–546.

de Chaisemartin, C. and D'Haultfoeuille, X. (2022). Difference-in-Differences Estimators of Intertemporal Treatment Effects. *SSRN Electronic Journal*.

Diaz, J. J. and Rosas, D. (2016). Impact Evaluation of the Job Youth Training Program Projoven Labor Markets and Social Security Division Inter-American Development Bank. Technical report.

Galitopoulou, S. and Noya, A. (2016). Understanding Social Impact Bonds. *Oecd*, pages 1–36.

González-Velosa, C., Ripani, L., and Rosas Shady, D. (2012). How Can Job Opportunities for Young People in Latin America be Improved? *IADB Technical Note*, (IDB-TN-345).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Groh, M., Krishnan, N., McKenzie, D., and Vishwanath, T. (2016). 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).

Gustafsson-Wright, E. and Gardiner, S. (2016). Using Impact Bonds to Achieve Early Childhood Development Outcomes in Low-and Middle-Income Countries Using Impact Bonds to Achieve Early Childhood Development Outcomes in Low-and Middle-Income Countries. *Global Economy and Development Program – BROOKINGS*.

Hirshleifer, S., McKenzie, D., Almeida, R., and Ridao-Cano, C. (2014). The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey. *Economic Journal*, 126(597):2115–2146.

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

Intiglio and GPOBA (2018). A Guide For Effective results-based financing strategies. pages 1–90.

Kugler, A., Kugler, M., Saavedra, J. E., and Herrera-Prada, L. O. (2022). Long-Term Educational Consequences of Vocational Training in Colombia. *Journal of Human Resources*, 57(1):178–216.

Osikominu, A. (2013). Quick job entry or long-term human capital development? The dynamic effects of alternative training schemes. *Review of Economic Studies*, 80(1):313–342.

Osikominu, A. (2016). The dynamics of training programs for the unemployed. *IZA World of Labor*.

Reis, M. (2015). Vocational training and labor market outcomes in Brazil. *B.E. Journal of Economic Analysis and Policy*, 15(1):377–405.

Sant'Anna, P. H. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122.

Silliman, M. and Virtanen, H. (2022). Labor Market Returns to Vocational Secondary Education. *American Economic Journal: Applied Economics*, 14(1):197–224.

Spinnewijn, J. (2013). Training and search during unemployment. *Journal of Public Economics*, 99:49–65.

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

7 Appendix

Table 7: *Empleando Futuro.* Canonical TWFE estimation

	Estimate	Std.Error	t-value	p-value	
Probability of getting a formal job	0.06778	0.01926	3.51947	0.07211	
City of residence (Cali)	-0.0445	0.00071	-62.470	0.00026	*
City of residence (Pereira)	-0.0661	0.00671	-9.8524	0.01015	*
SISBEN score	0.00111	0.00023	4.91043	0.03906	*
SISBEN missings	0.02307	0.01038	2.22349	0.15621	
Age	0.00430	0.00057	7.50937	0.01728	*
Gender (Females)	-0.0564	0.00946	-5.9600	0.02702	*

Note: This table reports the estimation of Equation 1. The star (*) reports the significance at the 95% level. The regression includes the following controls: residency city, age, gender, and the duo formed by being a SISBEN beneficiary and the SISBEN score. The TWFE standard errors were clustered at the city level.