Public Disclosure Authorized

Public Disclosure Authorized

Policy Research Working Paper 11117

Promoting Innovative Startups

Quasi-Experimental Evidence from Tunisia

Nadia Ali
Massimiliano Cali
Bob Rijkers



WORLD BANK GROUP
Economic Policy Global Department &
Development Research Group
May 2025

Policy Research Working Paper 11117

Abstract

This paper evaluates Tunisia's "Startup Act," a policy initiative to foster innovative firms through a "start-up" label and a bundle of incentives, including reduced social security contributions, corporate tax exemptions, easier access to foreign exchange, and simplified customs procedures. Detailed data on the program's selection process allow identifying marginal entrants and rejects, and hence limit selection on unobservables. Using a difference-in-differences strategy, the program is shown to increase survival and promote job creation. A back-of-the-envelope cost-benefit calculation suggests that the program is cost effective.

This paper is a product of the Economic Policy Global Department and the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at nali5@worldbank.org; mcali@worldbank.org; and brijkers@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Promoting Innovative Startups: Quasi-Experimental Evidence from Tunisia*

Nadia Ali* Massimiliano Calì† Bob Rijkers‡§

*Columbia University, nadia.ali@columbia.edu

†The World Bank, mcali@worldbank.org

‡The World Bank and Utrecht School of Economics, Utrecht University, brijkers@worldbank.org.

§This paper has been supported by the World Bank's Global Tax Program (GTP), Umbrella Facility for Trade, Labor and Gender Research Program of the Chief Economist's Office of the Middle East and Northern Africa region, Knowledge for Change Program (KCP), and Research Support Budget. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank of Reconstruction and Development, the World Bank, and their affiliated organizations or those of the Executive Directors of the World Bank, or the countries they represent. We thank Smart Capital and the Tunisian Institute of Statistics for generously sharing their data with us and hosting us, particularly Ramzi Channouf and Fathi El Hajamara. Jawhar Abidi and Mhamed Ben Saleh provided excellent research assistance. The paper also benefitted from useful comments from seminar audiences at the World Bank and Columbia University, and feedback from Miriam Bruhn, Michael Best, Fernando Blanco, Irene Botosaru, Bev Dhalby, David McKenzie, Simon Lee, Daniel Lederman, Nelly ElMallakh, Patrick Farell, Suresh Naidu, Bernard Salanie, Eric Verhoogen, and Erik Stam. All errors are our responsibility.

1 Introduction

Start-up firms disproportionately drive job creation and productivity growth (Haltiwanger et al., 2013), despite their high failure rates. To promote the creation, survival and expansion of start-ups, especially in high-tech sectors, both developed and developing countries are increasingly implementing targeted policies (Juhász et al., 2023). Yet evidence on their effectiveness remains limited, particularly for developing countries. This is unfortunate, because firms in developing countries tend to grow at much lower rates compared to their peers in developed nations and well-paid jobs in productive firms are scarce (Hsieh and Klenow (2014), Hsieh and Olken (2014)).

In this paper, we help fill this knowledge gap by examining the impact of a recent flagship program supporting start-ups in Tunisia, the Startup Act, which was introduced in 2019 to promote the creation and growth of innovative firms. We leverage the multi-stage nature of the program's selection process to identify firms that were marginally accepted or rejected and use a combination of program and administrative firm-level data to provide rigorous estimates of the impact of program participation on firms' survival and performance.

Our analysis focuses on the program's start-up label initiative, which awards selected firms a special start-up designation that grants them access to a range of benefits. These benefits include tax incentives in the form of coverage of social security contributions and profit tax exemptions, eased foreign currency restrictions, and simplified customs procedures. For firms less than a year old, the initiative may also grant a stipend to up to three founders without other employment.

To estimate the impact of the program, we use a difference-in-differences design, comparing changes in performance between firms that received the label and those that applied but were not selected. Our sample consists of 466 firms that applied between March 2019 and December 2021 (henceforth the Full Sample), with performance measured from 2018 through 2022.

The main empirical challenge is selection bias: firms that received the label may differ systematically from firms that were not selected in ways that also affect outcomes. This concern is partly mitigated by the design of the assessment process. Judges evaluate firms based on product innovation and scalability, without explicitly considering expected commercial success. As a result, selection may not be strongly related to current or future firm performance, which is consistent with the lack of significant differences between treated and control firms across observable variables. To address potential selection bias,

we control for a rich set of observable characteristics. In addition, we exploit detailed administrative data on the program's multi-stage evaluation process. Each application is reviewed by a nine-member committee of judges, with each member voting to accept, reject, or invite the applicant to pitch online. Firms that received five or more approval (or rejection) votes were immediately selected (or rejected), while those with fewer than five votes for or against proceeded to a pitch stage for a further evaluation. These "pitch-stage" firms—118 in our sample—are arguably more comparable to each other than firms accepted or rejected outright. Our preferred specification compares outcomes between pitch-stage firms that did and did not ultimately receive the label. We also report estimates using the Full Sample, which offers greater statistical power but may be more susceptible to selection bias.

Our main finding is that program participation promotes both survival and job creation. Labeled firms are 18 percentage points (pp) more likely to survive until the end of the sample period relative to a baseline of 53%. On average, treated firms increase their employment by 2.0 workers relative to a mean of 1.8 workers and their wage bill by 69,000 Tunisian Dinars (TND) (~ 23,000 USD) relative to a mean of 61,000 TND (~ 20,333 USD).[1] We find positive but imprecisely estimated effects on sales and cannot reject the null that program participation does not impact profits. These results are robust to various checks, including using the Poisson estimator, excluding peak COVID-19 quarters, and various placebo tests. A back-of-the-envelope calculation suggests that the program has been cost effective thus far.

Although our data do not allow us to quantify the impact of each component of the program (due to a lack of experimental variation in their intensity), an analysis of heterogeneity in program impacts points to access to foreign currency, preferential customs treatment and lighter administrative procedures being enablers of firm growth. In qualitative interviews, participating entrepreneurs also highlighted the importance of reduced social security contributions, which is perhaps not surprising since these account for the bulk of the tax incentives the program provides.

This paper builds on and contributes to two strands of literature. First, it adds to the research on entrepreneurship and firm growth in developing countries (see reviews by Quinn and Woodruff (2019), Verhoogen (2023), McKenzie and Woodruff (2014)). Previous studies typically focused on subsistence entrepreneurship and examined interventions like training (Chioda et al., 2021), capital provision (De Mel et al., 2008;

[1] Throughout the paper, we use an exchange rate of 3 TND/USD.

McKenzie and Woodruff, 2008, 2014; Fafchamps et al., 2014), business services (Anderson and McKenzie, 2022), and coaching (Anderson et al., 2021; Iacovone et al., 2022; Cusolito et al., 2023; Bruhn et al., 2018). These interventions are typically low-cost, and short-term, making them well-suited for randomized evaluations.[2] In contrast, our study focuses on a program designed for growth-oriented, technology-focused startups. Such interventions are generally more costly and multifaceted. Due to the smaller pool of eligible firms, these types of interventions are less suited for experimental evaluation, and we are unaware of randomized control trials assessing similar policy initiatives. Our study provides novel quasi-experimental evidence on the effectiveness of such programs.

Second, we add to the nascent literature evaluating support programs for high-tech start-ups. Previous evaluations have studied seed accelerators, which provide education and mentorship for start-up founders (Hochberg and Fehder, 2015), as well as incubators, angel investors and venture capitalists, which offer funding and advisory services (Lerner et al., 2015). The program we study differs from these venture support structures by offering a unique state certification, fiscal and trade advantages, and subsidized wage contributions that aim to promote job creation among innovative start-ups.[3] While wage subsidies have often, but not always, been shown to boost employment (Betcherman et al., 2010; Bruhn et al., 2018; De Mel et al., 2019; Huttunen et al., 2013), there is little evidence on how they function within specialized programs aimed at high-tech or growth-oriented firms in developing countries. To the best of our knowledge our paper provides the first assessment of a government program designed to promote innovative start-ups in a developing country context.

The paper proceeds as follows. Section 2 describes the institutional background. Section 3 discusses our data sources. Section 4 presents the empirical strategy and balance tests, and section 5 the main results, including a back-of-the-envelope cost-benefit analysis of the program. Section 6 concludes.

2 Institutional background

In 2018, Tunisia's Parliament passed the "Startup Act" –a set of regulations and fiscal incentives to encourage the development of start-ups, with the goals of promoting job

[2]For instance, McKenzie and Paffhausen (2023) list 35 randomized evaluations of such interventions as of 2017.

[3]It is perhaps most similar to the Italian Start-up Act, which has been shown to increase firms' employment and access to equity (Menon et al., 2018; Biancalani et al., 2022).

creation and innovation. The Act is coordinated and implemented by Smart Capital, a public-private management company, which generously shared its data with us.[4] One of the objectives of the Startup Act is to generate 10,000 jobs within 5 years.[5] Similar programs in Europe include the Start-up Act in Italy and The New Companies Promotion Act in Austria.

The Startup Act provides benefits to investors (a tax rebate, exemptions from the capital gains tax), entrepreneurs (subsidized leave from their current jobs for start-up creation), and start-ups. We focus on the latter part of the Startup Act that labels firms as start-ups and aims to promote their growth. The label allows the firms to access the advantages set by the Start-up Act and lasts until the 8th year since the start-up creation. The program defines a start-up as a "temporary organization designed to search for a repeatable and scalable business model."

2.1 The advantages of the Start-up Label

The program grants labeled firms a set of advantages to assist them in this search. We describe in turn the main incentives and the relevant Tunisian regulations. We briefly return to the different incentives in section 5.4 when we discuss potential mechanisms.

1. Coverage of Social Security contributions For labeled firms, the government covers the full social security contribution for the duration of the label. These contributions are shared between employer and employee, each party paying 16.57% and 9.18% of the employee's monthly gross salary, respectively.

2. Corporate Tax Exemptions Labeled firms are exempt from corporate taxes, which were equivalent to 25% of corporate profits in 2019-20 and to 15% thereafter.

3. Special Account in Foreign Currencies In Tunisia, almost all firms need approval from the central bank to access foreign currency in line with the foreign exchange code, and importers must be issued a letter of credit in order to pay for their imports in foreign currency. The program allows resident labeled firms to open a bank account in foreign currency for the duration of the label.

4. Technology Card Every resident firm with activities connected to telecommunication, information technology, education, consulting or research is provided with a "Technology Card", a prepaid international card available to resident

[4]The program is governed by Law No. 2018-20 of April 17, Decree No. 2018-840 of October 11, and Circulars of the Central Bank.

[5]Other objectives include issuing 1000 labels, and generating a cumulative turnover of 1 billion Tunisian Dinars (over 300 million USD) within 5 years.

firms, capped at TND 10,000 (~3,000 USD). The card can be used to pay for various products and services from overseas providers, such as memberships on business social media platforms, cloud storage, software licenses, costs for training overseas, or online advertisements. For labeled firms, the cap is raised tenfold to TND 100,000 (~30,000 USD) per year.

5. Authorized Economic Operator Status In Tunisia, the "Authorized Economic Operator" status is a legal status under the Customs Code. This status gives firms with trade operations access to simpler customs formalities, faster administrative approval, and certain exemptions from technical control on import operations. Labeled firms can apply to become Authorized Economic Operators.

In addition to these advantages, labeled firms have access to an online portal where they can apply for these various advantages.

2.2 Eligibility and selection

Appendix Figure A1 summarizes the selection process. Firms are invited to apply online to participate in the Start-up Label program on a rolling basis. As it is apparent from the eligibility criteria and program conditions, the program has an explicit focus on innovation, high-tech activity, and scalability. It is meant to select firms that aspire and have the potential to become Silicon-Valley-type successes.[6]

In the first round, applicants submit an online application, which consists of a short video that introduces the firm and a demo video showcasing a prototype of the product. Applicants are screened for eligibility based on age, size, and capital structure. Eligible firms must be younger than eight years old, employ fewer than 100 people, and have annual sales below TND 15 million (~ 5 million USD). Additionally, at least two-thirds of the firm's capital must be owned by individuals, investment funds, or foreign start-ups.

Eligible firms meeting these criteria advance to a second round in which their application is evaluated by a 9-member committee of judges, known as the 'College', which decides on who receives the label.[7] The College consists of individuals with expertise within the startup ecosystem, hailing from various backgrounds, including

[6]According to program reports, 77% of program firms are working on a technological innovation, and 13% are developing a hybrid (hardware and software) innovation. The goals of these innovations range from early detection of illnesses to digitization of government services. In order to innovate, nearly a fifth of all program firms are relying on artificial intelligence, the internet of things, or big data.

[7]Members of the College and selected by the prime minister from 25 public and private sector candidates recommended by the Minister of Communication Technologies and Digital Economy. A College president and eight members are appointed for a mandate of three years.

academia, corporate finance, non-governmental organizations, and business consulting.

In monthly online sessions, the College evaluates the written applications firms submit based on degree of innovation (does the firm's business model bring a new solution to a problem?) and scalability (does the firm have the potential to meet demand from a sizable market?). Judges are not remunerated and assess approximately 40 projects each month, a volume that several judges, in interviews, characterized as a substantial workload.[8] Each committee member independently casts one vote in favor of selecting, rejecting, or inviting the applicant to pitch the idea for their firm online. Votes are visible to others, and judges can justify their decisions with comments. Judges may recuse themselves in cases of conflict of interest.[9] Applicants that receive five or more approval (rejection) votes in this round are directly selected (rejected).

Firms that do not receive five votes in either direction proceed to a pitch stage in which they present their application online to the College for five minutes. The presentation is followed by a five-minute question and answer session and judges' deliberation. Selection again requires at least five approval votes. In interviews, judges mentioned that reaching this stage does not necessarily indicate a less innovative or scalable product. Rather, it indicates that the application did not sufficiently clarify these aspects of the startup's products. The fact that judges were undecided in the first round and required more information suggests that pitch stage firms may be more similar to each other than to those accepted or rejected directly, a feature of the selection procedure that we leverage in our empirical strategy to limit selection bias (as will be elaborated upon in Section 4). Rejected firms can reapply after six months, with no apparent stigma within the startup eco-system associated with the rejection.

In their evaluation, judges focus primarily on assessing innovation and scalability, which they interpret as the firm's potential to reach new clients. What qualifies as innovation varies depending on whether applicants pursue local innovation—introducing a novel product within the domestic market—or global innovation targeting international markets. In assessing these features, judges reported that they do not assess the likelihood of commercial success per se, noting that they lack both the time and the mandate to conduct such assessments. As such, they do not evaluate the applications based on the startup's potential for job creation, financial performance, or long-term viability. Rather

[8]Half of these projects are from non registered firms seeking a pre-label, which could eventually grant them a label upon the firm's registration. We do not include them in the analysis as they are not actual firms.

[9]Conflicts of interest include scenarios where a judge holds a financial stake in a candidate company, or has a link to a company in the same product market as a candidate.

than operating as venture capitalists, they focus on verifying that applicants have taken basic steps toward developing a viable prototype. To that end early signs of product adoption, such as informal usage or free trials, are often as valuable to secure approval as actual sale of the product.

The program's emphasis on innovation and scalability is reflected in the activities that firms that apply to receive the label undertake, as shown in Table A1. The table reports the sector classification from program data for the sample that includes all successful applicants as well as rejected firms. Applicants span a range of innovative industries, including e-commerce (25%, treated; 24%, control), artificial intelligence (AI) (12%, treated; 9%, control), business software (9%, treated; 14% control), education tech (9% of treated; 6% control), healthcare and wellness (9%, treated; 12% control), finance (7% of treated, 1% control), and communication (5%, treated; 12%, control).[10] Generally, sectoral representation is balanced across the two groups though we do reject the null hypothesis that they have the same sectoral composition. That is mainly because finance features more prominently as a sector in the selected pool, and communication is more heavily represented among control firms. In the Pitch sample only finance appears to be marginally more represented among treated firms and we cannot reject the null of having the same composition (see Table A2). Examples of projects submitted to the program include e-learning platforms, a digital banking application, a logistics platform connecting carriers and shippers, a virtual reality-based rehabilitation solution, a marketplace for renewable energy systems, and agricultural-waste-based textile dyes.

2.3 Economic context

During the period of analysis Tunisia faced a relatively high level of unemployment (15.3% of the labor force in 2022), particularly among youth, whose unemployment exceeded 37% (INS, 2022). In the same year, Tunisia had 825,707 registered firms, including 56,296 new entrants. However, the overwhelming majority of these firms are very small, with 726,940 employing only a single worker (INS, 2024a). Only a small share of new firms operate in high-tech or innovation-intensive sectors. The 48 sectors with the largest numbers of startup firms, which account for 90% of the startup applicant firms, saw a total of 8,220

[10] Agriculture and food (6%, treated; 4%, control), entertainment/travel (6%, treated; 10%, control), environment and energy (3%, treated; 2%, control), advanced manufacturing and robotics (2% treated and control), real estate (1% of treated firms), and other sectors making up each less than 1% account for the remainder.

new entrants during the period when our sample of firms applied to the program (2019-21), and 13,676 during the overall period of analysis (2018-22).[11] As such, applicants to the startup label comprise a relatively small but non negligible share of all startup firms in Tunisia.

The program focuses on such high-tech and innovation firms and was implemented during a period marked by sluggish economic growth, relatively high inflation, and rising debt. Between 2019 and 2022, the period of our analysis, Tunisia's gross domestic product (GDP) contracted by 2.2% in real terms, primarily due to the severe impact of the COVID-19 pandemic, which coincided with a contraction of GDP of 9% in 2020 (INS, 2024b).

The program period overlapped with the COVID-19 pandemic, raising the possibility that the pandemic altered the program's effects.[12] In Section 5.3, we test for potential COVID-19 impacts and find no strong evidence of substantial effects on the program outcomes within our sample impacts on program effects within our sample.

The subsequent recovery was moderate, hampered by external shocks, and included a hike in commodity prices in 2022. While unemployment is substantial, informality, such as unregistered employment, is limited (see Rijkers et al. (2014)), which helps assuage concerns that our estimates may be biased upwards because firms that are not in the program under-report employment (more than participating firms, which have an incentive to meet the program's job creation targets).[13]

Tunisia has an "offshore" tax regime that promotes exports by offering firms that export at least 70% of their production exemptions from customs duties, VAT, and consumption taxes on inputs. Offshore firms are also exempt from foreign exchange restrictions. These favorable conditions may attenuate the advantages such firms derive from participating in the start-up program. To account for this, we control for offshore status in all our regressions and distinguish between onshore and offshore firms in our heterogeneity analysis, to understand how program impact may vary across these groups.

[11] Sectors are identified according to the NAT 4-digit classification, the most disaggregated available in our data.

[12] See Dingel and Neiman (2020), Gottlieb et al. (2020), and Garrote Sanchez et al. (2021) for analyses of which types of jobs were most susceptible to the COVID-19 pandemic.

[13] One might be concerned that the program's impact on employment may be overestimated as firms outside the program may opt not to declare workers and pay social security contributions, but this is unlikely since being registered with the social security administration, CNSS, is important to be able to claim health insurance and other social insurance benefits.

3 Data

Our study compiles and harmonizes data from several sources.[14] The first is the program data. This consists of (i) votes data across the different selection stages, and (ii) a database of applicants. We construct (i) by webscraping the results of the College's monthly voting sessions, which are available online. This dataset provides information on firm name, founder name(s), vote distribution, and the College's decision (approval or rejection) in the second round and pitch stage, if applicable. Using a fuzzy match on firm and founder names, we link vote outcomes to (ii), the database of applicants shared by the program and containing the firm's unique tax identifier.

To evaluate the impact of the program on firm performance, a significant challenge is obtaining outcomes for both selected and rejected firms. Young firms in our sample face high failure rates, making them harder to track through surveys compared to established firms.[15] We address this challenge by utilizing an administrative database of all formally registered private sector firms in Tunisia, the Repertoire National des Entreprises (RNE). This database includes records for every registered firm, encompassing all program applicants. Additionally, it integrates data from various government sources, providing three key indicators of firm activity —tax data from the ministry of finance, employment data from the social security administration, and trade data from customs—which enable us to measure performance and track firm survival and exits. We use the firm's unique tax identifier to match the program data with this dataset. The data includes information on firms' economic activity (by 4-digit sector), location, age, legal form, founding date, firm origin, and trade regime. Importantly, we are able to identify both selected and rejected applicants in this database.[16]

We perform extensive cleaning of the RNE data to maximize data availability while ensuring the reliability of the data. Missing quarterly employment and wages are imputed by interpolating between quarters with positive data. In our main analysis we use annual averages so that the wage and employment data are measured for the same time frame as output and profits data. We consider a firm as active in year t if it has a nonzero variable from any of the administrative sources used in the RNE (customs, tax, or social security data). Inactive firms have either all missing or a mix of zero and missing administrative outcomes in a given year. To obtain a balanced sample for all regressions, we impute

[14] Data cleaning procedures are discussed in detail in Appendix A.1

[15] Surveys of entrepreneurs often encounter issues such as attrition and nonresponse, particularly for sensitive metrics like profits and revenues (McKenzie and Woodruff, 2014).

[16] We conducted analysis using this data at the office of the National Institute of Statistics (INS).

zeros for outcomes in the post-program period for firms that are inactive in a given year. In our analysis of survival, we consider a firm as active if it is inactive in one year but returns to activity in the next. For sales, profits, employment, and the total wage bill we exclude the top and bottom 5% of the distribution to purge the data of potential outliers. We deflate monetary values using the 2022 Consumer Price Index from the INS. Additional information on our data cleaning procedures is provided in Appendix A.1. 2.

The Start-up label program, which began in March 2019 and is ongoing, is analyzed using a sample of firms that applied between March 2019 and December 2021, ensuring at least one year of post-application data, as we have access to RNE data until the end of 2022. Our analysis involves two specific sample restrictions: first, we focus on firms that first appear in the RNE dataset either in the year they apply to the program or earlier. The 'baseline year' is defined as the application year for firms created in the same year they apply, or the year prior to the application for the others. Second, we further restrict our sample to firms with available sales information from tax records and employment data from the social security agency in both the baseline year and the post-treatment period. As a result of these restrictions, we are able to identify 466 firms that applied to the program between 2019 and 2021, which is around 70% of the universe of startup applicants during the period.

4 Empirical Strategy

This section presents our identification strategy and balance tests.

4.1 Identification and Estimation Strategy

To assess the impact of the program on firm survival and performance, we employ a difference-in-differences strategy, comparing changes in the performance of firms that received the start-up label with those of unsuccessful applicants. Given that many of the firms in our sample are new start-ups and firms older than 8 years are not allowed to participate in the program, we do not have extensive data on the pre-treatment period. We restrict the pre-treatment period to the baseline year (to maximize the number of observations).[17]

[17]Due to the fact that many of the firms in our sample are young we also do not have sufficient power to study how the expiration of eligibility for program benefits when firms mature beyond 8 years of age impacts their performance.

The first and foremost concern is that firms are selected into the program based on characteristics that may also influence outcomes, potentially introducing selection bias. However, the fact that judges evaluate applications based on innovation and scalability, rather than commercial viability, partially mitigates this issue, as these criteria are less likely to directly correlate with current or future firm performance. Remaining endogeneity concerns are addressed in two ways. First, we control for a rich set of firms' observable characteristics. Second, we leverage the multi-stage selection process and estimate program impacts not only on the Full Sample, but also on the more restricted Pitch Sample, which consists of firms that were marginal accepts and rejects.

The fact that judges were unable to make a decision in the initial round indicates that additional information was needed to evaluate these firms. In principle, if judges have more information at the pitch stage, selection bias could be exacerbated. However, we do not believe this to be the case. From qualitative interviews with judges about the specifics of voting in this round, it seems unlikely that in a 5-minute Q&A judges can collect a substantial amount of additional information. Judges reported using the Pitch Stage to mainly ask for clarifications, for example on the product demo. We rather view firms being invited to this stage as suggesting they were closer to the threshold for inclusion and thus more comparable to each other. That being said, while using the Pitch Sample reduces concerns about selection on unobservables by limiting (unobserved) heterogeneity, it also reduces the sample size, which may in turn lower statistical power.

Survival We start by examining the impact of the program on firm survival by estimating the following OLS regression in the post period:[18]

$$Y_{it} = \beta_0 + \beta_1 \cdot \text{Labeled}_i + \beta_2 \cdot X_{i,baseline} + \theta_t + \epsilon_{it} \tag{1}$$

where Y_{it} is an indicator of survival, notably a dummy that takes value 1 if the firm is active next year or a dummy that takes value 1 if the firm was active at the end of our sample period, 2022. Labeled_i is an indicator equal to 1 if the firm received the label. Standard errors are clustered at the firm-level. $X_{i,baseline}$ is a vector of controls that include an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, age dummies, an indicator for having at least one female founder, an indicator for at least one judge abstaining from voting due to a conflict of interest, and industry dummies.

[18]Given that our pre-treatment period comprises one year only, which coincides with the first year we observe the firm, we cannot study differential survival in this pre-period.

Firm performance To assess the impact of the program on firm performance and address treatment heterogeneity arising from staggered selection into the program, we use the estimator developed by Callaway and Sant'Anna (2021). Due to sample size constraints, we do not include a full set of time-interacted controls, as this would be computationally infeasible. Instead, to retain a sufficiently large sample in each cohort-by-time cell, we follow Borusyak et al. (2024) and residualize the outcomes of interest. Specifically, we regress each outcome of interest on a subset of the covariates discussed above using the sample in the pre-period. We then predict the residuals for the Full Sample and use them as the outcome variable in our difference-in-differences regression. The full set of controls includes dummies for being located in Tunis, being foreign owned, having at least one judge declare a conflict of interest in the evaluation of the firm's application, having a female founder, as well as dummies for age at baseline.

To account for treatment heterogeneity associated with staggered treatment timing, the following two-way fixed effects (TWFE) model is estimated:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it} \tag{2}$$

where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on the controls discussed above. α_i denotes firm fixed effects, λ_t represents year fixed effects, and Labeled_{it}^k is an indicator variable for being k periods away from treatment (e.g., $k = 0$ for the first treatment period, that is the year firms first receive the label, $k = 1, 2, \ldots$ for post-treatment periods, and $k = -1$ as the omitted reference period. $\text{Labeled}_{it}^k = 0$ for firms that never get treated). ϵ_{it} is the error term. The coefficients δ_k capture the treatment effects at different time periods relative to the reference period. Specifically, δ_0 estimates the treatment effect at the time firms first receive the label, while $\delta_1, \delta_2, \ldots$ measure the effect in the subsequent periods.

The Average Treatment Effect on the Treated (ATT) for a specific group g at a specific time t is given by:

$$\text{ATT}_{g,t} = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid D_i = 1, G_i = g, t]$$

where $Y_{it}(1)$ is the observed outcome for treated units (those that received the treatment in period g) at time t, $Y_{it}(0)$ is the counterfactual outcome for these same units at time t if they had not received the treatment, $D_i = 1$ indicates that the unit i is

treated, $G_i = g$ indicates that the unit i was treated in group g, t is the specific time period of interest.

The average ATT, or overall ATT, is computed by averaging the individual ATTs across all treated groups and post-treatment periods. The formula for the average ATT is:

$$\text{Average ATT} = \frac{\sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}_g} N_{g,t} \cdot \text{ATT}_{g,t}}{\sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}_g} N_{g,t}}$$

where $N_{g,t}$ is the number of treated firms in group g at time t, $\text{ATT}_{g,t}$ is the ATT for firms treated in period g at time t, and $\mathcal{G}$ and $\mathcal{T}_g$ represent the set of treatment groups and post-treatment periods, respectively. The average ATT provides a weighted average of the individual treatment effects, accounting for group and period-specific heterogeneity, and is our main estimand of interest.

Another econometric challenge stems from the skewed distributions of three of our main outcomes of interest, i.e. sales, profits and employment. Many of the young firms in our samples report zero sales and/or zero or even negative profits, and the employment distribution is also severely right skewed. To address this skewness and limit the impact of outliers we winsorize these outcomes at the 5% level. In addition, in robustness tests we estimate the proportional treatment effect on the treated using a Poisson quasi-maximum likelihood estimator (QMLE) difference-in-differences specification (Silva and Tenreyro, 2010; Correia et al., 2020) following the recommendations of Chen and Roth (2024). The identifying assumption is that, absent the program, the proportional change in mean outcomes in treated and untreated firms would have been the same.

Finally, we also conduct some heterogeneity analysis to provide indirect tests for the impact of specific components of the program. To that end, we assess whether program impacts are different for firms that are highly dependent on imports, engage in manufacturing, operate in the "onshore" export-regime, or have a female founder.

4.2 Balancing tests between treated and non-treated firms

Voting outcomes Table A3 reports summary vote statistics from the monthly voting sessions for all applicants for whom we have complete data, the Full Sample. In total, 70 percent of applicants (326 out of 466) receives the label. Seventy-five percent of applicants (348 firms) receive their decision in the second round (that is on the basis of their application only), during which 20% of all applicants (91 firms) are summarily

rejected and 55% (257 firms) are accepted right away. The remaining quarter (118 firms) proceeds to the pitch stage and constitutes the Pitch Sample. In this third stage, another 69 firms, or 14.8% of the total sample, receive the label and 49 (10.5% of the total sample) are rejected.

Differences at baseline Treated and control firms are not different in other dimensions, as is shown in Table 1, which presents descriptive statistics on general characteristics and outcomes for the Full Sample (columns 1-3) and the Pitch Sample (columns 4-6) using RNE data and program data from the baseline year, prior to the program's start. Columns 1 and 4 display averages and standard deviations (in parentheses) for firms that were rejected, whereas columns 2 and 5 provide these statistics for firms that received the label. Columns 3 and 6 report p-values from t-tests of differences in means between treatment and control groups in the Main and Pitch Samples respectively. Reassuringly, all variables, with the exception of "conflict declared" (which we discuss below), are strongly balanced. We fail to reject the null that the means are equal in the baseline year for both the Full Sample and the Pitch Sample for any of them. The F-test of the joint null of no treatment-control differences for all variables excluding "conflict declared" has a p-value of 0.83 for the Full Sample; 0.89 for the Pitch Sample.

The only unbalanced variable between treated and control groups is the probability of declaring a conflict of interest compelling judges to abstain from participating in the voting sessions. For 25% of program participants, at least one of the judges declared a conflict of interest, whereas this was the case for only 6% of control firms. This difference could reflect a correlation between connections to industry insiders such as College members and applicant's perceived quality.[19] To the extent that this positive correlation is driven by unobservable factors that could independently affect performance (e.g. founder's abilities, political connections), this could bias our estimates of interest. Hence in the analysis below we control for the declaration of a conflict of interest, among other firm characteristics, unless otherwise indicated. We conduct an additional robustness check by omitting firms with a declared conflict of interest in Section 5.3.

[19]We can only speculate as to why firms that end up being selected are more likely to be ones in which at least one of the judges has a personal interest. Since judges are experts in their respective domains, it is plausible that they are more inclined to invest in promising startups. This might explain the positive correlation between declared conflicts of interest and receiving the startup label. Note that a priori one may have anticipated a negative correlation since the threshold for obtaining the label is higher for firms where conflicts of interest are present, as fewer judges are available to cast the required five positive votes.

The summary statistics for the Full Sample suggest that firms are young and small on average. The average treated firm is 1.46 years old at the time of application and has a staff of 1.09 persons (0.92 for control firms). As we noted previously, because of the very young age of firms in our sample we cannot control for a large set of pre-periods in our analysis. Firm sales are on average TND 146,970 (USD 48,990) at baseline for firms that receive the label and TND 157,160 (USD 52,386) for control firms, but these averages mask sizeable heterogeneity across firms, as is evidenced by the large standard deviations. The relatively sizable sales figures confirm the stark contrast between entrepreneurs in our context and the microentrepreneurs typically studied in the literature on entrepreneurship in developing countries. For example, the majority of subsistence firms in these studies have no paid employees, with average monthly revenues of less than 100 USD (McKenzie and Woodruff, 2014). As is usual in the early years of a start-up, firms in our sample are not profitable at the time they apply to the program, on average. Firms that are selected into the program are making a loss of TND 12,180 (USD 4,060) at baseline while rejected applicants make a loss of TND 6,500 (USD 2,166), although the deviations from the means are very large. 48% of treated applicants (41% of control firms) are located in the capital, Tunis. Foreign-owned firms account for a small share of both the treated (6%) and the control samples (4%). Labor costs amount to roughly a fifth of total sales on average. The average annual wage bill of program participants at baseline is very similar to that of non-participants: TND 26,870 (USD 8,956) for treated and TND 26,270 (USD 8,756) for control firms. 28% of treated firms have at least one female founder, relative to 34% of control firms, and the difference between these means is not statistically significant. The fact that it is hard to predict which firms will receive an acceptance or rejection based on observable characteristics in both the Main and the Pitch samples is consistent with the idea that judges are not evaluating the potential for success of these firms. This also helps assuage concerns about program impacts being driven by selection bias.

As an auxiliary test for potential selection on unobservables, we also check the predictive power of the judges regarding the future success of firms, independent of the label effect. To that end we regress growth in sales, profits, employment and the total wage bill on the share of yes votes in the first round, whether or not at least one of the judges declared a conflict of interest, and the years since application. To isolate the effect of the yes votes from that of the label, we run the regressions separately for firms that received the label and those that did not. If the share of yes votes is a strong marker of the quality of firms we would expect it to be correlated with subsequent

improvements in performance irrespective of whether they obtained the label. Table A4 presents the results, which reveal that we cannot reject the null that the voting outcomes have no predictive power for any of the outcomes. This result applies both to the treated (panel B) and the control group (panel A). This is also aligned with the similarity in observables between the treated and control samples and further mitigates concerns about selection on unobservables.

The inability to predict the success of the firms is again consistent with the selection criteria applied by judges. As explained in Section 2, judges had relatively little time to assess each application and were not asked to evaluate financial viability or expected returns as venture capitalists would do. Instead, they focused on assessing innovation and scalability. This finding also dovetails with evidence from other business plan competitions. For example, McKenzie and Sansone (2019) conduct a business plan competition in Nigeria and show that judges' scores do not predict firm survival, employment, sales, or profits.

5 Results

We first present the results on firm survival and performance, followed by robustness checks, an exploration of potential mechanisms, and a back-of-the-envelope calculation of the returns to the program.

5.1 Survival

Table 2 reports the impacts of the program on survival. Columns 1-2 and 5-6 present results for the Full Sample, columns 3-4 and 7-8 present results for the Pitch Sample. Odd columns do not include any controls (other than year dummies in columns 1 and 3), whereas even columns control for being located in Tunis, being foreign-owned, offshore status, age dummies, a dummy for having had at least one jury member profess a conflict of interest during the selection process, industry dummies, and an indicator for having at least one female founder. Panel A (columns 1-4) examines the determinants of short-term survival, using as outcome variable a dummy indicating whether firms are active in a given year, whereas Panel B (columns 6-8) examines the determinants of long-run survival, notably surviving until at least 2022.

The effects of program participation on survival are large and economically significant. Firms in the Full Sample that receive the startup label are 18 percentage points (pp)

more likely to be active in a given year than firms whose application was rejected, an effect significant at the 1% level. Controlling for observable characteristics reduces this differential to 16 pp but the effect remains significant at the 1% level. When we use the Pitch Sample without controlling for observable characteristics, we find estimates of program impact that are very similar to those in the Full Sample (see column 3). Once we include controls, the estimated effect of participating in the program drops to 11 percentage points and is statistically significant at the 10% level. The effects on survival are economically meaningful since each year 3 of every 10 firms exit on average.

The impact of the program on longer-run survival is presented in Panel B. Firms in the Full Sample that received the label are 26 pp more likely than rejected applicants to be active in 2022, the last year in our data (see column 5). Including controls reduces this effect to 22 pp, as is shown in column 6. Estimates for firms in the pitch sample (presented in columns 7 and 8) are very similar. According to our best estimates which control for firm characteristics at baseline (column 8 that is), firms in the pitch sample that received the label are 18 pp more likely to be active at the end of our sample period than firms that applied but did not. Again these are large effects given that nearly half of all firms in the Main and Pitch samples were not active any longer by 2022.

5.2 Firm performance

Next we turn to the estimation of the impacts of the Start-up Label program on firm performance, including sales, profits, employment and the wage bill. The results are presented in Table 3. Panels A and B report regression 2 using the Main and the Pitch Samples, respectively. Exiting firms are included in both samples but counted as having zero sales, profits or employment. Our preferred estimates are based on the Pitch Sample, as this group is likely to exhibit the least selection bias, as discussed above.

The estimated impact of the program on sales is positive, sizable and statistically significant at the 10% level in the Full Sample (column 1) but very imprecisely estimated in the Pitch Sample (column 6). By contrast, the program's impact on profits is negative. In the Full Sample this effect is statistically significant (see column (2), while in our preferred Pitch Sample it is not (column (7)). This counterintuitive result appears in part driven by the differential propensity of program firms to survive. Recall that firms in our sample are very young and on average report losses.[20] In these regressions, however,

[20]During the early stages of a firm's lifecycle, sales are typically below potential as firms have not yet fully established their market, and they can be insufficient to cover fixed and variable production costs. Firms may also deliberately choose to operate at a loss as a strategic move to gain market share. Hence

exiting firms are included as firms with zero profits, which may contribute to the negative program impact on profits. Another possible explanation is that labeled firms invest more—perhaps facilitated by greater access to credit—which would impact their reported profits.

The program had a sizable and strongly significant impact on employment (columns 3 and 8). Labeled firms in the Full Sample hired on average 1.7 more workers than control firms after receiving the label, while those in the pitch sample hired 2.0 more workers on average. These employment increases are considerable, given that the average firm size in the main and pitch samples is 1.6 and 1.8 workers, respectively. Thus, the program has been successful in achieving one of its key objectives, that of creating jobs.

Yet, these employment increases are generally insufficient for firms to comply with the program's eligibility criteria, which require labeled firms to achieve a target of 10 wage employees within three years of participation. Receiving the start-up label indeed increases the likelihood that firms reach this target in any given year, notably by 6 pp in the Full Sample (column 4) and the pitch sample (column (9)), though only the former effect is significant at the 10% level. These impacts are sizable, given that just 6% of firms in our sample employ 10 or more workers. In fact, only a small proportion of participating firms (17.4%) achieved this employment threshold within three years. This outcome suggests that the 10-employee target may be overly ambitious for most firms, at least in the medium term. In short, while the Start-up program does create jobs, it may still fall short of its ambitious goal of generating 10,000 new jobs by 2025.

Mirroring the increase in employment documented in columns (3) and (8), specifications (5) and (10) show that the program also generates a significant average increase in the wage bill—an additional 54,000 Tunisian Dinars for firms in the Full Sample and 69,000 TND for firms in the pitch sample. These are large increases, equivalent to 116% of the control mean for the overall sample and 113% for the pitch sample.

Overall, the results we obtain using the pitch sample are similar to, and statistically indistinguishable from those obtained using the Full Sample. This is consistent with the fact that entry into the program was difficult to predict, as was shown in section 4.2, further assuaging concerns about selection bias.

We exploit the more frequent availability (quarterly) of employment data to assess the dynamic impacts of the program. Figure 1 plots the coefficients from the quarterly

profits and sales may be too noisy to serve as informative performance metrics for start-ups.

analog of equation 2 on employment using quarterly data for the Full Sample. The coefficient in each period is relative to the quarter before application. We see no evidence of pre-trends, which is further reassurance of limited selection bias. The point estimates starting in the quarters following application are positive and large in magnitude. The timing of the uptick in employment is consistent with a positive impact of the program on job creation.

Overall, the results point to a sizable impact of the program on job creation, even though many participants fail to achieve the employment targets the program aimed for. Selection bias appears limited since results for the main and Pitch Samples are consistently very similar.

5.3 Robustness

We perform several robustness tests to assess the sensitivity of our results. First, we re-estimate the regressions using Poisson quasi-maximum likelihood estimation (QMLE), following Silva and Tenreyro (2010) and Correia et al. (2020), to account for the skewed distribution of many outcome variables. The results, reported in Table A5, are consistent with our baseline estimates.[21]

Second, we re-estimate the model using non-winsorized data to assess the influence of outliers. As shown in Table A6, the results are again broadly consistent with our baseline. The negative effect on profits disappears for the Full Sample, while the estimated impacts on employment and the wage bill become even larger.

Third, we address potential selection bias related to conflicts of interest. While our regressions control for declared conflicts, judges may have informally supported firms they knew, which could have increased those firms' likelihood of selection. To assess the robustness of our results to this potential bias, we re-estimate equation 2 excluding all firms with declared conflicts. As shown in Table A7, the point estimates of program impact are all smaller. However, the employment and wage bill effects in the Full Sample remain positive and statistically significant. In the Pitch Sample, the estimated program impacts on employment and the wage bill are of comparable magnitude to those in the Full Sample, but no longer significant. This suggests heterogeneity in treatment effects, with stronger impacts among more connected firms. One possibility is that connected firms are simply more capable, which could explain both why judges invested in them in

[21] In this specification, we do not use the estimator of Callaway and Sant'Anna (2021), as a Poisson QMLE implementation of this method is, to the best of our knowledge, not yet available in standard software packages.

the first place, and why they benefit more from treatment. Given that the Full Sample results are largely robust to omitting these firms, we interpret this as suggesting that any endogeneity from this variable, if present, does not drive the main results.

Fourth, as another check for selection bias, in Table A8, we conduct a placebo test by comparing firms selected in the second round with those selected in the Pitch stage. Consistent with the evidence we have presented for limited selection bias with the Full Sample in section 4.2, we find that second round and pitch stage firms are no different in terms of their performance as a result of the program.

Finally, because our study period overlaps with the COVID-19 pandemic, we assess heterogeneity in treatment effects using three sector-level exposure measures from Le et al. (2024). The first classifies sectors as "essential" or "non-essential" based on labor supply disruptions. The second captures sectoral demand shocks, measured by the average change in realized and forecasted firm revenues pre- and post-pandemic onset. The third captures intermediate input supply shocks, defined as the share of firms in a given sector reporting input constraints in the World Bank's COVID-19 Enterprise Follow-up Survey.[22] For the latter two, sectors are split into high and low exposure groups relative to the sample median.

Table A9 reports the results. Panels A and B compare essential and non-essential sectors. Only 8% of firms fall into the "essential" category—mainly food and health services— while the majority of firms in our sample operate in non-essential tech-related sectors such as e-commerce, AI, and business software. Estimates of program impacts on essential firms are imprecise and indistinguishable from zero, likely due to the small sample size (35). For non-essential sectors, which were arguably more exposed to the COVID-19 pandemic, results closely resemble those for the baseline Full Sample findings presented Table 3. Panels C and D split firms by exposure to demand shocks, with most firms operating in sectors experiencing relatively low shocks. Results are similar across both groups with, if anything, higher program impacts in the most exposed group. Panels E and F examine input supply shocks, with smaller but imprecisely estimated effects for more exposed sectors, again likely due to limited sample size in this category. Overall, these results suggest no consistent evidence of differential treatment effects by pandemic exposure.

We also explore whether program impacts varied across phases of the pandemic.

[22]Firms report whether their input supply in the last completed month decreased relative to the same month in 2019. A response of "decreased" is classified as constrained input. The authors aggregate the share of firms with constrained inputs at the 3-digit ISIC industry level across 33 countries (2020–2022). This measure is then mapped to 3-digit NAT codes and standardized (Le et al., 2024).

Using the Oxford COVID-19 Government Response Stringency Index (Hale et al., 2021), we exclude quarters with the highest lockdown intensity in a quarterly difference-in-differences framework. Table A10 shows results for the Main and Pitch samples, both for the full period and excluding high-stringency quarters. Findings are consistent with our main results, with slightly larger estimates for employment and wage bill outcomes when excluding the worst-hit periods. In other words, COVID-19 may have somewhat attenuated the beneficial effects of the program on firms.

5.4 Potential mechanisms

Next we investigate the mechanisms behind the Start-up Label program's impact on employment. We first examine whether the program's effects are driven by increased firm survival or by job creation within surviving firms. To assess this, we focus on the intensive margin—restricting the sample to firms with strictly positive outcomes—and then examine firms that survive at least until 2022. We also investigate how the program's impact varies across different types of firms, sectors, and founder characteristics, and consider the potential role of the program's various components.

The results regarding the intensive margin are presented in Table A11. The estimated effects on employment and the wage bill are significantly larger in magnitude than those in Table 3 and remain economically meaningful. Among firms with strictly positive outcomes in the pitch sample, the program's impact on employment (3.43) is 70% greater than the effect observed in Table 3 (2.03) for the sample that includes firms with non-positive outcomes. The impact on the wage bill is 24% larger. Since the intensive margin sample represents 78% of the overall pitch sample, these findings suggest that most of the program's effect is driven by the intensive margin. This aligns with the substantially larger overall impact of the program on employment (110%) and the wage bill (113%) reported in Table 3 compared to its effect on long-term survival (18%) in Table 2.

Table A12 uses the sample of firms still active as of 2022, comprising 320 firms out of 466 in the Full Sample and 79 out of 118 in the pitch sample. The results are qualitatively similar. For the pitch sample, the employment and wage bill coefficients (3.14 and 114.0) are higher than the corresponding coefficients (2.03 and 69.2) for the overall sample of 118 firms in Table 3, providing further suggestive evidence that the program's impact is primarily driven by the intensive margin rather than by firm survival.

Which of the advantages offered by the program is most impactful? Since there is no (experimental) variation in the intensity of each component within the bundle of

advantages offered by the program, we cannot precisely identify which element(s) drive the observed effects. To shed some light on this issue, Table A13 examines how the program's impacts vary across distinct types of firms, each likely to be differently influenced by various program components. Specifically, we distinguish the program impact according to the degree of reliance on imports; whether the firm is in the manufacturing sector, and whether it operates in the onshore or offshore regime. The first two groups could help us isolate the impact of preferential customs treatment, as import-dependent and manufacturing firms are more reliant on international trade than other firms. The onshore/offshore analysis sheds light on the relevance of foreign exchange account access, as offshore firms already benefit from this advantage.

We use the Full Sample in order to maximize power given that we are slicing our data into various sub-groups to perform heterogeneity analysis. The results are presented in table A13. Panel A confines attention to firms operating in import-dependent sectors, defined as those with import shares above the median, calculated over the entire universe of registered firms. Panel B presents results for firms in non import-dependent sectors. The program impacts on employment and the wage bill are larger in import-dependent sectors than in non-import-dependent ones. Panels C and D present results for manufacturing and non-manufacturing sectors, revealing that program impacts are much more positive for the former than the latter. Combined, these findings suggest that preferential customs treatment may be an important enabler of firm growth and job creation.

Next, we examine the impacts on firms operating in the onshore and offshore regimes separately in panels E and F, respectively. The program has a robust, positive, and significant impact on sales, job creation, and the wage bill for onshore firms. Impacts for offshore firms, while positive, are slightly smaller and statistically insignificant, possibly due to the limited sample size. Nonetheless, the larger effects observed for onshore firms suggest that access to foreign currency accounts along with preferential customs treatment contribute positively to firm performance.

Finally, founder characteristics may influence how much firms benefit from the program. Unfortunately we have limited information on these characteristics, except for gender. In all our regressions, we control for having at least one female founder. In Table A14, we explore heterogeneity along this dimension. While estimated effects on sales and the wage bill are somewhat larger for firms with all-male founders, the impacts on employment are very similar across groups. Notably, the program is associated with

a slightly less negative impact on profits and slightly more positive impact on employment and wage bill for firms with at least one female founder. However, the overall patterns of heterogeneity are modest, and there is no strong evidence that gender composition materially modulates program impact. The last row of Table A14 shows that for each of the outcomes, we reject the null of equality of coefficients.

We complemented these results with interviews with a few firms. They claimed that the coverage of social security contributions, access to foreign exchange and to preferential customs' treatment were among the most useful features of the program. The program's robust impact on employment aligns with the supposition that reduced labor costs are a significant benefit. Although access to financing was not a direct benefit of the program, obtaining the label may have provided some signaling value, potentially easing access to finance in both credit and investor markets. However, our interviews suggest that, while this effect may have been present, its magnitude is likely to be marginal.

5.5 Cost-benefit analysis

Is the program worth it? To explore this, we conduct a back-of-the-envelope cost-benefit analysis. We compare the net annual cost per firm borne by the government with the average annual benefits the program has generated per firm thus far. This calculation is crude and inevitably somewhat ad hoc: it focuses on short-term, partial equilibrium effects and abstracts from spillovers, behavioral responses, and non-economic benefits of employment. It also relies on estimates subject to statistical and economic uncertainty. As such, results should be interpreted with caution and viewed as indicative rather than definitive. Taking this into account, we provide a range for our benefit-cost ratio, using estimates for the Full Sample and the Pitch Sample as a lower and an upper bound, respectively, in our calculations. Table 4 summarizes the results. For brevity, we describe in this section the estimates from the Pitch Sample, which provides our preferred estimates.

Costs Panel A of Table 4 shows the annual costs of the program per firm. These include (i) direct costs and (ii) fiscal incentives. The direct costs (i) consist of monetary expenditures disbursed to run the program (administrative costs) and to provide the stipends to the eligible founders (stipend costs), both of which we obtain from Smart Capital. The former account for TND 1,523 (USD 508), while stipend costs account for TND 10,181 (USD 3,394) per participating firm per year on average.[23] The main costs

[23]These figures are calculated by dividing the total amount of stipends received by all eligible firms in

of the fiscal incentives (ii) that the program offers are (1) foregone social security contributions and (2) foregone corporate tax revenue, but these are partially offset by (3) increased personal income tax revenue from new hires. Foregone social security contributions (1) amount to TND 23,415 (USD 7,805) per firm on average and are the biggest cost of the program.[24] Foregone corporate tax revenue (2) is a much smaller expense, amounting to TND 659 (USD 220) per participating firm per year.[25] Fiscal expenses in the form of foregone corporate taxes are limited because many firms report making losses. These tax expenditures are partially offset by (3) increased personal income tax revenue from new hires, which amounts to TND 6,136 (USD 2,045) per participating firm per year.[26] Net annual government costs per firm thus amount to TND 29,693 (USD 9,898).

Benefits Our estimates of short-term program impacts point to an increase in survival, employment, and the wage bill, with no significant impact on sales and profits. We define the net economic benefit of the program in the short term as the sum of the average yearly increase in the wage bill and the average stipend that firms receive. Given that we construct a balanced sample (we impute 0 outcomes for firms that exit), the survival benefits are taken into account in the regressions on employment and the wage bill reported in Table 3. The average added annual wage bill per program participant is TND 69,179 (USD 23,060) and average annual stipend spending is TND 10,181 (USD 3,394).[27] The total average annual benefits the program generates per participating firm thus amount to TND 79,360 (USD 26,453).

Returns The short-term return of the program is the ratio of benefits to costs. Our back-of-the-envelope calculation suggests that, so far, the program has yielded an estimated TND 2.68 in benefits for each TND spent. The program cost estimate also implies a short-term cost of TND 14,606 (USD 4,869) per full-time job created.[28] Using

a given year by the total number of labeled firms in a given year.

[24]They are calculated by multiplying the average annualized wage bill of program participants in the posttreatment period by τ_{ss}, the tax rate of 25.75%.

[25]Foregone tax revenue is calculated by taking the average of the hypothetical tax revenue across labeled firms in the post-treatment period. Firms with negative profits are assumed to pay zero taxes, firms with positive profits are assumed to be taxed at τ_c, 15%.

[26]To calculate this component, we use the results from our regression analysis. We calculate the increased personal income tax revenue as the (taxable) average annualized earnings per worker for firms participating in the program with employees in the posttreatment period, times the personal income tax rate of 19.5%, multiplied by the number of new workers (implied by the coefficient estimated in col (3) of our main Table 3). The tax schedule in Tunisia is progressive. Up to 5,000 TND is tax free. We apply τ_p to the taxable balance.

[27]Note that stipend spending is included both as a cost and as a benefit; it effectively is a transfer from the government. This accounting choice may overstate the program's net economic value.

[28]In comparison, Biancalani et al. (2022) calculate that the Italian Start-up Act cost EUR 32,000 per

estimates from the Full Sample instead of the pitch-stage sample yields a benefit-to-cost ratio of 1.99.[29] While these results indicate that the program appears to generate net social benefits, they must be interpreted with caution given the simplifying assumptions and uncertainties inherent in the analysis. Moreover, given the program's targeted focus on high-tech entrepreneurship, the estimated returns may not generalize to a broader population of firms.

6 Conclusions

Governments across the globe are committing considerable resources to programs that spur the development of companies with high potential for growth, innovation and job creation. Salient selection issues and data constraints often prevent rigorous evaluations of such programs. This paper is one of the first studies to examine high-tech entrepreneurship in a developing country context, by examining the impacts of Tunisia's Startup Act – a new form of institutional support for young ventures and the first program in Africa to offer a label of merit and several administrative advantages to spur innovative startups.

To address the concern that our estimates are biased due to selection on unobservables, we tested the robustness of our analysis to restricting it to a sample of firms for which the judges that select program participants are undecided in the first round of voting, arguing that unobservables are likely to be similar for this set of firms, irrespective of whether they are eventually accepted into the program. We also show that selection into the program is difficult to predict, and that the share of yes votes that firms receive in the first voting round does not predict subsequent growth, neither for program participants nor for unsuccessful applicants.

Our findings point to positive and significant effects on firm survival and employment, one to three years after program delivery. We find imprecisely estimated effects on sales, and no detectable effects on profits. A crude cost-benefit calculation suggests that the program has thus far been cost-effective, despite being implemented during a period marked by the COVID-19 pandemic and political upheaval. Given the program's novelty

full-time equivalent job created, conditional on the job existing for at least five years. Italy's GDP per capita is almost ten times higher than Tunisia's.

[29] As a robustness check, we replicate the analysis using the Marginal Value of Public Funds framework (see Appendix Table A15). This alternative approach—accounting for foregone fiscal revenues as benefits and evaluating the additional wage bill against its opportunity cost—produces returns ranging from 1.69 to 2.86.

and policy relevance, we set out to assess its mid-course impacts. While its relatively recent launch precludes an evaluation of long-term outcomes, this is a valuable topic for future research.

Bibliography

Anderson, Stephen J and David McKenzie, "Improving business practices and the boundary of the entrepreneur: A randomized experiment comparing training, consulting, insourcing, and outsourcing," *Journal of Political Economy*, 2022, *130* (1), 157–209.

Anderson, Stephen J., Pradeep Chintagunta, Frank Germann, and Naufel Vilcassim, "Do marketers matter for entrepreneurs? Evidence from a field experiment in Uganda," *Journal of Marketing*, 2021, *85* (3), 78–96.

Auriol, Emmanuelle and Michael Warlters, "The marginal cost of public funds and tax reform in Africa," *Journal of Development Economics*, 2012, *97* (1), 58–72.

Betcherman, Gordon, N Meltem Daysal, and Carmen Pagés, "Do employment subsidies work? Evidence from regionally targeted subsidies in Turkey," *Labour Economics*, 2010, *17* (4), 710–722.

Biancalani, Francesco, Dirk Czarnitzki, and Massimo Riccaboni, "The Italian start up act: A microeconometric program evaluation," *Small Business Economics*, 2022, *58* (3), 1699–1720.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 2024, p. rdae007.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico," *Journal of Political Economy*, 2018, *126* (2), 635–687.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Chen, Jiafeng and Jonathan Roth, "Logs with zeros? Some problems and solutions," *Quarterly Journal of Economics*, 2024, *139* (2), 891–936.

Chioda, Laura, David Contreras-Loya, Paul Gertler, and Dana Carney, "Making entrepreneurs: Returns to training youth in hard versus soft business skills," Working Paper 28845, National Bureau of Economic Research May 2021.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin, "Fast poisson estimation with high-dimensional fixed effects," *The Stata Journal*, 2020, *20* (1), 95–115.

Cusolito, Ana P., Ornella Darova, and David McKenzie, "Capacity building as a route to export market expansion: A six-country experiment in the Western Balkans," *Journal of International Economics*, 2023, *144* (1), 103794.

Dingel, Jonathan I and Brent Neiman, "How many jobs can be done at home?," *Journal of Public Economics*, 2020, *189*, 104235.

Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff, "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana," *Journal of Development Economics*, 2014, *106*, 211–226.

Gottlieb, Charles, Jan Grobovšek, and Markus Poschke, "Working from home across countries," *Covid Economics*, 2020, *1* (8), 71–91.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar et al., "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)," *Nature Human Behaviour*, 2021, *5* (4), 529–538.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hendren, Nathaniel and Ben Sprung-Keyser, "A unified welfare analysis of government

policies," *The Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.

Hochberg, Yael V and Daniel C Fehder, "Accelerators and ecosystems," *Science*, 2015, *348* (6240), 1202–1203.

Hsieh, Chang-Tai and Benjamin A. Olken, "The missing 'missing middle'," *Journal of Economic Perspectives*, 2014, *28* (3), 89–108.

_ and Peter J. Klenow, "The life cycle of plants in India and Mexico," *Quarterly Journal of Economics*, 2014, *129* (3), 1035–1084.

Huttunen, Kristiina, Jukka Pirttilä, and Roope Uusitalo, "The employment effects of low-wage subsidies," *Journal of Public Economics*, 2013, *97*, 49–60.

Iacovone, Leonardo, William Maloney, and David McKenzie, "Improving management with individual and group-based consulting: results from a randomized experiment in Colombia," *The Review of Economic Studies*, 2022, *101* (4), 346—-371.

INS, "Indicateurs de l'emploi et du chômage au quatrième trimestre 2022," 2022.

_ , "Evolution des entreprises privées par secteur d'activité," 2024.

_ , "Les comptes de la nation 2019-2023," 2024.

Juhász, Réka, Nathan J Lane, and Dani Rodrik, "The new economics of industrial policy," *Annual Review of Economics*, 2023.

Le, Minh-Phuong, Lisa Chauvet, and Mohamed Ali Marouani, "The great lockdown and the small Business: Impact, channels and adaptation to the covid pandemic," *World Development*, 2024, *182*, 106673.

Lerner, Josh, Antoinette Schoar, Stanislav Sokolinski, and Karen Wilson, "The globalization of angel investments: Evidence across countries," Technical Report, National Bureau of Economic Research 2015.

McKenzie, David and Anna Luisa Paffhausen, "Small firm death in developing countries," *The Review of Economics and Statistics*, 2023, *101* (4), 645—-657.

_ and Christopher Woodruff, "Experimental evidence on returns to capital and access to finance in Mexico," *The World Bank Economic Review*, 2008, *22* (3), 457–482.

_ and _ , "What are we learning from business training and entrepreneurship evaluations around the developing world?," *The World Bank Research Observer*, 2014, *29* (1), 48–82.

_ and Dario Sansone, "Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria," *Journal of Development Economics*, 2019, *141*, 102369.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "Returns to capital in microenterprises: evidence from a field experiment," *The Quarterly Journal of Economics*, 2008, *123* (4), 1329–1372.

_ , _ , and _ , "Labor drops: Experimental evidence on the return to additional labor in microenterprises," *American Economic Journal: Applied Economics*, 2019, *11* (1), 202–235.

Menon, Carlo, Timothy DeStefano, Francesco Manaresi, Giovanni Soggia, and Pietro Santoleri, "The evaluation of the Italian "Start-up Act"," 2018, (54).

Quinn, Simon and Christopher Woodruff, "Experiments and entrepreneurship in developing countries," *Annual Review of Economics*, 2019, *11* (1), 225–248.

Rijkers, Bob, Hassen Arouri, Caroline Freund, and Antonio Nucifora, "Which firms create the most jobs in developing countries? Evidence from Tunisia," *Labour Economics*, 2014, *31*, 84–102.

Sanchez, Daniel Garrote, Nicolas Gomez Parra, Caglar Ozden, Bob Rijkers, Mariana Viollaz, and Hernan Winkler, "Who on earth can work from home?," *The World Bank Research Observer*, 2021, *36* (1), 67–100.

Silva, JMC Santos and Silvana Tenreyro, "On the existence of the maximum likelihood estimates in Poisson regression," *Economics Letters*, 2010, *107* (2), 310–312.

Verhoogen, Eric, "Firm-level upgrading in developing countries," *Journal of Economic Literature*, 2023, *61* (4), 1410–1464.

7 Tables and Figures

Table 1: Firm Outcomes and Characteristics at Time of Application

	Full Sample			Pitch Sample		
	Control Mean (SD) (1)	Treatment Mean (SD) (2)	P-value (1)-(2) (3)	Control Mean (SD) (4)	Treatment Mean (SD) (5)	P-value (4)-(5) (6)
Located in Tunis	0.41	0.48	0.23	0.43	0.35	0.38
	(0.49)	(0.50)		(0.50)	(0.48)	
Foreign	0.04	0.06	0.26	0.04	0.06	0.68
	(0.19)	(0.24)		(0.20)	(0.24)	
Onshore	0.84	0.83	0.70	0.86	0.88	0.67
	(0.37)	(0.38)		(0.35)	(0.32)	
Age at baseline	1.46	1.46	0.97	1.51	1.64	0.70
	(1.73)	(1.64)		(1.83)	(1.71)	
Local Sales (1k 2022 dinars)	63.15	61.79	0.96	94.83	39.66	0.26
	(294.66)	(245.08)		(392.90)	(85.02)	
Exports (1k 2022 dinars)	85.01	76.72	0.85	199.93	146.93	0.70
	(460.28)	(412.74)		(757.45)	(733.59)	
Post-tax Sales (1k 2022 dinars)	157.16	146.97	0.85	309.83	193.39	0.43
	(571.79)	(492.30)		(876.72)	(731.53)	
Profits (1k 2022 dinars)	-6.50	-12.18	0.20	-10.47	-14.47	0.68
	(41.34)	(44.94)		(51.53)	(53.34)	
Employment	0.92	1.09	0.53	1.12	1.30	0.76
	(2.54)	(2.82)		(2.91)	(3.19)	
Total wagebill (1k 2022 dinars)	26.27	26.87	0.95	38.81	35.45	0.88
	(94.37)	(92.24)		(122.03)	(116.47)	
Conflict declared	0.06	0.25	0.00***	0.14	0.38	0.00***
	(0.25)	(0.43)		(0.35)	(0.49)	
At least one female founder	0.34	0.28	0.25	0.29	0.25	0.64
	(0.47)	(0.45)		(0.46)	(0.43)	
F-test of joint significance (p-value)			0.00***			0.16
F-test (p-value), excluding Conflict declared			0.83			0.89
F-test, number of observations			466			118
N	140	326		49	69	

Notes: Table presents descriptive statistics on general characteristics and outcomes for the Full Sample (columns 1-3) and the Pitch Sample (columns 4-6) using RNE data and program data from the baseline year, prior to the program's start. Columns 1 and 4 display averages and standard deviations (in parentheses) for firms that were rejected, whereas columns 2 and 5 provide these statistics for firms that received the label. Columns 3 and 6 report p-values from t-tests of differences in means between treatment and control groups in the Main and Pitch Samples respectively. The Full Sample comprises firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table 2: Effects of the Start-Up Label Program on Survival

	Full Sample		Pitch Sample	
	A. **Active this year**			
	(1)	(2)	(3)	(4)
Labeled	0.18***	0.16***	0.17**	0.11*
	(0.04)	(0.04)	(0.06)	(0.06)
R2	0.09	0.15	0.12	0.25
N	1,243	1,243	333	333
Firms	466	466	118	118
Mean in the control group	0.67	0.67	0.70	0.70
Controls		X		X
	B. **Survives until end of the sample period (2022)**			
	(5)	(6)	(7)	(8)
Labeled	0.26***	0.22***	0.24***	0.18*
	(0.05)	(0.05)	(0.09)	(0.10)
R2	0.06	0.15	0.06	0.21
N	466	466	118	118
Firms	466	466	118	118
Mean in the control group	0.51	0.51	0.53	0.53
Controls		X		X

Notes: Panel A reports results from estimating $Y_{it} = \beta_0 + \beta_1 \cdot \text{Labeled}_i + \beta_2 \cdot X_{i,\text{baseline}} + \theta_t + \epsilon_{it}$ in the post period. Panel B reports results from estimating $Y_i = \beta_0 + \beta_1 \cdot \text{Labeled}_i + \beta_2 \cdot Y_{i,\text{baseline}} + \epsilon_i$ in the year 2022. Y_{it} is an indicator of survival, notably a dummy that takes value 1 if the firm is active (in columns 1, 2, 3 and 4) or a dummy that takes value 1 if the firm was active at the end of our sample period, 2022 (in columns 5,6,7 and 8). Labeled_i is an indicator equal to 1 if the firm received the label. Standard errors are clustered at the firm-level. θ_t are year dummies. Controls ($X_{i,baseline}$) include an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, age dummies, an indicator for having at least one female founder, an indicator for at least one judge abstaining from voting due to a conflict of interest, and industry dummies. In panel B, we control only for age dummies, an indicator for having at least one female founder, an indicator for at least one judge abstaining from voting due to a conflict of interest, and industry dummies. We consider a firm as active in year t if it has a nonzero variable from any of the administrative sources used in the RNE (customs, tax, or social security data). Inactive firms have either all missing or a mix of zero and missing administrative outcomes in a given year. The Full Sample (columns 1-2, 5-6) comprises firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (columns 3-4, 7-8) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table 3: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level outcomes, Callaway and Sant'Anna (2021) Estimator

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **Full Sample**		
	(1)	(2)	(3)	(4)	(5)
ATT	101.8*	-13.1**	1.67***	0.063*	54.0***
	(60.4)	(6.41)	(0.48)	(0.033)	(18.1)
N	1665	1352	1709	1709	1709
Clusters	466	466	466	466	466
Mean in the control group	241.8	-8.45	1.57	0.060	46.5
			B. **Pitch Sample**		
	(6)	(7)	(8)	(9)	(10)
ATT	70.9	-17.5	2.03**	0.056	69.2**
	(109.5)	(13.0)	(0.80)	(0.047)	(34.8)
N	435	363	451	451	451
Clusters	118	118	118	118	118
Mean in the control group	350.5	-17.4	1.84	0.061	61.1

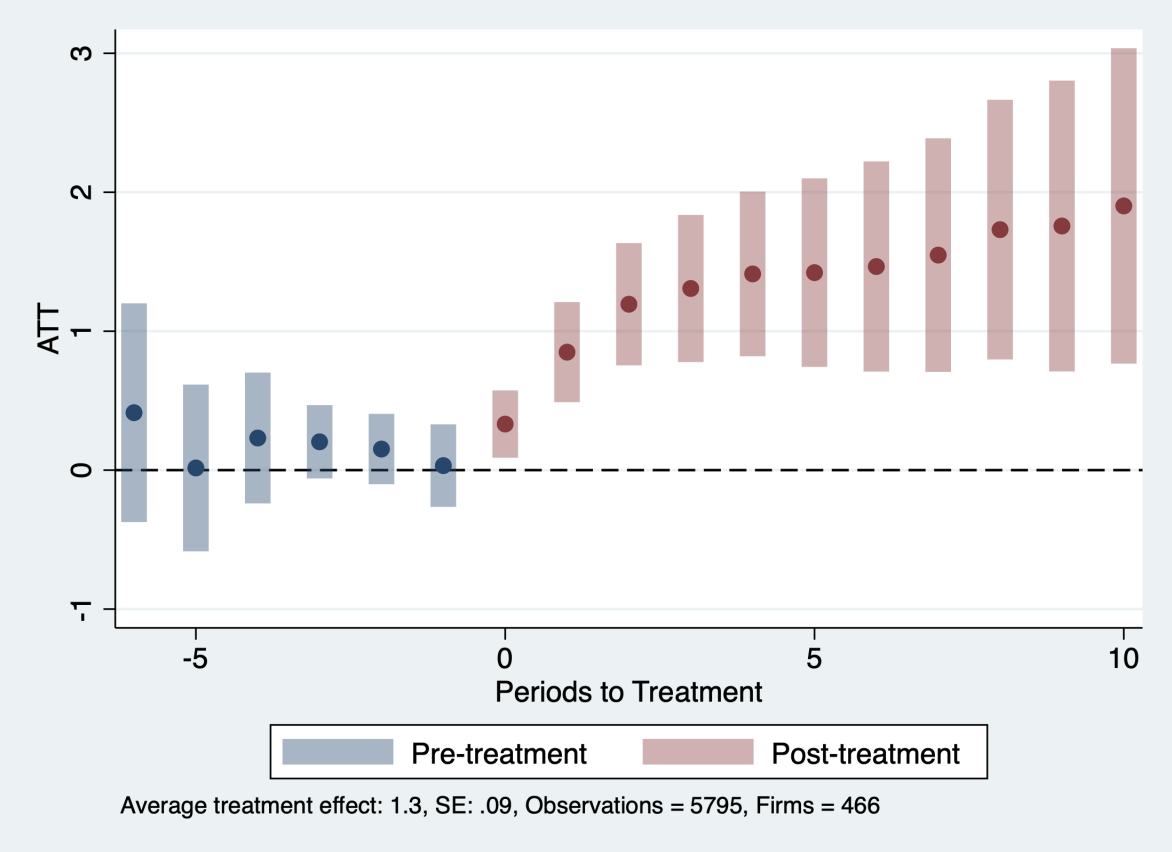
Notes: Panels A and B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample (Panel A) is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (Panel B) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table 4: Cost-Benefit Analysis

	Full Sample	Pitch Sample
	A. **Direct costs and fiscal incentives**	
	(1)	(2)
+ Average yearly administrative spending	1,523	1,523
+ Average yearly stipend spending	10,181	10,181
+ Foregone social security contributions	23,110	23,415
+ Foregone corporate tax revenue	364	659
- Added personal income tax revenue	5,460	6,136
Net cost	29,874	29,693
	B. **Benefits**	
	(3)	(4)
+ Average yearly stipend spending	10,181	10,181
+ Added wage bill	54,042	69,179
Net benefit	64,223	79,360
	C. **Returns**	
	(5)	(6)
Short-term return	2.16	2.68

Notes: Table reports results from our cost-benefit analysis. The parameters used in the analysis include the corporate tax rate (τ_c,15%), the social security contribution rate (τ_{ss}, 25.75%), and the personal income tax rate (τ_p, 19.5%). Panel A itemizes direct costs and fiscal incentives. Panel B itemizes economic benefits. Panel C calculates the benefit to cost ratio. Average yearly administrative spending is equal to the sum of (1) 10% of the cumulative program cost to date (including staff salaries, estimated at TND 1,113,828.3) and (2) the development cost of the online platform (TND 470,000), divided by the total number of labeled firms to date (1040). Average yearly stipend spending is equal to $\frac{\text{total amount of stipends received by all eligible firms}}{\text{total number of labeled firms}}$. Foregone social security contributions is. equal to average total wage bill in the post period for labeled firms $\times$ τ_{ss}. Foregone corporate tax revenue is equal to average yearly positive profits in the post period for labeled firms $\times$ τ_c. Personal income tax revenue from new hires is equal to average yearly earnings per worker in the post period for labeled firms$\times$average number of new workers$\times\tau_p$. New workers is equal to the coefficient in Column (3) of Table 3. Added wage bill is equal to the coefficient in Column (4) of Table 3. Monetary outcomes in 2022 TND. Calculations in (1), (3), (5) use the Full Sample, in (2), (4), (6) use the Pitch Sample.

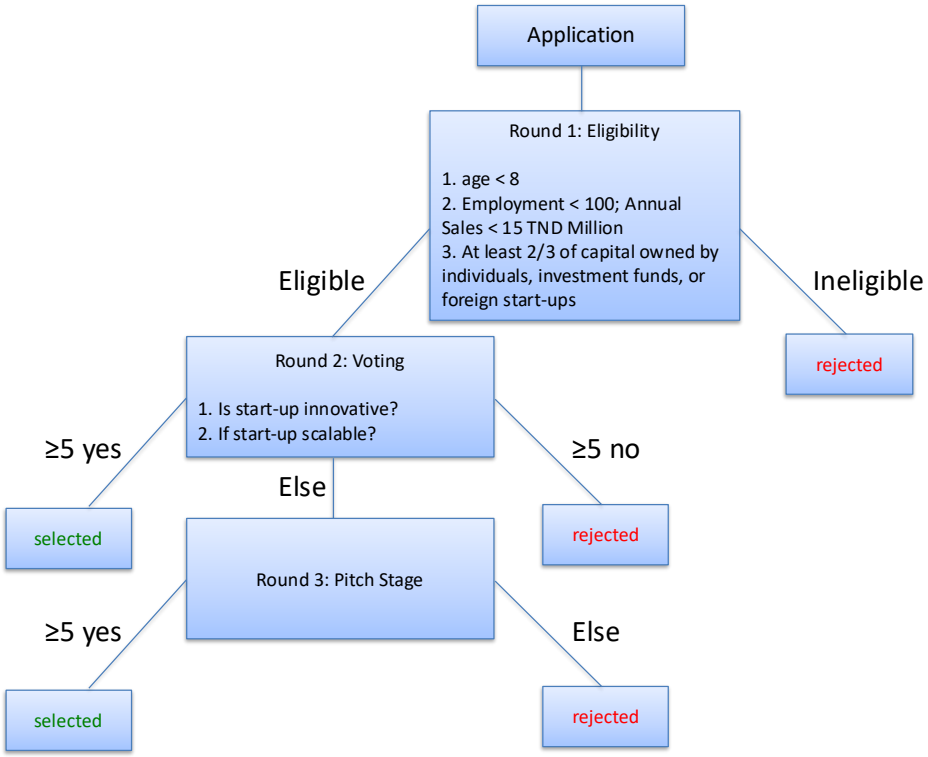
Figure 1: Dynamic Effects of the Start-Up Label Program on Employment, Full Sample, Callaway and Sant'Anna (2021) Estimator



Notes: Figure reports coefficients from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is quarterly employment, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, age dummies, and an indicator for having at least one female founder) in the pre-treatment sample. α_i are firm fixed effects, λ_t are quarter fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event quarter is the quarter of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. Employment is winsorized at the 5% level.

A Appendix

Figure A1: Selection Process



Notes: The figure outlines the three-stage process for the labeling program. In Round 1, firms are screened for eligibility. Firms must be younger than eight years, employ fewer than 100 people, generate under TND 15 million (≈ USD 5 million) in annual sales, and have at least two-thirds of their capital held by individuals, investment funds, or foreign start-ups. Those that meet these criteria advance to Round 2, where a nine-member committee—the "College"—convenes in monthly online sessions to evaluate written applications on the basis of innovation (whether the business model offers a novel solution) and scalability (whether the firm can address demand from a sizable market) before awarding the label through a vote. To be selected outright, a firm must secure at least five approval votes; those that neither achieve five approvals nor five rejections enter a pitch stage, delivering a five-minute presentation followed by a five-minute Q&A before the College votes again—requiring at least five approvals to receive the label.

Table A1: Sectoral Composition, Full Sample

	Control Mean (SD) (1)	Treatment Mean (SD) (2)	P-value (1)-(2) (3)
AI	0.09	0.12	0.25
	(0.28)	(0.33)	
Advanced Manufacturing and Robotics	0.02	0.02	0.83
	(0.15)	(0.13)	
Agriculture And Food	0.04	0.06	0.22
	(0.19)	(0.25)	
Business Software	0.14	0.09	0.16
	(0.34)	(0.29)	
Communication	0.12	0.05	0.01***
	(0.33)	(0.22)	
E-Commerce	0.24	0.25	0.84
	(0.43)	(0.43)	
Ed Tech	0.06	0.09	0.21
	(0.23)	(0.29)	
Entertainment/Travel	0.10	0.06	0.11
	(0.30)	(0.23)	
Environment and Energy	0.02	0.03	0.48
	(0.15)	(0.18)	
Finance	0.01	0.07	0.01**
	(0.12)	(0.26)	
Healthcare and Wellness	0.12	0.09	0.23
	(0.33)	(0.28)	
Other	0.04	0.05	0.77
	(0.20)	(0.22)	
Real Estate	0.00	0.01	0.26
	(0.00)	(0.10)	
F-test of joint significance (p-value)			0.02**
F-test, number of observations			466
N	140	326	

Notes: Table presents descriptive statistics on sectoral composition for the Full Sample using program data from the baseline year, prior to the program's start. Column 1 displays averages and standard deviations (in parentheses) for firms that were rejected, whereas column 2 provides these statistics for firms that received the label. Column 3 reports p-values from t-tests of differences in means between treatment and control groups in the Full Sample. The Full Sample comprises firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table A2: Sectoral Composition, Pitch Sample

	Control Mean (SD) (1)	Treatment Mean (SD) (2)	P-value (1)-(2) (3)
AI	0.06	0.16	0.11
	(0.24)	(0.37)	
Advanced Manufacturing and Robotics	0.02	0.03	0.77
	(0.14)	(0.17)	
Agriculture And Food	0.04	0.06	0.68
	(0.20)	(0.24)	
Business Software	0.20	0.10	0.12
	(0.41)	(0.30)	
Communication	0.12	0.06	0.22
	(0.33)	(0.24)	
E-Commerce	0.31	0.30	0.98
	(0.47)	(0.46)	
Ed Tech	0.04	0.06	0.68
	(0.20)	(0.24)	
Entertainment/Travel	0.06	0.01	0.17
	(0.24)	(0.12)	
Environment and Energy	0.02	0.04	0.50
	(0.14)	(0.21)	
Finance	0.00	0.06	0.09*
	(0.00)	(0.24)	
Healthcare and Wellness	0.10	0.06	0.38
	(0.31)	(0.24)	
Other	0.02	0.06	0.32
	(0.14)	(0.24)	
Real Estate	0.00	0.00	
	(0.00)	(0.00)	
F-test of joint significance (p-value)			0.30
F-test, number of observations			118
N	49	69	

Notes: Table presents descriptive statistics on sectoral composition for the Pitch Sample using program data from the baseline year, prior to the program's start. Column 1 displays averages and standard deviations (in parentheses) for firms that were rejected, whereas column 2 provides these statistics for firms that received the label. Column 3 reports p-values from t-tests of differences in means between treatment and control groups in the Full Sample. The Pitch Sample comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. "Yes votes", "No votes", and "Pitch votes" are the number of yes, no, and invitation to the pitch votes in round 2, respectively. "Advances to Pitch stage" is an indicator equal to 1 for firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round).

Table A3: Vote Statistics, Full Sample

	Control Mean (SD) (1)	Treatment Mean (SD) (2)	P-value (1)-(2) (3)
Yes votes, round2	0.84	5.63	0.00***
	(1.16)	(2.03)	
No votes, round2	5.09	0.65	0.00***
	(2.48)	(0.94)	
Pitch votes	1.74	1.40	0.06*
	(1.95)	(1.71)	
Advances to Pitch stage	0.35	0.21	0.00***
	(0.48)	(0.41)	
F-test of joint significance (p-value)			0.00***
F-test, number of observations			466
N	140	326	

Notes: Table presents descriptive statistics on voting outcomes for the Full Sample using program data from the baseline year, prior to the program's start. Column 1 displays averages and standard deviations (in parentheses) for firms that were rejected, whereas column 2 provides these statistics for firms that received the label. Column 3 reports p-values from t-tests of differences in means between treatment and control groups in the Full Sample. The Full Sample comprises firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table A4: The Predictive Power of Yes Votes for Change in Balance Sheet and Labor Outcomes

	Δ sales	Δ profits	Δ employment	Δ total wage bill
			A. **Unlabeled**	
	(1)	(2)	(3)	(4)
Share Yes Votes	-94.77	-3.87	0.95	9.71
	(142.48)	(3.40)	(1.21)	(49.59)
Conflict declared	-25.20	0.90	-0.02	-1.95
	(101.35)	(2.51)	(0.81)	(33.30)
Yrs since application	-7.03	-0.19	-0.65***	-26.09**
	(29.64)	(0.74)	(0.25)	(10.09)
R2	0.06	0.67	0.11	0.07
N	133	64	140	140
Mean	13.33	1.65	-0.03	-3.17
			B. **Labeled**	
	(5)	(6)	(7)	(8)
Share Yes Votes	99.53*	7.74	-0.04	3.03
	(59.13)	(10.55)	(0.43)	(12.87)
Conflict declared	22.01	1.73	0.34	11.51
	(36.78)	(6.75)	(0.27)	(8.04)
Yrs since application	-22.59	2.04	-0.38***	-6.61
	(18.56)	(3.07)	(0.14)	(4.12)
R2	0.30	0.15	0.23	0.07
N	300	75	326	326
Mean	38.62	5.51	0.55	13.72

Notes: Panels A and B show results from a regression of Y_{it} on share of yes votes (the number of yes votes divided by the number of non-abstaining judges), an indicator for having a judge declare a conflict of interest in the vote, years since application, and 2-digit sector dummies in the last year of the data, 2022. Y_{it} is an outcome of interest, notably the first difference in sales, profits, employment, or the total wage bill. Sample is Full Sample, for which sales information and employment are available in RNE in the baseline year and post period. Panel A is unlabeled firms and Panel B is labeled firms. The "Mean" row reports the mean first difference in the outcome of interest in 2022. Monetary outcomes in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table A5: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Poisson Quasi-Maximum Likelihood Estimator (QMLE) Estimator

	sales	profits	employment	total wage bill
			A. **Full Sample**	
	(1)	(2)	(3)	(4)
Labeled	0.23	-0.03	0.66***	0.65***
	(0.16)	(0.02)	(0.17)	(0.16)
Pseudo R2	0.91	0.47	0.72	0.92
N	1,665	1,352	1,709	1,709
Firms	466	466	466	466
Mean in the control group	241.75	-8.45	1.57	46.48
			B. **Pitch Sample**	
	(5)	(6)	(7)	(8)
Labeled	-0.02	0.06	0.81***	0.47
	(0.45)	(0.06)	(0.30)	(0.34)
Pseudo R2	0.93	0.38	0.74	0.94
N	435	355	451	451
Firms	118	112	118	118
Mean in the control group	350.51	-17.43	1.84	61.07

Notes: Panels A and B show results from equation $Y_{it} = \exp(\beta_0 + \beta_1 Labeled_{it} + \psi_{x(i),post} + \delta_i + \theta_t)\epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, or the total wage bill. $\psi_{x(i),post}$ is a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) interacted with the post dummy. γ_i are firm fixed effects, θ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. Regressions use the Poisson quasi-maximum likelihood estimator (QMLE) difference-in-differences specification (Silva and Tenreyro, 2010; Correia et al., 2020) following the recommendations of Chen and Roth (2024). The Full Sample (Panel A) comprises firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (panel B) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table A6: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level Outcomes, Callaway and Sant'Anna (2021) Estimator, Non-Winsorized data

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **Full Sample**		
	(1)	(2)	(3)	(4)	(5)
ATT	172.4	56.0	2.79***	0.063*	96.3***
	(108.5)	(106.3)	(0.72)	(0.033)	(28.9)
N	1665	1352	1709	1709	1709
Clusters	466	466	466	466	466
Mean in the control group	344.1	9.97	1.90	0.060	58.7
			B. **Pitch Sample**		
	(6)	(7)	(8)	(9)	(10)
ATT	43.3	-58.6*	3.99***	0.056	156.4**
	(240.3)	(35.6)	(1.51)	(0.047)	(66.8)
N	435	363	451	451	451
Clusters	118	118	118	118	118
Mean in the control group	584.4	10.5	2.40	0.061	83.3

Notes: Panels A and B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample (Panel A) is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (Panel B) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are not winsorized. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table A7: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level outcomes, Callaway and Sant'Anna (2021) Estimator, Omitting Firms with Declared Conflict of Interest

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **Full Sample**		
	(1)	(2)	(3)	(4)	(5)
ATT	58.6	-14.5**	1.29***	0.046	41.4***
	(54.5)	(5.72)	(0.44)	(0.030)	(14.8)
N	1318	1076	1342	1342	1342
Clusters	375	375	375	375	375
Mean in the control group	227.7	-7.13	1.44	0.053	42.8
			B. **Pitch Sample**		
	(6)	(7)	(8)	(9)	(10)
ATT	32.8	-21.5**	1.17	0.012	39.5
	(114.3)	(10.3)	(0.81)	(0.042)	(32.2)
N	313	262	321	321	321
Clusters	85	85	85	85	85
Mean in the control group	330.7	-13.9	1.55	0.039	51.9

Notes: Panels A and B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample (Panel A) is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (Panel B) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. The samples omit firms where a judge declares a conflict of interest during the vote.

Table A8: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level outcomes, Callaway and Sant'Anna (2021) Estimator, Comparing Round 2 and Pitch Stage Firms

	sales	profits	employment	Has 10+ emp.	total wage bill
	(1)	(2)	(3)	(4)	(5)
ATT	7.27	-11.1	0.70	0.021	3.29
	(56.1)	(7.96)	(0.54)	(0.037)	(13.0)
N	1187	948	1222	1222	1222
Clusters	326	326	326	326	326
Mean in the control group	272.7	-19.7	2.75	0.11	71.9

Notes: Table shows results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Pitch}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Pitch_{it} is the interaction of an indicator equal to 1 if the firm received the label in the pitch stage and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. The sample includes only selected firms, comparing those selected in Round 2 to those selected in the Pitch Stage.

Table A9: Heterogeneous Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes by Covid Exposure, Level Outcomes, Callaway and Sant'Anna (2021) Estimator

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **low labor input shock (essential)**		
	(1)	(2)	(3)	(4)	(5)
ATT	58.0	-5.18	0.63	-0.025	8.42
	(151.1)	(13.5)	(1.25)	(0.11)	(17.5)
N	117	95	121	121	121
Clusters	35	35	35	35	35
Mean in the control group	222.1	-4.52	1.90	0.059	25.8
			B. **high labor input shock (non-essential)**		
	(6)	(7)	(8)	(9)	(10)
ATT	117.2*	-13.3**	1.81***	0.073**	58.4***
	(60.5)	(6.68)	(0.48)	(0.033)	(18.3)
N	1548	1257	1588	1588	1588
Clusters	431	431	431	431	431
Mean in the control group	243.2	-8.74	1.55	0.060	48.0
			C. **low demand shock**		
	(11)	(21)	(13)	(14)	(15)
ATT	116.8	-8.57	1.68**	0.038	63.7**
	(91.2)	(8.03)	(0.69)	(0.046)	(27.2)
N	1275	1038	1307	1307	1307
Clusters	360	360	360	360	360
Mean in the control group	288.5	-6.27	1.99	0.084	60.9
			D. **high demand shock**		
	(16)	(17)	(18)	(19)	(20)
ATT	137.6**	-19.7*	2.23***		48.7***
	(63.9)	(11.0)	(0.48)		(13.2)
N	390	314	402		402
Clusters	106	106	106		106
Mean in the control group	127.8	-14.1	0.56		11.8
			E. **low intermediate input supply shock**		
	(21)	(22)	(23)	(24)	(25)
ATT	127.9*	-12.7*	1.92***	0.068*	63.2***
	(74.2)	(6.78)	(0.58)	(0.039)	(22.6)
N	1411	1148	1449	1449	1449
Clusters	393	393	393	393	393
Mean in the control group	243.7	-6.55	1.68	0.067	52.2
			F. **high intermediate input supply shock**		
	(26)	(27)	(28)	(29)	(30)
ATT	39.5	-2.95	0.70	0.037	15.0
	(89.2)	(15.9)	(0.78)	(0.069)	(11.4)
N	251	202	257	257	257
Clusters	72	72	72	72	72
Mean in the control group	241.3	-14.9	1.15	0.032	24.1

Notes: Panels A-F show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. We incorporated measures from Le et al. (2024). In A-B, we distinguish between "essential" and "non-essential" sectors. In C-D, we capture sectoral demand shocks. In E-F, we capture intermediate input supply shocks. In col (19), we are unable to residualize the outcome because no firms in this subsample have an outcome of 1 in the pre-period.

Table A10: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level outcomes, Callaway and Sant'Anna (2021) Estimator, Omitting High-Stringency Quarters

	employment	Has 10+ emp.	quarterly wage bill
		A. **All Quarters**	
	(1)	(2)	(3)
ATT	1.36***	0.078**	12.3***
	(0.47)	(0.037)	(4.47)
N	4502	4502	4502
Clusters	466	466	466
Mean in the control group	1.67	0.071	12.1
		B. **Omitting High-Stringency Quarters**	
	(4)	(5)	(6)
ATT	1.39**	0.096**	13.6**
	(0.57)	(0.047)	(5.59)
N	2201	2201	2201
Clusters	466	466	466
Mean in the control group	1.70	0.070	12.3

Notes: Panels A and B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably employment, a dummy for employing at least 10 workers, or the quarterly wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are quarter fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event quarter is the quarter of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. Monetary outcomes are presented in 2022 TND 1k. The wage bill is winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. Panel A includes all quarters. Panel B omits high-stringency quarters, defined as those where the Oxford COVID-19 Government Response Stringency Index (Hale et al., 2021) for Tunisia is higher than than the median over the period 2020q1-2023q1.

Table A11: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Intensive Margin, Callaway and Sant'Anna (2021) Estimator

	sales	profits	employment	total wage bill
			A. **Full Sample**	
	(1)	(2)	(3)	(4)
ATT	117.0	1.90	2.73***	110.0***
	(95.7)	(4.05)	(1.01)	(39.9)
N	1023	186	885	872
Clusters	357	119	304	300
Mean in the control group	469.7	25.5	4.37	131.6
			B. **Pitch Sample**	
	(5)	(6)	(7)	(8)
ATT	129.0	3.34	3.43***	136.4**
	(133.4)	(3.92)	(1.22)	(54.4)
N	281	60	218	214
Clusters	92	34	74	72
Mean in the control group	666.7	36.9	4.65	158.4

Notes: Panels A and B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is a (strictly positive) outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample (Panel A) is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (Panel B) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively.

Table A12: Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level Outcomes, Callaway and Sant'Anna (2021) Estimator, Surviving Firms

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **Full Sample**		
	(1)	(2)	(3)	(4)	(5)
ATT	126.5	-14.5	2.33***	0.085*	79.7***
	(90.4)	(10.8)	(0.69)	(0.050)	(26.4)
N	1166	856	1201	1201	1201
Clusters	320	320	320	320	320
Mean in the control group	450.4	-14.6	2.90	0.11	86.8
			B. **Pitch Sample**		
	(6)	(7)	(8)	(9)	(10)
ATT	136.1	-1.63	3.14***	0.092	114.0**
	(127.4)	(22.9)	(0.98)	(0.065)	(45.9)
N	295	223	308	308	308
Clusters	79	79	79	79	79
Mean in the control group	635.2	-31.7	3.30	0.11	111.1

Notes: Panels A and B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample (Panel A) is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. The Pitch Sample (Panel B) comprises firms that compete for program entry in the pitch stage (i.e. whose application was not immediately accepted or rejected in the second selection round). Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. We use the samples of firms active in 2022.

Table A13: Heterogeneous Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level Outcomes, Callaway and Sant'Anna (2021) Estimator

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **Import-dependent**		
	(1)	(2)	(3)	(4)	(5)
ATT	139.0	-31.1**	2.33**	0.16**	53.4**
	(121.7)	(12.1)	(0.92)	(0.073)	(24.3)
N	418	333	427	427	427
Clusters	116	116	116	116	116
Mean in the control group	204.0	-15.0	1.22	0.043	30.6
			B. **Non-Import-dependent**		
	(6)	(7)	(8)	(9)	(10)
ATT	47.2	-3.49	1.30**	0.022	46.8**
	(60.8)	(4.99)	(0.55)	(0.035)	(22.8)
N	1247	1019	1282	1282	1282
Clusters	350	350	350	350	350
Mean in the control group	253.2	-6.43	1.68	0.065	51.4
			C. **Manufacturing**		
	(11)	(12)	(13)	(14)	(15)
ATT	258.5***	-28.5**	3.11***	0.19***	70.5***
	(77.7)	(12.6)	(0.61)	(0.066)	(17.3)
N	233	182	238	238	238
Clusters	63	63	63	63	63
Mean in the control group	43.6	-5.24	0.28	0	3.67
			D. **Non-Manufacturing**		
	(16)	(17)	(18)	(19)	(20)
ATT	85.6	-10.4	1.49***	0.047	49.2**
	(64.8)	(6.80)	(0.52)	(0.035)	(19.7)
N	1432	1170	1471	1471	1471
Clusters	403	403	403	403	403
Mean in the control group	261.8	-8.76	1.71	0.066	50.9
			E. **Onshore**		
	(21)	(22)	(23)	(24)	(25)
ATT	120.1**	-17.3**	1.56***	0.060*	43.3***
	(60.2)	(7.06)	(0.45)	(0.031)	(13.3)
N	1368	1120	1411	1411	1411
Clusters	388	388	388	388	388
Mean in the control group	175.2	-8.55	1.12	0.032	27.3
			F. **Offshore**		
	(26)	(27)	(28)	(29)	(30)
ATT	-37.5	3.40	1.78	0.032	73.1
	(123.1)	(11.4)	(1.25)	(0.079)	(52.5)
N	297	232	298	298	298
Clusters	78	78	78	78	78
Mean in the control group	563.2	-7.92	3.80	0.20	141.0

Notes: Panels A-F show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, an indicator for having at least one female founder, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. In A-B, we distinguish firms operating in import-dependent sectors, defined as those with import shares above the median, calculated over the entire universe of registered firms. In C-D, we present results for manufacturing and non-manufacturing sectors. In E-F, we examine the impacts on firms operating in the onshore and offshore regimes.

Table A14: Heterogeneous Effects of the Start-Up Label Program on Balance Sheet and Labor Outcomes, Level Outcomes, Callaway and Sant'Anna (2021) Estimator

	sales	profits	employment	Has 10+ emp.	total wage bill
			A. **No female founders**		
	(1)	(2)	(3)	(4)	(5)
ATT	123.2*	-13.2*	1.70***	0.072*	56.8***
	(70.8)	(7.76)	(0.55)	(0.040)	(20.4)
N	1213	990	1246	1246	1246
Clusters	327	327	327	327	327
Mean in the control group	281.6	-12.0	1.75	0.065	51.6
			B. **At least one female founder**		
	(6)	(7)	(8)	(9)	(10)
ATT	77.2	-12.3**	1.90**	0.060	54.7*
	(56.6)	(5.97)	(0.75)	(0.038)	(32.2)
N	452	362	463	463	463
Clusters	139	139	139	139	139
Mean in the control group	150.3	-0.55	1.16	0.047	34.8
P-value for null of coefficient equality	0.67	0.93	0.82	0.83	0.96

Notes: Panels A-B show results from estimating equation $Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \delta_k \cdot \text{Labeled}_{it}^k + \epsilon_{it}$, where Y_{it} is an outcome of interest, notably sales, profits, employment, a dummy for employing at least 10 workers, or the total wage bill, residualized on a vector of firm characteristics in the baseline year (an indicator for being located in Tunis, an indicator for being foreign, an indicator for being onshore, an indicator for having a judge declare a conflict of interest in the vote, and age dummies) in the pre-treatment sample. α_i are firm fixed effects, λ_t are year fixed effects. Labeled_{it} is the interaction of an indicator equal to 1 if the firm received the label and indicator equal to 1 for the post period. Standard errors are clustered at the firm-level. The event year is the year of application to the program. Regressions use the Callaway and Sant'Anna (2021) estimator. The overall ATT is computed by averaging the individual ATTs across all treated groups and post-treatment periods. Full Sample is firms for which sales information and employment are available in RNE in the baseline year (year of application for newly created firms, year before application for older firms) and post period. Monetary outcomes are presented in 2022 TND 1k. Our measure for sales is post-tax sales. Sales, profits, and the wage bill are winsorized at the 5% level. *, **, and *** denote significance at the 10%, 5% and 1% significance level respectively. In A-B, we distinguish firms with and without at least one female founder. Founder gender is deduced from name in the voting data. The p-value for the test of coefficient equality is obtained by first estimating the standard error of the difference via the delta method.

A.1 Data Sources and Cleaning

This section provides an overview of our data sources and cleaning procedures.

A.1. 1 Program Data Cleaning

We digitized the outcomes of monthly voting sessions using reports published online by Smart Capital.[30] For each applicant, the data list the firm name, the founder names, whether the firm is applying for a pre-label or a label, the voting breakdown for Round 2 and Round 3 (if applicable), whether a judge declared a conflict of interest, and the voting outcome. The data also list the firm that transition from the pre-label to the full label. Our analysis focuses on firms that applied between March 2019 and December 2021. For firms that were initially rejected and later reapplied, we retained the outcome of their most recent application.

We excluded from our sample firms that were rejected during the initial screening stage, as well as those that applied only for the pre-label and never advanced. We retained firms that first received the pre-label and subsequently transitioned to the label.

We also obtained from the Smart Capital a list of all applicants as of December 2022, which included firm names and unique tax identifiers. Additionally, we received a list of firms that were awarded stipends, along with the number of eligible founders for each recipient. After extensive cleaning/harmonizing of firm names, we used a fuzzy matching algorithm to link firms across the voting data, the applicant list, and the stipend recipient list. We then used a unique identifier derived from the tax ID to match firms to the RNE database.

Finally, we constructed an indicator for founder gender based on first names, and another indicator for declared conflicts of interest.

A.1. 2 RNE Data Cleaning

The Repertoire National des Entreprises (RNE) is a register maintained by the National Institute of Statistics (INS) of all formally registered private-sector firms in Tunisia. It integrates quarterly firm-level employment and wage data from the Caisse Nationale de Sécurité Sociale (CNSS), annual local, export, and total (pre- and post-tax) sales figures from the Direction Générale des Impôts (DGI), and yearly import and export flows from the Direction Générale des Douanes (DGD). Covering 1996 through 2021, the RNE identifies each firm by the first seven digits of its unique tax number.

[30] These reports are available for download at https://startup.gov.tn/fr/startup_act/results

We began by cleaning the RNE employment and wage data. For quarters with missing values, we applied interpolation using the immediately preceding and following quarters with positive data. If a variable was missing for three consecutive quarters, we set its value to zero. We then constructed annual averages for both employment and the wage bill based on the quarterly data.

Next, we created an "active" indicator variable. In each year following a firm's application, we set this indicator to 1 if at least one of the following variables was non-missing: sales, profits, employment, wage bill, imports, or exports. Our goal was to identify firms that had not permanently exited. Accordingly, if a firm was inactive in year t but returned in a subsequent year, we considered it active in year t.

To construct a balanced panel for our regressions, we imputed zeros for sales, profits, employment, and wages in any post-treatment year where a firm was classified as inactive (i.e., permanently exited). In the reference year, if profits were missing and the firm had no recorded sales or employment, we also set profits to zero.

All monetary variables were deflated to 2022 dinars using the July Consumer Price Index (CPI) from the INS. We winsorized sales, profits, employment, and the wage bill at the 5th and 95th percentiles. Finally, we multiplied the average quarterly wage bill by four to compute the annual wage bill.

Using data from the RNE, we computed the weighted ratio of import revenues to gross total sales at the 2-digit NAT (Nomenclature d'Activités Tunisienne) level for the 2017–2018 period. We winsorized the resulting ratios at the 5% level to limit the influence of outliers. Sectors with an import share above the sample median were classified as import-dependent for the heterogeneity analysis.

A.1. 3 Measures of COVID-19 Exposure

We also incorporated three measures of sector-level exposure to COVID-19, provided by the authors of Le et al. (2024), at both the 4-digit and 2-digit NAT levels. The first measure distinguishes between "essential" and "non-essential" sectors based on labor supply disruptions. The second captures sectoral demand shocks, measured as the average change in firm revenues—realized and forecast—before and after the onset of the pandemic. The third reflects intermediate input supply shocks, defined as the proportion of firms in a sector reporting input constraints in the World Bank's COVID-19 Enterprise Follow-up Survey. For the demand and supply shock measures, we classified a sector as highly exposed if its shock value exceeded the sample median.

A.2 Cost-Benefit Analysis Using the Marginal Value of Public Funds

We complement the cost-benefit analysis presented in the main text (in sub-section 5.5 with an alternative approach based on the marginal value of public funds (MVPF). As noted by Hendren and Sprung-Keyser (2020), the MVPF is useful because it measures the welfare delivered per unit of government spending on a policy. For our case, this calculation requires assumptions about the opportunity cost of gainful employment, which we do not directly observe. Consequently, these estimates are subject to measurement error and serve as a robustness check on the results presented in Table 4).

The MVPF results are presented in Table A15. The program costs are computed as in Table 4, but the calculation of benefits differs in two key ways. First, foregone fiscal revenues (both social security and corporate taxes) are included as benefits, as they eventually accrue to society and contribute to overall welfare. Second, the additional wage bill from new recruits is evaluated against its opportunity cost, i.e., the income the individual would have earned if not hired by the start-up. Since we lack individual-level data to estimate alternative expected incomes, we consider three scenarios: 25%, 50%, and 75% of the actual wage in the start-up. The MVPF results are similar to the cost-benefit calculations, with returns ranging between 1.72 and 2.82.

The estimated range of program returns appears relatively high, irrespective of the calculation method used. However, to assess whether the program delivers a net social benefit, these returns must be compared against the marginal cost of public funds (MCF)—the social cost of raising an additional unit of tax revenue to fund the program. While calculating the MCF for Tunisia is beyond the scope of this paper, as it requires a general equilibrium model and detailed data on the informal economy (which are not readily available), benchmarks from existing studies can provide useful guidance. For instance, Auriol and Warlters (2012) estimate the MCF for 38 sub-Saharan African countries, with a range of 1.05 to 1.72. These values fall below the program's estimated returns, suggesting that the start-up program in Tunisia likely generates net social benefits for the country.

Table A15: Cost-Benefit Analysis, (MVPF)

	Full Sample	Pitch Sample
	A. **direct costs and fiscal incentives**	
	(1)	(2)
+ Average yearly administrative spending	1,523	1,523
+ Average yearly stipend spending	10,181	10,181
+ Foregone social security contributions	23,110	23,415
+ Foregone corporate tax revenue	364	659
- Added personal income tax revenue	5,460	6,136
Net cost	29,874	29,693
	B. **benefits**	
	(3)	(4)
+ Higher earnings for existing workers	12,737	13,772
+ Average yearly stipend spending	10,181	10,181
+ Foregone social security contributions	23,110	23,415
+ Foregone corporate tax revenue	364	659
- Added personal income tax revenue	5,460	6,136
	C. **return (wa=.25w)**	
	(5)	(6)
+ Benefits to new workers	30,979	41,556
Net benefit	71,910	83,446
MVPF	2.42	2.82
	D. **return (wa=.5w)**	
	(7)	(8)
+ Benefits to new workers	20,652	27,704
Net benefit	61,583	69,594
MVPF	2.07	2.35
	E. **return (wa=.75w)**	
	(9)	(10)
+ Benefits to new workers	10,326	13,852
Net benefit	51,257	55,742
MVPF	1.72	1.88

Notes: Table reports results from a cost-benefit analysis using the MVPF methodology (Hendren and Sprung-Keyser, 2020). The parameters used in the analysis are the corporate tax rate (τ_c,15%), the social security contribution rate (τ_{ss}, 25.75%), and the personal income tax rate (τ_p, 19.5%). Panel A itemizes direct costs and fiscal incentives. Panel B itemizes economic benefits. Panel C calculates the benefit to cost ratio. Average yearly administrative spending is equal to the sum of (1) 10% of the cumulative program cost to date (including staff salaries, estimated at TND 1,113,828.3) and (2) the development cost of the online platform (TND 470,000), divided by the total number of labeled firms to date (1040). Average yearly stipend spending is equal to $\frac{\text{total amount of stipends received by all eligible firms}}{\text{total number of labeled firms}}$. Foregone social security contributions is equal to average total wage bill in the post period for labeled firms $\times$ τ_{ss}. Foregone corporate tax revenue is equal to. average yearly positive profits in the post period for labeled firms $\times$ τ_c. Personal income tax revenue from new hires is equal to average yearly earnings per worker in the post period for labeled firms $\times$ average number of new workers $\times$ τ_p. New workers is equal to the coefficient in Column (3) of Table 3. Higher earnings for existing workers is equal to the change in mean earnings $\times$ baseline employment. "wa" is the alternative wage rate. Benefits to new workers is equal to (alternative wage rate - mean firm wage) $\times$ New workers. Added wage bill is equal to the coefficient in Column (4) of Table 3. Monetary outcomes in 2022 TND. Calculations in the first set of columns use the Full Sample, in the second set of columns use the Pitch Sample.