Startups and the State

Raman Singh Chhina*
University of Chicago

Live Document. This Version: November 29, 2025. Please click here for the latest version

Job Market Paper

Abstract

Can bureaucrats in developing countries identify and selectively promote high-growth startups? How much additional gain does selective targeting provide over uniform startup subsidies? I develop a quantitative endogenous growth model of selective targeting in which the entry and growth of different startup types responds heterogeneously to the government's ability to filter and select startups. To estimate this selection ability and quantify the resulting gains, I build a novel dataset of startup selections and rejections from an online startup registry and bureaucratic board meeting minutes, as well as novel patent application data and hand-collected income statements, in the context of the Startup India Program—one of the largest such policies, launched in 2016. I find substantial variation in selection ability across components of the program: startup labeling selects below average-quality firms and distorts exit decisions, whereas provisions of R&D benefits and tax-holiday approvals by a bureaucratic board successfully identify innovative, high-growth startups. The latter double the benefits-to-cost ratio relative to uniform subsidies. I also derive implications for optimal program design by evaluating counterfactual policies that vary the duration and composition of subsidies.

Keywords: Startups, Entrepreneurship, Growth, Development Policy
JEL codes: D22, O31, O38, O47, J24.

*Department of Economics, University of Chicago (email:rschhina@uchicago.edu). I would like to thank my advisors Ufuk Akcigit, Greg Kaplan and Mike Golosov for continued support and guidance. This paper also gained from comments and discussions with Esteban Rossi-Hansberg, Josh Lerner, Nicolas Serrano-Velarde, Emanuele Colonnelli, Sina Ates, Craig Chikis, Furkan Kilic, Varun Kapoor, seminar participants at Macro Lunch Seminar at UChicago, and a number of policy makers and practitioners, especially at the India Policy Forum. I would also like to thank UChicago Economics Department, Michael and Ling Markovitz Dissertation Award for financial support and Stanford University for providing access to Redivis and Pitchbook datasets (and Adrien Auclert for hosting me as a PhD student). Finally, thanks to the Tracxn India team for patiently answering all my questions and the support in gathering firm level data.

1 Introduction

There is a new wave of industrial policies in developing economies where the government plays an active role in identifying and selectively promoting high-potential startups[1]. Selective targeting is arguably valuable because only a small minority of high-potential entrants—such as space technology startups in India[2] or companies like BYD in China—contribute disproportionately to local productivity, growth, and knowledge spillovers. By contrast, a large majority of other entrants remain small and lack any growth ambitions (Hsieh and Klenow, 2014; Hurst and Pugsley, 2011).

Can such targeted startup policies generate any gains as compared to uniform entrant subsidies? While a substantial theoretical and empirical literature examines the tools of industrial policy targeted towards large incumbent firms, policy focus on startups is recent and much less explored. Developing countries often have low state capacity, which makes discretionary policy interventions vulnerable to poor targeting, indiscriminate giveaways, rent seeking, and even regulatory capture (Krueger, 1974; Reinikka and Svensson, 2004; Khwaja and Mian, 2005). Moreover, the high churn in entry and exit among young firms implies that weak state capacity or bad targeting not only supports unproductive firms but also actively promotes their entry and survival. The gains from such programs, if any, therefore depend on how effectively the government is able to select high-potential startups without misallocating resources, and on how different types of startups respond to both this selection ability and the benefits provided.

In this paper, I empirically estimate the targeting ability from a large startup policy program in India and embed it in a model where multiple startup types endogenously respond to the policy— allowing me to map these decisions to evaluate aggregate gains. In doing so, I make three contributions. First, I develop a quantitative endogenous growth model in which multiple ex-ante startup types enter and respond heterogeneously to the probability of being selected or rejected by the program and to the types of subsidies they receive. This implies that, given the government's selection ability and subsidy design, the model generates rich predictions about the behavioral differences between selected and rejected startups, as well as overall changes in entry and exit in response to the policy. Second, I collect new data and document empirical findings on differences in growth between startups selected and rejected by the policy, as well as on the entry and exit of limited liability firms in response to the program. Third, I discipline the model to match these empirical findings and recover the government's ability to select among different startup types and to promote or distort various margins of startup dynamics. Using the quantified model, I evaluate the overall gains from targeting and explore counterfactual designs.

[1] In 2016, India launched the Startup India program (Government of India, 2016) and China launched the Little Giants program (Brown, Chimits, and Sebastian, 2023). In 2018, Tunisia and Senegal adopted 'Startup Acts' followed by sixteen other African countries (Wolken, 2020; ICR Report, 2021). Similar such programs were also launched across Latin America. World Bank identified 54 such programs across developing countries in 2019 (Goswami, Medvedev, and Olafsen, 2019). Horizon Europe and UK SEIS are similar examples from high-income countries.

[2] See The Economist (2022) and The New York Times (2023)

I study the bureaucratic selection ability and the aggregate gains in the context of the Startup India Program—one of the largest such policies. Formally launched in 2016, it serves as an umbrella-term for several targeted benefits provided to innovative and high-growth startups. The program operates in two main stages. In the first stage, any incorporated firm less than ten years old, regardless of sector, is eligible to apply for government recognition via an online application. If selected, a startup gets a suite of benefits, including exemptions from labor and environmental regulations, interest and lease subsidies, eligibility for government seed funding, and preferential access to public procurement. Since 2016, approximately 160,000 startups have been recognized under the program. In second stage, recognized startups can apply for and get additional benefits like tax holidays, incubation, and venture funds. For profit tax holidays, their applications are considered by a bureaucratic board made up of high-level technocrats from the government department of industrial policy, science and technology, biotechnology etc. Conditional on selection they get a three year tax break on their corporate profit income. Similarly, startups engaged in R&D activities receive patent-filing subsidies and may apply for incubation support at government research institutions, which are further screened by the host institutions. All of the above benefits are subject to a sunset clause and the startup loses them at the age of ten. Similar multi-stage structures—combining startup labelling with additional selective subsidies—are common in startup policy programs across developing countries[3].

Theory I embed this policy—with selective targeting and multiple instruments—into an endogenous growth model featuring ex-ante heterogeneous types in order to evaluate its impact on firm dynamics and aggregate growth. At the core of the model is the government's "selection ability" function and a set of "effective subsidies," which jointly determine each type's decisions of entry, growth, and probability of survival. A marginal entrant of each type decides to enter based on its probability of being selected for the program and the additional benefits it may receive if selected. Thus, the model generates a rich set of predictions about the policy's impact on different margins of startup dynamics which I use to back-out the selection ability, effective subsidy parameters, and evaluate aggregate gains.

The model adds two layers of ex-ante firm heterogeneity in a quality ladder growth model[4] to quantitatively match the firm and growth dynamics in a developing economy. Only a small number of firms engage in frontier innovation in developing economies, while a vast majority either make no productivity improvements or adopt technologies already existing in advanced economies. Thus, in the first layer, I introduce firms which are ex-ante heterogeneous in the type of activities they do to expand — innovative (R&D) firms and non-innovative (non-R&D firms). I assume that all the R&D firms are alike but introduce a second layer of heterogeneity in the type of non-R&D firms. Each non-R&D firm is endowed with a heterogeneous search ability, determined at birth. Firms with high search ability are able to more efficiently find and adopt ideas

[3]See Appendix A
[4]Klette and Kortum (2004); Acemoglu, Aghion, and Zilibotti (2006); Acemoglu et al. (2018)

to make productivity improvements thus expanding quickly. By contrast, firms with the lowest search ability never make productivity improvements after entry but still use a portion of the economy's resources for production. This is in accordance with the vast literature documenting large ex-ante differences in firms where a vast majority enter with no growth ambitions (Hurst and Pugsley, 2011; Guzman and Stern, 2020).

The main externality in the model is an intertemporal knowledge spillover that depends on the overall composition of entrant and incumbent firms. Each successive productivity improvement by a firm builds on the prevailing level of productivity in the economy, which itself reflects past improvements made by other firms. However, firms do not internalize that their investments in R&D or search activities contribute to these intertemporal spillovers. Since innovative and high-type R&D firms generate larger spillovers, shifting the composition of entry or reallocating more resources toward these firms can lead to aggregate productivity gains. Moreover, as the economy transitions towards the frontier — where only productivity improvements can be made by innovative firms — the aggregate gains on the path also depend on which type of firms are subsidized.

On the economy's transition path towards the frontier, I model the startup policy as an unanticipated shock with three parameter blocks: (i) government selection ability[5], (ii) subsidies that act as effective wedges on firms' different margins and choices (profits, fixed operating costs, R&D investment costs, and exit probabilities), and (iii) an age-based limit on benefits—capturing all the key policy design features. Given complete information about government selection ability across types and different subsidy wedges, entrepreneurs of different types in the entering cohort optimally choose their entry decisions. Hence, a startup policy not only influences overall entry but also the *composition* of the entrants. This composition depends on selection ability: a marginal entrepreneur weighs the probability of receiving program benefits upon entry. If the program is highly selective, only high-type entrepreneurs find it worthwhile to enter. If selectivity is low, entry remains heterogeneous but now varies with the types of benefits offered and how much each benefit raises the value of entry for different types. In addition, because subsidies depend on a startup's age, post-entry outcomes for selected and rejected firms depend on both the composition of selected startups and the effective subsidy wedges they receive.

The gains from the startup policy depend on how effectively the government can target innovative and high-growth startups without misallocating resources to low-type entrepreneurs. The extent of such misallocation depends on the share of low-type entrepreneurs that receive benefits and on how the policy affects their entry decisions and probability of survival. I identify the government's selection ability, as well as the resulting gains and distortions, by matching the model's predictions to the empirically observed startup dynamics in response to the policy, which I describe next.

[5] Rather than modeling selection as an explicit screening technology, I treat selection ability as the reduced-form outcome of the process—bureaucratic screening, ordeals, application costs, political connections, etc.—and measure it directly in the data.

Data Collection and Empirical Findings One of the main contributions of the paper is to construct a new dataset that tracks startup entry, survival, and selection or rejection under the policy, along with the subsequent outcomes of selected and rejected startups. Such detailed, program-linked data are rarely available in less developed economies, where the implementation of these policies faces the greatest challenges. I use this dataset to document the program's impact on startup entry, growth, and survival. Wherever possible, I empirically separate the effects of selection from those of subsidies. Ultimately, my main identification strategy relies on using these empirical findings to estimate the policy parameters by matching model-generated moments to their counterparts in the data.

To construct the dataset of applications and selections into the program, I use data from an online startup registry and meeting minutes of a bureaucratic board. I begin by assembling a dataset of all limited liability firms in the economy, using the official Indian business registry to track their entry and survival. Among these firms, I identify those that applied for and received a "Startup Recognition" through an online government portal, which publishes basic information on all registered startups and their recognition status. For the second stage of the program—profit tax holidays—the portal only discloses the firms that were selected. To recover the full universe of applicants and decisions, I digitize the meeting minutes of the bureaucratic board responsible for evaluating tax holiday applications. These minutes provide the list of all applicants, along with the decisions, reasons for approval or rejection, and relevant dates.

To measure startups' outcomes, I assemble data from two new sources: firm-level patent filings and hand-collected income statements and balance sheets. Because a key goal of the policy is to promote innovative startup activity, I first construct a dataset of all patent applications filed with the Indian Patent Office between 2012 and 2022, identify applicant types, and merge these with the universe of limited-liability firms in India. To evaluate growth outcomes for selected and rejected startups that do not engage in R&D, I then hand-collect income statements and balance sheets for all firms that applied for tax exemptions by partnering with a Tracxn, a data collection startup in India. These data are high quality—firms are legally required to file them—and I address sample selection concerns by collecting financial statements for every firm that applied for tax exemptions between 2016 and 2019.

Using this data, I document three sets of empirical results, corresponding to the policy's three main instruments—tax holidays, R&D benefits, and recognitions. For each instrument, I show how entry responds to the program and how selected and rejected startups differ in their growth, composition, and survival. These responses are shaped jointly by the types of startups that enter post-policy, the cumulative effects of different instruments, the general equilibrium effects of the program, and the counterfactual no-policy transition of the economy. Thus, the goal of the empirical analysis is to document the key patterns in startup behavior around the introduction of the policy, which then serve as important calibration targets for the model.

First, using a differences-in-differences strategy—comparing patent applications filed by startups to those filed by older firms over the sample period—I show that post-policy patent applications by startups increase by nearly 50%, with no significant pre-trends. Splitting the analysis by recognized and non-recognized startups reveals that the effect is entirely driven by newly formed, recognized startups, while patenting by non-recognized startups remains relatively unchanged before and after the policy. Moreover, a firm fixed-effects specification among startups with at least one pre-policy patent shows no intensive-margin effect of the policy on patenting. Thus, the policy primarily operates by increasing the entry of startups engaged in R&D.

For tax holidays, I show that the bureaucrats have a strong ability to select high-growth firms. On average, firms selected by the bureaucratic board increase their revenue by a factor of 11 by age eight, relative to entry, while rejected firms grow by only a factor of 1.7. To disentangle selection ability from the causal impact of tax holidays, I compare outcomes of selected firms with two sets of rejected firms. In the first set, I match selected firms with otherwise similar firms that were rejected solely due to a date-of-formation cutoff. In the second set, I match selected firms with rejected firms that exhibited similar growth and other characteristics prior to the exemption decision. I find that most of the differences between selected and rejected startups are driven by the board's selection ability.[6] The causal effect of tax holidays, estimated using the matched firms, explains only a small fraction of the observed differences.[7]

Lastly, for recognitions, I show that recognized startups—though similar to other firms in terms of equity capital—are 50 percentage points more likely to survive to age ten. These differences persist even after controlling for sectoral and regional variation across selected and rejected startups. Finally, exploiting variation in selection probabilities across Indian districts, I find that a one–percentage-point increase in recognition probability in a given year is associated with a 2.5% rise in overall firm formations in the subsequent year and a 9.7% increase in the number of firms that attain recognition within a year of incorporation.

Quantification and Main Results To quantify the model and estimate the government's selection ability and subsidy wedges, I jointly match the model's long-run transition and its response to an unanticipated policy shock—while still on the transition path—to the data. Since the policy primarily affects entry, its composition, and the allocation of resources among young firms, the resulting gains depend on parameters that determine the pre-policy allocation of resources between entry and incumbents of different types. I estimate these parameters—namely, the costs of entry and expansion for different types of firms—by matching the model to (i) the levels and transition path of entry by innovative and non-innovative firms, (ii) the evolution of the average firm life cycle between 1990 and 2010, and (iii) the survival probabilities of firms born in

[6] I provide evidence that the board relies on observable signals—such as VC investment, patent filings, and incubation status—to make its decisions. For example, having VC investment before the meeting increases the probability of selection by 15 percentage points, while having been incubated increases it by 6 percentage points.

[7] Analysis of profits shows that most startups are not profitable within their first ten years, which limits the impact of the subsidy. However, conditional on earning profits, the subsidy reduces effective tax rates by 4 percentage points in the first eight years.

each cohort over the transition. Together with these fundamental parameters, I then estimate the selection ability and subsidy wedges by matching ten dynamic model moments of the policy response to the empirical findings documented earlier. These include moments such as the entry responses, differences in growth and survival of selected and rejected startups for each of the policy instrument.

The calibration exercise reveals strong targeting ability for R&D benefits and tax holidays. On average, startups selected for tax holidays are 2.2 times more efficient at searching for expansion ideas than the mean pre-policy entrant. In contrast, startup recognitions perform poorly: the average selected firm is only 0.9 times as efficient as the pre-policy entrant. Why does the exercise finds adverse selection for startup recognitions? Intuitively, the model gives a relationship between survival and growth of each firm type. As startups selected for recognitions survive a lot more but don't grow as much, the model implies that they have a larger share of low type firms which never grow after entering. Simultaneously, the calibration exercise finds that recognition not only selects firms adversely but also significantly distorts their survival, as reflected in a large exit wedge. Finally, although startup entry responds strongly to the policy, the selected firms do not grow faster. This implies a large operating cost wedge that makes entry more attractive, but only a small effective search subsidy wedge, since it doesn't translates into faster post-entry growth.

With the selection ability and effective subsidy wedges separated, I next examine whether there are any gains from selective targeting at the observed levels of selection ability. I keep the effective subsidies for tax holidays and recognitions fixed but assume that each entrant type has an equal probability of being selected. The results reveal a clear pecking order in the benefits-to-cost ratio (BCR) across policies: targeted R&D benefits have the highest BCR of roughly 7, followed by targeted tax holidays at 5, and recognitions—with their observed low targeting ability—at just 2.84. By contrast, removing targeting from tax holidays halves the BCR, while doing so for recognitions only reduces it by 0.04 points. Why is there still a drop? Even with weak targeting, some low-type firms were previously screened out. Without any targeting, more of these low-type firms would enter, increasing misallocation and reducing overall policy effectiveness. Thus, overall, selective targeting doubles the BCR when the policy effectively targets, and in cases where targeting is weak, the results are nearly indistinguishable from having no targeting at all.

Would the gains from targeting be higher if the policy provided different types or lengths of subsidies? The answer depends on the subsidy design, as different startup types respond heterogeneously. For fixed operating-cost subsidies—the main feature of recognitions—the entry of the lowest-type startups responds the most. These firms do not grow, so a fixed subsidy disproportionately raises their value relative to high-growth startups. By contrast, entry by high-potential firms responds more strongly to profit tax holidays—but only when those holidays extend over longer sunset periods. Since low-type firms' profits do not grow with age, the value of the tax exemption flattens quickly, whereas high-type firms continue to grow and value the exemption more later in their life cycle. Longer sunset durations, however, come with costs. In the final part of the paper, I show that extending the sunset beyond twenty years would overturn the welfare gains from recognitions and make them negative. Thus, sunset clauses act as a form of insurance: they reduce the upside of effective targeting but also limit the downside risk of poor targeting.

Literature This paper contributes to two main emerging literature: (i) high-growth startups and startup policy in developing economies; and (ii) ex-ante entrepreneurial heterogeneity and the role of policy in shaping entry composition. The first literature has documented the historic lack of high-growth firms in developing economies (Goswami, Medvedev, and Olafsen, 2019; Eslava, Haltiwanger, and Pinzón, 2022; Eslava, Haltiwanger, and Urdaneta, 2024; Peters and Zilibotti, 2021) and recent rise of such firms alongside the emergence of venture capital markets (Lerner et al., 2024; Colonnelli, Li, and Liu, 2024; Colonnelli et al., 2025). I contribute by collecting new data and facts about high-growth startups in India — which is the third largest economy by VC investment after US and China but has received limited attention. A large body of work that studies Indian firms mostly relied on census or survey data focused on manufacturing (Hsieh and Klenow, 2014; Martin, Nataraj, and Harrison, 2017; Rotemberg, 2019; Bertrand, Hsieh, and Tsivanidis, 2021; Bau and Matray, 2023). I provide new data sets that help in documenting the overall business dynamism and prevalence of high growth firms in the economy.

In the context of startup policies, the literature has mostly focused at firm-level empirical evaluations of specific instruments implemented at small scale such as incubators (Gonzalez-Uribe and Leatherbee, 2018; González-Uribe and Reyes, 2021; Nejad, 2024), business competitions (McKenzie, 2017) or startup labeling (Ali, Calì, and Rijkers, 2024). I contribute by evaluating both, the firm level impact of such policies as well as the long-run aggregate impact. Moreover, by explicitly incorporating selection ability and the impact of different instruments as effective subsidies in the firm problem I am able to jointly study the diverse set of tools available to the policy makers.

The second strand of the literature uses data from developed economies to document large ex-ante differences in growth ambitions among entrepreneurs (Hurst and Pugsley, 2011; Sterk, Sedláček, and Pugsley, 2021; Guzman and Stern, 2020). Recently, the literature has also started exploring the role of policies in shifting the entry composition of firms such tax policy (De Haas, Sterk, and Van Horen, 2022) or education policy (Akcigit et al., 2025). I contribute both empirically and theoretically to this literature. Empirically, I document the differing entry response of different startup types in response to policies that selectively promote either R&D by startups or other other type of firms. Theoretically, I build and calibrate a model in which different startup types respond endogenously to both: different policies as well as their expectations about getting the subsidy when the policy makers are not able to target perfectly. In other contexts, Alatas et al. (2016), Adda and Ottaviani (2024) and Kleven and Kopczuk (2011) show that self-selection of applicants can play an important role in allocation of benefits when the governments screen the applicants for the grant of subsidies, I rather introduce "selection ability" as an outcome of the selection process that includes self-selection, screening, ordeals etc and directly estimate it in the data.

2 Institutional and Policy Background

The new wave of Startup Policies Startup policies have been adopted by multiple low- and middle-income economies, as well as in developed countries, over the past decade. In this paper, I focus on the Startup India Program, but in Appendix A, Table 12, I provide examples from Nigeria, Tunisia, Senegal, Brazil, Italy, Spain, Portugal, and France, each of which implements targeted benefits for young startups selected through a discretionary process.[8] The first stage of most of these programs involves the selection and labeling of high-potential startups to grant them targeted benefits. In India, this is known as "Startup Recognition"; in Nigeria and Tunisia, as a "Startup Label"; and in France, firms can obtain the Jeune Entreprise Innovante (Young Innovative Firm) status. Once selected, firms become eligible for benefits such as corporate tax holidays, exemptions from employer social contributions, regulatory waivers, access to special foreign-currency accounts, and dedicated funding or incubation support. Most of these programs also impose an age-based cutoff for benefits. In India, Nigeria, and Senegal, firms lose their status after ten years, while in France, tax exemptions last for seven years. Next, I discuss the specifics of the Startup India program in detail.

Startup India Program The Indian government launched the Startup India program in April 2016 (Government of India, 2016) with the aim of promoting startup activity. Although India has been relatively successful in attracting venture-capital investment and fostering startup activity given its per capita income (Lerner et al., 2024), overall business formation and formal entrepreneurship rates remain strikingly low (Appendix Figure 18a). Motivated by the rapid success of a small set of high-growth startups and a growing policy consensus around supporting high-potential firms, the government introduced a suite of high-profile initiatives under the Startup India scheme.

The program is administered by the Department for Promotion of Industry and Internal Trade (DPIIT)[9] and operates in successive stages. In the first stage, as in similar programs, any incorporated firm less than ten years old may apply for "Startup Recognition." Applications are submitted online through a government portal and must include a description of the business idea, founder background, and supporting documents. After review, the startup is either approved for recognition or rejected. As of December 2024, roughly 160,000 startups had been recognized. Once recognized, startups receive a de facto bundle of benefits, including regulatory

[8] These programs often include ex-ante eligibility criteria; for example, under the Nigerian Startup Act, only "technology-enabled" firms may apply.

[9] Previously named the Department of Industrial Policy and Promotion

exemptions from labor and environmental laws, relaxed public procurement norms, and simplified procedures for winding up. Recognition is also a prerequisite for major government funding schemes[10], access to incubators at public institutions, and various state-level programs that offer benefits such as interest rate subsidies and lease rental support. In addition to estimating the value of this bundle, I focus on two major benefits for which recognized startups become eligible in the next stages: R&D benefits and tax holidays.

Once recognized, a startup automatically becomes eligible for the Startup Intellectual Property Protection scheme. This provides fast-track processing for patent and trademark applications, access to facilitators who assist with filing, and an 80 percent rebate on patent fees. The program has also launched a number of incubators at government institutions and universities across India. While each incubator operates under its own set of rules, firm incorporation and recognition are typically core eligibility requirements. However, recognition does not guarantee research or funding support at these incubators, as further screening is usually required. According to policy documents, as of December 2024, startups supported by the program have filed approximately 13,000 patents and 49,000 trademarks.

Table 1: Summary of the Main Policy Instruments

	Startup Recognition	Profit Tax Holidays	R&D Benefits
Eligibility	Age less than ten years Any sector	Recognized Startup	Recognized Startup
Selection Process	Online application	Review by a bureaucratic board	-Patent Filing, -Selection by Incubator
Benefits	- Regulatory exemptions. - Relaxed public procurement. - Eligibility for credit guarantee, seed funds, and state level lease subsidies/incubators - Eligibility for Government VC	Corporate income tax holiday	- Patent filing subsidies - Research Incubators - Special Funds
Sunset Criteria	Ten years of age	Any three consecutive years within first ten years	Ten years of age
Total Counts	164,000 Startups Selected	4,100 Startups Selected	13,000 Patents by Selected Startups

The table provides the summary of the main instruments used by the policy and the corresponding number of startups selected. Appendix A.2 provides additional details on the selection process of the firms for each stage. The firm selection numbers correspond to the end of 2024. Section 4.1 provides the details of the samples used for the analysis.

[10]These include a Seed Fund Scheme, a Credit Guarantee Scheme, and a Fund of Funds through which the government provides capital to private VCs to invest in startups.

Similarly, recognition makes a startup eligible to apply for a tax holiday. However, the startup must submit a separate, detailed application that includes information on funding, income statements, recognitions, and awards to an Inter-Ministerial Board. The Board, comprising members from the Department for Promotion of Industry and Internal Trade, the Department of Biotechnology, and the Department of Science & Technology, evaluates each application before granting the tax holiday. The exemption allows the startup to forgo profit taxes for any three consecutive years within the first ten years of incorporation. As of October 2025, approximately 4,100 startups have been granted profit tax exemptions. Table 1 summarizes the three main instruments I analyze in the paper—recognitions, R&D benefits, and tax holidays—along with other key policy details.

3 Model

To study the impact of three key policy features: 1) recognitions, tax holidays and R&D benefits, 2) with discretionary selection of young startups, and 3) age limits, this section builds a novel endogenous growth framework. The model builds on the Klette and Kortum (2004); Acemoglu, Aghion, and Zilibotti (2006); Acemoglu et al. (2018) models of innovating firms but augments in multiple directions relevant to studying the policies in a developing country.

In order to capture the large ex-ante heterogeneity among startups and growth dynamics of a developing country, where a large portion of growth comes from adopting technologies rather than innovating, the model adds multiple types of ex-ante heterogeneous firms. The firms are of two broad types: R&D firms and non-R&D firms. All the R&D firms are of ex-ante same type but each non-R&D firm belongs to an ex-ante different type which determines their average efficiency of searching and adopting new technologies. R&D firms increase their size by creative destruction i.e. by doing productivity improvements on existing products. The role of non-R&D firms, on the other hand, depends on the distance of the economy from the frontier. If the economy is fully developed, these firms do not contribute to aggregate growth but grow and shrink themselves by searching for existing ideas in the economy and replacing other R&D firm and non-R&D firms who exit due to exogenous shocks. If the economy is behind the frontier, however, these firms can contribute to growth by adopting technologies from the world frontier and replacing firms with previous vintages[11].

Lastly, I incorporate government "selection ability" into the model, which determines the probability that a given type of firm will receive a subsidy. I then model startup policies as "effective subsidies" that affect key margins of the firm's problem, including profits, fixed operating costs, R&D and search investment costs, and the likelihood of a destruction shock. Since these policies have age limits and selection is imperfect, each firm type forms expectations about its chances of

[11] These assumptions capture two key phenomenon. First, it is no longer true that a high growth firm contributes more to growth but rather it would depend on its activity type and state of the economy. Second, it allows to study targeting of different firms: R&D and multiple ex-ante types of non-R&D firms depending on the state of the economy.

being selected and receiving benefits upon entry, making the entry decision endogenous to the policy design.

3.1 Environment and Preferences

Time is continuous. I consider a single-country economy that is closed to trade in goods and labor markets but interacts with the rest of the world through a global knowledge frontier. It is populated by a single representative consumer that derives logarithmic utility from consumption. The discounted sum of the lifetime utility at time t is given by

$$U_t = \int_t^\infty e^{-\rho(s-t)} \ln C_s ds$$

where C_s is the consumption at time $t = s$ and ρ is the discount rate. The budget constraint of the representative household is given by

$$C_t + \dot{B}_t = w_t L_t + r_t B_t - T_t$$

where L_t is the labor (supplied inelastically and set equal to 1), w_t and r_t are the equilibrium wages and interest rates, B_t is their total aggregate asset holding and T_t is the total lump sum taxation used by the government to fund its policies.

3.2 Final Output

The final output good which is used for consumption is produced by a perfectly competitive aggregator with the following production technology

$$\ln Y_t = \int_0^1 \ln y_{jt} \, dj \tag{1}$$

where y_{jt} is the variety of good $j \in [0, 1]$ at time t. Each variety is produced by a monopolist firm that owns the production line at that instant of time.

3.3 Intermediate Production

Each intermediate variety is produced by a firm using a single input factor: labor l_{ijt}. Only the firm with the latest vintage of productivity produces in a given line which is denoted by q_{jt}. The production function, which is same for each type of firm, is thus given by $y_{ijt} = q_{jt} l_{ijt}$ for a firm i in line j.

3.3.1 Frontier and Laggard Product Lines

Productivity in each line changes over time as, in addition to producing, the firms also invest in adding new product lines to their firm by making productivity improvements over the incumbent producer of a given variety. These individual productivity improvements lead to aggregate productivity growth in the economy and the process is described in Section 3.3.3 and 3.7.2. However, as the economy grows different lines have heterogeneous productivity levels with some very far behind the frontier and other lines leading the world economy.

Thus, I classify each line into one of the two status types: Frontier and Laggard. The status of a line, at time t, depends on the productivity level of the line compared to the line specific world frontier productivity. Denote by $\bar{q}_{j,t} \quad \forall j$ the world frontier productivity in line j at time t. Given $\bar{q}_{j,t}$ and the current domestic productivity of the line q_{jt}, it is grouped as either frontier or laggard i.e. $\mathbb{Q}(j) \in \{F, L\}$ as follows

$$\mathbb{Q}(j) = \begin{cases} F & \text{if} \quad q_{j,t} = \bar{q}_{j,t} \\ L & \text{if} \quad q_{j,t} < \bar{q}_{j,t} \end{cases}$$

i.e. the line is a frontier line if its productivity is equal to the world frontier productivity and laggard if its productivity is less than the world frontier at time t.

The total mass of the lines that are at the frontier is given by M_t^f and the mass of laggard lines is denoted by M_t^l. For simplicity, I assume that the world frontier productivity in a line remains constant until domestic productivity catches up. After that point, both move together, driven solely by domestic productivity improvements.

3.3.2 Static Production Choice and Profits

In equilibrium, each line is owned by an entrepreneur who has the highest productivity q_{jt} (a single entrepreneur can hold multiple product lines at the same time). Given the final good production technology in Eq. 1, the demand for each variety is unit-elastic and and the revenue share of each variety is the same, equal to the aggregate output $p_{jt}y_{jt} = Y_t$ (aggregate price index normalized to 1). Once the firm starts operating, it becomes an immediate monopolist in that line and I assume that each firm charges the same markup μ, which I take as exogenous. This implies that the profits of the monopolist producing in the line are given by

$$\pi_{jt} = \frac{\mu - 1}{\mu} Y_t = \pi_t \tag{2}$$

This also implies that the total employment in a line is $l_{jt} = l_t = \frac{1}{\mu}\frac{Y_t}{w_t}$ — both of which are independent of the level of productivity in the line[12]. Thus a firm operating n frontier lines and l laggard lines makes $(n + l)\pi$ flow of profits.

[12]This assumption simplifies the analysis and is common in the growth literature (Peters and Zilibotti, 2021)

3.3.3 Intermediate Firms: R&D and Non-R&D Entrepreneurs

I assume two classes of firms (each a collection of product lines): R&D (innovative) firms and non-R&D. A firm's class is fixed at birth and cannot change after entry. In addition to producing, R&D firms invest in research to expand their size by pushing the frontier on randomly drawn lines and adding those improved lines to the firm. Non-R&D firms expand by searching for and adopting *existing* frontier technologies; they differ in their search capacity, defined below.

R&D Firms hire workers to do research, in addition to hiring workers for production. If the firm i has n products and hires h_{it} workers for researching productivity improvements, it successfully adds a product line to its portfolio at flow rate $x_{it} = n\left(\psi_t^R h_{it}\right)^{\frac{1}{\zeta_R}}$. Hence, the cost function of successfully generating innovations at rate x_{it} for a R&D firm with n products is given as

$$C_R(n, x_{it}) = n\frac{x_{it}^{\zeta_R} w_t}{\psi_t^R}$$

If the firm is successful in generating a productivity improvement, it immediately replaces the incumbent firm and becomes the monopolist producer. What happens to the productivity of the line? It depends on the status of the line. If the line was laggard, with productivity below frontier, an innovation by a R&D firm brings it immediately to the frontier productivity level $\bar{q}_{jt}$. If the line was already frontier, the productivity improves multiplicatively by a factor λ. Thus, all the products of a R&D firm are always on frontier lines.

Each product line held by R&D firms can also be lost due to two events: i) creative destruction, at flow rate τ_t, when another R&D firm innovates on that line, it displaces the incumbent holder through a quality-improving step, ii) exogenous replacement, denoted by $\kappa\Psi_t$, when the firm is hit by a "bad-luck" shock on one of its lines and loses the product to another firm. In the second case, other firms do not need to make productivity improvements to displace the firm and instead replace it at the same level of productivity.

Lastly, I also assume that the firm is required to pay a constant c_R flow operating costs. Thus, a firm with n product lines makes $n\pi_t$ amount of flow profits, pays c_R operating costs and optimally chooses x_{it} by weighing the cost of R&D against the potential gains from expanding through the addition of new product lines. As firms are forward-looking, they also internalize the hazard of losing product lines due to the events described above. However, they do not internalize that their R&D investments contribute to aggregate productivity growth in the economy or that they shift product lines from laggard to frontier status.

Non-R&D Firms also hire workers to invest in searching for new ideas. However, they can only make productivity improvements by searching for ideas *already existing* in the world economy i.e. they cannot push the frontier. Their search costs have a similar structure to R&D firms but with different parameters. Moreover, I assume that non-R&D firms have heterogeneous search capacity $z \in \tilde{Z}$ which is determined at birth and stays constant over time. Specifically, for a firm of type z to successfully find ideas at rate x_{it}, the cost function of firm which has n products in the frontier lines and l products in the laggard lines, is given by

$$C_N(n+l, x_{it}, z) = (n+l)\frac{x_{it}^{\zeta_N} w_t}{z\psi_t^N}$$

i.e. the high z firms have lower search costs to get to the same rate of getting ideas as compared to low type firms. The search for all type of firms is un-directed and leads to one of the following outcomes

1. The firm finds a new idea for a production technology in a laggard line with productivity $\tilde{q}_{jt}$ between $(q_{jt}, \bar{q}_{jt})$, i.e. the current incumbents productivity and the world frontier productivity. I term this as *technology adoption*.
2. The firm finds a non-R&D firm operating in a frontier line that is experiencing a "bad luck" shock and replaces it at the same level of productivity.
3. The firm finds a R&D firm operating in a frontier line that is experiencing a "bad luck" shock and replaces it at the same level of productivity. This, however, happens only with probability $\kappa < 1$.

Intuitively, outcomes 2 & 3 capture the churn in the economy whereby a large majority of firms in the economy get replaced by other firms due entrepreneurs life-cycle exit and other negative idiosyncratic events[13]. Further, I assume that with probability γ, a technological adoption is large and brings the line to the frontier and in the other case, i.e. with probability $1-\gamma$, the resulting productivity improvement is given by $\ln\tilde{q}_{jt} = \ln q_{jt} + \theta[\ln\bar{q}_{jt} - \ln q_{jt}]$ with $\theta \sim \text{Uniform}(0,1)$ i.e. the resulting productivity improvement is a random draw between the current incumbent and the frontier productivity. As a result, in the case of Outcome 1, the firm gains a frontier line product with probability γ i.e. $((n,l) \to (n+1,l))$ and a laggard line product with probability $1-\gamma$, moving the product portfolio to $(n, l+1)$. In the case of outcomes 2 & 3, the firm always gains a laggard product.

Only total number of product lines matter for profits $(n+l)\pi_t$, constant flow operating costs c_N, and search costs $C_N(n+l, x_{it}, z)$. However, as the firms are forward looking, the product mix impacts their overall value as different kind of product lines face different hazards of being hit by creative destruction, adoption of replacement shocks. The firm loses frontier and laggard products with same rate τ_t (creative destruction) plus Ψ_t (replacement). However, laggard line products of firm face an additional hazard of being lost to technological adoption A_t done by other firms in the economy.

[13] $\kappa < 1$ captures the intellectually property rights of firm with R&D whereby their replacement might be harder than replacement of non-R&D firms

3.4 Startup Subsidies and Bureaucratic Selection

I assume that the startup policies provide a set of subsidies to selected firms. Each firm gets these subsidies once, at birth, and they expire as the firm reaches a certain age[14]. I do not model an explicit application process but rather assume that the selection process leads to a government "selection ability" which is described below along with the details of tax exemptions, R&D benefits and recognitions.

Firm Recognitions and R&D Benefits Both R&D and non-R&D firms can be recognized/labeled by the government. $h_i \in \{0,1\}$ is an indicator which denotes whether the startup is recognized or not. Recognized firms get a three kinds of effective subsidies: i) operating cost subsidy which reduces their constant operating costs[15], ii) R&D and search cost subsidies which depend on the type firm[16], and iii) a distrotionary exit wedge[17]. For simplicity, I assume that all of the effective subsidies/wedges are constant between the age $(0, \bar{A})$ after which they go to zero. Specifically, the subsidies for recognized R&D and non-R&D startups are given by

$$\left(s^i_{c,t}(a,h), s^i_{x,t}(a,h), s^i_{\tau,t}(a,h)\right) := \begin{cases} = \begin{cases} (\bar{s}^j_c, \bar{s}^j_x, \bar{s}^j_\tau) & \text{if } a \leq \bar{A} \\ (0,0,0) & \text{if } a > \bar{A} \end{cases} & \text{if recognized, } h_i = 1 \\ 0 & \text{otherwise, } h_i = 0 \end{cases}$$

where $j \in \{R, N\}$ denotes whether the firm is a R&D firm or a non-R&D firm.

Startup selection for Recognition: The selection of startups for recognition benefits is controlled by three parameters: $\phi^R_t \in [0,1]$ and $\phi^N_t \in [0,1]$ which control the overall mass of R&D and non-R&D firms that the government decides to subsidize. And, $\phi^{h,z}_t(z) := Pr[h = 1|z]$ which denotes the conditional probability that type z firm would get a recognition. This selection ability captures the realized outcome after filtering, self-selection, ordeals, bureaucratic selection done by the government to achieve final selection.

Tax Exemptions Conditional on getting a recognition, I assume that the startups can also get a profit tax exemption. If a startup gets selected, its profits get a proportional subsidy between

[14]In the description of the firm problem above, age didn't matter as the profit, costs and hazard rates of losing products only depends on total products and firm type.

[15]This captures policy instruments like lease rental subsidies, interest rate subsidy, relaxed procurement norms and reduced regulatory compliance.

[16]These capture patent filing subsidies, research incubators for R&D firms and seed funds, grants, venture funds for non-R&D startups

[17]This captures the cumulative impact of all the policy instruments on startups survival

ages $\underline{A}$ and $\bar{A}$ given as follows

$$\left(\bar{s}^{i}_{\pi,t}(a,d)\right) := \begin{cases} = \begin{cases} \bar{s}_{\pi} & \text{if } \underline{A} \leq a \leq \bar{A} \\ 0 & \text{if } a < \underline{A} \text{ or } \bar{A} < a \end{cases} & \text{Given tax exemption, } d_i = 1, h_i = 1 \\ 0 & \text{if not selected, } d_i = 0 \end{cases}$$

I assume that the total R&D startups getting tax exemptions is controlled by a scalar $\phi^{R,d}_{t}$ whereas for non-R&D startups the selection function for the tax exemption is given by $\phi^{d,z}_{t}(z) := Pr[d = 1|z]$.

3.5 Firm Value Functions

Given the description of the firm problem above, I now describe the firm value functions for the R&D and non-R&D firms respectively.

R&D Firm Value Function has five state variables: number of products n, age a, time t and whether the firm has recognition subsidies or not $h \in \{0,1\}$ and whether it has tax holidays or not $d \in \{0,1\}$. And its value function $V^{R}_{t}(n,a,h,d)$ is given by the following HJB

$$\begin{aligned} r_t V^R_t = \max_{x_R} \Big\{ & (1 + s_{\pi,t}(a,d)) n\pi_t - n \frac{x_R^{\zeta_R} w_t}{\psi^R_t (1 + s_{x,t}(a,h))} - c_R (1 - s_{c,t}(a,h)) + n x_R \left[V^R_t(n+1) - V^R_t(n) \right] \\ & + (1 - s_{\tau,t}(a,h)) \Big[\tau_t \left(n[V^R_t(n-1) - V^R_t(n)] \right) + \kappa \Psi_t \left(n[V^R_t(n-1) - V^R_t(n)] \right) \Big] \\ & + \partial_t V^R_t + \partial_a V^R_t \Big\} \end{aligned} \tag{3}$$

The first three terms in the first line represents the flow profits that a firm generates from operating n product lines accounting for the R&D costs and fixed operating costs along with the respective subsidies. The fourth term captures the fact that the firm gains a product if it is successful at innovation which happens at rate nx. The second line capture the two cases with which the firm can lose a product: i) through creative destruction at rate τ_t i.e. another pioneer firm innovates on that product line, 2) through replacement, $\kappa\Psi_t$, which leads to a follow-up firm replacing the pioneer firm in that product line. The wedge on this term captures the policies impact on distorting the creative destruction and churn in the economy. Last line captures the evolution of the firms value over age (due to age dependent subsidies) and calendar time (due to change in aggregate variables and policies over time).

Non-R&D Firm Value Function has seven state variables: number of products in the frontier line n, number of products in the laggard lines l, firm type z, firm age a, time t and whether the firm has recognition subsidy or not $h \in \{0,1\}$ and whether it has tax holidays $d \in \{0,1\}$. The firm value function is thus given as follows:

$$
\begin{aligned}
r_t V_t^N(n,l,z,a,h,d) = \max_{x_N} \Big\{ & (1+s_{\pi,t}(a,d))(n+l)\pi_t - (n+l)\frac{x_N^{\zeta_N} w_t}{\psi_t^N(z)(1+s_{x,t}(a,h))} - c_N(1-s_{c,t}(a,h)) \\
& + (n+l)x_N \Big[M_t^f \left(m_R^f \kappa + (1-m_R^f) \right) \left[V_t^N(n+1) - V_t^N(n) \right] \\
& + (1-M_t^f) \Big(\gamma [V_t^N(n+1) - V_t^N(n)] + (1-\gamma)[V_t^N(l+1) - V_t^N(l)] \Big) \Big] \\
& + (1-s_{\tau,t}(a,h)) \Bigg[\tau_t \Big(n[V_t^N(n-1) - V_t^N(n)] + l[V_t^N(l-1) - V_t^N(l)] \Big) \\
& + \Psi_t \Big(n[V_t^N(n-1) - V_t^N(n)] \Big) + A_t \Big(l[V_t^N(l-1) - V_t^N(l)] \Big) \Bigg] \\
& + \partial_t V_t^N + \partial_a V_t^N \Big\}
\end{aligned}
\tag{4}
$$

The first line corresponds to the flow profits of having (n, l) products accounting for both profits, search costs and fixed operating costs along with the subsidies. The second and third line captures the gains from the firm investing in idea search: it can either gain a frontier line product through replacement or a large adoption event or it can gain a laggard line product by regular adoption. The probability that the firm will randomly find a frontier line is given by M_t^f (mass of frontier lines at time t). Conditional on frontier line, it is successful with probability κ if the line is operated by a R&D firm (mass m_R^f) and with probability one otherwise. On the other hand, it can find a laggard line with probability $1 - M^f$ and end up gaining a product by successfully adopting a new technology; with probability γ the resulting adoption is large and the line converts to a frontier line. Fourth line capture the fact that the firm might lose either the frontier or the laggard product line due to creative destruction. Similarly, fifth line denotes the probability of losing a frontier and a laggard product line to other firms randomly replacing (due to a "bad luck" shock) and adoption respectively. Last line captures the evolution of the firms value over age (due to age dependent subsidies) and calendar time (due to change in aggregate variables and policies over time).

3.6 Entry

There is a unit mass of potential entrants of each type of firm, R&D and a continuum of non-R&D type z's, each instant of time. Similar to incumbent firms they also hire workers to do research or to find existing ideas in the economy. Upon a successful technological innovation or search they enter the economy with a single product by replacing incumbent producer in a random line.

Entry of R&D Startups A unit mass of potential entrants hire labor to do technological innovations. Similar to incumbents, their cost function (to get an arrival rate of $x_{e,R,t}$) is given by

$$C_R^e(x_{e,R,t}) = \frac{1}{\psi_{e,t}^R} x_{e,R,t}^{\zeta_e} w_t$$

If successful they enter as a one product startup firm in any random line. However, their value from entering also depends on whether they get selected for a subsidy or not. Thus, a potential entrant at time t solves the following problem

$$\max_{x_{e,R,t}} \left[x_{e,R,t} \Big(\phi_t^R \left[\tilde{V}_t^R(1,0,h=1) \right] + (1-\phi_t^R) \left[V_t^R(1,0,d=0,h=0) \right] \Big) - \frac{1}{\psi_{e,t}^R} x_{e,R,t}^{\zeta_e} w_t \right] \tag{5}$$

where the first two state variables of the value function denote a single frontier line product and age equal to zero upon entry. And $\tilde{V}_t^R(1,0,d=1,h=1) := \phi_t^{R,d} V_t^R(1,0,d=1,h=1) + (1-\phi_t^{R,d}) V_t^R(1,0,d=0,h=1)$ i.e. expected value with a tax holiday conditional on getting a recognition.

Entry of non-R&D Startups The cost function of a potential non-R&D entrant of type z, to get an arrival rate of $x_{e,N,t}$ is given by

$$C_R^e(x_{e,N,t}, z) = \frac{1}{\psi_{e,t}^N(z)} x_{e,N,t}^{\zeta_e} w_t$$

If successful, i.e. at rate $x_{e,N,t}$ they find a product line and replace the existing entrepreneur exactly like the incumbent entrepreneur i.e. they enter with a frontier product with probability $(\kappa m_R^f + (1-m_R^f))M_t^f + \gamma(1-M_t^f)$ and with a laggard product with probability $(1-\gamma)(1-M_t^f)$. Thus, given the overall startups selected for subsidies and targeting across types, the potential entrants of type z solve the following optimization problem

$$\begin{aligned} \max_{x_{e,N,t}} \Bigg[& \phi^N \left[\Big(\phi_t^{h,z}(z) \tilde{V}_t^N(z,0,h=1) + (1-\phi_t^{h,z}(z)) \tilde{V}_t^N(z,0,h=0) \Big) \right] \\ & + (1-\phi^N) \tilde{V}_t^N(z,0,h=0) - \frac{1}{\psi_{e,t}^N(z)} x_{e,N,t}(z)^{\zeta_e} w_t \Bigg] \quad \text{where} \end{aligned} \tag{6}$$

$\tilde{V}^N(z,0,h) = ((1-(1-\kappa)m_R^f)M_t^f)\hat{V}_t^N(1,0,z,0,h) + (1-M_t^f)[\gamma V_t^N(1,0,z,0,h) + (1-\gamma)\hat{V}_t^N(0,1,z,0,h)$ is the expected value of entering in either the laggard or the frontier line. And $\hat{V}_t^N(1,0,z,0,h=1) = \phi^{d,z}(z) V_t^N(1,0,z,0,h=1,d=1) + (1-\phi^{d,z}(z)) V_t^N(1,0,z,0,h=1,d=0)$. The entry of each type of startup firm, thus, responds heterogeneously to startup subsidies given the government's selection ability and the type of benefits provided by the subsidy — which change the firm value function conditional on getting a subsidy.

3.7 Aggregate Rates

Aggregate growth and firm dynamics in the model depend on levels of aggregate creative destruction, adoption and replacement in the economy. In equilibrium, as the economy is closed, all of these rates are, in turn, the sum of individual incumbent and entrant firm behaviors. Let $g_{R,t}(n,a,h)$ and $g_{N,t}(n,l,z,a,d,h)$ be the total mass of R&D and non-R&D firms in each state variable at a given time t. And let $g^S_{R,t}$ and $g^S_{N,t}$ be the corresponding mass of only subsidized startups.

Aggregate Creative Destruction in the economy is the sum of the R&D done by the entrants $x_{e,R,t}$ and R&D done by the incumbents $x_{R,t}(n,a,h)$.

$$\tau_t \left(1 - s_\tau \left(\int n g^S_{R,t} + \int (n+l) g^S_{N,t}\right)\right) = x_{e,R,t} + \int n x_{R,t}(n,a,d,h) g_{R,t}(n,a,d,h) \tag{7}$$

where the aggregate creative destruction is adjusted for exit wedge that the policy introduces in the subsidized firms.

Aggregate Adoption in the economy is the rate at which non-R&D incumbent and entrant firms do productivity improvements on laggard lines by adopting. It is given by

$$A_t \left(1 - s_\tau \int l g^S_{N,t}\right) = (1 - M^f_t) \int x_{e,N,t}(z) dz + (1 - M^f_t) \int (n+l) x_{N,t}(n,l,z,a,d,h) g_{N,t}(n,l,z,a,d,h) \tag{8}$$

Aggregate Replacement happens by both entrants and incumbent adopter firms, the aggregate value is given by

$$\begin{aligned}\Psi_t \left(1 - s_\tau \left(\int n g^S_{R,t} + \int (n+l) g^S_{N,t}\right)\right) &= ((1 - (1-\kappa) m^f_R) M^f_t) \int x_{e,N,t}(z) dz \\ &+ ((1 - (1-\kappa) m^f_R) M^f_t) \int (n+l) x_{N,t}(n,l,z,a,d,h) g_{N,t}(n,l,z,a,d,h)\end{aligned} \tag{9}$$

3.7.1 Catch-up Rate

The distance of the economy from the frontier is summarized by M^f_t i.e. the share of product line that are at the frontier. Till the time $M^f_t < 1$, adoption of new technologies by the non-R&D firms plays a key role in generating growth by imitating frontier technologies. However, once $M^f_t = 1$, the only growth in the economy is generated by the innovative firms. As noted above, a product line transitions from laggard to frontier when an innovative firm innovates on that line for the first time or a large adoption happens by a non-R&D firm. Thus, if there are $1 - M^f_t$ laggard lines at time t, the rate at which they transition to frontier lines is simply given by

$$\dot{M}_t^f = (\tau_t + \gamma A_t)(1 - M_t^f) \tag{10}$$

Thus, if the economy begins at time $t = 0$ if M_0^f frontier lines, their evolution over time is given by $M^f = 1 - \left(1 - M_0^f\right) \exp\left(- \int_0^t [\tau(s) + \gamma A(s)]\, ds\right)$

3.7.2 Aggregate Growth

Each successful innovation in a frontier line increases productivity proportionally by step size λ. In a laggard line, however, the productivity can jump to any value between the current productivity of the line the world frontier productivity depending on whether it is hit by an innovating firm or a non-R&D firm. As noted earlier, the productivity in a line jumps directly to world frontier $\bar{q}_{jt}$ if it gets hit by creative destruction τ_t. If it gets hit by non-R&D firm, who imitate already existing technologies, it's productivity jumps directly to the frontier with probability γ or to a random draw between its current value and the frontier with probability $1 - \gamma$. Specifically, for the latter case, if the log productivity of a line is $\ln q_{jt}$, then it jumps to a new value given by $\ln q'_{jt} := \ln q_{jt} + \theta[\ln \bar{q}_{jt} - \ln q_{jt}]$, where $\theta \sim \text{Uniform}[0, 1]$. Thus, in a frontier line with productivity q_{jt}, after an interval Δt, the new productivity equals

$$\ln q_{j,t+\Delta t} = \begin{cases} \ln \lambda + \ln q_{jt} & \text{with probability } \tau_t \Delta_t \\ \ln q_{jt} & \text{with probability } (1 - \tau_t \Delta_t) \end{cases}$$

and in a laggard line with productivity q_{jt}, the productivity evolves as

$$\ln q_{j,t+\Delta t} = \begin{cases} \ln \bar{q}_{jt} & \text{with probability } (\tau_t + \gamma A_t)\Delta_t \\ \ln q_{jt} + \theta[\ln \bar{q}_{jt} - \ln q_{jt}] & \text{with probability } (1 - \gamma) A_t \Delta_t, \quad \theta \sim \text{Uniform}[0, 1] \\ \ln q_{jt} & \text{otherwise} \end{cases}$$

The parameter γ, thus controls the average size of the productivity improvements made via imitation. If γ equal zero, the average productivity gain in a line j is $\frac{\ln \bar{q}_{jt} + \ln q_{jt}}{2}$. And as γ rises, the expected step size is strictly larger than this value. Hence, by aggregating over all the lines, when the mass of frontier lines is M_t^f and the mass of laggard lines is $1 - M_t^f$, the aggregate productivity evolves as

$$\ln Q_{t+\Delta t} - \ln Q_t = M_t^f \tau_t \ln \lambda + (1 - M_t^f)\left[(\tau_t + \gamma A_t) m_t + \tfrac{1}{2}(1 - \gamma) A_t\, m_t\right] \tag{11}$$

where $m_t := \mathbb{E}[\ln \bar{q}_{jt} - \ln q_{jt}]$ is the average productivity gap across the lines from the frontier and evolves as[18]

$$\dot{m}_t = -\frac{1-\gamma}{2} A_t \, m_t. \tag{12}$$

Thus Equations 10, 11 and 12 jointly determine the productivity evolution of the economy on the transition path — along with the aggregate creative destruction and adoption rate. Equation 11 also clearly shows that the as $M_t^f \rightarrow 1$, the economy converges to having productivity gains completely determined by creative destruction. However, $M_t^f < 1$, productivity growth is in the high (catch-up) phase i.e. $> \tau_t \ln \lambda$ and decreases slowly over time as the economy catches up with the frontier.

World Frontier After the domestic productivity in the economy catches-up with the world frontier, the frontier evolves solely driven by the productivity improvements made by the domestic firms. Hence, the frontier is given by $\bar{q}_{jt} = \max(\bar{q}_j, q_{jt})$

3.8 Government Budget and Market Clearing

Government finances the total subsidies provided to the startups by raising lump taxes from the households. The total government subsidy expenditure and the total corresponding taxes raised are thus given by

$$T_t = S_{\pi,t} + S_{x,t}^R + S_{x,t}^N + S_{\tau,t} + S_{c,t} \tag{13}$$

where $S_{\pi,t}$ is the total expenditure on tax holidays for all the overlapping cohorts of startups with subsidies at time t, $S_{x,t}^R$ is the total cost of R&D subsidies, $S_{x,t}^N$ is the total cost of search subsidies, $S_{\tau,t}$ is the total implied value of exit wedges and $S_{c,t}$ is the total cost of operating cost subsidies. In equilibrium, the total expenditures are computed using the optimal R&D and search decisions of startups, $x_{R,t}(n,a,d,h)$ and $x_{N,t}(n,l,z,a,d,h)$, the implied profits, $\pi_t^R(n,a,d)$ and $\pi_t^N(n,l,z,a,d,h)$, and the additional flow values that the firms receives from creative destruction, replacement and adoption wedges generated by the policy defined as $\nu_t^R(n,a,h) := V_t^R(n,a,h,\cdot) - V_t^R(n-1,a,h,\cdot), \nu_t^{N,F} := V_t^N(n,l,\cdot) - V_t^N(n-1,l,\cdot), \nu_t^{N,L} := V_t^N(n,l,\cdot) - V_t^N(n,l-1,\cdot)$.

[18]To see this, note that drastic adoption and innovation hit the lines randomly so they leave the average gap unchanged (as lines transition to becoming frontier lines randomly). However, when a non-drastic adoption hits the line at rate $(1-\gamma)A_t$, it shrinks the average gap to half as the innovations are uniformly distributed between the current value and the frontier.

Hence, the total expenditures on these subsidies are given as

$$
\begin{aligned}
S_{\pi,t} &= \iint \phi^N \phi_t^{d,z}(z) s_{\pi,t}(a,d) \pi_t^N(n,l,z,a,d,h) g_{N,t}^S dadz + \int \phi^R s_{\pi,t}(a,d) \pi_t^R(n,a,d) g_{R,t}^S da \\
S_{c,t} &= \iint \phi^N \phi_t^{h,z}(z) s_{c,t}(a,h)(c_N g_{N,t}^S dadz + \int \phi^R s_{c,t}(a,h) c_R g_{R,t}^S \, da \\
S_{x,t}^R &= \int \phi^R \frac{s_{x,t}(a,h)}{1+s_{x,t}(a,h)} \Big[n \frac{x_{R,t}(n,a,h)^{\zeta_R} w_t}{\psi_t^R} \Big] g_{R,t}^S da \\
S_{x,t}^N &= \iint \phi^N \phi_t^{h,z}(z) \frac{s_{x,t}(a,h)}{1+s_{x,t}(a,h)} \Big[(n+l) \frac{x_{N,t}(n,l,z,a,d,h)^{\zeta_N} w_t}{\psi_t^N(z)} \Big] g_{N,t}^S dadz \\
S_{\tau,t} &= \iint \phi^N \phi_t^{h,z}(z) s_{\tau,t}(a,h) \Big((\tau_t + \Psi_t) n \nu_t^{N,F} + (\tau_t + A_t) l \nu_t^{N,L} \Big) g_{N,t}^S dadz \\
&\quad + \int \phi^R \, s_{\tau,t}(a,h) \, (\tau_t + \kappa \Psi_t) \, n \nu_t^R \, g_{R,t}^S \, da
\end{aligned}
\tag{14}
$$

where the total amount of the subsidies is controlled by overall size ϕ^R, ϕ^N as well as targeting ability across types which control the conditional distribution of type z having a subsidy.

Market Clearing The goods, labor and assets market clear at all times. As the only input used for production, R&D and search is labor, all the aggregate output in the economy is used for consumption and to fund the government subsidies. The aggregate amount of labor, on the other hand, is fixed to unity, so the labor employed in production ($L_{P,t}^R$, $L_{P,t}^N$), in incumbent R&D ($L_{x,t}^R$), entrant R&D ($L_{e,t}^R$), in incumbent and entrant search ($L_{x,t}^N$, $L_{e,t}^N$) sum's to 1 at all the time Additionally, all the firms are held by the households and, hence, their total asset holdings equal the total aggregate value of all the firms in the economy.

$$
Y_t = C_t + T_t \tag{15}
$$

$$
1 = L_{x,t}^R + L_{x,t}^N + L_{e,t}^R + L_{e,t}^N + L_{P,t}^R + L_{P,t}^N \tag{16}
$$

$$
B_t = \sum_{p \in \{R,N\}} \int V_t^p(\cdot) d g_{p,t}(\cdot) \tag{17}
$$

Definition 1 (Equilibrium)**.** *Given the paths of age-based subsidies* $\{s_{c,t}(a,h), s_{x,t}(a,h), s_{\tau,t}(a,h), s_{\pi,t}(a,h)\}_{t\geq 0}$*, program sizes* $\{\phi_t^R, \phi_t^N\}_{t\geq 0}$ *and the government selection abilities* $\{\phi_t^{d,z}(z), \phi_t^{h,z}(z)\}_{t\geq 0}$ *the dynamic general equilibrium in the economy is given by the path of productivity, output and labor employed in each line* $\{y_{jt}, l_{jt}, q_{jt}\}_{t\geq 0}^{j\in[0,1]}$ *firm value functions* $\{V_t^R(n,a,h,d), V_t^N(n,l,z,a,h,d)\}_{t\geq 0}$*, policy functions* $\{x_{R,t}(n,a,h,d), x_{N,t}(n,l,z,a,h,d)\}$*, entry rates* $\{x_{e,R,t}, \{x_{e,N,t}(z)\}_{z\in\bar{Z}}\}_{t\geq 0}$*, profits* $\{\pi_{R,t}(n,a,h,d), \pi_{N,t}(n,l,z,a,h,d)\}_{t\geq 0}$*, wages* $\{w_t\}_{t\geq 0}$*, interest rate* $\{r_t\}_{t\geq 0}$*, aggregate rates of creative destruction* $\{\tau_t\}_{t\geq 0}$*, replacement* $\{\Psi_t\}_{t\geq 0}$*, adoption* $\{A_t\}_{t\geq 0}$*, mass of frontier lines* $\{M_t^f\}_{t\geq 0}$*, average log-productivity gap in the laggard lines* $\{m_t\}_{t\geq 0}$*, aggregate productivity* $\{Q_t\}_{t\geq 0}$*, total labor demands* $\{L_{x,t}^R, L_{x,t}^N, L_{e,t}^R, L_{e,t}^N, L_{P,t}^R, L_{P,t}^N\}_{t\geq 0}$*, total expenditure on subsidies* $\{S_{\pi,t}, S_{x,t}^R, S_{x,t}^N, S_{\tau,t}, S_{c,t}\}_{t\geq 0}$*, total taxes* $\{T_t\}_{t\geq 0}$*, aggregate consumption* $\{C\}_{t\geq 0}$*, output* $\{Y_t\}_{t\geq 0}$*, assets* $\{B_t\}_{t\geq 0}$ *and the distributions* $\{g_{R,t}(n,a,h,d), g_{N,t}(n,l,z,a,h,d)\}_{t\geq 0}$

such that i) $\{y_{jt}, l_{jt}, q_{jt}\}_{t\geq 0}^{j\in[0,1]}$ solve the intermediate production problem given the aggregate production in Eq 1, ii) $\{V_t^R(n,a,h,d), V_t^N(n,l,z,a,h,d)\}_{t\geq 0}$ and $\{x_{R,t}(n,a,h,d), x_{N,t}(n,l,z,a,h,d)\}$ solve the firm HJBs in Eq. 3 and 4, iii) entry rates $\{x_{e,R,t}, \{x_{e,N,t}(z)\}_{z\in\bar{Z}}\}_{t\geq 0}$ solve Eq. 5 and 6, iv) productivity Q_t, mass of frontier lines M_t^f and average gap M_t evolve according to Eqs. 10, 11 and 12, v) $\{\tau_t, \Psi_t, A_t\}_{t\geq 0}$ satisfy Eqs. 7, 8 and 9, vi) total taxes and subsidies satisfy Eqs 13 and 14, vii) wages are such that the labor market clears at all times according Eq. 16, viii) firm distributions of subsidized and non-subsidized firms $\{g_{R,t}(n,a,h,d), g_{N,t}(n,l,z,a,h,d)\}_{t\geq 0}$ evolve according to the Kolmogrov-Forward Equations , ix) $\{C_t, Y_t, B_t\}_{t\geq 0}$ satisfy the goods and labor market clearing conditions in Eqs 15 and 17.

3.9 Model Summary and Discussion

The goal of the model is to capture the key mechanisms through which heterogeneous startup firms contribute to aggregate growth and knowledge spillovers in a developing economy—and to assess how startup policies affect these mechanisms. To this end, allowing for meaningful heterogeneity across firm types, including R&D startups and ex-ante distinct non-R&D firms, captures for large ex-ante variation in firm types. In addition, incorporating the possibility of adopting technologies from the world frontier enables the model to realistically reflect the growth dynamics of a developing economy.

In terms of the policy, the model flexibly captures the bureaucratic selection through the selection ability functions and parameters $\phi^{z,d}(z), \phi^{z,h}(z), \phi^N, \phi^R, \phi^{R,d}$, multiple different margins of the policy as effective subsidies and it captures the sunset clauses by making the subsidies age-dependent. At the same time, the model is flexible enough so that the long run and the short run dynamics can be calibrated using empirical data. For instance, the impact of the policy could go in the either direction depending upon the selection ability of the government and the type of subsidies and distortions that the policy introduces.

Sources of Sub-Optimality To conclude the description of the model, I discuss the main externalities and market failures in the model that make the competitive equilibrium inefficient and allow for a meaningful role for a startup policy

1. Inter-temporal Knowledge Spillovers: When the firms in the model do innovations or technological adoption they increase the overall productivity in the economy (Eq 11). This increases the absolute productivity gains from future innovation but the firms do not internalize this channel. Thus, there could be potential under-investment in R&D and adoption which the government can promote by re-allocating labor from production to R&D and search.

2. Reallocation: All firms, however, do not contribute uniformly to the knowledge spillover. Consider a low type firm with $z \to 0$ and a high type firm with $z \to \infty$. A production

line operated by the latter is much more socially valuable as it does the same amount of production but also makes large spillovers. Thus, a policy that could re-allocate resources towards high-type firms or increase their entry can generate larger spillovers. However, the government's ability to do so would depend on how well the government can re-allocate resources and by how much it can increase the spillovers.

3. Monopolistic Competition and Business Stealing: Lastly, firms in the model have monopolistic power because of which they potentially under-produce and they also face a business stealing externality. This externality may lead them to over-invest in R&D and search as transfers the production line to their firm but from the aggregate planner's perspective firm shares do not directly matter.

The extent of each channel depends on the equilibrium labor allocation to each firm type and activity—which are pinned down by the model and policy parameters. Next section provides the details of data and empirical exercises that I use to calibrate the model. Section 5 estimates the main parameters that govern the overall transition of the economy as well as the aggregate impact of the policy.

4 Data Construction, Measurement and Empirical Findings

This section presents empirical results on the government's ability to target different types of startups, as well as the impact of the policy on key margins of firm dynamics such as entry, exit, and growth. I begin by describing the data collection procedures used to assemble the new datasets required for this analysis. I then outline the empirical specifications and discuss the corresponding results.

4.1 Data

The datasets I construct fall into two broad categories: (i) firm applications to the program and the corresponding selection or rejection decisions, and (ii) outcome data used to measure firm dynamics such as entry, exit, and growth. The first dataset allows me to document the program's ability to target different types of firms, while the second enables me to evaluate the program's impact on firm outcomes. I describe each of these datasets in detail below.

4.1.1 Universe of Firms and Selections by the Policy

I focus on limited liability firms incorporated in India, as this is one of the program's eligibility requirements. Within this group, I first introduce the dataset that covers the universe of entry and exit among private limited liability firms. I then describe the data collection process used to assemble information on firms' applications to the program and the corresponding decisions across various policy stages and instruments.

This cumulative dataset includes all private limited firms[19] registered with regional Registrar of Companies (ROC) offices under the Indian Companies Act. The dataset contains approximately 2.2 million registered private limited firms, with the earliest registration dating back to 1857. Of these, roughly 1.4 million were active as of 2024. Each firm is identified by a unique Corporate Identification Number (CIN), which I use for disambiguation and to merge with other datasets. The dataset also includes key firm characteristics such as legal name, date of incorporation, address, authorized and paid-up capital[20], sector, and operating status. Firms that cease operations in a given tax year must file for closure, allowing the dataset to classify firms as either "Active" or "Strike-Off." I use the "Active" status as an indicator of firm survival. Appendix B.1 provides summary statistics and describes the temporal and geographic distribution of firm formations in India. Importantly, because this dataset is not a panel, I supplement it with additional collections to construct firm-level outcomes. The primary outcomes I draw from this dataset are firm entry, survival, and paid-up capital.

Applications, Selections and Rejections by the Startup India Policy: From the universe of firms, any firm less than ten years old is eligible to apply for benefits under the Startup India policy. Since its launch in 2016, the Department for Promotion of Industry and Internal Trade (DPIIT) has certified approximately 160,000 firms as recognized startups, with yearly recognitions steadily increasing. To obtain recognition, firms must apply through the online Startup India portal[21]. The application requires the firm to be a registered private limited company or a limited liability partnership and to submit details including a brief description of the business model, granular geographic and sectoral information, founder education and background, funding raised, patents or awards, and responses to self-reported questions on innovation and scalability. The government typically reviews applications and issues decisions within ten days. Once a firm registers on the portal, part of its information becomes publicly available on its profile page. I scraped the profile pages of approximately 300,000 such firms, extracting legal names, business model descriptions, district/city, and sector. I then matched these firms to the MCA dataset using name and location. Each profile page also includes a badge indicating whether the firm is recognized by DPIIT, along with the date of recognition. I classify firms with a profile page and a recognition badge as "applied and approved," and those with a profile page but no badge as "applied and rejected." Appendix B.2 provides summary statistics on recognized startups, including their stage at recognition and sectoral distribution. At the time of data collection, I identified 112,950 recognized startups, of which 94,648 were private limited firms. I successfully matched 71,890 of these to the MCA dataset.

To analyze the allocation of tax holidays, I begin by extracting text from the minutes of seventy Inter-Ministerial Board meetings responsible for evaluating startups' applications for profit tax

[19] Limited Liability Partnerships (LLPs) are another major form of limited liability organization in India, introduced under the Limited Liability Partnership Act, 2008. However, LLP data is not centralized in the same manner as private limited firms, so I abstract from them in the main analysis.

[20] Paid-up capital refers to the funds received from shareholders in exchange for shares issued by the company.

[21] `https://www.startupindia.gov.in/content/sih/en/startupgov/startup_recognition_page.html`

exemptions. As described in Section 2, a startup recognition provides firms with certain de facto benefits, but additional benefits require further applications. I use patent filings to identify recognized startups eligible for R&D benefits, and the meeting minutes to identify startups applying for tax holidays. After recognition, a startup can apply through the same portal for an income tax exemption under Section 80IAC of the Indian Income Tax Code. While the portal displays firms approved for exemptions, it does not list rejected applicants, preventing construction of a full comparison group. The meeting minutes thus provide a crucial supplement by offering complete application outcomes. These meetings, held monthly since April 2016, record the government's decision—accepted, rejected, or ineligible—along with the reason for each outcome. Appendix B.3 provides an example of the original format. Accepted startups are typically approved for being innovative or scalable, while rejections often cite lack of innovation or violations of pre-established rules, such as eligibility cut-off dates, subsidiary status, or joint ventures. From these minutes, I extract each firm's name, decision, reason, and the meeting date. I match the firms identified in the meeting minutes to the recognition data using the DPIIT-provided identification number, and then to the MCA data using each firm's Corporate Identification Number (CIN). To ensure a sufficiently long panel to track firm outcomes, I focus on meetings held between 2016 and the end of 2019. This yields a sample of 2,981 startups that applied for tax holidays, of which 233 were selected.

4.1.2 Novel Data on Firm Outcomes: Patent Applications and Annual Income Filings in India

To measure the innovation and growth outcomes of startups in India, I collect data from two main sources: firm-level income and balance sheet filings, and patent application records. A major challenge in studying the performance of young firms in less developed economies is the scarcity of reliable, high-quality data. One of the central contributions of this paper is the construction of a new dataset that enables analysis of firm performance and the impact of targeted startup policies.

Each private limited firm in India is required to file their annual income statements and balance sheets with the Ministry of Corporate Affairs. While the whole dataset is not harmonized and not provided to researchers for use, each firm's annual statements are publicly available on the MCA website for a payment. I work with Tracxn, a data collection startup in India which provides VCs with information on startups, to collect these income filings and annual statements for all the recognized startups that apply for tax holidays between 2016 and 2019. For these, 2,981 startups we construct a harmonized panel dataset of firm level income and balance sheet items such as revenues, profits, assets, liabilities and equity since their incorporation to up till 2024. Each firm's annual filings contain multiple forms but we focus on the AOC-4. This form is mandatory for a for a private limited firm to file with their Registrar of Companies within a few days of their Annual General Meeting. It is divided into three parts which separately provide information on the firm's ownership structure, flow income statements and balance sheet information. The income statements contain information on revenue, cost of materials, employment benefits, profits/losses and tax expenses. The balance sheets contain information on share capital, reserves and surpluses, liabilities such as short-term borrowings, trade payables, current and non-current assets. The final dataset and its construction is described in Appendix B.4. It contains 2,215 unique firms with a total of 12,143 firm-year observations. There is a significant heterogeneity in the applicant startups with median revenue of roughly 16,000 USD and the 99th percentile of roughly 4.3 million USD.

I also construct a new dataset on firm-level patenting activity in India. To build this dataset, I scrape data on all patent applications filed with the Indian Patent Office between 2012 and 2022 from the Indian Patent Advanced Search System (InPASS) portal of the Office of the Controller General of Patents, Designs, and Trademarks. The portal provides full text and details of all patent applications filed and granted in India. In addition, the Indian government publishes a weekly update containing application numbers, categories, and office details for all patents filed or granted during that week. I use these application numbers to retrieve and scrape the corresponding full records from the InPASS portal. This yields a dataset of 434,262 patent applications over the decade. However, the vast majority of these applications are filed by individual or academic inventors. To identify firm-level patenting, I perform a keyword search to isolate applications filed by corporate entities and then fuzzy-match the applicant names to the universe of private limited firms in the MCA business registry. This procedure results in 47,668 matched patent applications filed by 4,864 unique firms in the MCA dataset. Appendix B.5 provides further details on the data construction and summary statistics of the final dataset.

Other Firm and Entrepreneur Characteristics:—To study which observable characteristics predict selection by the bureaucratic board for tax holidays, and to match selected firms with similar rejected firms, I augment my dataset with information on startup and entrepreneur backgrounds from two additional sources. First, using Tracxn, I compile company profiles for all firms in the tax holiday application dataset that are also tracked by Tracxn. These profiles contain details on startup and founder characteristics, including funding rounds (venture capital, angel, and seed), incubation history, team composition, and founder education and experience. Second, I supplement this information with firm-level venture capital data from PitchBook, a global database covering over 600,000 firms and 1.2 million deals. I restrict this dataset to firms incorporated in India that have received at least one round of venture capital investment, yielding approximately 8,400 VC-financed firms. I then fuzzy-match these firms to the applicant and recognized firm datasets described above.

4.2 Empirical Findings

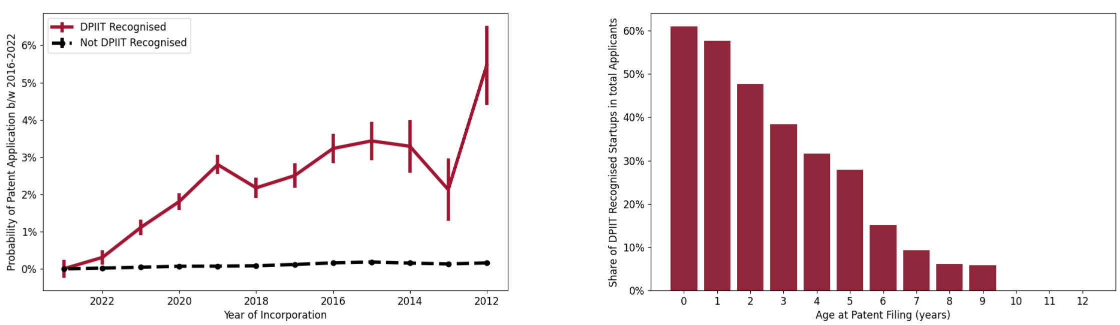
This section documents empirical findings on the governments ability to identify high potential startups and to selectively promote their entry and growth. I document five main empirical results about the three broad policy instruments R&D benefits, tax holidays and firm recognitions.

I show that i) entry and patenting activity of innovative startups increased substantially after the policy launch, ii) the bureaucratic selection process for tax holidays is able to selectively target high growth firms, iii) tax holidays have a marginal impact on startup growth, iv) recognized startups have average growth but are substantially less likely to exit, and v) a one percentage point increase in recognitions predicts a 2.5% rise in overall entry in the subsequent year and 9.7% rise in entrants which get recognized within a year of incorporation.

1. Policy Selects Innovative Firms and Patent Applications by Startups Rise by 50% Post-Policy

I first document the representation of innovative firms among the startups recognized by the policy. Fig. 1a plots the share of firms incorporated in each year that eventually filed a patent application between 2012 and 2022. The dashed black line shows the probability that an average limited liability firm in the economy files a patent, which is close to zero. The red line shows the same probability, conditional on being recognized under the Startup India policy. About 3-4% of the firms recognized under the Startup India program have filed for a patent application depending on the year of incorporation. This implies that either through self-selection or through incubators, filtering etc the policy is substantially more like to target innovative firms in equilibrium as compared to uniform targeting. Moreover, about 50% of all the patent applications by firms in the age category between 0-3 in the years 2012-22 were filed by recognized startups, as shown in Fig. 1b which plots the fraction of applications in each age bin that were filed by recognized startups

Figure 1: Innovation of Recognised vs Average Firm



(a) Pr. of filing a patent of Recognized vs Non-Recognized Startups by Year of Formation

(b) Fraction of total patent applicants which are Recognized Startups by Age

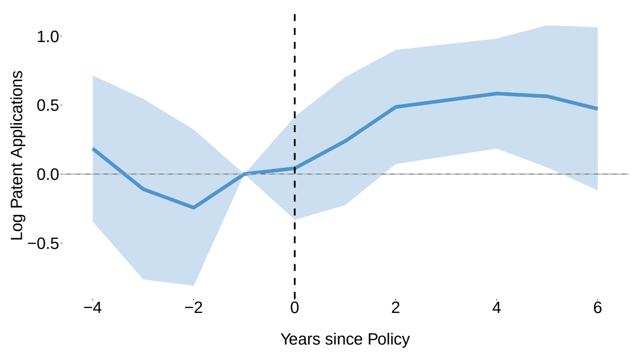
Next, to understand the impact of the policy on the entry and R&D of innovative startups, I document the evolution of the patenting activity of startups in India pre-and-post the policy launch. The provisions of R&D benefits such as incubators at universities/research institutions, special venture funds, patent filing subsidies, patent facilitators were targeted towards startups instead of the older firms. Thus, I use older firms as a comparison group and study the overall patenting activity of startup firms vs the older firms. In each year, pre-and-post the policy, I classify each firm as a startup as opposed to an old firm if it's age at the time of the patent application is less than ten year. I then aggregate the patent applications for each year at a age group $\times$ sector $\times$ state level. Specifically, I run the specification

$$\text{Log Patent Applicants}_{\text{sijt}} = \alpha_{\text{i}} + \gamma_{\text{j}} + \omega_s + \lambda_t + \sum_{\tau \neq -1} \beta_\tau (D_t^\tau \times \mathbf{1}[\text{Startup}_{sijt}]) + \epsilon_{sijt} \tag{18}$$

where the dependent variable, Log Patent Applicants$_{\text{s i j t}}$, denotes the log of total patent applications filed in year t, by firms in sector i, state j and firm category, s, equal to 1 or 0 (startup is denoted by 1). The specification includes state/sector fixed effects along with year and age-category fixed effects. The coefficient β_k estimates the growth in R&D activity done by startups relative to the older firms controlling for unobserved regional, sectoral and time effects.

Comparison with the patenting activity of old firms in the economy allows me to control for time-varying evolution or the growing use of intellectual property protection in the Indian economy, assuming that it effects both the categories equally. However, age-based categorization is dynamic by definition and the same startup can transition between categories. Especially, if the policy has persistent effect on innovation for a startup then the effects might persist even as the policy benefits phase-out after the age ten. This would likely downward-bias the coefficient. To address these concerns, I document the intensive effect of recognition on startup patenting at the end of this section and use the same regression as a target in the model.

FIGURE 2: PATENT APPLICATIONS OF STARTUPS RELATIVE TO OLD FIRMS



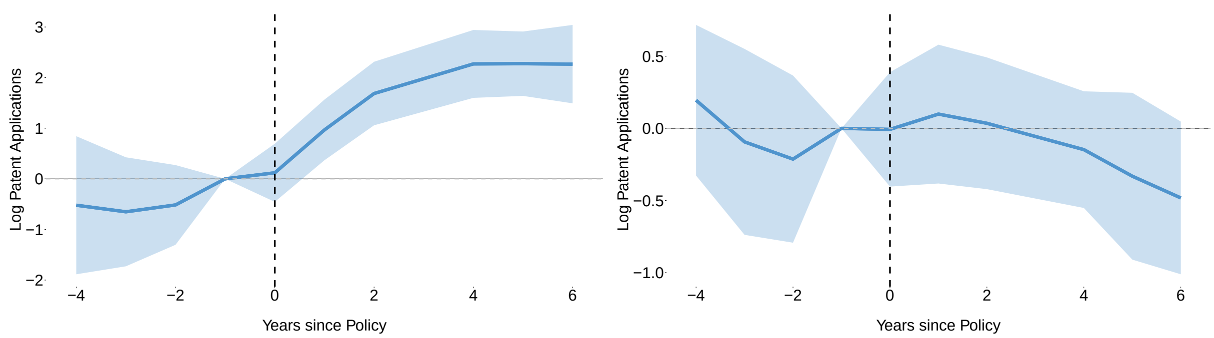
Notes: The figure plots the coefficient $beta_k$ from the following specification $\text{Log Patent Applicants}_{\text{sijt}} = \alpha_{\text{i}} + \gamma_{\text{j}} + \omega_s + \lambda_t + \sum_{\tau \neq -1} \beta_\tau (D_t^\tau \times \mathbf{1}[\text{Startup}_{sijt}]) + \epsilon_{sijt}$ where i denotes sector, j state, t time and s denotes whether the firm is a startup or not i.e. less than ten years of age. $k = 0$ denotes 2016 when then policy was launched. Thus the coefficient estimates the percent change in patent applications by startups pre and post the policy relative to to total patent applications by all the other firms in the economy controlling for sector, state, time and firm category fixed effects.

The coefficients β_k from the specification Eq. 18 are plotted in Figure 2 where $k = 0$ denotes the policy launch year. The results show a substantial rise in the patenting activity of startups after the policy launch. Prior to the policy launch, the estimates of β_k are statistically insignificant and close to zero implying that startups held a relatively stable share of patenting activity prior to the policy launch[22]. However, after the policy launch, the coefficient of $\beta_k = 0.5$ corresponds to a 50% larger rise in the startup patenting relative to old firm patenting rise in the same period. The rise also happens gradually after the policy launch in 2016, peaking around 2020.

Is this rise in patenting activity of startup firms associated with startups supported by the policy through recognitions or does it reflects an overall rise in startup activity since 2016? To isolate the rise associated with the policy, I run the same specification as in Eq. 18 but by splitting the startups into recognized and non-recognized startups, denoted by $R \in \{0, 1\}$. All the patent applications by a startup are classified as $R = 1$ if a startup gets recognized at any point of time. Intuitively, the coefficient β_τ^R, for $k \in \{-4, \cdots, 4\}$, captures both the intensive and extensive growth in patent applications by startups who eventually get recognized or never get recognized as compared to old firms.

$$\text{Log Patent Applicants}_{\text{sRijt}} = \alpha_{\text{i}} + \gamma_{\text{j}} + \lambda_t + \omega_{\text{sR}} + \sum_{\tau \neq -1} \beta_\tau^R (D_t^\tau \times \mathbf{1}[\text{Startup}_{sRijt}]) + \epsilon_{sRijt} \tag{19}$$

Figure 3: Recognized Startups and Never-Recognized Startups Patenting Activity



(a) Young Recognized Firms

(b) Young Non-Recognized Firms

Notes: This figure splits the results for startups in Fig. 2 by recognized and non-recognized startups. Panel (A) plots the coefficient for the recognized startups pre-and-post the policy. Firms that got recognized after the policy launch but had patent applications before the policy show no pre-trends but there is large growth in this category by the entry of new firms post the policy. Panel (B) plots the results for the firms which were eligible but never got recognized. Their patenting activity remains stable as compared to other firm albeit some crowding out towards the end of the sample.

Figs. 3a and 3b reveal a striking pattern. The rise in startup patenting is driven entirely by recognized firms, while patent applications by non-recognized startups remain statistically indistinguishable from those of older firms before beginning to decline three years after the policy launch. As Fig. 2 shows, however, this later decline does not completely offsets the gains. Thus, rather than merely re-labeling existing R&D activity, the policy is associated with an overall increase in patent filings among small firms, driven by the surge in recognized startups, while activity among non-recognized firms remains largely unchanged.

[22]The young firm share of patenting activity is remarkably stable in other economies as well. Goldschlag and Perlman (2017) document a very stable share of startups among patenting firms in the US even though the overall shares have substantially declined since the 80s

Lastly, the trend in Fig. 3a captures both the intensive and extensive effect on patent filings associated with the policy. The rise could either be driven by patent filing subsidies which incentivize existing firms to file more patents or by newly formed startups which are incubated at government institutions and require recognitions for incubation application. In Appendix C.2.3, I show that for startups that had a patent application before the policy launch, getting recognized doesn't leads to additional patent filing. Overall, this implies a small effect of recognition on the intensive innovation effort of the firms and the gains are primarily driven by the entry of new startups in the economy post policy.

2. Bureaucratic Board is Able to Select High Growth Startups for Tax Holidays

A large share of firms in the economy do not directly engage in innovation. For these non-R&D startups, can the government effectively distinguish between different types, especially when the applicants are still young? To assess the government's selection ability, I use data from bureaucratic board meetings that award tax holidays to startups. The median startup age at the time of the board's decision is about two years. In this section, I examine whether the board is able to identify "scalable" or high-growth startups when granting tax holidays.

Specifically, I run the specification in Eq. 20, where i denotes a startup, a age and Tax Holiday$_i$ is dummy which is equal to one if the startup was approved for a tax holiday and zero other wise. The coefficient β_a capture the percentage difference in the growth of selected and rejected startups as compared to the revenue at entry.

$$\text{Log Revenue}_{i,a} = \alpha + \sum_{a \neq 1} \beta_a \, \mathbf{1}[\text{Age}_i = a] \times \text{Tax Holiday}_i + \varepsilon_i \tag{20}$$

Table 2 reports the estimated coefficients at ages four and eight. The results reveal large and statistically significant differences in growth rates between selected and rejected startups. The coefficient of 2.45 at age eight implies that selected startups have revenues roughly eleven times larger than at age one, relative to rejected startups. By contrast, the baseline average growth of rejected startups over the first eight years is just 1.7. These figures highlight substantial differences between the two groups. As a useful benchmark, Hsieh and Klenow (2014) document that the average life-cycle growth for an average US manufacturing plant is 2 by the age of ten. Thus, the average rejected firm, even though a recognized startup by the government policy, has a life cycle growth of slightly below the average US manufacturing plant. For selected firms, the life-cycle growth is of the similar level as for the VC financed firms in the US (Puri and Zarutskie, 2012).

Table 2: Differences in Average Growth

	Age 4	Age 8
Log Revenue	1.480***	2.452***
	(0.305)	(0.622)

The table presents the estimates of β_a for $a = 4$ and $a = 8$ from the regression Log Revenue$_i = \alpha + \sum_{a \neq 1} \beta_a \mathbf{1}[\text{Age}_i = a] \times \text{Tax Holiday}_i + \varepsilon_i$. Here i denotes a startup and Tax Holiday$_i = 1$ if the startup was granted a tax holiday.The coefficient β_a capture the percentage difference in the growth of selected and rejected startups as compared to the revenue at age 1.

What explains the selection of high-growth firms by the board?:— In Appendix C.2.2, I collect additional information on the ex-ante characteristics of the applicant firms and study their relative ability to predict selection by the bureaucratic board. I find that controlling for the city and sector fixed effects, the bureaucratic board is much more likely to select firms that have secured venture capital funding before the time application, or have been a part of an incubator or if the founders have longer experience. Having VC funding increases the probability of selection by 15 percentage points whereas being a part of an incubator increases it by 5.8 percentage points. Each additional year of founder experience increases the likelihood of selection by 0.6 percentage points. Thus, the board partially relies on the signals from the private market as well as from other branches of the government (selection for incubation) in selecting the firms for tax holidays.

3. Profit Tax Holidays have a Marginal Impact on Startup Growth

The ex-post differences in growth between selected and rejected firms documented above reflect both underlying ex-ante differences in firm types (which are not directly observable) and the effects of the profit subsidies provided by the policy. While the raw differences are informative about the government's selection ability, they do not disentangle these two channels. To separate them, I implement a difference-in-differences strategy that exploits variation in the age at which different startups receive a tax holiday, while matching treated firms to rejected firms to control for observable pre-selection differences.

The main concern in comparing selected firms with rejected firms is that the bureaucrats select already high growing firms and thus the differences completely reflect selection bias. To address this issue, I compare the selected startups performance with a set of two "placebo" firm groups. In the first group, I exploit a cut-off in the policy design which mandated that only startups incorporated after April 2016 would be eligible for a tax exemption. Thus, otherwise similar, and recognized, startups which also qualified for all the other benefits of the policy could not get selected for tax exemptions. I observe these startups[23] in the meeting minutes and thus they serve as my first control group. Additionally, I also match the selected startups to rejected

[23]See Fig. 14 for an example of the firms rejected due to date of formation cut-off from the meeting minutes.

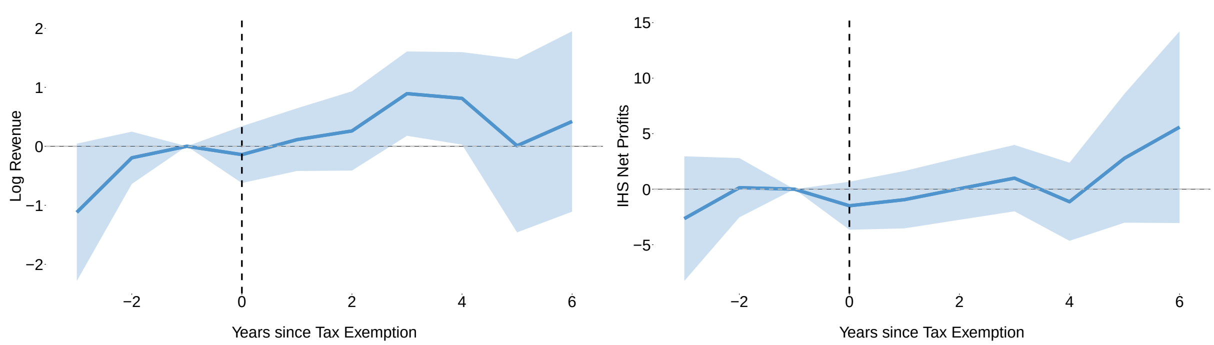
startups based on their pre-decision characteristics such as levels and growth of revenues, assets, equity, whether they were incubated[24] or not and whether they had VC financing before the meeting. These matched startups serve as my second control group. Appendix C.1 and C.1.1 report the summary statistics of the selected and rejected firms and the details of the matching.

I estimate the following specification using a sample of treated firms matched to rejected firms—either excluded due to the age-of-formation cutoff or rejected for idiosyncratic board reasons, but otherwise observationally similar. Y_{ia} denotes the outcomes for firm i at age a which includes revenues, assets, profits and equity. D_{ia}^{τ} is a dummy for time since exemption and δ_i and δ_a denote the firm and age specific fixed effects.

$$\ln Y_{ia} = \delta_i + \delta_a + \sum_{\tau \neq -1} \alpha_\tau (D_{ia}^{\tau} \times \mathbf{1}[\text{Tax Holiday}_i]) + \epsilon_{ia} \tag{21}$$

Fig. 4 presents the results from the above specification. Revenues of selected firms begin to rise about three years after the tax holiday is granted, before reverting to baseline levels. In contrast, net profits after taxes show no discernible effect of the policy. While the revenue trajectory exhibits an upward trend, the overall impact is not statistically significant. Columns 2 and 3 of Table 3 report the average policy effects when treated firms are matched to those rejected due to ineligibility and to all other rejected startups, respectively. The point estimates suggest an economically meaningful increase—roughly 30–40 percent in post-holiday revenues—but these estimates are not statistically significant.

FIGURE 4: EFFECT OF TAX HOLIDAYS ON STARTUP OUTCOMES



The figures plot the coefficient α_τ for $\tau \in \{-3, -2, \cdots, 6\}$ for Log Revenues and IHS Profits respectively from the following regression specification: $\ln Y_{ia} = \delta_i + \delta_a + \sum_{\tau \neq -1} \alpha_\tau (D_{ia}^{\tau} \times \mathbf{1}[\text{Tax Holiday}_i]) + \epsilon_{ia}$. Here i denotes the startup, a age and $\mathbf{1}[\text{Tax Holiday}_i$ denotes if the startup was treated i.e. it was approved for a tax holiday. D_{ia}^{τ} is a dummy variable which is equal to 1 if $a - a_0^i = \tau$ i.e. if the time since exemption is equal to τ for $\tau \in \{-3, -2, \cdots, 6\}$. The sample for the regression is constructed by matching the selected startups to all rejected startups based on pre-meeting variables. The matching process is described in Appendix C.1.

[24] The policy documents (Government of India, 2016) also show a high preference for incubated startups before the selection criteria was ultimately relaxed.

Table 3: Average Diff-in-Diff Estimates

	Baseline	Matched w/ Ineligible	Matched w/ All Rejected
Log Revenue	0.908***	0.473	0.315
	(0.216)	(0.323)	(0.278)
Unique Firms	2,215	522	312

The table provides the average post tax holiday coefficient from the regression in Eq. 21 i.e. Log Revenue$_{ia} = \delta_i + \delta_a + \sum_{\tau \neq -1} \alpha_\tau (D^\tau_{ia} \times \mathbf{1}[\text{Tax Holiday}_i]) + \epsilon_{ia}$. The first column presents the results from a sample of all selected, rejected and ineligible firms. Second column presents results from a sample where the selected firms are matched with a set of startups which were ineligible to get the tax holiday due to a date-of-formation cutoff. And the third column presents the results from the sample where the selected startups are matched with all rejected startups regardless of eligiblity cut-off.

Why do tax holidays not have a significant impact on startup growth?:— Table 18 reports the average profits of startup firms in the sample. On average, firms make negative profits, reflecting the sizable costs associated with early-stage expansion. As a result, a corporate profit tax holiday alone does not significantly improve firm performance. Table 20 provides an estimate of the average value of the profit subsidy offered by the policy. Selected firms pay, on average, 4% less in taxes on their profits during the first eight years—conditional on survival and on having positive profits or tax liabilities in a given year. The effective tax rate is calculated as tax expense divided by profits before taxes.[25]

4. Average Recognized Startup Not High-growth but has Significantly Lower Rate of Exit

Startup recognitions comprise the largest component of the policy and serve as a prerequisite for accessing the tax holidays and R&D benefits discussed above. In this section, I examine the impact of recognitions on the growth, entry, and exit of startups. I find large and significant differences in exit behavior conditional on recognition.

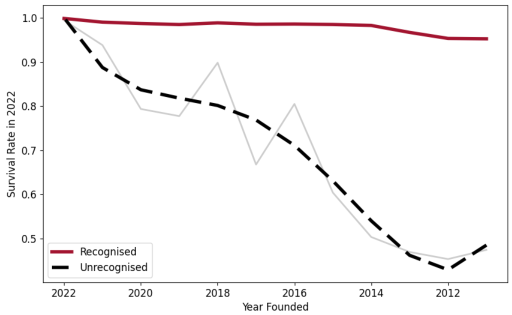
Figures 5a and 5b plot the average survival probability and average paid-up capital in 2024 for recognized and non-recognized firms, respectively. By age ten, recognized firms are 50 percentage points less likely to have exited and have 22% higher paid-up capital.[26] A firm is classified as surviving if its status in the Ministry of Corporate Affairs (MCA) database is listed as "Active" in 2024 and Table 22 in Appendix C.3.1 details the distribution of MCA status codes. For instance, among firms founded in 2012, the average survival rate is around 40%, compared to nearly 95% for recognized startups.[27]

[25]This approach smooths over year-to-year tax variation arising from loss carryforwards or deferred tax liabilities. In some cases, firms are also subject to the Minimum Alternate Tax, so this method captures the effective tax reduction rather than the statutory rate specified in policy documents.

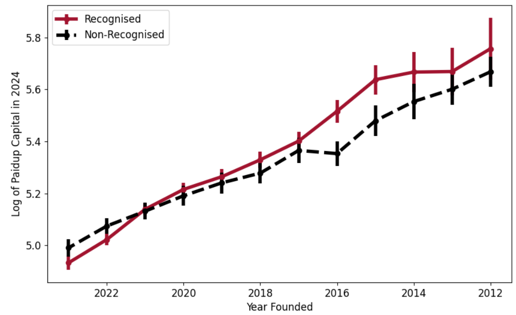
[26]Paid-up capital is defined as the total amount of funds that shareholders have contributed to a company in exchange for its shares, and thus represents the firm's equity financing.

[27]Exit in the MCA is recorded when a firm reports winding up, merger, acquisition, etc. While some exits may go

Figure 5: Characteristics of Recognised vs Average Firm



(a) Survival of Recognized vs Non-Recognized by Formation Year



(b) Size of Recognized vs Non-Recognized by Formation Year

Note: Panel (A) plot the average probability of survival in 2024 (y-axis) against the year of formation (x-axis). The red line denotes startups recognized by the policy (anytime between 2016-24, N = 78,919) and the dashed black line denotes the average non-recognized startup (N = 1.34 million). Panel (B) plots the average log paidup capital for the two set of firms.

Conditional on survival, recognized firms formed in 2012 are about 22% larger in terms of paid-up capital. Fig. 5b plots the raw means of the log of paid-up capital for both recognized firms and the average firm in the economy, by year of incorporation. Notably, recognized firms founded in 2022 have lower paid-up capital than the average firm. Although paid-up capital is an imperfect proxy for firm size, Appendix C.3.2 shows—using a sample of firms merged with Prowess and PitchBook data—that it correlates strongly with both sales and venture capital financing. Since equity financing is a primary source of capital for high-potential startups, paid-up capital provides a reasonable measure to distinguish between the two subgroups of firms. Furthermore, this finding is consistent with the empirical results in Section 4.2, where the average recognized startup applying for a tax holiday exhibited relatively modest growth.

Entry Response to Firm Recognitions Are the benefits provided by recognition associated with an increase in startup formation? In this section, I document the relationship between the expectation of receiving recognition and subsequent changes in startup entry. Specifically, I estimate the empirical counterpart of model equation 6 using aggregate data for all firm types. I merge business registry and recognition data at the district level and exploit cross-sectional variation in the likelihood of obtaining recognition benefits to establish the relationship between recognitions and entry. To do this, I construct a measure of the probability that a potential entrant in period $t + 1$ expects to receive recognition benefits. I define the exposure of a potential entrant to the

unreported, such under-reporting is likely to affect both recognized and non-recognized firms similarly.

policy as

$$\text{Exposure}_{d,t} = \frac{\text{Recognized Startups}_{d,t}}{\sum_{s=t,t-1,t-2} \text{Firm Formations}_{d,s}} \tag{22}$$

i.e., the total number of startups recognized in a given period divided by the total number of startups formed in the previous three years.[28] This measure serves as a proxy for potential entrants' expectations of receiving recognition, based on the observed share of entrants in the same district that were recognized in the previous year. Using this measure, I estimate the relationship in Eq. 23, exploiting both geographical and temporal variation in exposure to the program across districts in India.

$$\log z_{d,t+1} = \delta_d + \delta_t + \beta \times \text{Exposure}_{d,t} + \epsilon_{d,t} \tag{23}$$

where the dependent variable is the log of new firm formations in district d in year $t+1$. The specification includes fixed effects for district (δ_d) and year (δ_t). The model in Eq. 23 assumes that the entrepreneurs in period t form expectations about receiving the benefits of the startup recognition policy based on fraction of the eligible firms that got recognized in the previous period, $t-1$. Thus β captures the relationship between the percentage growth in new firm formations and a one percentage point change in the potential entrants expectation of receiving the recognition benefits. The specification controls for time-invariant district specific factors via the district fixed effects and common temporal shocks over the years using year fixed effects.

Table 4: Effect of Startup Recognitions on Firm Entry

Dependent Variable:	Log New Firm Formations$_t$		
Model:	Overall	Large Districts (Top 10 percentile)	Small Districts (Bottom 90 percentile)
Variables			
$\frac{\text{Recognitions}_{t-1}}{\sum_{k\in\{t-1,t-2,t-3\}} \text{Formations}_k} \times 100$	-0.0094	0.0255*	-0.0024
	(0.0074)	(0.0131)	(0.0051)
Observations	4,553	540	4,013
R^2	0.96547	0.98009	0.93958

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The table presents the estimates from the regression specification in Eq. 23 i.e. $\log z_{d,t+1} = \delta_d + \delta_t + \beta \times \text{Exposure}_{d,t} + \epsilon_{d,t}$ where $z_{d,t}$ denotes the new firm formations in a district d at time t, and exposure to the policy is as defined in Eq. 22. The first column presents the results for all the districts in India while the second and third columns split by pre-policy levels of business formations.

[28]Although firms formed in the last ten years are eligible for recognition, I focus on the previous three years because the median age at recognition is 7 months, and the 90th percentile is about 3.7 years. The results are robust to alternative definitions.

Results from the regression in Eq. 23 i.e. the estimates of coefficient β are reported in Table 4. Column 1 presents the estimate from a regression including all the 572 districts in the sample. The resulting coefficient, -0.0094, is statistically insignificant and small in magnitude. The average effect, however, masks significant heterogeneity. As firm formations and startup activity are concentrated in urban districts with large population, the next two columns split the sample based total firm formations in 2015 i.e. prior to the policy. While the coefficient is again statistically insignificant and small in smaller districts, in the top ten percentile districts it takes a statistically significant value[29]. Economically, the affect is substantial. A one percentage point increase in the recognition probability is associated with a 2.5% growth in new firm formations.

5 Model Estimation

The model calibration strategy has three steps. First, I calibrate the terminal Balanced Growth Path (BGP) of the model to match the aggregate and firm dynamic moments of the advanced economies. Then I calibrate the path of cost parameters in the model such that the model is able to match the level and changes in the key empirical trends in the Indian economy since late 1990s. And lastly, I calibrate the "selection ability" and other policy related parameters to match the cross-sectional responses documented in Section 4.2.

5.1 Final BGP Calibration Strategy and Results

Thirteen parameters, $\{\rho, \lambda, \zeta_R, \psi_{R,T}, \kappa, \zeta_N, \psi_{N,T}, \zeta_R^e, \psi_{R,T}^e, \zeta_N^e, \psi_{N,T}^e, \omega_0, \alpha\}$, govern the final balanced growth path of the model. The subscript T denotes the terminal values. I chose the values of the six parameters directly from the prior literature and calibrate the remaining seven to match key empirical firm dynamics moments from the US and other advanced economies.

Externally Calibrated Parameters:— I set the discount rate, ρ, to 0.05 per year in line with the prior endogenous growth literature. I also directly set ζ_R, ζ_N i.e. the curvature parameters the R&D and search cost function to be equal to 2 i.e. a quadratic cost function in line with the vast growth literature (Peters and Zilibotti, 2021; Acemoglu et al., 2018; Akcigit, Alp, and Peters, 2021) [30]. I also directly set the curvature parameters on the entry cost functions to be also 2 which is consistent with the prior literature estimates of the entry elasticity of the firms with respect to the value of entry (De Haas, Sterk, and Van Horen, 2022; Sedláček and Sterk, 2017; Gutiérrez, Jones, and Philippon, 2019). And lastly, I set κ equal to 0.4.

Internally Calibrated:— I calibrate the remaining seven parameters by jointly matching seven firm dynamics moments in the model's terminal BGP with the data from US and other advanced economies. While all the parameters are jointly estimated, each moment is chosen to identify

[29] The results are robust to alternate percentile cut-offs and are reported in the appendix.

[30] This is also consistent with empirical literature estimating the R&D elasticity with respect to taxes, see Hall and Ziedonis (2001); Bloom, Griffith, and Van Reenen (2002) etc.

a corresponding parameter in the model. The innovation step size maps the rate of creative destruction in the economy to rate of aggregate productivity growth which I target to be 2%. The entry cost parameter for the R&D firms, $\psi^e_{R,T}$, determines there share of R&D firms in the balanced growth path relative to the non-R&D firms. I match it to be 3%, consistent with share of innovative firms in the Business Dynamics Statistics of Innovative firms Goldschlag and Perlman (2017) Goldschlag and Miranda (2020). I match the life cycle growth rate of R&D firms to average size at ten years of age of R&D firms identified in Akcigit et al. (2025). The moment is governed by the R&D firm search cost parameter $\Psi_{R,T}$.

While the entry and life cycle growth of non-R&D firms is governed by only two parameters, the corresponding values for the non-R&D firms are determined by the distribution of entry costs $\psi^e_{N,T}(z)$ and average search cost parameter $\Psi_{N,T}$. Type specific entry costs determine the share of entrants of each type in the economy as well as the final firm type distribution in the BGP — as each type of firm grows at a different rate depending on their type z. I, first, assume that the entry costs for different types are such that the resulting distribution of entrants follows a Pareto distribution with a mass at $z = 0$. The mass $z = 0$ type firms have infinite search cost and never grow after entering the economy.. The CDF of the resulting entrant distribution is governed by two parameters ω_0 and the Pareto shape parameter α —

$$F(z) = \begin{cases} \omega_0 & \text{for } z = 0 \\ \omega_0 & \text{for } z \in (0, 1) \\ \omega_0 + (1 - \omega_0)(1 - z^{-\alpha}) & \text{for } z > 1 \end{cases} \tag{24}$$

Table 5: Terminal BGP Moments

Moment	Data	Model
Entry Rate of Normal Firms	0.13	0.13
Share of R&D Firms	0.04	0.03
Average R&D Firm Size at 10 years	5.8	5.3
Average Non-R&D Firm Size at 10 years	2.3	2.3
Mass of one product firms	0.55	0.55
Average Size of the non-R&D top 5^{th} percentile at 10 years	7.0	7.3
Aggregate Productivity Growth	0.02	0.02

I discretize this distribution on grid of $z \in \tilde{Z} := \{z_0, z_1, \cdots, z_N\}$ values and and estimate the parameters, $\Psi^e_{N,T}(z)$, $\forall z \in \tilde{Z}$ such that the resultant entrant distribution follows the CDF defined in Eq. 24. This pins down the distribution of entry costs. To pin down the level, I calibrate the mean of the search costs to match the overall entry rate of non-R&D firms in the model to be equal to 13% (Decker et al., 2014; Peters and Zilibotti, 2021). I calibrate the parameters ω_0 such that in the final stationary distribution, the share of one product firms is equal to 55% corresponding to the distribution of Corporation firms in the US (Census Bureau, 2021). And, finally, I calibrate

α to match the share of high-growth firms in the US data as documented by Kim et al. (2024). Table 5 summarizes the terminal BGP moments that I targeted and the resulting values from the model.

Table 6 reports that parameter values resulting from the calibration exercise. Intuitively, $\psi^e_{R,T} < \int \psi^e_{N,T}(z)dz$ captures the fact the R&D firms have much higher entry costs which is required to match their small share in the economy. While $\psi_{R,T}$ is also lower as compared to $\psi_{N,T}$, the actual average search costs for non-R&D firms is much higher as 68% (ω_0) of the non-R&D entrants have an infinite search cost. There is also a substantial amount of re-allocation in the terminal balanced growth path. This is reflected by the difference between ω_0 and the targeted mass of one product firms (55%). This implies, that although low type entry is substantial in the economy, in the stationary distribution there share is almost 13% lower as more productive firms grow and capture the larger market share.

Table 6: Internally Calibrated Terminal BGP Parameters

Parameter	Description	Value
λ	Innovation Step Size	6.5
$\int \psi^e_{N,T}(z)dz$	Average Entry Cost Parameter	0.0716
$\psi^e_{R,T}$	R&D Entry Cost Parameter	0.00001
$\psi_{N,T}$	Search Cost Parameter	0.333
$\psi_{R,T}$	R&D Cost Parameter	0.236
ω_0	Non-R&D Entrants with $z = 0$	0.680
α	Pareto Shape Parameter for entrants with $z > 0$	3.22

5.2 Transition and Policy Parameters Calibration Strategy

The model economy eventually converges to the above balanced growth path. However, the transition path of the economy depends on, i) the long-run dynamic evolution of the entrant and incumbent R&D/search costs over time and ii) endogenous response of the entrepreneurs to the startup policy. In this section, I jointly estimate the transition and policy parameters by matching the transition path of the Indian economy since 1990 (defined as $t = 0$) along with the empirical estimates of the policy response documented in Section 4.2.

The initial distance of the economy from the frontier is governed by initial productivity differences: $\{\ln \bar{q}_{j0} - \ln q_{j0}\}_{j\in[0,1]}$. Thereafter, the economy converges endogenously towards the frontier at rate governed by terminal BGP parameters and the path of parameters which determine the evolution of long-run supply of entrants and the costs of expansion for incumbent firms: $\{\Psi^e_{N,t}(z), \Psi^e_{R,t}, \Psi_{R,t}, \Psi_{N,t}\}^{z\in\tilde{Z}}_{t\in[0,\infty)}, \gamma\}$. As the economy transitions on this path, I introduce a policy shock in 2016 governed by the parameters $\{\{\phi^R, \phi^N, \phi^{h,z}(z), \phi^{d,z}(z), \bar{s}^R_c, \bar{s}^R_c, \bar{s}^R_x, \bar{s}^N_x, \bar{s}_\tau, \bar{s}_\pi\}$ which applies to the cohorts of firms that enter in 2016 and after. The policy shock impacts overall entry, exit and growth of different types of firms heterogeneously while the parameters governing the long-run evolution of the economy continue on their trend path. I estimate the parameters jointly by simulating the model path and cross-sectional distribution given the following set of parameters

$$\{\{\ln \bar{q}_{j0} - \ln q_{j0}\}_{j\in[0,1]}, \{\Psi^e_{N,t}(z), \Psi^e_{R,t}, \Psi_{R,t}, \Psi_{N,t}\}^{z\in\tilde{Z}}_{t\in[0,\infty)}, \gamma\}, \{\{\phi^R, \phi^N, \phi^{h,z}(z), \phi^{d,z}(z), \bar{s}^R_c, \bar{s}^R_c, \bar{s}^R_x, \bar{s}^N_x, \bar{s}_\tau, \bar{s}_\pi\}$$

and matching corresponding moments in the data. I next describe the set of moments of that I target and the corresponding parameters that each moment identifies.

5.2.1 Transition Parameters Identification

The pre-policy and counterfactual-no-policy transition of the economy is governed by the following parameters: $\{\{\ln \bar{q}_{j0} - \ln q_{j0}\}_{j\in[0,1]}, \{\Psi^e_{N,t}(z), \Psi^e_{R,t}, \Psi_{R,t}, \Psi_{N,t}\}^{z\in\tilde{Z}}_{t\in[0,\infty)}, \gamma\}$ where $\{\ln \bar{q}_{j0} - \ln q_{j0}\}_{j\in[0,1]}$ denotes the initial productivity distance from the frontier across different lines, $\{\Psi^e_{N,t}(z), \Psi^e_{R,t}, \Psi_{R,t}, \Psi_{N,t}\}^{z\in\tilde{Z}}_{t\in[0,\infty)}$ denotes the path of entry and incumbent R&D/search costs, and γ controls the average productivity gains that come from technological adoption by non-R&D firms. The corresponding empirical moments that each of these parameters identify are given in Table 7 and their relationship with each of the parameters is described below.

For initial productivity gaps, I assume that the gaps from the frontier are normally distributed with mean m_0 (which evolves as in Eq. 12) and standard deviation 1 i.e. $\ln \bar{q}_{j0} - \ln q_{j0} \sim \mathcal{N}(m_0, 1)$. I also assume that initially most of the lines are laggard i.e. $M^f_0 \to 0$[31]. The average gap is identified using the GDP per capita differences between US and India in 1990. I directly calibrate m_0 by matching the log differences in per-capita GDP between India and the US in 1990. Using the constant-2015 US dollars, Indian GDP per capita in 1990 was 537.9 and US was 39,200[32]. This gives me an initial log GDP per capita difference target of 1.86. The evolution of the productivity gaps are determined by the path of aggregate technological adoption and innovation on the transition path. The relative contribution of the two, in turn depends on innovation step size $\{\lambda_t\}_{t\geq 0}$ at which adoption of frontier technologies shift the lines from laggard to frontier: γ. I assume that the innovation step size remains constant and equal to the value calibrated in Sec 5.1 i.e. $\lambda_t = \lambda \, \forall t$. I calibrate γ to match the path of GDP per capita growth rate in the economy relative to the balanced growth path. Given the initial productivity differences, the parameter γ captures the speed at which the aggregate adoption efforts in the economy translate into aggregate productivity gains.

[31] I do not set it exactly equal to zero for numerical convenience

[32] `https://data.worldbank.org/indicator/NY.GDP.PCAP.KD?locations=US-IN`

Table 7: Transition Moments

Moment	Model	Data
Δ Log GDP per Capita between India and US in 1990	1.86	1.86
Entry Path of R&D Firms 1990-2016	Fig. 6	
Entry Path of non-R&D Firms 1990-2016	Fig. 6	
Δ Life Cycle Growth 1990-2000 Cohort	Fig. 19	

The path of entry cost parameters, $\{\Psi^{e}_{N,t}(z), \Psi^{e}_{R,t}\}^{z\in\tilde{Z}}_{t\in[0,\infty)}$ is identified by matching the path of aggregate entry rates of R&D and non-R&D firms in India and their levels relative to the US. When the economy is far away from the frontier and the overall mass of firms is low, there are large gains from entering and the entrant firms can capture a large share of market quickly. Hence, in a calibration where entry costs are same as the BGP, entry rates jump in the beginning and eventually converge down to the BGP levels (Appendix Figure 21). Empirically, however, the entry rates (per capita) of limited liability firms have been extremely low and and have only risen slowly post liberatization in 1991 (Appendix Figures 18a and 18b). There are large and persistent differences in limited liability formation rates between India and developed and other middle income economies. Appendix D.2 summarizes the differences using which I target the initial business formation rates in 1990 India to be $1/27^{th}$ of the balanced growth path and target $1/18^{th}$ as the corresponding number for R&D firms. Hence, I calibrate the path of the entry costs such that the model matches the observed entry rates in the economy.

Table 8: Policy Moments in the Model and the Data

Moment	Model	Data
Tax Exemptions		
Fraction of Startups with Tax Exemptions	0.001	0.001
Selected Firm Growth - Rejected Firm Growth	10.9	10.9
Tax Expense Selected Firm - Tax Expense Rejected Firm	0.046	0.046
R&D Effective Subsidies		
Fractions of R&D Startups with Recognition	0.55	0.55
Δ R&D Share of Startups Post-Policy	0.5	0.5
Impact of Recognition on Patenting of existing R&D Startups	0.0	0.0
Startup Recognitions		
Fraction of Total Startups Recognized	0.06	0.06
Survival Rate of Recognized - Average Startup	0.25	0.25
Growth of Recognized - Average Startup	0.12	0.12
Entry Response to Recognition Probability	0.025	0.025

Similarly, on the transition path, different cohorts of firms which enter at different points in time grow at different rates which is governed by the evolution of incumbent R&D and search costs. I assume that the path of R&D and Search cost parameters for the incumbent change by the same amount in relative terms and estimate them by matching the evolution of the firm life cycle over time and its value in 2016. Specifically, I match the difference in the life cycle growth of two cohorts of firms (1990 and 2000) from the Prowess data[33] Given the path, I estimate the level by matching the firm survival rates match with the average survival rates reported in Fig. 5a.

Lastly, I assume that the entry costs of different types move in tandem to the BGP levels which gives me three effective cost paths to estimate $\{\bar{\Psi}^e_{N,t}, \Psi^e_{R,t}, \Psi_{N,t}\}_{t\in[0,\infty)}$[34]. I parametrize each of these three cost paths by introducing a wedge over the terminal values that slowly declines over time, ultimately converging to one at the terminal balanced growth path. The path of the wedge is governed by two parameters S_{ψ^j} and ν_j for each $j \in \{N, N^e, R^e\}$. The first parameter governs the initial size of the wedge and the second parameter governs the speed at wedge exponentially declines over.[35]

5.2.2 Policy Parameters Identification

The endogenous response of startups of different types to the startup policy is governed by the two set of policy parameters: the selection abilities $\{\phi^{z,h}(z), \phi^{z,d}(z)\}$ and the effective wedges that the policy introduces in the different margins of the startup behavior $\{\bar{s}^R_c, \bar{s}^N_c, , \bar{s}^R_x, \bar{s}^N_x, \bar{s}_\tau, \bar{s}_\pi\}$. In the baseline, I keep the age limits on sunset clauses as 10 years. Each of these parameters is identified by matching with moments shown in Table 8.

For the selection abilities, I first parameterize the selection function (conditional distribution of selected firms) using a beta distribution whose shape is governed by a single parameter η. I map the set of different ex-ante types, z to a unit interval between i.e. $\tilde{z} \in [0, 1]$, where $\tilde{z} := \frac{z}{\max(z)}$[36] Then I use a single parameter $\eta \in \mathbb{R}$, to control the two parameters of the Beta distribution depending on its sign i.e. I define the two parameters of the beta distribution a, b as follows

$$(a, b) = \begin{cases} (1+\eta,\ 1), & \eta \geq 0, \\ (1,\ 1-\eta), & \eta < 0. \end{cases}$$

[33]Prowess Data provides firm level data for medium and large sized Indian firms. It contains data for around 70,000 firms beginning in 1988. I compute the life cycle growth as the ratio of sales at age 10 relative to sales at age for the cohort of firms born between 1990-95 and 2000-2005.

[34]As $\Psi^e_{N,t}(z) = \bar{\Psi}^e_{N,t}$ and $\Psi_{N,t} = \Psi_{R,t}\ \forall t$

[35]The exact functional form is given as follows, where t denotes time and t_0 denotes the initial period

$$\psi_j(t) = \frac{\psi_j^T}{1 + S_{\psi_j}\, e^{-[\nu_j\,(t-t_0)]^k}}$$

The parameters S_{ψ_j} and ν_j are calibrated to match the target moments. I set $k = 1.5$ to get a shape that converges slowly at the beginning but eventually grows exponentially.

[36]Note that the minimum z is zero.

Given these two parameters, I define the density over types for the selected firms as[37]

$$\tilde{\phi}(z) = \frac{\tilde{z}^{\,a-1}\,(1-\tilde{z})^{\,b-1}}{\int \tilde{z}'^{\,a-1}\,(1-\tilde{z}')^{\,b-1}dz'},$$

When $\eta = 0$, each firm type has an equal probability of getting selected. On the other hand, when $\eta < 0$, low types have a higher conditional probability of getting selected and vice versa. I use η^h and η^d to denote the selection function parameters for startup recognition and tax holidays respectively.

Tax holiday parameters, η^d and $\bar{s}_\pi$, are estimated directly by matching the differences in the growth of selected vs rejected firms and the effective profit gains provided by the policy described in Section 4.2. As different firm types grow at different rates, the parameter η^d, by controlling the composition of selected startups, directly controls the difference between the average selected and rejected startup. If $\eta^d \to -\infty$, the selected startups have zero growth post-entry as $z = 0$ type have infinite search costs. On the other hand, as $\eta \to \infty$, the selected startups grow at extremely high rate as they have very high search ability.

While the effective value of profit subsidies introduced by the policy is directly measurable in the data, the corresponding values are for R&D benefits and recognitions are not directly observed. I identify the three R&D benefits related parameters, $\{\phi^R, \bar{s}_c^R, \bar{s}_x^R\}$, of the policy by matching three moments: fraction of R&D startups which were recognized, change in entry of R&D startups and intensive growth in the patent filing of R&D startups. While ϕ^R is directly estimated from the corresponding moment in the data, $\bar{s}_c^R$ and $\bar{s}_x^R$ control the extensive and intensive margins of R&D by startups in the economy. A fixed cost subsidy increases the value from entering for the startups but doesn't affects their intensive decisions post-entry. An R&D subsidy, on the other hand, jointly impacts both the value from entering and dynamic decision to invest in R&D post entry. In Section 4.2, I isolate the intensive effect of the policy and the overall change in entry which, thus, pins down both the policy parameters.

Lastly, firm recognition parameters, $\{\phi^N, \eta^h, \bar{s}_c^N, \bar{s}_x^N, \bar{s}_\tau\}$, are estimated by matching five moments: number of entrants recognized by the policy, survival of recognized vs non-recognized startups, average size of selected vs rejected startups conditional on survival, average entry response to 1pp change in probability of selection, and the intensive effect of firm recognitions on growth. While, ϕ^R is directly observed in the data, other parameters are jointly estimated. Intuitively, the entry response of the startups captures the combined additional value that the policy provides to an average startup given the selection probabilities across types — as percent changes in entry are directly proportional to the percent changes in firm value as in Eq. 5 and 6. This overall change

[37] This implies that the conditional probability of getting selected for a type z, i.e. $\Pr(d = 1 \mid z)$, is

$$\phi(z) = \Pr(d = 1 \mid z) = s(z) = \tilde{\phi}^N \frac{\tilde{\phi}(z)}{f_Z(z)}, \qquad \tilde{\phi}^N := \Pr(d = 1),$$

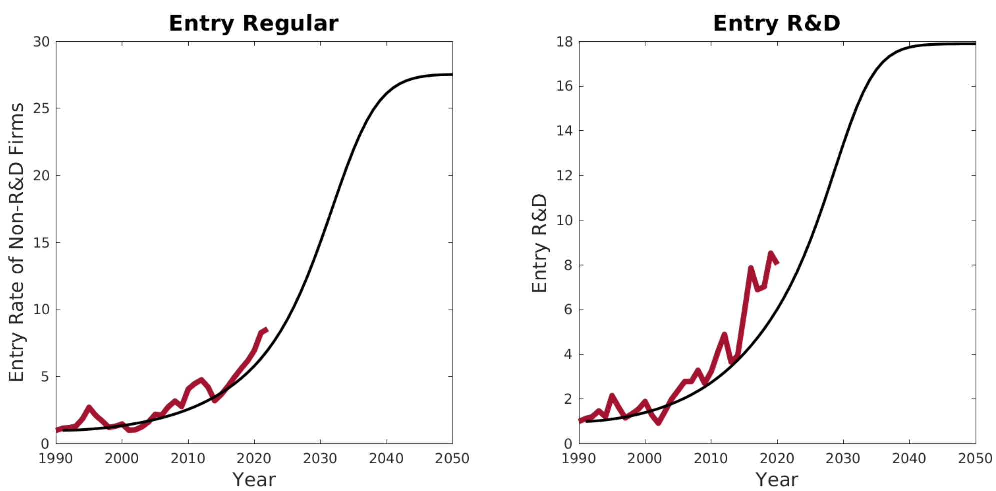
where $f_Z(z)$ is the unconditional pre-policy type distribution.

in value is comprised of $\{\bar{s}_c^N, \bar{s}_x^N, \bar{s}_\tau\}$ of which, s_x^N is directly pinned down by the intensive effect of recognition on incumbent firms. $\bar{s}_\tau$ and η^h are pinned down by the difference in survival of recognized vs non-recognized startups and differences in their growth (which are governed by composition i.e. η^h). Lastly, s_c^N captures the remaining additional value required to match the original entry response.

5.3 Transition and Policy Parameters Calibration Results

Long Run Transition:— Figures 6 and Appendix D.2.2 Fig 19 plot the long run evolution of the entry and life-cycle of firms in the model and the data. The red line in Figure 6 correspond the growth in entry of non-R&D and R&D firms in data which have steadily grown over time from their low values in 1990s. The model is able to closely match this transition as well as the life cycle growth of different cohorts in the data.

FIGURE 6: MATCHING THE TRANSITION



The figures compare the entry rates of non-R&D and R&D firms in the data vs the model. The red lines represent the data and the solid black lines represent the calibrated transition in the model.

The estimated wedges in the four cost paths are plotted in Appendix Figure 20 which provides two main insights. First, the estimated wedge in the cost for entrants is almost an order of magnitude higher than the incumbent costs. This implies that while the incumbent firms in the Indian economy do face larger frictions to growth but as compared to entrants these frictions are much lower — resulting in a low supply of entrepreneurs. Intuitively, this is also reflected in the firm size distribution of the Indian limited liability firms. While the entry rates are extremely low as compared to the developed economies, the firm size distribution of Indian limited liability firms is almost identical to US corporations — with the share of one product firms being close to 55% in both the economies. Crucially, this fact implies that a policy that incentivizes entry without bringing the overall entry costs down might substantially misallocate labor as the same worker is much less efficient while working for the entrants as compared to the incumbents.

Second, the slow convergence of entry costs substantially slows down the catch-up of the economy (Fig 22). In the absence of aggregate cost wedges, almost eighty percent of the product lines transition into frontier lines within ten years. With the estimated cost wedges, this process is almost six times slower. This implies that there could be large potential gains from incentivizing high growth firm entry as these firms can accelerate the path of technological adoption which generates large inter-temporal knowledge spillovers by increasing the overall base productivity in the economy.

5.3.1 Selection Ability and Effective Subsidy Wedges

The calibration results for policy parameters are shown in Table 9. The first part of the table reports the average type composition of the firms selected for recognition and for tax holidays as compared to the average pre-policy entrant type in the economy. The second panel reports the effective subsidy wedges introduced by the policy.

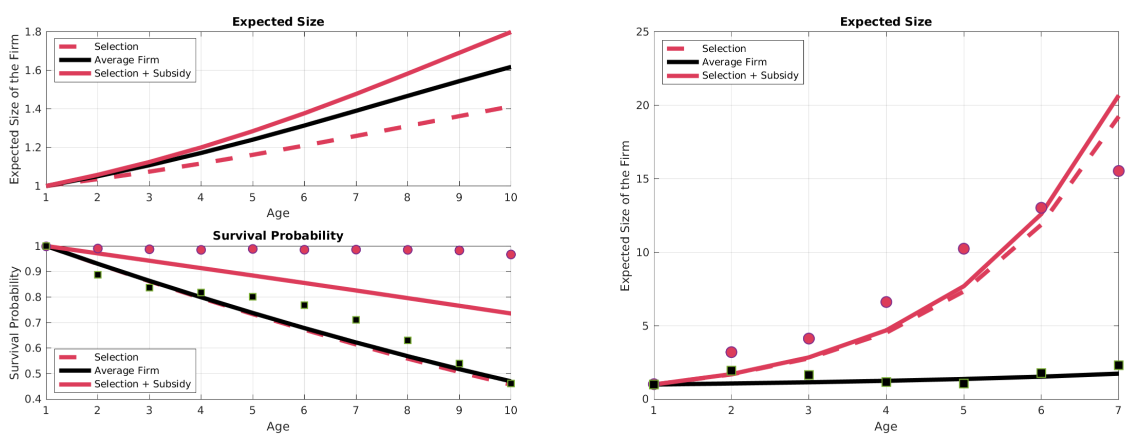
Table 9: Calibrated Policy Parameters

Variable	Value
Targeting: Overall and across Types	
$\mathbb{E}[z]$ Baseline Pre-policy	1
$\mathbb{E}[z]$ Tax Exempt Firms	2.23
$\mathbb{E}[z]$ Recognized Firms	0.91
Fraction R&D Firms Targeted in Cohort, ϕ^R	0.56
Fraction Non-R&D Firms Recognized in a Cohort, ϕ^N	0.12
Effective Subsidies	
Tax Exemption: Profit Subsidy, $\bar{s}_\pi$	0.04
R&D Firm: R&D Subsidy, $\bar{s}_x^R$	0.0
R&D Firm: Operating Cost Subsidy R&D Firms, $\bar{s}_c^R$	0.89
Recognition: Exemptions, $\bar{s}_\tau$	0.72
Recognition: Search subsidy, $\bar{s}_x^N$	0.0
Recognition: Operating Cost Subsidy Non-R&D Firms, $\bar{s}_c^N$	0.05

The average firm selected for tax holidays is nearly 2.2 times more efficient at searching for ideas and expanding product lines than the average firm in the economy. However, as shown in Section 4.2, the direct impact of profit subsidies on firm growth is minimal. This is illustrated in Figure 7b, where dots represent the empirical data and the lines depict the corresponding model-implied growth paths. The dashed red line isolates the selection effect—i.e., the growth trajectory of the selected firms based solely on their higher underlying type distribution. Because the effective subsidy wedge for tax holidays is small, the policy's incremental impact on post-entry growth, beyond this compositional effect, remains marginal. This is consistent with the empirical evidence showing that tax holidays do not significantly alter firm behavior after entry.

For startup recognitions, however, the calibration exercise reveals evidence of adverse selection. The expected type of recognized startups, $\mathbb{E}[z]$, is roughly 10% lower than that of the average entrant in the pre-policy economy. This implies that the policy disproportionately selects entrepreneurs who, upon entry, are less likely to grow—a stark contrast to the distribution of entrants in the absence of the program. What drives this result? Figure 7a plots the model-implied growth and survival rates of recognized startups relative to rejected ones. Given the substantial exit distortion created by the policy, the model predicts that, if recognitions targeted the average entrant, the resulting growth outcomes should have been significantly higher. Yet empirical evidence shows that recognized startups have only about 20% higher paid-up capital than rejected firms. To reconcile this with the observed dynamics, the model assigns a negative selection wedge: the dashed red line in the top-left panel reflects the lower underlying type distribution of selected entrants.

Figure 7: Selection vs Effective Subsidies



(a) Model vs Data Firm Recognitions

(b) Model vs Data Tax Exemptions

Recognition subsidies generate a significant distortion in the process of creative destruction for selected startups. In the calibrated model, conditional on being selected, a recognition reduces the exit probability of a product line by approximately 72% relative to the baseline. This decline in exit risk raises the continuation value of the firm, thereby inflating its post-selection value. As a result, the model attributes a relatively small implicit operating cost subsidy to recognitions: much of the value transferred to firms comes instead from the distortion to exit and replacement dynamics, rather than from direct fiscal support.

6 Policy Design and Counterfactuals

Given the selection ability and effective wedges estimated in the previous section, how much more effective is a targeted program compared to a uniform subsidy? And how do different types of startups respond to variations in selection ability, the composition of subsidies, and their duration and size? In this section, I use the calibrated model to answer these questions.

6.1 Aggregate Returns from Targeted vs Uniform Subsidy

Net welfare gains vary substantially across policy instruments. Table 10 (top panel) reports the aggregate welfare gains, fiscal costs, and gross benefits for the three main instruments in the calibrated model. In each case, I simulate the effects of subsidizing a single cohort of entrants using the estimated selection ability and subsidy wedge specific to that policy. The corresponding impulse response functions—shown in Appendix D.3—track the deviation of key variables from the baseline transition path. While all three instruments stimulate startup entry, the magnitude and composition of the response differ sharply. Tax holidays and R&D benefits induce only a modest increase in the entry of non-R&D firms, as both policies are effectively targeted toward high-type, innovative firms. Under tax holidays, the government's high selection ability means that only high-growth entrepreneurs expect to qualify and thus change their entry decisions. Similarly, the embedded operating cost subsidies in the R&D benefit are fully targeted toward innovative firms, leaving marginal non-R&D entrants largely unresponsive. By contrast, startup recognitions increase aggregate entry by roughly 5%, primarily by inducing entry among average-quality entrepreneurs. Because recognitions are less selective, a broader set of marginal entrants expect to qualify for benefits—and respond accordingly. This distinction in the quality composition of new entrants plays a central role in driving the relative welfare effects across the policies.

The increase in entry—together with changes in post-entry R&D and search behavior—raises the economy's overall rate of productivity growth. This, in turn, boosts output and consumption, generating welfare gains. The second column of Table 10 expresses these gains in terms of equivalent units of period-0 consumption that would leave households indifferent between the policy and the baseline. Since each policy remains active for ten years following the cohort's entry, the associated fiscal costs are also accrued over that horizon. The discounted value of these costs is reported in the third column, while the final column shows the benefit-to-cost ratio—that is, the amount of welfare gained per unit of cost—for each instrument.

Targeted tax exemptions and R&D benefits have a BCR of 5.14 and 6.69 respectively whereas startup recognitions have a much lower benefits-to-cost ration of 2.84.These differences stem from the composition of startups that the policies induce to enter. For a given fiscal cost, the benefits of tax holidays and R&D subsidies are concentrated among innovative, high-growth startups—the firms that generate the largest productivity spillovers. As a result, these policies

Table 10: Aggregate Policy Returns to Different Instruments

Policy	Welfare (% change)	Cost (% of GDP)	BCR
Non-R&D Recognitions	0.0707 %	0.0248 %	2.84
R&D Benefits	0.0281 %	0.0042 %	6.69
Tax Exemptions	0.0036%	0.0007 %	5.14
Counterfactual Uniform Policy			
Uniform Non-R&D Recognitions	0.9598 %	0.34 %	2.81
Uniform Tax Exemptions	0.0254 %	0.0086 %	2.95

disproportionately promote the entry of high-type firms, yielding high benefit-to-cost ratios. In contrast, recognitions are not well targeted: they tend to support average- or below-average-type entrants. Moreover, recognitions also distort exit behavior by prolonging the survival of low-type firms. Since these firms do not invest in search or innovation after entry, their continued presence leads to a misallocation of labor and dampens aggregate productivity growth. This post-entry misallocation ultimately reduces the overall welfare gains from recognitions, even when their entry effects are positive.

How do these gains compare to a policy that does not target benefits across different types? The bottom panel of Table 10 presents the welfare effects of a counterfactual policy in which the subsidies are allocated uniformly across types—i.e., every entrant has an equal conditional probability of being selected. Under this untargeted scheme, the benefits-to-cost ratio is only slightly lower than in the targeted program for recognitions, but it is substantially lower for tax holidays. The difference arises from the nature of the two subsidies. For tax holidays, the effectiveness of the program is highly dependent on selecting high-growth firms; removing targeting therefore significantly reduces the gains. In the case of recognitions, however, even the targeted policy selects many low-type firms. The model implies that an untargeted scheme would have selected marginal entrants of even lower quality, reducing the benefits, but only modestly. Thus, while government selection marginally improves the effectiveness of recognitions, the gains from targeting are much more pronounced for policies that rely on post-entry scaling and innovation, such as tax holidays.

6.1.1 Gains from Entry and Post-Entry Behavior

A substantial share of the net gains from recognition subsidies stems solely from the productivity improvements made by startups at the time of entry; absent this entry margin, recognitions generate an overall welfare loss. Table 11 decomposes the welfare effects into total gains and those arising exclusively from post-entry dynamics. The results show that recognitions deliver positive net gains only because they induce additional entry. Once the startups have entered, the misallocation of labor caused by recognitions—due to the subsidization of firms that are not necessarily high-growth—results in a net welfare loss. Appendix Figure 23 illustrates that recognition subsidies marginally increase search activity post-entry. However, the fiscal costs of providing these subsidies outweigh the modest consumption gains from the resulting productivity improvements. As a result, the post-entry contribution of recognitions is negative, and the policy's overall welfare effect is strictly positive only to the extent that it boosts entry.

Table 11: Entry vs Post-Entry

Policy	Net Δ Welfare	Δ Welfare without R&D/Search to Enter
Non-R&D Recognitions	0.0459 %	-0.0199 %
R&D Benefits	0.0339 %	0.0239 %
Tax Exemptions	0.0024 %	0.0013 %

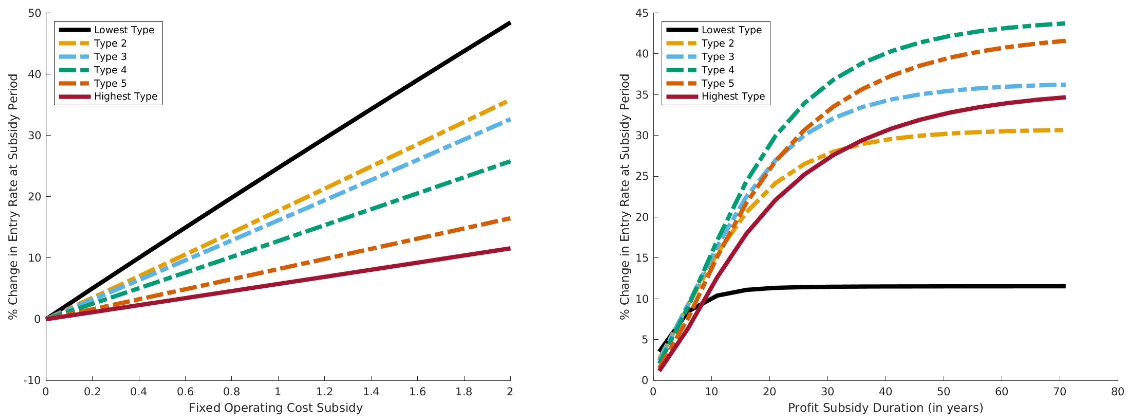
In contrast to fixed cost subsidies, a substantial share of the welfare gains from R&D benefits and tax holidays—roughly 30% and 50%, respectively—comes from the R&D and search investments made by startups at the time of entry. Since innovative and high-growth startups continue to expand their product portfolios through ongoing technological improvements after entry, a large fraction of the gains also derive from post-entry innovation. Importantly, these gains persist well beyond the duration of the subsidy: once the subsidized firms enter, they continue innovating until they exit, even after the policy expires.

Moreover, the composition of surviving firms further amplifies the benefits of targeting. Because low-type firms fail to grow and eventually exit, the pool of older firms consists disproportionately of higher-type incumbents. Targeted subsidies, therefore, shift resources toward firms with strong growth potential, whereas untargeted entry subsidies tend to reallocate resources toward lower-type entrants on average. By selectively promoting high-type startups, R&D benefits and tax holidays generate net welfare gains both at entry and through their dynamic effects on the distribution of firm types over time.

6.2 Heterogeneous Entry Response to Subsidy Type and Duration

Figures 8a and 8b illustrate the heterogeneous entry responses of different startup types to two policy instruments: operating-cost subsidies and profit subsidies with varying sunset durations. As shown in Figure 8a, which plots the percent change in entry against the size of the operating-cost subsidy, fixed cost subsidies disproportionately promote the entry of the lowest-type firms. This is because these firms never scale beyond their initial product; a fixed cost subsidy therefore represents a large share of their total value, making entry highly attractive. In contrast, growing firms also benefit from a cost subsidy but to a much smaller degree. Their long-run value is driven primarily by the expected gains from adding new products as they age, so a fixed subsidy affects only a small fraction of their overall growth potential. This differential response highlights the policy trade-off: while operating-cost subsidies can effectively boost entry, especially among low-type firms, they do not necessarily target high-growth firms that contribute more to aggregate productivity.

Figure 8: Endogenous Type Specific Entry Response



(a) Entry vs Fixed Cost Subsidy Size

(b) Entry vs Length of Profit Subsidy

The entry response to profit subsidies, however, crucially depends on the length of the sunset clause. In the calibrated model, a 4% profit subsidy with a sunset clause length of upto 7 years leads to the highest entry growth of lowest type firms. However, as the length of the sunset clause is extended the entry response of the lowest types flattens out quickly. The high type entry response, while small at short sunset durations, increases sharply as the durations are extended. Intuitively, this happens as a result of the fact that all the firm types except for the lowest type increase their profits by adding more products as they grow older conditional on survival. As a result, the value of a profit subsidy increases over the life cycle for these firms, making longer subsidy durations more attractive. Low-type firms, by contrast, do not expand and almost surely lose their single product to creative destruction as they age; for them, late-life subsidies have little marginal value. Thus, the interaction between firm growth potential and subsidy duration drives the heterogeneity in entry responses across types.

7 Conclusion

This paper studies the design and welfare implications of targeted startup policies in developing economies. I develop a quantitative framework in which heterogeneous startups respond endogenously to different policy instruments and to the government's ability to select high-growth firms. These micro-level responses aggregate up to determine the policy's overall economic impact. To calibrate the model, I assemble new firm-level data on selections and post-entry outcomes from India's Startup India program. The results show that well targeted subsidies can double the aggregate returns as compared to uniform support. However, when the incentives for awarding subsidies are not well structured, such as startup labeling, the policy targets badly and introduces distortions by prolonging the survival of low-performing selected firms. The composition of subsidies and the duration of support emerge as key design elements that shape the entry response and long-run effectiveness of the program.

References

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti. 2006. “Distance to frontier, selection, and economic growth.” *Journal of the European Economic association*, 4(1): 37–74.

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. 2018. “Innovation, reallocation, and growth.” *American Economic Review*, 108(11): 3450–3491.

Adda, Jérôme, and Marco Ottaviani. 2024. “Grantmaking, grading on a curve, and the paradox of relative evaluation in nonmarkets.” *The Quarterly Journal of Economics*, 139(2): 1255–1319.

Akcigit, Ufuk, Harun Alp, and Michael Peters. 2021. “Lack of selection and limits to delegation: firm dynamics in developing countries.” *American Economic Review*, 111(1): 231–275.

Akcigit, Ufuk, Harun Alp, Jeremy Pearce, and Marta Prato. 2025. “Transformative and Subsistence Entrepreneurs: Origins and Impacts on Economic Growth.” *Working Paper*.

Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna. 2016. “Self-targeting: Evidence from a field experiment in Indonesia.” *Journal of Political Economy*, 124(2): 371–427.

Ali, Nadia, Massimiliano Calì, and Bob Rijkers. 2024. “Promoting Innovative Startups: Quasi-Experimental Evidence from Tunisia.” *Available at SSRN 5078981*.

Banerjee, Abhijit, Rema Hanna, Benjamin A Olken, and Diana Sverdlin Lisker. 2024. “Social protection in the developing world.” *Journal of Economic Literature*, 62(4): 1349–1421.

Bau, Natalie, and Adrien Matray. 2023. “Misallocation and capital market integration: Evidence from India.” *Econometrica*, 91(1): 67–106.

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis. 2021. “Contract labor and firm growth in india.” National Bureau of Economic Research.

Bloom, Nick, Rachel Griffith, and John Van Reenen. 2002. “Do R&D tax credits work? Evidence from a panel of countries 1979–1997.” *Journal of Public Economics*, 85(1): 1–31.

Brown, Alexander, François Chimits, and Gregor Sebastian. 2023. “Accelerator state: How China fosters “little giant” companies.” *Merics, Aug*, 3.

Census Bureau, U.S. 2021. “Statistics of U.S. Businesses (SUSB) Annual Data Tables 2018.” *https://www.census.gov/data/tables/2018/econ/susb/2018-susb-annual.html*.

Colonnelli, Emanuele, Bo Li, and Ernest Liu. 2024. "Investing with the government: A field experiment in China." *Journal of Political Economy*, 132(1): 248–294.

Colonnelli, Emanuele, Mario Cruz, Mariana Pereira-Lopez, Tommaso Porzio, and Chun Zhao. 2025. "Startups in Africa." Work in Progress.

De Haas, Ralph, Vincent Sterk, and Neeltje Van Horen. 2022. "Start-up types and macroeconomic performance in Europe."

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The secular decline in business dynamism in the US." *Unpublished draft, University of Maryland*, 3(2).

Eslava, Marcela, John Haltiwanger, and Álvaro Pinzón. 2022. "Job Creation in Colombia Versus the USA:'Up-or-out Dynamics' Meet 'The Life Cycle of Plants'." *Economica*, 89(355): 511–539.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta. 2024. "The size and life-cycle growth of plants: The role of productivity, demand, and wedges." *Review of Economic Studies*, 91(1): 259–300.

Goldschlag, Nathan, and Elisabeth Perlman. 2017. "Business dynamic statistics of innovative firms."

Goldschlag, Nathan, and Javier Miranda. 2020. "Business dynamics statistics of High Tech industries." *Journal of Economics & Management Strategy*, 29(1): 3–30.

Gonzalez-Uribe, Juanita, and Michael Leatherbee. 2018. "The effects of business accelerators on venture performance: Evidence from start-up Chile." *The Review of Financial Studies*, 31(4): 1566–1603.

González-Uribe, Juanita, and Santiago Reyes. 2021. "Identifying and boosting "Gazelles": Evidence from business accelerators." *Journal of Financial Economics*, 139(1): 260–287.

Goswami, Arti Grover, Denis Medvedev, and Ellen Olafsen. 2019. *High-growth firms: Facts, fiction, and policy options for emerging economies.* World Bank Publications.

Government of India. 2016. "Gazette Notification G.S.R. 180(E): Definition of "Startup"." *Gazette of India, Extraordinary, Part II—Section 3—Sub-section (i)*, Ministry of Commerce and Industry, Department of Industrial Policy and Promotion (DIPP); F. No. 5(91)/2015-BE. I. Published in the Official Gazette on February 18, 2016.

Government of India. 2017. "Gazette Notification G.S.R. 501(E): Revised Definition of "Startup"." *Gazette of India, Extraordinary, Part II—Section 3—Sub-section (i)*, Ministry of Commerce and Industry, Department of Industrial Policy and Promotion (DIPP). Issued in supersession of G.S.R. 180(E) dated February 17, 2016.

Government of India. 2018. "Gazette Notification G.S.R. 364(E): Definition of "Startup"." *Gazette of India, Extraordinary, Part II—Section 3—Sub-section (i)*, Ministry of Commerce and Industry, Department of Industrial Policy and Promotion (DIPP). Issued in supersession of G.S.R. 501(E) dated May 23, 2017.

Government of India. 2019. “Gazette Notification S.O. 127(E): Definition and Recognition of “Startup”.” *Gazette of India, Extraordinary, Part II—Section 3—Sub-section (i)*, Ministry of Commerce and Industry, Department for Promotion of Industry and Internal Trade (DPIIT). Amends earlier notifications including G.S.R. 364(E) dated April 11, 2018 and S.O. 34(E) dated January 16, 2019.

Gutiérrez, Germán, Callum Jones, and Thomas Philippon. 2019. “Entry costs and the macroeconomy.” National Bureau of Economic Research.

Guzman, Jorge, and Scott Stern. 2020. “The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US States, 1988–2014.” *American Economic Journal: Economic Policy*, 12(4): 212–243.

Hall, Bronwyn H, and Rosemarie Ham Ziedonis. 2001. “The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995.” *rand Journal of Economics*, 101–128.

Hsieh, Chang-Tai, and Peter J Klenow. 2014. “The life cycle of plants in India and Mexico.” *The Quarterly Journal of Economics*, 129(3): 1035–1084.

Hurst, Erik, and Benjamin Wild Pugsley. 2011. “What do small businesses do?” National Bureau of Economic Research.

ICR Report. 2021. “Startup Acts: An Emerging Instrument to Foster the Development of Innovative High-Growth Firms.” ICR Facility ICReport.

Khwaja, Asim Ijaz, and Atif Mian. 2005. “Do lenders favor politically connected firms? Rent provision in an emerging financial market.” *The quarterly journal of economics*, 120(4): 1371–1411.

Kim, J Daniel, Joonkyu Choi, Nathan Goldschlag, and John Haltiwanger. 2024. “High-growth firms in the United States: key trends and new data opportunities.”

Klette, Tor Jakob, and Samuel Kortum. 2004. “Innovating firms and aggregate innovation.” *Journal of political economy*, 112(5): 986–1018.

Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. “Transfer program complexity and the take-up of social benefits.” *American Economic Journal: Economic Policy*, 3(1): 54–90.

Krueger, Anne. 1974. “The political economy of the rent-seeking society.” *American Economic Review*, 64(3): 291–303.

Lerner, Josh, Junxi Liu, Jacob Moscona, and David Y Yang. 2024. “Appropriate entrepreneurship? the rise of china and the developing world.” National Bureau of Economic Research.

Martin, Leslie A, Shanthi Nataraj, and Ann E Harrison. 2017. "In with the big, out with the small: Removing small-scale reservations in India." *American Economic Review*, 107(2): 354–386.

McKenzie, David. 2017. "Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition." *American Economic Review*, 107(8): 2278–2307.

Nejad, Mohaddeseh Heydari. 2024. "A structural model of mentorship in startup accelerators: Matching, learning, and value creation." Univerity of Toronto Working Paper.

Peters, Michael, and Fabrizio Zilibotti. 2021. "Creative destruction, distance to frontier, and economic development." National Bureau of Economic Research.

Puri, Manju, and Rebecca Zarutskie. 2012. "On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms." *The Journal of Finance*, 67(6): 2247–2293.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local capture: evidence from a central government transfer program in Uganda." *The quarterly journal of economics*, 119(2): 679–705.

Rotemberg, Martin. 2019. "Equilibrium effects of firm subsidies." *American Economic Review*, 109(10): 3475–3513.

Sedláček, Petr, and Vincent Sterk. 2017. "The growth potential of startups over the business cycle." *American Economic Review*, 107(10): 3182–3210.

Sterk, Vincent, Petr Sedláček, and Benjamin Pugsley. 2021. "The nature of firm growth." *American Economic Review*, 111(2): 547–579.

The Economist. 2022. "Indian startups join the space race." *The Economist*.

The New York Times. 2023. "The Surprising Striver in the World's Space Business." *Alex Travelli*.

Wolken, Jordan. 2020. ""Startup Acts are the next form of policy innovation in Africa"."

A Policy Appendix

This Appendix documents additional details of the Startup India Program and similar programs in other developing economies.

A.1 Startup Policies in other Countries

Table 12 documents comparable startup support programs in both developing and developed economies that share key design features with the Startup India Program. A common pattern across developing countries is the use of startup labelling, a government certification, to identify high-growth and innovative entrepreneurs. These labels are then used as a gateway to allocate targeted resources, such as tax exemptions, subsidized credit, or fast-track regulatory support, to the certified firms. This approach reflects a broader trend of selective promotion in startup policies: governments attempt to identify and promote firms with high growth potential rather than subsidizing all entrants uniformly.

Table 12: Examples of other countries with Startup Labelling/Recognition Programs

Country	Program	Link
Low and Middle Income Country Examples		
Nigeria	Startup Label (Nigeria Startup Act)	startup.gov.ng
Tunisia	Startup Label (Startup Act)	startup.gov.tn
Senegal	Startup Label	der.sn
Brazil	Marco Legal das Startups	planalto.gov.br
Developed Country Examples		
Italy	Innovative Startup (Startup innovativa)	startup.registroimprese.it
Spain	Empresa emergente (Ley 28/2022)	enisa.es
Portugal	Startup/Scaleup Status Certificate	startupportugal.com
France	Jeune Entreprise Innovante (JEI)	service-public.fr

The table provides the details of similar programs in other developing and developed economies.

Several high-income countries, especially in Europe, where business dynamism has slowed, have introduced targeted startup programs over the past decade. In France, for example, the government classifies eligible firms into three categories: young innovative companies (JEI), young university companies (JEU), and young growth companies (JEC). Once designated under one of these labels, startups become eligible for a range of targeted benefits, including exemptions from profit, property, and business taxes, as well as waivers on employer social security and family allowance contributions.

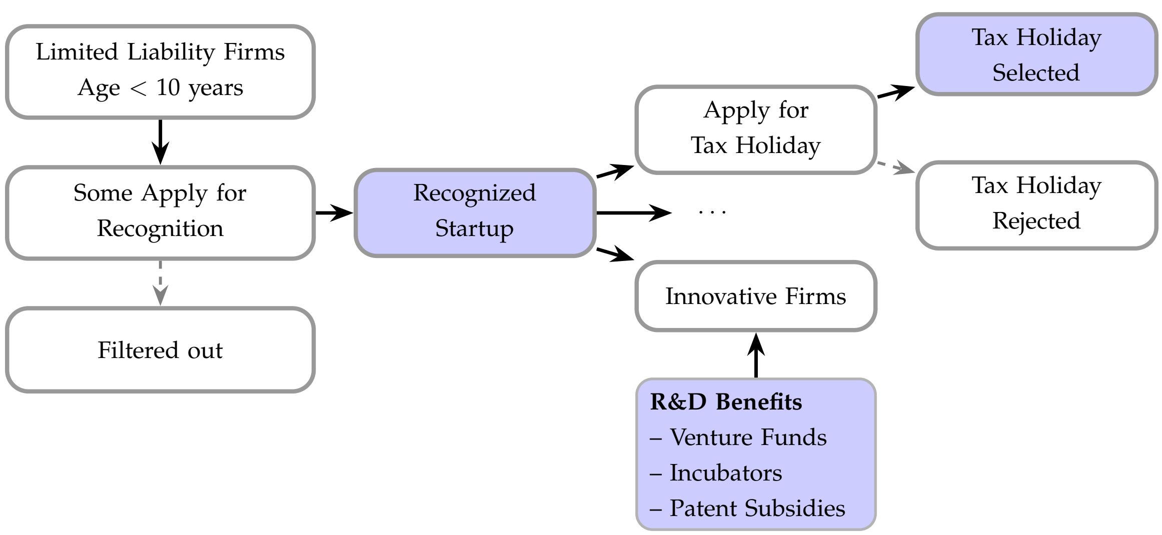
A.2 Startup India Policy Additional Details

The main criteria for selection into the program, as a recognized startup, outlined in the policy documents (Government of India, 2016, 2017, 2018, 2019) is as follows

> The startup is engaged in the innovation, development, or improvement of products, processes, or services or a scalable business model with a high potential for employment generation or wealth creation.

However, rather than applying this definition once and automatically extending all benefits to recognized startups, the policy is implemented in a multi-stage manner. Recognition makes startups eligible for a basic set of de facto benefits, but access to additional benefits requires them to undergo further screening and approval. Figure 9 summarizes the application process for the startups selected under the program. For the recognition step, the startup has to apply on an online portal and has to fill an online form providing a brief description about the business model of the firm, it granular geographic and sectoral information, education and background of all the firm founders, funding raised, any other achievements such as patents, competition and multiple questions about the self-reported innovation and scalability of the startup. These applications are then reviewed by the government and the decision is provided in roughly ten days from application. Conditional on selection in this stage, the startup can then apply to the successive stages of the policy which include tax holidays, incubation benefits, patent holidays and seed funds etc.

Figure 9: Application Process for the Startup India Program



Note:This figure summarizes the multi-stage screening process and the associated benefits under the Startup India Policy. The blue-shaded boxes indicate the core components of the program for which I collect data and use to estimate both the government's selection ability and the effective subsidy wedges.

B Data Appendix

This Appendix section provides the details of the data construction. I first describe the universe of limited liability firms master data. Then I describe the details of the data that I use to identify startups selected and rejected by the policy. Finally, I lay out the details of the process that I use to get the outcome variables for the firms such as income statements and patent applications.

B.1 Private Limited Firms Master Data

Indian Ministry of Corporate Affairs consolidates the identifying information, along with a few key variables, of all private limited firms in a publicly accessible dataset called the Master data[38]. The dataset is released as a cumulative file. Thus, the panel information for each firm is not directly available in the dataset. Table 13 provides the summary statistics for the main variables used in the paper from the 2024 vintage of the dataset[39]. There are about 1.88 million unique firms in the data, with about 63.3% being active in 2024. Large fraction of active firms results from the fact that historically firm formation rates were quite low and most of the firms in the dataset are formed after 2000 (Panel B). Majority of the firms in the data are unlisted (Panel C). Panel D summarizes the distribution of paidup capital.

[38]The data is available here: `https://www.data.gov.in/catalog/company-master-data`

[39]Each vintage overwrites the previous one

Table 13: Summary statistics of MCA Master Data

Panel A: Total and Active Firms	
Total Firm Observations	2,244,301
Unique firms (CIN)	1,886,069
Active firms	1,194,324
Active share	63.3%
Panel B: Incorporations by Year	
First incorporation year	1857
Before 1990	113,380
1990–1999	243,500
2000–2009	303,081
2010–2014	323,536
2015–2019	384,773
2020+	516,930
Panel C: Listed vs Unlisted Private Firms	
Unlisted	1,877,252
Listed Public Firms	6,189
Panel D: Paid-up Capital (INR)	
Mean	25,116,425
25^{th} percentile	100,000
50^{th} percentile (median)	100,000
75^{th} percentile	500,000
90^{th} percentile	3,992,910
95^{th} percentile	13,798,500
99^{th} percentile	120,000,000

Notes: This table provides the summary statistics of the Ministry of Corporate Affairs (MCA) Master data. The master data is a cumulative file released by the MCA with each vintage containing the latest status, address, paidup capital and other details of each firm. The summary statistics are from the 2024 snapshot of the data.

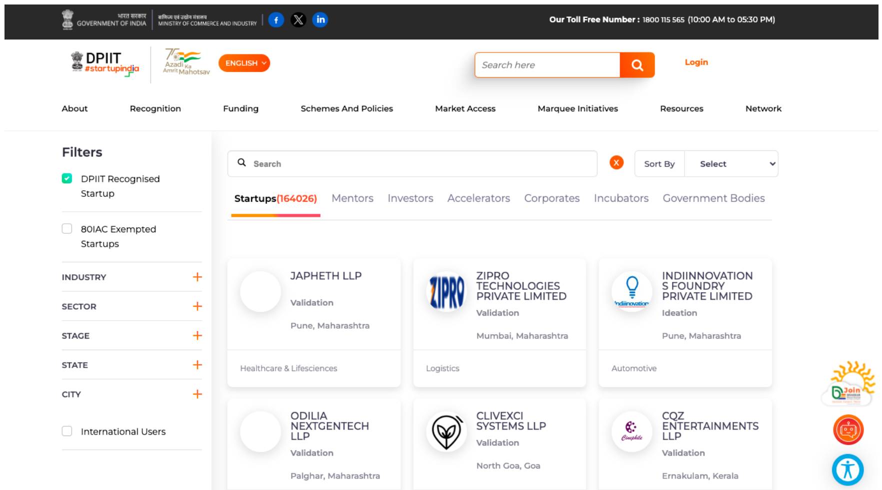
B.2 Startup Recognition Data

Under the Startup India Program, each private limited or limited liability partnership firm under ten years of age is eligible to apply for a recognition. The Department of Promotion of Industry and Internal Trade publishes a basic profiles of all the recognized startups on its website[40]. As the startups are required to register on this platform to be considered for recognition, they self report

[40] `https://www.startupindia.gov.in/content/sih/en/search.html?roles=Startup&page=0`

information about the founders and the details about the business idea and any achievements. Each startup profile also contains detailed sector information, stage of the startup at joining the platform and the date of recognition. I collect the information on all the startups on this platform with December 2024 as the latest recognition date in my dataset.

FIGURE 10: STARTUP RECOGNITIONS



Linking with Master Data and Summary Statistics The recognized firms can be either a LLP or a Private Limited firm. However, in the master data, I only observe private limited firms. Thus, I first exclude all the LLP firm from the recognition dataset and match the remaining firms with the master data. In addition to fuzzy matching on the startup name, I use the geographical information at the most granular available level (city or state) to narrow down the set of possible matches for each startup.

Dealing with the missing data The final data has 94,648 recognized firms over the period of 2016-2024. My matching exercise gives me a 70% match rate with a total of 66,890 firms matched with the MCA data. I manually check for a sample of un-matched firms in the MCA data and was unable to locate their data. To get their details, I then use the Tracxn data which accesses the same master data but is consistently reported using an API of the Ministry of Corporate Affairs. I take a random sample of 5,000 firms and get their details from Tracxn.

TABLE 14: SUMMARY STATISTICS OF DPIIT DATA AND MATCH WITH MCA MASTER DATA

Panel A: Total Firms and Firm Types	
Total Companies	112,950
LLPs	13,492
Private Ltd	94,648
Panel B: Startup Stage at Recognition	
Validation	40,316
Prototype	30,554
Early Traction	30,374
Scaling	11,880
Panel C: Recognition Status	
Recognised	112,898
Expired	50
Cancelled	2
Panel D: Top 10 Industries	
IT Services	12,701
Healthcare & Lifesciences	10,354
Education	7,180
Agriculture	6,418
Food & Beverages	6,088
Professional & Commercial Services	5,856
Construction	5,752
Technology Hardware	3,942
Finance Technology	3,709
Renewable Energy	3,228

Match Details	
Direct Matches with MCA Data	66,890
Matched through Tracxn Data	5,000
Total Firms	94,648

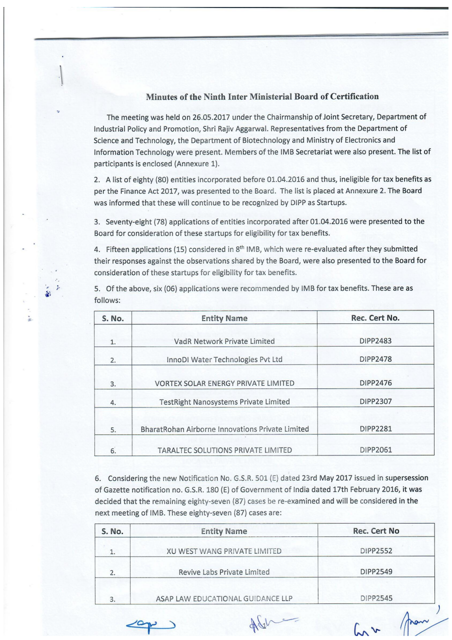
Notes: This table provides the summary statistics of the DPIIT Recognized firms data and its match with the MCA Master data. LLP stands for Limited Liability Partnerships and are not included in the matching due to inconsistent data availability.

B.3 Tax Exemption Data: Meeting Minutes of the Inter-Ministerial Board

Each recognized firm gets eligible to apply for a tax exemption. The tax exemption grants a holiday from corporate profit taxes for any three consecutive years within the first ten years. The applications for tax exemptions are considered by an Inter-Ministerial Board which meets monthly and deliberates on the profile of each startup and calling them for an interview or asking for additional details if required. Fig. 11 provides an example of the two minutes documents[41]. At the top each document notes the members of the board which are typically the secretary of the Industrial Policy Department and representatives from the technocratic government departments of biotechnology, science and technology, electronics and IT and National Research Development Corporation.

[41] These documents are publicly available at `https://www.startupindia.gov.in/content/sih/en/startupgov/imb.html`

Figure 11: Meeting Minutes Examples



Minutes of the Ninth Inter Ministerial Board of Certification

The meeting was held on 26.05.2017 under the Chairmanship of Joint Secretary, Department of Industrial Policy and Promotion, Shri Rajiv Aggarwal. Representatives from the Department of Science and Technology, the Department of Biotechnology and Ministry of Electronics and Information Technology were present. Members of the IMB Secretariat were also present. The list of participants is enclosed (Annexure 1).

2. A list of eighty (80) entities incorporated before 01.04.2016 and thus, ineligible for tax benefits as per the Finance Act 2017, was presented to the Board. The list is placed at Annexure 2. The Board was informed that these will continue to be recognized by DIPP as Startups.

3. Seventy-eight (78) applications of entities incorporated after 01.04.2016 were presented to the Board for consideration of these startups for eligibility for tax benefits.

4. Fifteen applications (15) considered in 8th IMB, which were re-evaluated after they submitted their responses against the observations shared by the Board, were also presented to the Board for consideration of these startups for eligibility for tax benefits.

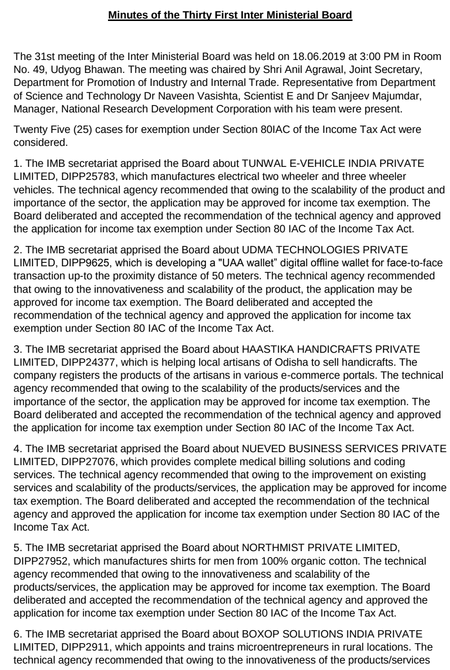
5. Of the above, six (06) applications were recommended by IMB for tax benefits. These are as follows:

S. No.	Entity Name	Rec. Cert No.
1.	VadR Network Private Limited	DIPP2483
2.	InnoDI Water Technologies Pvt Ltd	DIPP2478
3.	VORTEX SOLAR ENERGY PRIVATE LIMITED	DIPP2476
4.	TestRight Nanosystems Private Limited	DIPP2307
5.	BharatRohan Airborne Innovations Private Limited	DIPP2281
6.	TARALTEC SOLUTIONS PRIVATE LIMITED	DIPP2061

6. Considering the new Notification No. G.S.R. 501 (E) dated 23rd May 2017 issued in supersession of Gazette notification no. G.S.R. 180 (E) of Government of India dated 17th February 2016, it was decided that the remaining eighty-seven (87) cases be re-examined and will be considered in the next meeting of IMB. These eighty-seven (87) cases are:

S. No.	Entity Name	Rec. Cert No
1.	XU WEST WANG PRIVATE LIMITED	DIPP2552
2.	Revive Labs Private Limited	DIPP2549
3.	ASAP LAW EDUCATIONAL GUIDANCE LLP	DIPP2545

(a) Example 1 of Meeting Minutes



Minutes of the Thirty First Inter Ministerial Board

The 31st meeting of the Inter Ministerial Board was held on 18.06.2019 at 3:00 PM in Room No. 49, Udyog Bhawan. The meeting was chaired by Shri Anil Agrawal, Joint Secretary, Department for Promotion of Industry and Internal Trade. Representative from Department of Science and Technology Dr Naveen Vasishta, Scientist E and Dr Sanjeev Majumdar, Manager, National Research Development Corporation with his team were present.

Twenty Five (25) cases for exemption under Section 80IAC of the Income Tax Act were considered.

1. The IMB secretariat apprised the Board about TUNWAL E-VEHICLE INDIA PRIVATE LIMITED, DIPP25783, which manufactures electrical two wheeler and three wheeler vehicles. The technical agency recommended that owing to the scalability of the product and importance of the sector, the application may be approved for income tax exemption. The Board deliberated and accepted the recommendation of the technical agency and approved the application for income tax exemption under Section 80 IAC of the Income Tax Act.

2. The IMB secretariat apprised the Board about UDMA TECHNOLOGIES PRIVATE LIMITED, DIPP9625, which is developing a "UAA wallet" digital offline wallet for face-to-face transaction up-to the proximity distance of 50 meters. The technical agency recommended that owing to the innovativeness and scalability of the product, the application may be approved for income tax exemption. The Board deliberated and accepted the recommendation of the technical agency and approved the application for income tax exemption under Section 80 IAC of the Income Tax Act.

3. The IMB secretariat apprised the Board about HAASTIKA HANDICRAFTS PRIVATE LIMITED, DIPP24377, which is helping local artisans of Odisha to sell handicrafts. The company registers the products of the artisans in various e-commerce portals. The technical agency recommended that owing to the scalability of the products/services and the importance of the sector, the application may be approved for income tax exemption. The Board deliberated and accepted the recommendation of the technical agency and approved the application for income tax exemption under Section 80 IAC of the Income Tax Act.

4. The IMB secretariat apprised the Board about NUEVED BUSINESS SERVICES PRIVATE LIMITED, DIPP27076, which provides complete medical billing solutions and coding services. The technical agency recommended that owing to the improvement on existing services and scalability of the products/services, the application may be approved for income tax exemption. The Board deliberated and accepted the recommendation of the technical agency and approved the application for income tax exemption under Section 80 IAC of the Income Tax Act.

5. The IMB secretariat apprised the Board about NORTHMIST PRIVATE LIMITED, DIPP27952, which manufactures shirts for men from 100% organic cotton. The technical agency recommended that owing to the innovativeness and scalability of the products/services, the application may be approved for income tax exemption. The Board deliberated and accepted the recommendation of the technical agency and approved the application for income tax exemption under Section 80 IAC of the Income Tax Act.

6. The IMB secretariat apprised the Board about BOXOP SOLUTIONS INDIA PRIVATE LIMITED, DIPP2911, which appoints and trains microentrepreneurs in rural locations. The technical agency recommended that owing to the innovativeness of the products/services

(b) Example 2 of Meeting Minutes

Only firms incorporated after April 2016 and registered as private limited or limited liability partnerships are eligible for exemptions and the board sets the following broad outline as a criteria for getting selected

> The startup is engaged in the innovation, development, or improvement of products, processes, or services or a scalable business model with a high potential for employment generation or wealth creation.

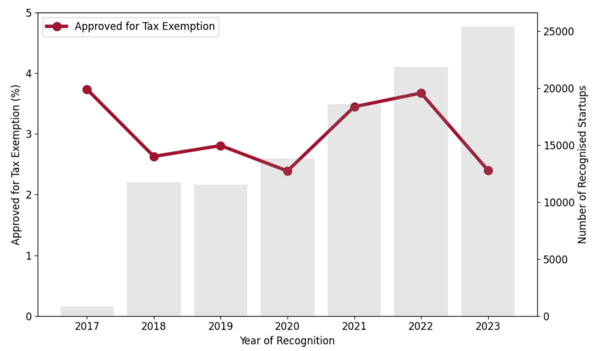
The first meeting was held in May 2016 and I digitize the meeting minutes from the first forty board meetings with the last meeting being in December 2019. This helps to avoid the Covid-19 related disruptions but more importantly allows to me to track firm outcomes for the subsequent years post-policy.

Extracting Firm Names and Decisions I use a simple OCR procedure to extract the firm names, DIPP Numbers, and decision from each document. I manually tweak the OCR process to adjust for different formats of different documents which vary substantially during the first forty meetings. In each meeting, a given firm gets one of the four decisions: Accepted, Rejected, Deferred, Ineligible. A deferral implies that the board contacted the firm for additional information and the decision was not granted in that particular meeting. Under this case, I track the firm in all the subsequent meetings and classify it as accepted if I find an acceptance decision and rejected if either there was an explicit rejection or if the firm is never mentioned again in the subsequent minutes. Similarly, a firm is classified as ineligible using the Annexure tables of each board meeting which list all the firms that were rejected because they were simply formed before April 2016 and thus ineligible for a tax exemption.

Matching with the Startup Recognition and MCA Data Each firm in the meeting data is associated with an Department of Industrial Policy's identification number called DIPP Number. I directly use this number to match the firms with the recognized firms and the MCA-Recgnition data matching to get their other identifying information.

Summary Statistics Table 12b provides the summary statistics of the firms extracted from the digitized meeting minutes. The final data has 2,981 firms out of which 233 were granted a tax exemption. Around 87% of these firms were formed in years 2016 and 2017 and hence almost all the startups are less than three years of age at the time of application. Fig. 12a plots the total recognized firms by each year (right y-axis) and a fraction of those firms which get tax exemptions (left y-axis). On average only around 3-4% of the recognized firms get selected for tax exemptions.

Figure 12: Summary Statistics for Tax Exempt Firms



(a) Fraction Granted Tax Holidays

(b) Summary statistics of DPIIT Startup Dataset

Panel A: Coverage	
Total Startups	2,981
Selected Startups (Decision=1)	233
Ineligible Startups	320
Panel B: Year of Formation	
2015	258
2016	1,204
2017	1,400
2018	102
2019	17
Panel C: Top 5 States	
Maharashtra	602
Delhi	469
Karnataka	421
Uttar Pradesh	209
Gujarat	170

Notes: This figure and table provide the summary statistics for the firms that applied for tax exemptions along with the decisions of the Inter-Ministerial Board. Ineligible means that the Startup was formed before 1^{st} April 2016 and is thus ineligible for a tax holiday.

B.4 Income Statements and Balance Sheet Data

Registered firms in India are legally required to file to annual documents with the Ministry of Corporate Affairs. These filings contain two Forms: AOC-4 and MGT-7[42].

Form AOC 4 According to Section 137(1) of the Companies Act, 2013, all companies must submit their financial statements, along with consolidated statements and approved documents from the annual general meeting (AGM), to the Registrar within thirty days after the AGM. It contains information on balance sheets, profit/loss income statement, cash flow etc.

Form MGT 7 this document summarises a company's shareholder makeup, directorship, capital structure, and important governance actions during the previous financial year. Unlike financial statements, which provide a breakdown of the company's financial performance. The annual report offers stakeholders a comprehensive overview of the company's activities, financial structure, ownership details, governance framework, executive compensations, and other important non-financial elements at the end of the fiscal year.

I collect these forms for all the firms in my data that applied for an income tax exemption regardless of their decision outcome. The data thus tackles the selection issues common in datasets with venture capital funding or startups from developing countries which tend to capture only the firms with ex-post positive outcomes. As one the main exercises in the paper is to understand how well the government is able to target high growth firms, it is essential to gather the full information about all the applicant firms.

Summary Statistics Out of the total 2,981 startups that applied for tax exemptions. We were able to get the income and balance sheet filings of 2,215 firms. The startups usually face large failure rates. While some firms make it to the stage where the start making revenues and thus have to file the income statements, others don't. As a result for around 24% of the firms we were not able to find any corresponding annual filings at the Ministry of Corporate Affairs portal. The remaining 76% of the startup firms, however, had corresponding income filings which were collected and harmonized to get the final dataset. Table 15 summarizes the key variables and their corresponding distributions in the two sub data samples. The mean revenue of the firms is around 23 million INR (roughly 260,000 USD) annually. The median, however, is much lower and close to 16,000 USD. This reflects the up-or-out dynamics of the young firms wherein a large majority of them fail or don't grow but a few are succesful and end up growing large. The average profit before tax, on the other hand, is negative and only the firm-year observations in the top 25^{th} percentile have positive profits. The top 1 percentile of the firms have profit before taxes equal to 19 million INR (roughly 214,000 USD) annually. It is an important outcome variable for the analysis as the tax holidays are provided on the profits earned by the startups but these can

[42] `https://www.datatracks.com/in/blog/understanding-aoc-4-and-mgt-7-filings/`

e quite low as reflected in the Table 15 as even high growth startups might retain their earnings within the firm for re-investment or to avoid tax expenses.

In terms of their balance sheets, the median startup firm in the sample has 1,772,000 INR assets (roughly 19,000 USD) in total assets, 1.7 million INR in liabilities, and 76,500 INR in equity. The assets of the firm in the balance sheet are further divided into two broad categories, current and non-current assets. Non-current assets include fixed assets (tangible, intangible), non-current investments, deferred tax and long term loans/advances; while current assets includes inventories, cash, trade receivables etc. To construct the sales per asset measure, I use the fixed asset as the denominator and the last row of Table 15 describes its summary statistics.

Table 15: Summary statistics of startup financial data

Variable	Obs	Mean	Median	p25	p75	p99
Panel A: Income Statement						
Revenue (Rs. 000s)	11,383	23,047.1	1,442.4	89.7	9,371.4	384,147.6
PBT (Rs. 000s)	11,299	-3,677.4	-53.0	-1,024.6	53.1	22,747.9
Net Profit (Rs. 000s)	10,824	-2,697.2	-32.5	-685.7	51.5	19,791.4
Unique Firms	2,053					
Panel B: Balance Sheet						
Total Assets (Rs. 000s)	12,143	26,694.2	1,772.5	287.2	9,359.2	324,767.9
Total Liabilities (Rs. 000s)	10,670	25,112.7	1,703.1	298.7	6,756.0	154,655.7
Total Equity (Rs. 000s)	11,839	10,213.9	76.5	-460.7	1,200.8	219,978.9
Fixed Assets (Rs. 000s)	10,880	7,903.4	416.3	36.8	2,830.6	106,445.2
Unique Firms	2,215					

Notes: All values are in INR thousands. Obs refers to firm-year observations.

B.5 Patent Application Data

The details of the patent applications and grants in India is released in two main formats: Weekly Patent Applications Published/Granted[43] and through the Indian Patent Advanced Search System (inPASS)[44]. The first format publishes the publication number, office location, field of invention and title of invention for each patent application filed and granted in a week. However, it does not provide any details on the inventors, assignees and their firms or institutions. The inPASS portal, on the other hand, provides the complete specification for each application/grant and can be searched by querying the applications through application number, date of application or inventors etc.

[43] This data is accessible here: `https://www.data.gov.in/resource/weekly-patent-application-published`
[44] This data can be accessed here: `https://iprsearch.ipindia.gov.in/PublicSearch/`

I use the inPASS portal to search for all the patent applications between 2012-22. I focus on patent applications as there is usually a significant processing time between the application of the patent and its eventual grant. Given the limited number of years since the policy launch in 2016, focusing on applications gives me a longer horizon for analysis. Moreover, the policy explicitly subsidizes patent applications thus making it a natural outcome variable for the analysis. Given the patent application numbers, I then scrape the entire patent application for each patent at a time and use it to construct my final dataset of 434,262 patent applications over the periods of 2012-22.

A vast majority of these patent applications are, however, by individual inventors, academic institutions or government organizations. In order to narrow down the patent applications to only those of the Indian firms and startups, I query the applicant names for firm identifiers such as 'PVT LTD', 'PRIVATE LIMITED', 'PRIVATE', 'PVT', 'LTD', 'LIMITED', 'LLP', 'LLC', 'INC', 'CORP', 'CORPORATION', 'CO', 'COMPANY' etc. I then use a fuzzy matching algorithm to match the applicant names with the MCA data described in Section B.1. This process gives me a match of total 47,668 patent applications which are filed by firms in the Ministry of Corporate Affairs database.

Table 16 provides the basic summary statistics of the matched patent-firm data. The total number of unique firms matched with the MCA data is 4,864 with on average about 9 patent applications per firm. As documented well in the US context (Goldschlag and Perlman, 2017), the patenting firms in India also are generally older. The median age of the firm computed using the MCA Corporate Identification Numbers is 23 years. However, young firms represent a significant share of the total applications. Firms less than ten years account for 27.6% of the total patent applications whereas firms less than 5 years of age account for 19.6% of the total applications. The fields of invention, in the sample period, for both young and small firms is, however, quite similar. Both the groups have mechanical and chemical fields in the top two, followed by computer science and electrical.

Table 16: Patents: Young vs. Old Firms — Coverage, Distributions, and Top Fields

	Obs	Mean	Median	p25	p75	p99
Panel A: Coverage						
Total applications	47,668					
Unique firms (CIN)	4,864					
Patents per firm	4,864	9.8	1.0	1.0	3.0	150.0
Age at patent	47,389	27.5	23.0	9.0	43.0	86.0
Share of patents by young firms ($\leq$10y)		27.6%				
Share of patents by very young firms ($\leq$5y)		19.6%				
Panel D: Top 5 Fields of Invention — Young Firms ($\leq$10y)						
CHEMICAL	2,631					
MECHANICAL ENGINEERING	2,075					
COMPUTER SCIENCE	2,019					
PHARMACEUTICALS	1,229					
COMMUNICATION	968					
Panel E: Top 5 Fields of Invention — Old Firms (>10y)						
MECHANICAL ENGINEERING	9,489					
CHEMICAL	6,676					
COMPUTER SCIENCE	4,024					
ELECTRICAL	2,921					
COMMUNICATION	1,888					

Notes: Young firms are defined by age at patent $\leq$ 10 years; old firms by $>$ 10 years. Shares are fractions of all applications. Counts are application-level unless noted.

C Empirical Appendix

This appendix section provides additional results and robustness checks for the empirical findings section of the paper. The results are organized by the three policy instruments tax holidays, R&D benefits and recognitions.

C.1 Tax Holidays

Conditional on application, the selection rates for tax holidays is 7.8% as shown in Table 17. There are also substantial differences in the startups selected and rejected by the bureaucratic board as reported in Table 18.

Table 17: Tax Exemption Selection Summary Statistics

Variable	Value
Total Applicants	2,981
Approval Rate	7.8%
Firms with Income Filings	2,279

This table reports the total number of startups that applied for tax holidays, the number of startups that were approved and the total number of startups for which income statements were collected.

Table 18: Summary Statistics by Decision (USD 000's)

	Selected			Rejected		
Variable	Mean	p25	p95	Mean	p25	p95
Revenue	423.2	9.7	1,866.0	255.5	0.8	983.9
PBT	-30.9	-19.1	180.7	-44.6	-11.4	35.6
Net Profit	-25.8	-13.7	171.5	-32.4	-7.7	28.7
Total Assets	664.8	16.0	1,782.2	279.2	2.9	859.5
Total Liabilities	186.3	11.2	761.8	305.3	3.2	545.6
Total Equity	414.0	-2.5	1,048.2	92.0	-5.7	318.3
Non-current Assets	174.6	2.2	423.0	84.3	0.4	329.6
Observations (Firm-Year)	1,084			10,584		

The table reports the summary statistics of the selected and rejected firms for tax holidays. Observations are at the firm-year level.

C.1.1 Matching Details

Fig. 13 summarizes the policy announcement and the meeting minutes that were digitized. The first meeting was held in May 2016 and in the subsequent meetings only startups formed after

April 2016 were eligible to get a tax holiday. I collect the data for all the firms that applied for tax holidays between 2016 and 2019.

Figure 13: Timeline of policy announcement and board meetings

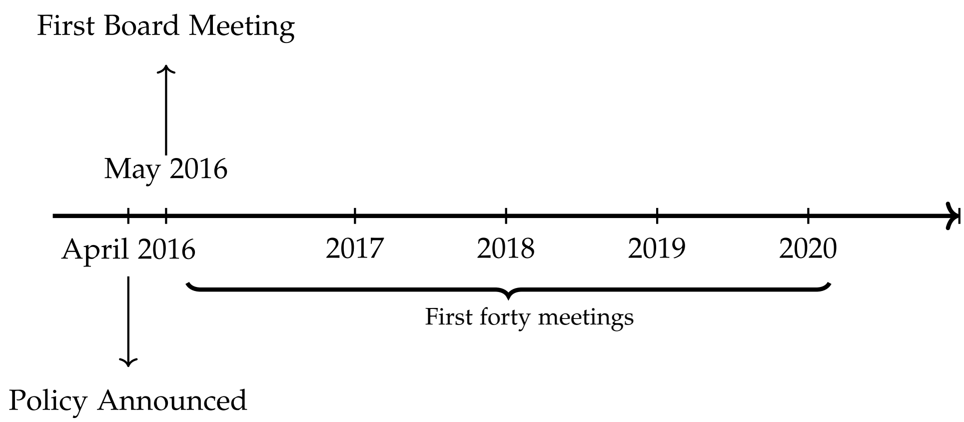


Table 19 provides the details of the variables used for matching startups and the corresponding balance. I match the startups based on i) the level and growth in revenues and assets prior to the meeting, ii) whether VC investment at the same age, and iii) whether they were incubated or not. I also similarly match the firms to a set of all rejected firms instead of just the ones rejected due to age-of-formation cutoff.

Table 19: Baseline Balance: Treated vs Control (first observation per firm)

Variable	Treated	Control		
Variable	Mean	Mean	Diff	Std. Diff
Log Revenue (Age 1)	13.680	13.749	-0.069	-0.031
Log Assets (Age 1)	14.055	13.990	0.065	0.034
Growth Log Revenue (2 yrs)	0.976	0.942	0.034	0.014
Growth Log Assets (2 yrs)	0.920	0.849	0.071	0.047
IHS(Equity, Age 1)	7.968	9.072	-1.103	-0.174
Age at Meeting	2.755	2.529	0.226	0.313
VC Investment (Pre-Exemption)	0.135	0.108	0.027	0.083
VC Investment (Post)	0.271	0.268	0.003	0.008
Incubated	0.271	0.210	0.061	0.142
Eligible	1.000	0.000	1.000	–
N (firms)	155	157		

The second and third column of the table report the means of the each of the variable in the treated and the control groups.

C.1.2 Firm Level Control Variables for Matching and Selection Regressions

I use Venture Capital funding and Incubator details to match startups. I also construct an additional set of explanatory variables to run the regression on the key determinants of selection. The additional variables constructed and their details are as follows

- **Paid-up Capital** is defined as the total shareholder equity capital in the firm reported to the Ministry of Corporate Affairs.
- **City** is determined from the granular address of the firm in the MCA database. I further classify the cities as Tier 1 if the city is among the ten largest cities in India.
- **State** is also determined from the address in the MCA data.
- **DPIIT Sector** is the sectoral classification used by the DPIIT. It differs from the NIC8 classification and provides more granular level of classification suited for the startups.
- **DPIIT Industry** is an aggregated version of the sector reported by the DPIIT
- **Total Funding Raised before Meeting** is defined as the sum total of of funding raised from all the funding rounds as reported by Tracxn.
- **Venture Capital Investment Before Exemption** denotes the total pre-meeting funding that is classified as VC funding by Tracxn.
- **Angel Investment Before Exemption** denotes the total pre-meeting funding that the startup raised from angel investors.
- **Private Incubated** is constructed from the incubator details for a startup reported in Tracxn. I manually classify the incubators as private based on their website.
- **Government Incubated** is constructed from the incubator details for a startup reported in Tracxn. I manually classify the incubators as government based on their website.
- **Size Founding Team** is the total founding team size as reported by Tracxn.
- **Education Level of Founding Team** is constructed using either the university names reported in Tracxn or from the self-reported details on the Startup Portal. I use LLM prompts to extract educational qualifications from the portal where ever reported and manually verify against a sample of firms.
- **Average Experience Founding Team** is also constructed from either Tracxn or from using LLM prompts to extract from the portal when reported.
- **Top School Founding Team** is classified as top school if the school features in top university rankings.

C.2 Rejected due to Age of Formation Cut-Off

Fig. 14 shows an example of the startups that were rejected for the tax holidays because they were formed after prior to the cut-off date of 1^{st} April 2016. The third column reports their actual date of registration and the last column provides the reason as reported by the bureaucratic board.

FIGURE 14: STARTUPS REJECTED DUE TO AGE OF FORMATION CUT-OFF IN THE MEETING MINUTES

28.	DIPP2827	Agrilife Technologies Private Limited	18-12-2015	bala@agribolo.com	Registered before 01.04.2016
29.	DIPP2833	Bellatrix Aerospace Private Limited	04-02-2015	yashas@bellatrixaerospace.com	Registered before 01.04.2016
30.	DIPP2839	Cleanse Solutions India Pvt Ltd	16-01-2015	arnavshah@bikecleanse.com	Registered before 01.04.2016
31.	DIPP2840	Wegot Utility Solutions Private Limited	01-04-2015	abilash@wegot.in	Registered before 01.04.2016
32.	DIPP2843	ZONTA ENVIRONMENT PRIVATE LIMITED	13-08-2013	compliance@zontainfratech.com	Registered before 01.04.2016

C.2.1 Reported Tax Expenses

The following table reports the differences in tax rates when the startups make positive profits. The selected startups have on average 4% lower tax rates in the first eight years conditional on making profits.

TABLE 20: EFFECTIVE PROFIT TAX SUBSIDY FOR SELECTED FIRMS

	Average Tax Rate on Profits		
	OLS		
	(1)	(2)	(3)
Tax Exempted	0.218***		
	(0.008)		
Rejected		0.264***	
		(0.003)	
Difference			-0.046***
			(0.009)
Observations	2,358	2,358	2,358
R^2	0.011	0.011	0.011
Note:	*p<0.05; **p<0.01; ***p<0.001		

C.2.2 What explains selection by government?

In this section, I use the startup characteristics created in Section C.1.2 to study the selection of firms by the bureaucratic board. Table 21 reports the coefficient estimates from a linear probability model of the selection by the bureaucratic board. Each column successively adds city, sector and board meeting fixed effects respectively. Venture Capital Investment before the meeting, Incubation and founder experience are the three variables which significantly predict selection by the bureaucratic board.

Table 21: Selection by Bureaucratic Board

Dependent Variable:	Decision			
Model:	(1)	(2)	(3)	(4)
Variables				
LLP	0.0985	0.1229	0.1164	0.1030*
	(0.0754)	(0.0789)	(0.0990)	(0.0543)
VC Investment Before Exemption	0.1743***	0.1933***	0.2116***	0.1519***
	(0.0537)	(0.0583)	(0.0550)	(0.0508)
Angel Investment Before Exemption	-0.0452	-0.0437	-0.0415	-0.0398
	(0.0471)	(0.0311)	(0.0271)	(0.0310)
Masters	-0.0306	-0.0138	-0.0127	-0.0055
	(0.0290)	(0.0333)	(0.0365)	(0.0227)
PhD	0.0266	0.0509	0.0123	-0.0013
	(0.0596)	(0.0684)	(0.0644)	(0.0666)
Founder Experience	0.0053***	0.0063*	0.0060*	0.0032
	(0.0020)	(0.0032)	(0.0033)	(0.0025)
Top School	0.0249	0.0274	0.0338	0.0082
	(0.0350)	(0.0319)	(0.0308)	(0.0247)
Incubated	0.0607*	0.0936***	0.0901**	0.0580**
	(0.0312)	(0.0351)	(0.0363)	(0.0227)
Environment	0.1933**	0.2066*		
	(0.0884)	(0.1078)		
Health	-0.0157	-0.0106		
	(0.0544)	(0.0736)		
Education	-0.0046	-0.0186		
	(0.0551)	(0.0632)		
Agri	0.0126	-0.0074		
	(0.0449)	(0.0527)		
Aero	0.3890	0.3830		
	(0.2705)	(0.2645)		
Non Major City	0.0412			
	(0.0370)			
Fixed-effects				
City		Yes	Yes	Yes
Sector			Yes	Yes
Board Meeting				Yes
Fit statistics				
Observations	861	861	861	861
R^2	0.04390	0.20249	0.25052	0.61118

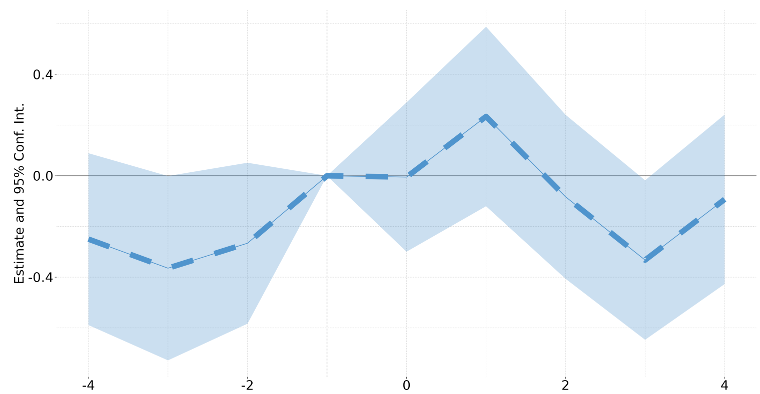
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

C.2.3 Intensive Effect of Recognition on Firms Patent Applications

The following provides the results from the following firm level specification

$$\text{Log Patent Applications}_{ia} = \delta_i + \delta_a + \sum_{\tau=-4}^{4} \alpha_\tau (D^\tau \times \mathbf{1}[\text{Startup Recognition}_{ia}]) + \epsilon_{ia} \tag{25}$$

FIGURE 15: PRE-POST POLICY WITH FIRM FIXED EFFECTS



Note: The figure plots the coefficient α_τ from the regression specification in Eq. 25. The coefficient measures the intensive effect of a recognition on startup's patents prior and post the policy.

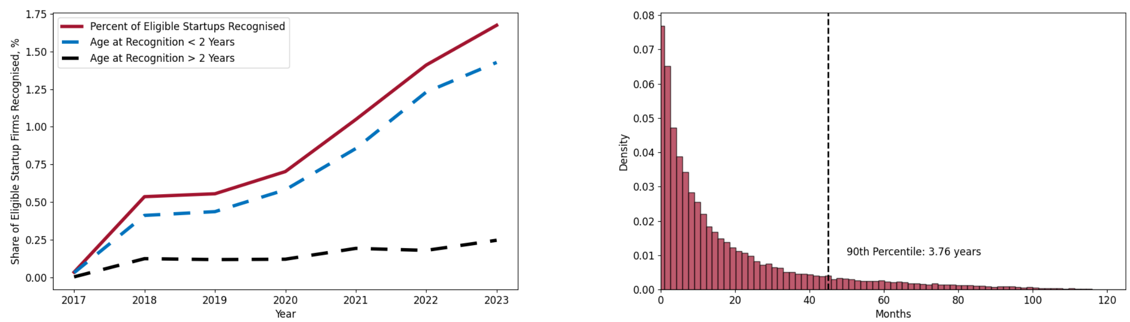
C.3 Recognitions

The aggregate impact of a policy that selectively promotes some firms depends on the total number of firms selected and the length/size of the benefits that the firms receive. Fig. 16a shows that around 1-1.5% of the eligible firms (age less than ten years) get DPIIT recognition each year and the rates have been increasing over time. A large majority of the firms that get recognized are aged less than two years (dashed blue line). As there is a strict age based phasing-out of the policy there should be no firms older than ten years who get recognized and the firms should be incentivized to get recognized early. Fig. 16b shows that this is indeed the case, with median age at recognition of seven months and the 90th percentile of 3.76 years and no firms over ten years of age.

Are the selection rates among eligible firms low due to strong filtering of the applicants or due to low application rates? There is a large degree of self-selection in the policy as only two percent of the eligible firms apply for recognition. [45]

[45] Some startups are either not aware or the effort costs for them are too high so they self-select themselves out of the policy. Indiscriminate and perpetual give-away of exemptions and policies could have large impacts on government expenditure. Hence literature (Alatas et al., 2016; Banerjee et al., 2024), for example, argue that adding small application costs could lead to large benefits to government by filtering out applicants. However, a startup could be non-applicant either because it got deterred by the small application cost or it didn't have the information about the policy to begin with.

Figure 16: Firm Selection Rates and Age at Selection



(a) Selection Rate over Time

(b) Selection by Age

C.3.1 Summary Statistics on Firm Survival Status

The following table provides the counts of different firm status of recognized and non-recognized firms in 2014. I use the firm status categories denoting dissolved, inactive, strike-off and liquidated to determine the survival of a firm i.e. if a firm status is in any of these categories, the firm is classified to have exited from the economy.

Table 22: Firm Status by Recognized and Non-Recognized Firms

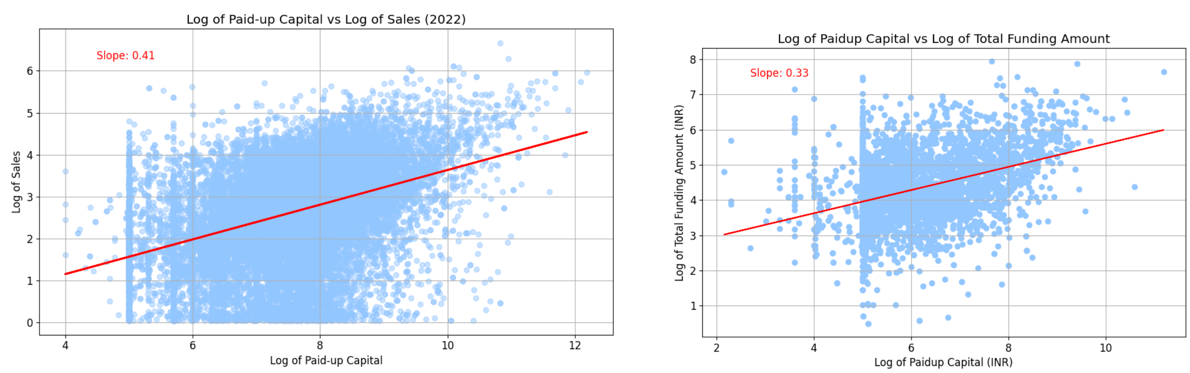
Firm Type Firm Status	Non-Recognized	Recognized
Active	1,340,705	78,919
Amalgamated	28,180	11
Converted to LLP	19,164	17
Dissolved (Liquidated)	12,304	3
Dissolved under section 54	74	0
Dissolved under section 59(8)	122	1
Dormant under section 455	2,209	29
Inactive for e-filing	9,124	12
Strike Off	720,502	509
Strike Off-Awaiting Publication	403	9
Unclassified	1	0
Under CIRP	1558	9
Under Liquidation	7508	25
Under process of striking off	22,758	148
Vanished	1	0

Note: The table provides the summary statistics for the firm status used to determine survival from the Ministry of Corporate Affairs database snapshot in 2024.

C.3.2 Paid Up Capital Summary Statistics

To establish the use of paidup capital as an informative measure of firm size, I compare the firm's paidup capital with their sales and venture capital funding levels using the Prowess and the PitchBook datasets. While Prowess data already contains the information on the paidup capital of each firm, I merge the firms in the PitchBook data to the MCA data using the same fuzzy matching process described in the Data Appendix A. Figures 17a and 17b plot the resulting correlation between the two sets of variables. The results suggest that both of the variables are strongly correlated with paidup capital, thus making it an informative measure of firm size for private limited firms.

FIGURE 17



(A) PAID-UP CAPITAL VS SALES

(B) PAID-UP CAPITAL VS TOTAL VC FUNDING

Notes: The panel A plots the correlation of sales with paid-up capital using the firms in the Prowess dataset. Panel B plot the relationship between venture capital financing raised by a startup and its paidup capital using the Pitchbook dataset.

C.4 Relationship between Recognition Probability and Entry

This appendix section provides additional details for the regressions in Section 4.2 which estimate the relationship between recognition probability and the entry in the subsequent period.

C.4.1 Summary Statistics

I use the MCA data to get the granular address of each firm in the data and use the pincode to map it to around 780 districts in India. Tables 23 and 24 provide the summary statistics and the distribution of the two main variables. The summary tables show the fact both the firm formations and recognitions are concentrated in larger urban districts with number of private limited formations in a district around 140 in 2017 and 238 in 2022.

Table 23: Summary statistics of the data for t = 2017

	N	mean	std	min	25%	50%	75%	max
Formations	575	137	448	0.00	13	27	82	6116
Recognitions	575	1	4	0	0	0	0	82
$\frac{\text{Recognitions}_t}{\sum_{k\in\{t,t-1,t-2\}}\text{Formations}_k}$	575	0.15%	0.45%	0%	0%	0%	0%	3.03%

The table provides the summary statistics of firm formations and recognitions across different districts in 2017.

Table 24: Summary statistics of the data for t = 2022

	N	mean	std	min	25%	50%	75%	max
Formations	572	238	660	0	35	68	165	8594
Recognitions	572	25	82	0	2	6	15	1203
$\frac{\text{Recognitions}_t}{\sum_{k\in\{t,t-1,t-2\}}\text{Formations}_k}$	572	2.93%	1.61%	0%	1.92%	3.02%	3.90%	6.78%

The table provides the summary statistics of firm formations and recognitions across different districts in 2017.

Table 25: Relationship between recognitions and entry

Dependent Variable:	Log New Firm Formations$_t$			
Model:	Overall	Large Districts (Top 5 percentile)	Large Districts (Top 10 percentile)	Small Districts (Bottom 90 percentile)
$\frac{\text{Recognitions}_{t-1}}{\sum_{k\in\{t-1,t-2,t-3\}} \text{Formations}_k} \times 100$	-0.0094	0.0688***	0.0255*	-0.0024
	(0.0074)	(0.0161)	(0.0131)	(0.0051)
Fixed-effects				
Year	Yes	Yes	Yes	Yes
District	Yes	Yes	Yes	Yes
Mean values				
$\frac{\text{Recognitions}_{t-1}}{\sum_{k\in\{t-1,t-2,t-3\}} \text{Formations}_k} \times 100$	1.35%	2.08%	1.99%	1.27%
New Firm Formations$_t$	180.10	1708.76	1061.39	61.69
Fit statistics				
Observations	4,553	280	540	4,013
R^2	0.97	0.99	0.98	0.94
Within R^2	0.00146	0.12454	0.01262	0.00010

Clustered (Year) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 26: Recognition probability and entry of new firms that apply for recognition

Dependent Variable:	Log New Recognized Firm Formation$_t$			
Model:	Overall	Large Districts (Top 5 percentile)	Large Districts (Top 10 percentile)	Small Districts (Bottom 90 percentile)
Variables				
$\frac{\text{Recognitions}_{t-1}}{\sum_{k\in\{t-1,t-2,t-3\}} \text{Formations}_k} \times 100$	0.0411	0.1595**	0.0973*	0.0135
	(0.0424)	(0.0544)	(0.0515)	(0.0273)
Mean values				
$\frac{\text{Recognitions}_{t-1}}{\sum_{k\in\{t-1,t-2,t-3\}} \text{Formations}_k} \times 100$	1.35%	2.08%	1.99%	1.27%
New Firm Formations$_t$	9.11	93.21	57.29	2.63
Fixed-effects				
Year	Yes	Yes	Yes	Yes
District	Yes	Yes	Yes	Yes
Fit statistics				
Observations	4,553	280	540	4,013
R^2	0.82935	0.97696	0.95511	0.73844
Within R^2	0.00751	0.09284	0.02768	0.00109

Clustered (Year) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

D Model Calibration and Quantification

D.1 Terminal BGP

- **Share of R&D firms** Goldschlag and Perlman (2017) report the number of patenting firms in the US to be 27,800 which amounts to be out 0.05% of the total firms in the US or around 2.79% of the C-corporations. To account for other kinds of research activity not captured by patents such as software copyrights, trademarks etc, I target the share of R&D firms to be 5%.

- **Average size at age 10 for R&D firms** Goldschlag and Perlman (2017) report that the R&D firms are substantially larger than other firms, with 2.1% of patenting firms having more than 5,000 employees compared to only 0.03% of other firms. Akcigit et al. (2025) report an average size of 7 relative to birth for firms that engage in R&D. I use this as an target for the model.

- **Average size at age 10 for non-R&D firms** Hsieh and Klenow (2014) report an average size of 2 relative to birth for the US manufacturing plants. Akcigit et al. (2025) use Danish micro data and also report a size of 2 for firms that don't hire any R&D workers.

- **Share of firms in smallest size bin** In order to make an exact comparison with Private Limited Firms I use the share of firms in the lowest size bin for the US C-Corporations from the Statistics of US Businesses.[46]. Using the 2015 and 2018 tables, the share of the C-Corporation firms in the lowest size bin (i.e. <5 employees) is 55%.

- **Share of high growth firms** Kim et al. (2024) use the US Census data to define high-growth firms as firms with DHS growth rate of greater than 0.8 or correspondingly a growth rate greater than 130% in a given year. Using this definition, they report a share of high growth continuing firms to be around 5% (though steadily declining over time). This figure is for all firm types and the data is not available split by legal form of organization. I thus use it directly as a target for share of high growth firms in the non-R&D category.

D.2 Transition Calibration

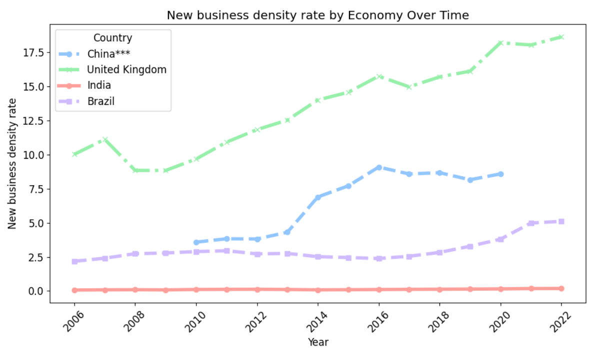
D.2.1 Transition Moments Construction

Entry Rates Limited Liability firm formation rates in India are extremely low. Fig. 18a in Appendix D.2.1 plots the formal firm formation rates for UK, China, Brazil and India from the World Bank Entrepreneurship Database. India's firm creation rate is around $1/60^{th}$ to that of China and $1/80^{th}$ to that of UK per 1000 people. Over time, however, entry rates have improved as shown in Fig 18b. Entry remained low throughout the early 90s after the liberalization and it

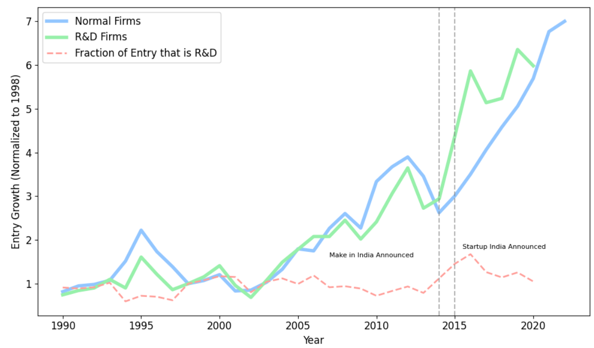
[46]The data split by Legal Form of Organization (LFO) is available here: `https://www.census.gov/data/tables/2015/econ/susb/2015-susb-annual.html`. Business Dynamics Statistic do not report data by LFO thus SUSB is the closest available substitute data

only started to increase in early 2000s. The share of R&D firms among entrants also stayed stable throughout this period but increased after 2016.

Figure 18: Entrepreneurship Rates and Entry Growth



(A) Business Formations per 1000 population

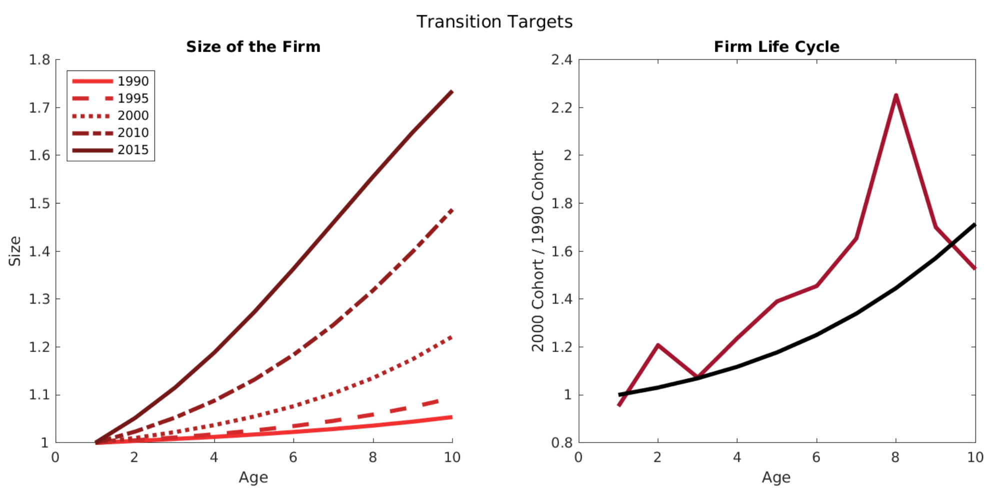


(B) Entry Growth of R&D and Non-R&D Firms

Notes: Panel (A) plots the new business formation (limited liability) density over time for China, India, Brazil and UK. The rates are drastically low in India Panel (B) plots the growth of regular and R&D firm entry over time in India since early 1990s. Relative to the pre-policy trend overall entry increases by a small amount but the less than the entry growth of R&D firms.

D.2.2 Life Cycle Growth Evolution Results

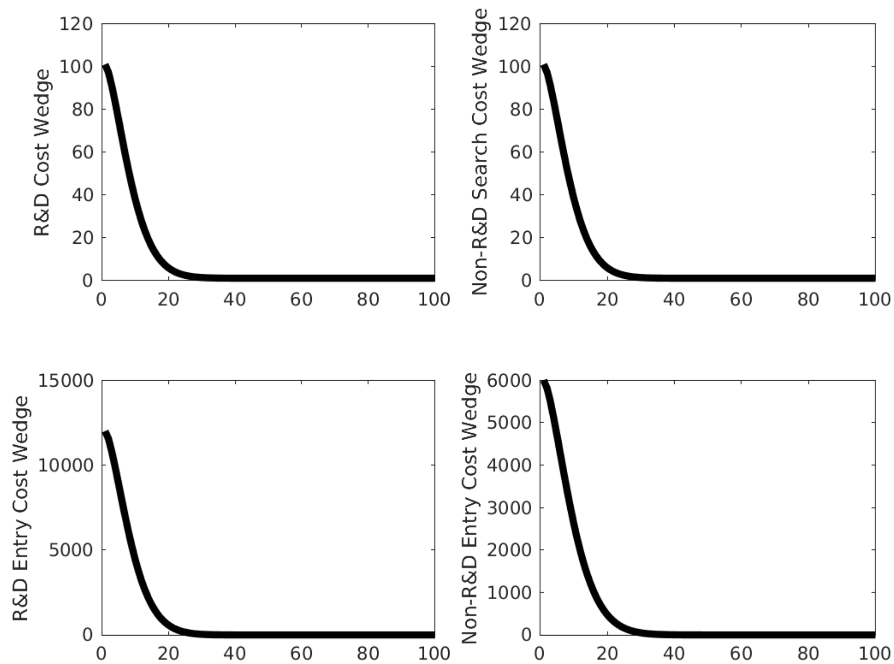
Figure 19: Matching the Transition



The panel (B) of the figure plots the change in life-cycle growth of the cohort born in 1990 vs 2000. The black line represents the ratio of the lifecycle growth in the model and the red line plots the corresponding ratio from the Prowess data. Panel (A) shows the implied evolution of the firm lifecycle in the model.

D.2.3 Transition Calibrated Wedges

Figure 20: Transition Calibrated Wedges



D.2.4 Transition with and without Aggregate Wedges

Figure 21: No friction transition

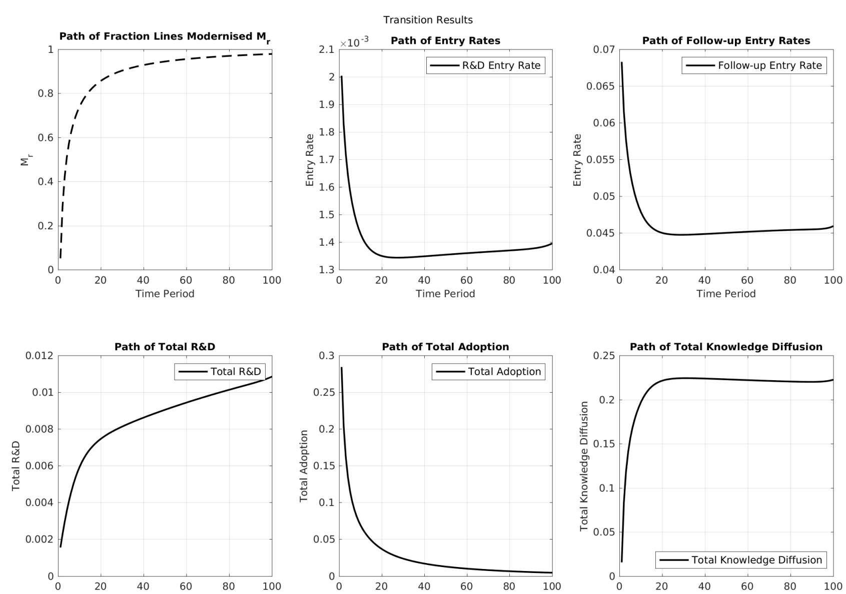
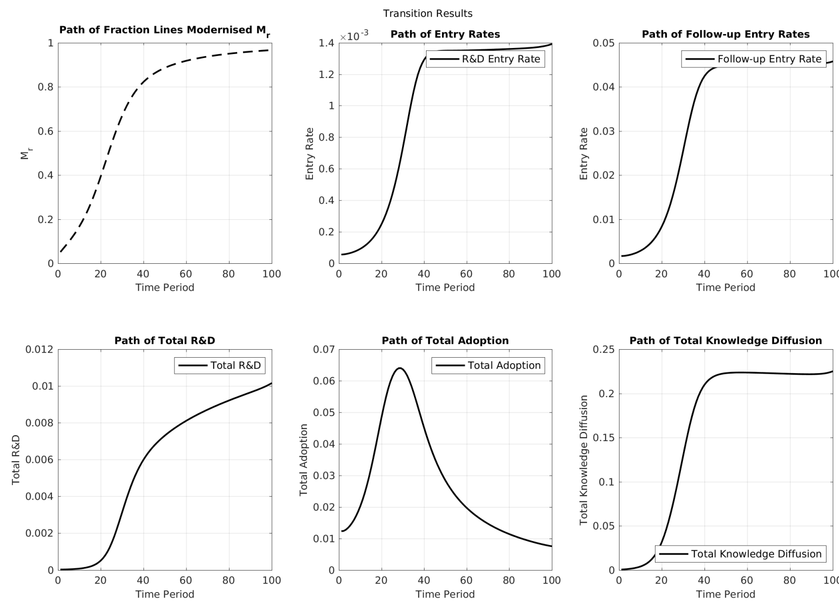


Figure 22: No friction transition



D.3 Policy IRFs in the Calibrated Model

Figure 23: IRF for targted recognition till age 10 for a cohort entering at time zero

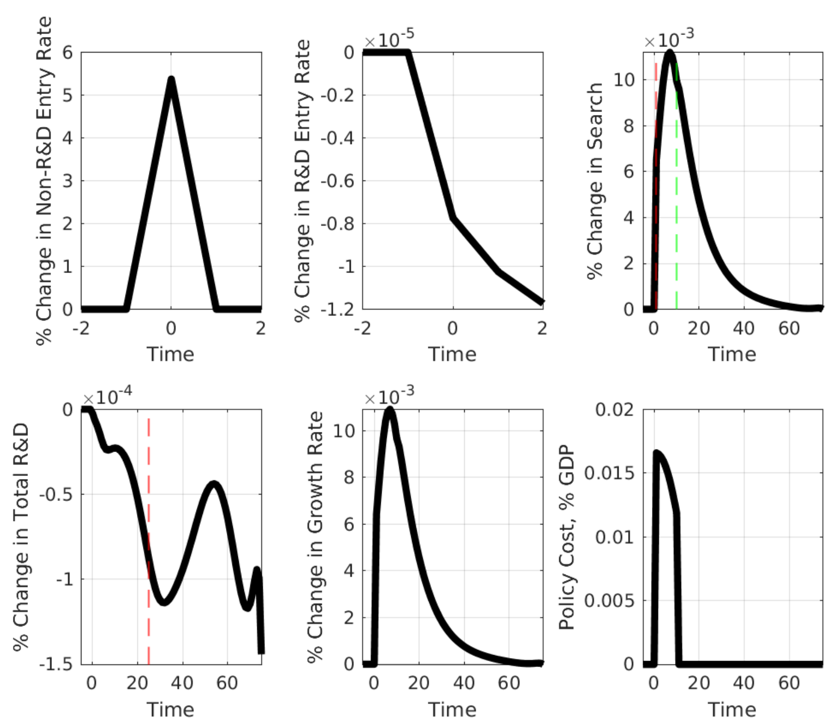


Figure 24: IRF for targeted R&D benefits for a cohort entering at time zero

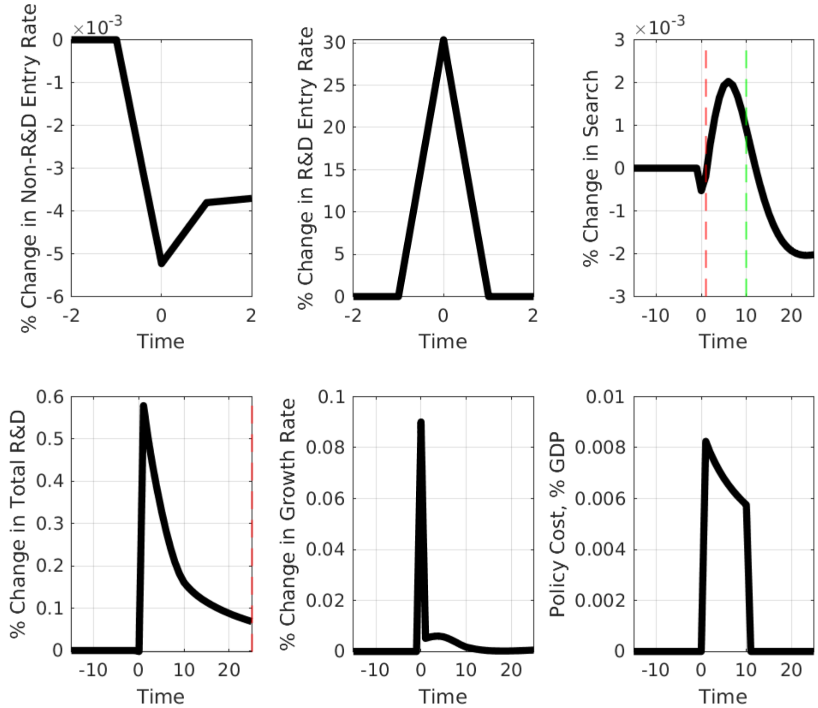


Figure 25: IRF for the targeted profit subsidy of 4% till age 10